

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC TÂY NGUYÊN



HỒ ĐÌNH BẢO

**ƯỚC LƯỢNG KHẢ NĂNG HẤP THỤ CO₂ CỦA KIỂU
RỪNG LÁ RỘNG THƯỜNG XANH TRÊN ĐỊA BÀN
TỈNH ĐẮK LẮK DỰA VÀO DỮ LIỆU VIỄN THÁM
VÀ KỸ THUẬT GIS**

LUẬN ÁN TIẾN SĨ LÂM SINH

ĐẮK LẮK - 2026

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC TÂY NGUYÊN**

HỒ ĐÌNH BẢO

**ƯỚC LƯỢNG KHẢ NĂNG HẤP THỤ CO₂ CỦA KIỂU
RỪNG LÁ RỘNG THƯỜNG XANH TRÊN ĐỊA BÀN
TỈNH ĐẮK LẮK DỰA VÀO DỮ LIỆU VIỄN THÁM
VÀ KỸ THUẬT GIS**

Ngành: Lâm sinh

Mã số: 9 62 02 05

LUẬN ÁN TIẾN SĨ LÂM SINH

Người hướng dẫn khoa học: PGS.TS. Nguyễn Thị Thanh Hương

ĐẮK LẮK - 2026

LỜI CAM ĐOAN

Luận án Tiến sĩ này được hoàn thành theo Chương trình đào tạo tiến sĩ tại Trường Đại học Tây Nguyên, khóa 3 (2021 – 2026). Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi. Các số liệu và kết quả trình bày trong luận án là trung thực và chưa từng được sử dụng để bảo vệ bất kỳ luận án nào khác. Các thông tin, tài liệu khoa học được trích dẫn trong luận án đều đã được ghi rõ nguồn gốc.

Một phần số liệu trong luận án được sử dụng từ công trình “*Điều tra phân bố, sinh thái của một số loài thực vật thân gỗ quý hiếm phục vụ công tác bảo tồn nguồn gen tại tỉnh Đắk Lắk*” do PGS.TS. Bảo Huy làm chủ nhiệm. Trong công trình này, tôi là một trong những thành viên nghiên cứu của đề tài. Toàn bộ các dữ liệu đều được trích dẫn và ghi rõ nguồn gốc trong luận án.

Đắk Lắk, ngày 04 tháng 03 năm 2026

Người hướng dẫn khoa học

Tác giả

PGS.TS. Nguyễn Thị Thanh Hương

Hồ Đình Bảo

LỜI CẢM ƠN

Trước hết, tôi xin bày tỏ lòng biết ơn sâu sắc đến PGS.TS. Nguyễn Thị Thanh Hương – người đã dành nhiều thời gian, tâm huyết và kiến thức quý báu để hướng dẫn tôi trong suốt quá trình thực hiện luận án tiến sĩ này. Những gợi ý sâu sắc, tinh thần nghiêm túc trong nghiên cứu và sự hỗ trợ tận tình của cô là nguồn động lực to lớn để tôi hoàn thành công trình nghiên cứu này.

Tôi xin chân thành cảm ơn Ban Giám hiệu Trường Đại học Tây Nguyên, cùng các thầy cô trong Bộ môn Lâm nghiệp, Khoa Nông nghiệp đã tạo điều kiện thuận lợi cho tôi trong suốt quá trình học tập, nghiên cứu và bảo vệ luận án.

Tôi cũng xin gửi lời cảm ơn tới các chuyên gia, đồng nghiệp, cơ quan, đơn vị đã hỗ trợ tôi trong việc thu thập số liệu, cung cấp tư liệu, chia sẻ kinh nghiệm và tạo điều kiện thực địa phục vụ nghiên cứu.

Đặc biệt, tôi xin bày tỏ lòng biết ơn sâu sắc đến gia đình, người thân và bạn bè, những người luôn đồng hành, động viên và là chỗ dựa tinh thần vững chắc giúp tôi vượt qua những khó khăn trong suốt hành trình nghiên cứu và học tập đầy thử thách này.

Mặc dù đã nỗ lực hết mình, luận án chắc chắn vẫn còn những thiếu sót nhất định. Tôi rất mong nhận được những ý kiến góp ý quý báu từ các thầy cô, nhà khoa học và độc giả để hoàn thiện hơn trong các nghiên cứu tiếp theo.

Xin trân trọng cảm ơn!

Đắk Lắk, ngày 04 tháng 03 năm 2026

Tác giả luận án

Hồ Đình Bảo

MỤC LỤC

DANH MỤC CÁC BẢNG.....	x
DANH MỤC CÁC HÌNH ẢNH.....	xi
MỞ ĐẦU.....	1
1. Tính cấp thiết của đề tài.....	1
2. Mục tiêu nghiên cứu	2
3. Phạm vi nghiên cứu	3
4. Ý nghĩa khoa học và thực tiễn	4
5. Những điểm mới của nghiên cứu.....	4
6. Cấu trúc của luận án.....	5
CHƯƠNG 1. TỔNG QUAN VẤN ĐỀ NGHIÊN CỨU.....	6
1.1. Nhu cầu xác định sinh khối, các-bon rừng.....	6
1.2. Ước tính sinh khối, các-bon cây rừng.....	8
1.3. Ứng dụng GIS và viễn thám trong ước tính sinh khối cây rừng.....	11
1.3.1. Nguyên tắc Ước tính AGB sử dụng dữ liệu Viễn thám	13
1.3.2. Nguồn dữ liệu ảnh viễn thám sử dụng ước lượng AGB.....	15
1.3.3. Các phương pháp ước tính AGB	21
1.4. Thách thức và triển vọng khi sử dụng ảnh viễn thám để ước tính AGB	27
1.5. Thảo luận.....	29
1.6. Khoảng trống nghiên cứu và lý do lựa chọn đề tài	31
1.6.1. Khoảng trống nghiên cứu	31
1.6.2. Lý do lựa chọn đề tài	32
CHƯƠNG 2. ĐỐI TƯỢNG, NỘI DUNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU.....	34
2.1. Đối tượng và địa điểm nghiên cứu.....	34
2.1.1. Đối tượng nghiên cứu.....	34
2.1.2. Địa điểm nghiên cứu	34
2.2. Nội dung nghiên cứu.....	35
2.3. Phương pháp nghiên cứu.....	36

2.3.1. Thu thập và xử lý dữ liệu thực địa	39
2.3.2. Dữ liệu viễn thám	44
2.3.3. Phân tích thay đổi thảm phủ rừng giai đoạn 2015-2025	49
2.3.4. Phân tích mối quan hệ giữa đặc trưng ảnh viễn thám và AGB	52
2.3.5. Xây dựng mô hình ước tính AGB từ dữ liệu viễn thám.....	55
2.3.6. Lập bản đồ AGB và lượng CO ₂ hấp thụ, đánh giá biến động AGB giai đoạn 2015-2025	61
CHƯƠNG 3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN	66
3.1. Đánh giá thay đổi RTX giai đoạn 2015 – 2025	66
3.1.1. Phân loại ảnh xác định kiểu RTX.....	66
3.1.2. Phân tích thay đổi diện tích RTX giai đoạn 2015 – 2025	71
3.2. Phân tích mối quan hệ giữa đặc trưng ảnh viễn thám và AGB.....	75
3.2.1. Chuẩn hóa dữ liệu sinh khối ô mẫu	75
3.2.2. Phân tích tương quan Pearson	79
3.2.3. Phân tích thành phần chính (PCA)	82
3.2.4. Phân tích đa cộng tuyến (VIF - Variance Inflation Factor).....	86
3.3. Xây dựng mô hình ước tính AGB từ dữ liệu viễn thám	87
3.3.1. Ước tính AGB sử dụng mô hình hồi quy tuyến tính	87
3.3.2. Ước tính AGB rừng sử dụng mô hình hồi quy phi tuyến.....	92
3.3.3. Ước tính AGB rừng sử dụng mô hình RF	95
3.3.4. Đánh giá độ chính xác của các mô hình ước lượng	104
3.4. Lập bản đồ AGB rừng và khả năng hấp thụ CO ₂ giai đoạn 2015 – 2025	111
3.4.1. Xây dựng bản đồ AGB giai đoạn 2015 - 2025.....	111
3.4.2. Phân tích biến động AGB giai đoạn 2015 - 2025	113
3.4.3. Kiểm định xu thế AGB giai đoạn 2015-2025	122
3.4.4. Ước tính khả năng hấp thụ CO ₂ của rừng	124
KẾT LUẬN VÀ KIẾN NGHỊ.....	129
Kết luận	129
Tồn tại	130

Kiến nghị	131
DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ.....	133
TÀI LIỆU THAM KHẢO.....	134
PHỤ LỤC	P-1

DANH MỤC TỪ VIẾT TẮT

Từ viết tắt	Đầy đủ	Nguyên nghĩa
AGB	Aboveground Biomass	Sinh khối trên mặt đất
ANN	Artificial Neural Network	Mạng nơ-ron nhân tạo
ARVI	Atmospherically Resistant Vegetation Index	Chỉ số thảm thực vật kháng khí quyển
BGB	Belowground Biomass	Sinh khối dưới mặt đất
CA	Crown Area	Diện tích tán lá
CD	Crown Diameter	Đường kính tán lá
DBH	Diameter at Breast Height	Đường kính ngang ngực
DL	Deep Learning	Học sâu
DVI	Difference Vegetation Index	Chỉ số khác biệt thực vật
EVI	Enhanced Vegetation Index	Chỉ số thực vật cải tiến
GAM	Generalized Additive Model	Mô hình cộng tính tổng quát
GEE	Google Earth Engine	Nền tảng điện toán đám mây của Google
GEDI	Global Ecosystem Dynamics Investigation	Nhiệm vụ điều tra động thái hệ sinh thái toàn cầu (LiDAR trên ISS)
GIS	Geographic Information System	Hệ thống thông tin địa lý
H	Height	Chiều cao cây
HH	Horizontal-Horizontal polarization	Phân cực ngang – ngang (radar)
HV	Horizontal-Vertical polarization	Phân cực ngang – dọc (radar)
LAI	Leaf Area Index	Chỉ số diện tích lá
LiDAR	Light Detection and Ranging	Công nghệ quét laser
LR	Linear Regression	Hồi quy tuyến tính

Từ viết tắt	Đầy đủ	Nguyên nghĩa
MAE	Mean Absolute Error	Sai số tuyệt đối trung bình
MK	Mann-Kendall test	Kiểm định Mann-Kendall
ML	Machine Learning	Học máy
MR	Multiple Regression	Hồi quy đa biến
MSI	Moisture Stress Index	Chỉ số căng thẳng ẩm
NDVI	Normalized Difference Vegetation Index	Chỉ số khác biệt thực vật chuẩn hóa
NDWI	Normalized Difference Water Index	Chỉ số khác biệt nước chuẩn hóa
NIR	Near Infrared	Cận hồng ngoại
NQ/TW		Nghị quyết/Trung ương
OA	Overall Accuracy	Độ chính xác tổng thể
OLS	Ordinary Least Squares	Hồi quy bình phương tối thiểu thông thường
OOB	Out Of Bag	Dữ liệu ngoài túi
PA	Producer's Accuracy	Độ chính xác nhà sản xuất
PCA	Principal Component Analysis	Phân tích thành phần chính
PFES	Payment for Forest Ecosystem Services	Chi trả dịch vụ hệ sinh thái
QĐ-TTg		Quyết định – Thủ tướng
R ²	Coefficient of Determination	Hệ số xác định
REDD+	Reduce Emission from Deforestation and Forest Degradation	Giảm phát thải từ suy thoái và mất rừng
RF	Random Forest	Thuật toán rừng ngẫu nhiên
RMSE	Root Mean Square Error	Sai số trung phương bình phương

Từ viết tắt	Đầy đủ	Nguyên nghĩa
RNDVI	Renormalized Difference Vegetation Index	Chỉ số NDVI tái chuẩn hóa
RTX		Rừng thường xanh
SAR	Synthetic Aperture Radar	Radar khẩu độ tổng hợp
SAVI	Soil-Adjusted Vegetation Index	Chỉ số thực vật hiệu chỉnh đất
SE	Standard Error	Sai số chuẩn
SIPI	Structure Insensitive Pigment Index	Chỉ số sắc tố không nhạy cấu trúc
SR	Simple Ratio	Chỉ số tỷ số đơn giản
SRTM	Shuttle Radar Topography Mission	Nhiệm vụ radar địa hình tàu con thoi
UA	User's Accuracy	Độ chính xác người dùng
UAV	Unmanned Aerial Vehicle	Thiết bị bay không người lái
VH	Vertical-Horizontal polarization	Phân cực dọc – ngang (radar)
VI _s	Vegetation Index	Các chỉ số thảm thực vật
VIF	Variance Inflation Factor	Hệ số phóng đại phương sai
VV	Vertical-Vertical polarization	Phân cực dọc – dọc (radar)

DANH MỤC CÁC BẢNG

Bảng 2.1. Các chỉ số ảnh vệ tinh Landsat	48
Bảng 2.2. Giá trị và mức độ chặt chẽ theo chỉ số Kappa (Nguồn: Navulur, 2007) ..	51
Bảng 3.1. Kết quả tổng hợp độ chính xác phân loại các năm (OA, Kappa, PA, UA)	66
Bảng 3.2. Diện tích RTX các huyện giai đoạn 2015-2025	75
Bảng 3.3. Thống kê mô tả dữ liệu AGB	76
Bảng 3.4. Phân loại nhóm biến theo mối tương quan với logAGB	80
Bảng 3.5. Các biến lựa chọn có VIF thấp	86
Bảng 3.6. Kết quả hồi quy đơn biến giữa logAGB và các biến viễn thám	88
Bảng 3.7. Kết quả phân tích hồi quy tuyến tính đa nhân tố với logAGB	90
Bảng 3.8. Kết quả mô hình GAM	93
Bảng 3.9. Hiệu suất mô hình RF theo các cấu hình khác nhau.....	101
Bảng 3.10. So sánh hiệu suất của mô hình RF dựa trên 2 bộ dữ liệu khác nhau	102
Bảng 3.11. Hiệu suất mô hình trên tập huấn luyện	104
Bảng 3.12. Kết quả kiểm định các mô hình hồi quy bằng cross validation.....	107
Bảng 3.13. Kết quả phân tích định lượng thay đổi sinh khối RTX tỉnh Đắk Lắk giai đoạn 2015–2025).....	117
Bảng 3.14. Kết quả phân tích định lượng thay đổi sinh khối RTX theo huyện, giai đoạn 2015–2025 (tần AGB)	118
Bảng 3.15. Tổng sinh khối RTX giai đoạn 2015-2025	122
Bảng 3.16. Kết quả kiểm định xu thế AGB giai đoạn 2015–2025	124
Bảng 3.17. Thống kê lượng CO ₂ tích lũy RTX tỉnh Đắk Lắk giai đoạn 2015–2025	127

DANH MỤC CÁC HÌNH ẢNH

Hình 1.1. Minh họa ước tính AGB bằng kỹ thuật viễn thám.....	13
Hình 1.2. Sử dụng phổ điện từ ước tính sinh khối rừng trên mặt đất (<i>Xiao và cộng sự, 2019</i>).....	15
Hình 1.3. Dữ liệu ảnh viễn thám khác nhau trong ước tính AGB (<i>Nguồn: Tian và cộng sự, 2023</i>).....	20
Hình 1.4. Các thuật toán sử dụng trong các nghiên cứu gần đây.....	22
Hình 2.1. Vị trí khu vực nghiên cứu	34
Hình 2.2. Sơ đồ tiếp cận nghiên cứu.....	37
Hình 2.3. Bản đồ vị trí các ô mẫu	41
Hình 2.4. Phương pháp lập ô mẫu đo đếm.....	42
Hình 2.5. Minh họa đoạn mã code thu thập ảnh trên GEE	46
Hình 3.1. Ma trận nhầm lẫn phân loại.....	67
Hình 3.2. Diện tích rừng thường xanh theo huyện giai đoạn 2015-2025	68
Hình 3.3. Phân bố thảm phủ RTX tỉnh Đắk Lắk giai đoạn 2015 - 2025	70
Hình 3.4. Bản đồ thay đổi diện tích rừng thường xanh.....	72
Hình 3.5. Thay đổi diện tích RTX của các huyện giai đoạn 2015-2025.....	74
Hình 3.6. Phân bố giá trị AGB của ô mẫu theo năm.....	76
Hình 3.7. Biểu đồ phân bố AGB ô mẫu (dữ liệu gốc)	77
Hình 3.8. Biểu đồ phân bố AGB ô mẫu sau biến đổi lô-ga-rít	78
Hình 3.9. Ma trận tương quan giữa các biến viễn thám và logAGB	79
Hình 3.10. So sánh tương quan AGB và loAGB với các nhân tố ảnh.....	81
Hình 3.11. Tỷ lệ phương sai giải thích của các thành phần chính	83

Hình 3.12. Biểu đồ phân tích mức đóng góp của các biến	84
Hình 3.13. Đồ thị hàm trơn từ mô hình GAM	93
Hình 3.14. Tương quan logAGB thực tế và ước tính.....	96
Hình 3.15. Mức độ quan trọng của các biến trong mô hình RF.....	97
Hình 3.16. Tương quan logAGB thực tế và ước tính (chọn lọc biến)	99
Hình 3.17. Tương quan logAGB thực tế và ước tính (chọn lọc biến)	100
Hình 3.18. Mức độ phù hợp giữa dữ liệu thực tế và ước tính từ các mô hình.....	106
Hình 3.19. So sánh hiệu suất mô hình RF bằng kiểm định chéo k-fold	108
Hình 3.20. Đồ thị tương quan AGB ước lượng và thực tế (độc lập)	110
Hình 3.21. Phân bố sinh khối rừng qua các giai đoạn	112
Hình 3.22. Bản đồ biến động AGB giai đoạn 2015 – 2025	114
Hình 3.23. Bản đồ phân loại biến động AGB rừng tỉnh Đắk Lắk các giai đoạn	116
Hình 3.24. Biểu đồ thay đổi AGB theo huyện giai đoạn 2015-2025.....	119
Hình 3.25. Tổng sinh khối rừng giai đoạn 2015 – 2025	121
Hình 3.26. Đường LOESS mô tả xu thế phi tuyến của tổng sinh khối RTX giai đoạn 2015–2025.....	123
Hình 3.27. Bản đồ phân bố CO ₂ tích lũy giai đoạn 2015-2025	126
Hình 3.28. Biểu đồ tổng tích lũy CO ₂ theo huyện	128

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Biến đổi khí hậu đang là thách thức môi trường lớn nhất của thế kỷ 21, tác động mạnh đến các hệ sinh thái toàn cầu, đặc biệt tại vùng nhiệt đới – nơi có độ nhạy cao trước hạn hán, nắng nóng và các hiện tượng cực đoan [86]. Trong bối cảnh đó, rừng giữ vai trò thiết yếu trong điều hòa khí hậu khi lưu trữ từ 76% đến 98% lượng các-bon trên mặt đất [87], [148] và duy trì chu trình các-bon thông qua hấp thụ CO₂ từ khí quyển [134],[150],[177], [184]. Đối với vùng nhiệt đới ẩm, rừng thường xanh là một trong những hệ sinh thái có khả năng hấp thụ và lưu trữ các-bon lớn nhất, đóng góp trực tiếp vào giảm nhẹ biến đổi khí hậu.

AGB là tham số quan trọng phản ánh cấu trúc, năng suất và khả năng hấp thụ CO₂ của rừng [90],[114],[126]. Việc ước tính chính xác AGB và lượng CO₂ hấp thụ là cơ sở khoa học cho quản lý rừng bền vững, kiểm kê khí nhà kính và hoạch định chính sách ứng phó biến đổi khí hậu [163].

Tại Việt Nam, các chính sách lớn như Luật Lâm nghiệp năm 2017 (được Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam thông qua ngày 15/11/2017), Chiến lược phát triển lâm nghiệp giai đoạn 2021–2030, tầm nhìn đến năm 2050 (ban hành kèm theo Quyết định số 523/QĐ-TTg ngày 01/4/2021 của Thủ tướng Chính phủ), và Nghị quyết số 57-NQ/TW ngày 22/12/2024 của Bộ Chính trị đều nhấn mạnh yêu cầu hiện đại hóa công tác điều tra, giám sát tài nguyên rừng và nâng cao vai trò của rừng trong thực hiện mục tiêu phát thải ròng bằng “0” (Net-Zero). Điều này đòi hỏi các phương pháp định lượng các-bon rừng phải có độ chính xác cao, minh bạch và đáp ứng các tiêu chuẩn quốc tế.

Đối với khu vực Tây Nguyên, đặc biệt tại Đắk Lắk, rừng thường xanh có ý nghĩa quan trọng đối với điều tiết khí hậu và bảo vệ lưu vực. Tuy nhiên, khu vực này đang chịu áp lực từ suy thoái rừng, biến động sử dụng đất và hạn hán kéo dài [19]. Nhu cầu giám sát AGB và CO₂ hấp thụ một cách chính xác, liên tục và quy mô lớn

vì vậy trở nên đặc biệt cấp bách.

Trong những năm gần đây, sự phát triển mạnh mẽ của công nghệ viễn thám (bao gồm cả ảnh quang học và ra-đa) kết hợp với các thuật toán học máy như Random Forest (RF) đã mở ra các phương pháp hiện đại và hiệu quả để mô hình hóa sinh khối rừng và lượng CO₂ hấp thụ. Ví dụ, nghiên cứu tại rừng Amazon cho thấy khi kết hợp dữ liệu ra-đa và quang học với RF cho độ chính xác rất cao [33], và tổng quan gần đây cũng nhấn mạnh tầm quan trọng của việc sử dụng nhiều nguồn (multi-sensor) trong ước tính trữ lượng cacbon [142]. Ứng dụng tại các hệ rừng nhiệt đới và núi cao (như Myanmar) cũng cho thấy Sentinel-2 và RF hoặc mô hình máy học khác đem lại khả năng ước tính AGB rất tốt [179]. Những công cụ này đặc biệt phù hợp cho rừng nhiệt đới có điều kiện địa hình phức tạp và biến động cao, nơi việc khảo sát thực địa truyền thống gặp nhiều khó khăn.

Vì vậy, đề tài “Ước tính lượng CO₂ hấp thụ của rừng thường xanh bằng dữ liệu viễn thám và kỹ thuật GIS” mang tính cấp thiết trong bối cảnh hiện nay, góp phần cung cấp bằng chứng khoa học về sinh khối – các-bon rừng tại Đắk Lắk, đồng thời đóng góp vào việc hoàn thiện các phương pháp giám sát rừng hiện đại phục vụ quản lý bền vững và ứng phó biến đổi khí hậu.

2. Mục tiêu nghiên cứu

- Mục tiêu tổng quát: Ước tính lượng CO₂ hấp thụ từ sinh khối trên mặt đất (AGB) của kiểu rừng thường xanh trên địa bàn tỉnh Đắk Lắk thông qua tích hợp dữ liệu viễn thám đa nguồn, kỹ thuật GIS và số liệu điều tra thực địa.
- Mục tiêu cụ thể
 - Phân tích mối quan hệ giữa các chỉ số và đặc trưng ảnh viễn thám với AGB của rừng thường xanh.
 - Xây dựng và đánh giá mô hình ước tính AGB từ dữ liệu viễn thám, lựa chọn mô hình tối ưu cho khu vực nghiên cứu.
 - Lập bản đồ AGB và lượng CO₂ hấp thụ, đồng thời phân tích biến động

theo không gian – thời gian.

3. Phạm vi nghiên cứu

Trong khuôn khổ thời gian và nguồn lực cũng như khả năng tiếp cận dữ liệu, phạm vi nghiên cứu được xác định như sau:

- *Phạm vi không gian và đối tượng*

Nghiên cứu chỉ thực hiện trong phạm vi tỉnh Đắk Lắk (địa giới hành chính cũ) và tập trung vào một kiểu rừng là rừng lá rộng thường xanh, không xem xét các kiểu rừng khác. Nghiên cứu tập trung vào 5 huyện có diện tích rừng thường xanh lớn của tỉnh Đắk Lắk là Krông Năng, Ea Kar, M’Đrắk, Krông Bông và Lắk. Thành phần loài trong các ô mẫu được tiếp cận ở mức tổng hợp chung, không phân tích chi tiết theo từng loài hoặc nhóm loài.

- *Phạm vi thời gian*

Do giới hạn về dữ liệu ảnh ra-đa Sentinel-1 chỉ có từ năm 2015 nên đối với xây dựng mô hình ước lượng AGB, mô hình được phát triển từ các ô mẫu thực địa thu thập giai đoạn 2020–2024. Bộ ô mẫu năm 2013 từ nghiên cứu trước được sử dụng nhằm kiểm tra khả năng áp dụng của mô hình đối với dữ liệu quá khứ thông qua ảnh Landsat 8 OLI năm 2013. Đối với phân tích biến động rừng thường xanh và trữ lượng các-bon, nghiên cứu chỉ sử dụng chuỗi ảnh Sentinel-1 giai đoạn 2015–2025 và Landsat 8 OLI.

- *Phạm vi nội dung*

Luận án tập trung nghiên cứu các vấn đề chính như sau:

- ✓ Phân loại, lập bản đồ và đánh giá thay đổi RTX tỉnh Đắk Lắk giai đoạn 2015-2025.
- ✓ Phân tích mối quan hệ giữa đặc trưng ảnh viễn thám và AGB.
- ✓ Xây dựng mô hình ước tính AGB từ dữ liệu viễn thám.
- ✓ Lập bản đồ AGB và lượng CO₂ hấp thụ, đánh giá biến động theo thời gian.

Trong khuôn khổ nghiên cứu, luận án chỉ thử nghiệm một số nhóm mô hình chính, gồm: (i) các mô hình hồi quy tuyến tính và phi tuyến (hồi quy đơn biến, đa biến và mô hình cộng tính tổng quát (GAM), và (ii) mô hình học máy Random Forest (RF). Các thuật toán học máy khác (SVM, ANN, XGBoost...) cũng như các phương pháp mô hình hóa vật lý hoặc cơ học không nằm trong phạm vi thực hiện. Bên cạnh đó, do hạn chế về thời gian và nguồn lực, nghiên cứu chưa triển khai đánh giá đầy đủ sai số và độ bất định của mô hình bao gồm toàn bộ các nguồn sai số phát sinh từ điều tra ô mẫu, mô hình hóa AGB và quá trình lập bản đồ CO₂.

4. Ý nghĩa khoa học và thực tiễn

Ý nghĩa khoa học:

Góp phần làm rõ mối quan hệ giữa đặc trưng ảnh viễn thám và sinh khối rừng, từ đó phát triển các mô hình ước tính khả năng hấp thụ CO₂ dựa trên nền tảng thống kê hiện đại và nền tảng dữ liệu mở;

Đóng góp cơ sở lý luận và phương pháp luận cho việc tích hợp viễn thám đa nguồn và GIS trong đánh giá dịch vụ hệ sinh thái rừng, đặc biệt là dịch vụ hấp thụ các-bon và chương trình trung hòa các-bon.

Ý nghĩa thực tiễn:

Kết quả nghiên cứu hỗ trợ cơ quan quản lý nhà nước, chủ rừng và các bên liên quan trong việc xác định tiềm năng hấp thụ CO₂ của rừng theo thời gian, phục vụ hiệu quả chính sách chi trả dịch vụ các-bon rừng, quản lý rừng bền vững và trong các chương trình trung hòa các-bon;

Mô hình được xây dựng từ nghiên cứu có thể được áp dụng mở rộng cho các khu vực có điều kiện sinh thái tương đồng, cũng như các kiểu rừng khác nhằm góp phần nâng cao hiệu quả quản lý tài nguyên rừng trên quy mô rộng.

5. Những điểm mới của nghiên cứu

Nghiên cứu này đóng góp một số điểm mới nổi bật như sau:

- ✓ Xây dựng phương pháp tích hợp dữ liệu viễn thám đa nguồn (Sentinel-1, Landsat), GIS và mô hình học máy RF để ước tính sinh khối AGB và lượng CO₂ hấp thụ của rừng thường xanh trên địa bàn tỉnh Đắk Lắk.
- ✓ Xác định được một số đặc trưng ảnh viễn thám có mối quan hệ chặt với AGB và khả năng hấp thụ CO₂ của RTX tỉnh Đắk Lắk.
- ✓ Cung cấp bằng chứng khoa học hỗ trợ việc định giá dịch vụ hấp thụ carbon của rừng, phù hợp với định hướng chuyển đổi số và cam kết Net-Zero của Việt Nam.

6. Cấu trúc của luận án

Luận án gồm có các phần được bố cục như sau:

Mở đầu: Trình bày bối cảnh nghiên cứu, lý do lựa chọn đề tài, mục tiêu – câu hỏi – giả thuyết nghiên cứu, phạm vi, giới hạn và đóng góp khoa học của đề tài.

Chương 1. Tổng quan tài liệu: Tổng hợp các công trình nghiên cứu liên quan trong và ngoài nước về ước tính sinh khối rừng, khả năng hấp thụ CO₂, viễn thám và ứng dụng GIS trong ước tính sinh khối.

Chương 2. Cơ sở lý thuyết và phương pháp nghiên cứu: Trình bày đặc điểm của khu vực nghiên cứu; nguồn, loại và cách xử lý dữ liệu viễn thám, dữ liệu thực địa và dữ liệu phụ trợ khác; Khung lý thuyết, khái niệm, các công thức chuyển đổi AGB – carbon – CO₂, quy trình xử lý dữ liệu viễn thám và phương pháp xây dựng mô hình ước tính.

Chương 3. Kết quả và thảo luận: Trình bày các kết quả chính: mô hình ước tính AGB, bản đồ phân bố AGB, lượng CO₂ hấp thụ, phân tích và bàn luận kết quả.

Kết luận và kiến nghị: Tóm tắt các phát hiện chính, đóng góp thực tiễn – khoa học, hạn chế của nghiên cứu và hướng nghiên cứu tiếp theo.

CHƯƠNG 1. TỔNG QUAN VẤN ĐỀ NGHIÊN CỨU

1.1. Nhu cầu xác định sinh khối, các-bon rừng

Rừng là một trong những bể chứa các-bon lớn nhất trên lục địa và đóng vai trò quan trọng trong việc loại bỏ CO₂ khỏi khí quyển. Theo Báo cáo Tổng hợp lần thứ sáu của Tổ chức liên chính phủ về Biến đổi khí hậu [94], để đạt mục tiêu phát thải ròng bằng không, các quốc gia cần đồng thời giảm nhanh phát thải khí nhà kính và tăng cường bảo vệ, phục hồi các bể chứa các-bon tự nhiên, trong đó rừng là trụ cột. Điều này đòi hỏi phải nắm bắt chính xác phân bố sinh khối và năng lực tích lũy các-bon của rừng ở nhiều cấp độ nhằm hỗ trợ lập kế hoạch, theo dõi và đánh giá hiệu quả các hành động giảm phát thải [94].

Các đánh giá gần đây cho thấy lượng các-bon lưu trữ trong rừng toàn cầu vẫn thấp hơn tiềm năng, đặc biệt tại vùng nhiệt đới, nơi còn nhiều cơ hội phục hồi rừng và cải thiện cấu trúc quần xã. Đồng thời, các biến cố như cháy rừng, hạn hán kéo dài, suy thoái rừng và mất rừng đã làm suy giảm khả năng hấp thụ CO₂ của hệ sinh thái rừng trong một số giai đoạn, cho thấy sự dễ tổn thương của các bể chứa tự nhiên [82],[94]. Điều này nhấn mạnh nhu cầu theo dõi sinh khối và các-bon rừng liên tục, chính xác và ở độ phân giải cao.

Trong thực tiễn quản lý, việc ước lượng sinh khối đóng vai trò quan trọng để cân bằng giữa sản xuất gỗ và duy trì chức năng hấp thụ các-bon. Tuy nhiên, dữ liệu hiện trường truyền thống thường không đủ đại diện cho sự đa dạng về loài cây, tuổi rừng, tầng tán và điều kiện địa hình, gây khó khăn khi xây dựng các phương án quản lý tối ưu. Xu hướng nghiên cứu hiện nay hướng đến tích hợp dữ liệu đa nguồn, gồm đo đạc hiện trường, ảnh vệ tinh quang học, dữ liệu radar và LiDAR (mặt đất, UAV,...), nhằm khắc phục hiện tượng bão hòa tín hiệu quang học và mô tả tốt hơn cấu trúc thẳng đứng của rừng [110],[133].

Tại Việt Nam, sự đa dạng hệ sinh thái rừng cùng áp lực từ suy thoái rừng, chuyển đổi đất và biến đổi khí hậu làm cho nhu cầu xác định phân bố sinh khối và

động thái các-bon theo không gian trở nên đặc biệt quan trọng. Đây là cơ sở để quy hoạch sử dụng và quản lý rừng, phân vùng bảo vệ – phục hồi – khai thác và giảm đánh đổi giữa mục tiêu kinh tế và khí hậu. Đồng thời, dữ liệu sinh khối là thành phần quan trọng của hệ thống kiểm kê khí nhà kính quốc gia và báo cáo REDD⁺.

Đáng chú ý, Việt Nam đã tuyên bố cam kết đạt phát thải ròng bằng “0” vào năm 2050 tại COP26. Cam kết này được thể chế hóa trong Chiến lược quốc gia về biến đổi khí hậu đến năm 2050, ban hành theo Quyết định số 896/QĐ-TTg ngày 26/7/2022 và Quyết định số 888/QĐ-TTg ngày 25/7/2022 của Thủ tướng Chính phủ. Để triển khai các mục tiêu này, Chính phủ đã ban hành Nghị định số 06/2022/NĐ-CP ngày 07/01/2022 và Nghị định số 119/2025/NĐ-CP ngày 09/6/2025, quy định cụ thể về giảm phát thải khí nhà kính, phát triển thị trường các-bon và các cơ chế quản lý liên quan. Đồng thời, Chỉ thị số 13/CT-TTg ngày 02/5/2024 và Quyết định số 232/QĐ-TTg ngày 24/01/2025 tiếp tục nhấn mạnh vai trò then chốt của rừng trong tăng hấp thụ các-bon, giảm phát thải từ mất rừng và suy thoái rừng, cũng như yêu cầu cấp thiết về xây dựng cơ sở dữ liệu các-bon rừng minh bạch, chính xác để phục vụ thị trường các-bon và hoàn thiện hệ thống MRV (Measurement, Reporting and Verification) trong lĩnh vực sử dụng đất và lâm nghiệp.

Bên cạnh yếu tố chính sách, động lực kinh tế từ thị trường tín chỉ các-bon toàn cầu, cơ chế CORSIA (Chương trình Bù đắp và Giảm thiểu Carbon cho Hàng không Quốc tế), PFES (Chi trả dịch vụ môi trường rừng) và các dự án các-bon tự nguyện đang thúc đẩy nhu cầu có dữ liệu sinh khối tin cậy. Để phát triển tín chỉ các-bon chất lượng cao, các đơn vị quản lý rừng cần cung cấp bằng chứng định lượng về mức hấp thụ CO₂ và biến động sinh khối theo thời gian, dựa trên phương pháp minh bạch và có thể kiểm chứng độc lập [176]. Thực tiễn triển khai một số giao dịch tín chỉ các-bon rừng ở Việt Nam cho thấy tính chính xác của dữ liệu sinh khối quyết định trực tiếp tới chất lượng tín chỉ.

Ở quy mô quốc gia, thiếu dữ liệu chi tiết hoặc sai lệch trong ước lượng sinh khối có thể ảnh hưởng đến báo cáo kiểm kê khí nhà kính, kết quả REDD⁺ và tiến độ

thực hiện mục tiêu phát thải ròng bằng không. Do đó, các xu hướng hiện nay tập trung vào chuẩn hóa phương pháp ước lượng, ứng dụng dữ liệu viễn thám đa nguồn, tăng cường giám sát liên tục và kiểm định độc lập, đáp ứng yêu cầu của các thị trường các-bon quốc tế [116].

Tóm lại, trong bối cảnh yêu cầu giảm phát thải toàn cầu và mục tiêu phát thải ròng bằng không vào năm 2050 của Việt Nam, việc xác định sinh khối và các-bon rừng một cách chính xác, minh bạch và có thể kiểm chứng là yêu cầu cấp thiết. Phương pháp kết hợp dữ liệu đa nguồn và hệ thống giám sát đáng tin cậy là nền tảng quan trọng cho quản lý rừng bền vững, phát triển thị trường các-bon và thực thi cam kết khí hậu quốc gia.

1.2. Ước tính sinh khối, các-bon cây rừng

Các mô hình ước tính sinh khối thường được thực hiện một cách chính xác thông qua sinh khối thân cây, do đó, các yếu tố đo đếm được sử dụng trong quá trình tính toán trữ lượng cây đứng được coi là những nhân tố đáng tin cậy nhất [151]. Phương pháp chặt hạ cây mẫu (destructive sampling) vẫn được xem là cách tiếp cận chính xác nhất trong việc xác định sinh khối và lượng các-bon của cây rừng, vì cho phép đo đạc trực tiếp khối lượng khô và tỷ lệ các-bon trong từng bộ phận (thân, cành, lá, rễ). Tuy nhiên, phương pháp này bị hạn chế bởi chi phí cao, yêu cầu nhân lực lớn và tác động tiêu cực đến hệ sinh thái, do đó chỉ được áp dụng ở quy mô nhỏ để hiệu chỉnh các mô hình không chặt hạ (non-destructive models) [62].

Số lượng và cấu trúc mẫu chặt hạ phụ thuộc vào đặc trưng sinh thái và phạm vi ứng dụng mô hình. Dutcã và cộng sự (2020) khuyến nghị số lượng cây mẫu cần thiết nằm trong khoảng 20–150 cây để đảm bảo độ tin cậy thống kê từ 90–95%, trong khi Picard và cộng sự (2012) cho rằng khoảng 100 cây phân bố đồng đều theo các cấp kính là tối ưu. Đối với rừng thuần loài hoặc đồng tuổi, việc lấy mẫu đồng đều theo cấp kính giúp giảm sai số nội tại; trong khi đó, ở rừng khác tuổi hoặc đa loài, cần phân bố mẫu theo tiết diện ngang (basal area) hoặc mật độ loài chiếm ưu thế [38].

Các phép đo sinh khối tươi, tỷ lệ khô và hàm lượng các-bon của từng bộ phận cây được thực hiện theo quy trình chuẩn hóa, thường sấy ở nhiệt độ 80 - 105°C cho đến khi đạt khối lượng ổn định [83]. Những dữ liệu này là cơ sở để hiệu chỉnh và kiểm chứng mối quan hệ giữa sinh khối và các biến điều tra lâm phần như đường kính ngang ngực (DBH), chiều cao cây (H), diện tích tán lá (CA), đường kính tán (CD) và dung trọng gỗ (WD) [45].

Trong các mô hình sinh khối, DBH thường là biến độc lập có mức giải thích cao nhất vì có tương quan thuận mạnh với sinh khối cây rừng [45]. Biến H giúp cải thiện độ chính xác, đặc biệt trong các vùng có biến động lập địa lớn [38],[106], trong khi WD đóng vai trò chuyển đổi từ thể tích tươi sang sinh khối khô và phản ánh đặc tính cơ học – sinh lý của loài cây [83]. Một số nghiên cứu gần đây cũng cho thấy việc bổ sung biến CD hoặc CA có thể tăng độ chính xác của mô hình, vì hai cây có cùng DBH và H có thể có sinh khối khác nhau do khác biệt về tán lá và quang hợp.

Các mô hình sinh trắc thường được thể hiện dưới dạng tổng quát:

$$y_i = f(x_j) \quad (1.1)$$

Trong đó:

y_i : là sinh khối khô hoặc lượng các-bon

x_j : là các biến dự báo (predictors).

Các hàm thường dùng gồm hàm lũy thừa (Power), lô-ga-rít, hàm mũ, hoặc các hàm kết hợp. Trong số đó, hàm Power vẫn là lựa chọn phổ biến nhất cho vùng nhiệt đới nhờ tính đơn giản và khả năng mô tả quan hệ phi tuyến giữa sinh khối và DBH [45],[152].

Ở các quốc gia ôn đới, nơi tổ thành loài ít phức tạp, các mô hình sinh khối loài riêng biệt đã được thiết lập cho hầu hết cây bản địa [96]. Ngược lại, ở các khu rừng nhiệt đới đa dạng loài như Việt Nam, việc xây dựng mô hình cho từng loài riêng biệt là bất khả thi, nên hướng tiếp cận hiện nay tập trung vào các mô hình chung đa loài hoặc mô hình vùng sinh thái [45],[106].

Trong hơn một thập kỷ qua, Việt Nam đã đạt được nhiều tiến bộ trong phát

triển mô hình sinh khối – các-bon phục vụ cho chương trình REDD+. Các nghiên cứu trong nước ([2],[12], [13],[15],[16][17],[26]) đã thiết lập hệ thống mô hình cho RTX, có khả năng ước tính riêng biệt sinh khối phần trên mặt đất (AGB) và dưới mặt đất (BGB). Các công trình này sử dụng kết hợp DBH, H, WD cùng các biến dễ đo đạc tại hiện trường để tăng khả năng áp dụng cho cộng đồng và giám sát lâu dài. Gần đây hơn, các mô hình tích hợp dữ liệu viễn thám và biến sinh thái đã được chứng minh là cải thiện đáng kể độ chính xác và giảm sai số hệ thống trong ước lượng sinh khối rừng.

Việc ước tính BGB thường dựa vào hệ số chuyển đổi (R) từ sinh khối phần trên mặt đất, theo hướng dẫn của IPCC (2019):

$$BGB = R \times AGB \quad (1.2)$$

Trong đó, $R \approx 0,37$ đối với rừng mưa nhiệt đới và $0,20 - 0,24$ đối với rừng nửa rụng lá ẩm. Tuy nhiên, nghiên cứu khác cho thấy R có thể biến thiên mạnh theo kích thước và loài cây, đòi hỏi phải hiệu chỉnh theo vùng sinh thái[56]. Ở Việt Nam, mô hình BGB địa phương cho RTX đã được phát triển bởi Bảo Huy (2013) và mở rộng bởi Kralicek và cộng sự (2017) cho cả RTX và rừng khộp, tạo nền tảng cho việc đánh giá khả năng lưu trữ và hấp thụ CO₂ của rừng tự nhiên.

Đối với kiểu RTX ở Tây Nguyên, Bảo Huy và cộng sự (2012) đã thiết lập mô hình ước tính sinh khối của rừng thông qua các nhân tố DBH, H và WD. Kết quả cho thấy AGB được ước lượng tốt thông qua các biến số này với hệ số tương quan cao và cho kết quả ước lượng tốt hơn so với mô hình chung cho rừng nhiệt đới của Brown và Chave. Các mô hình được đề xuất sử dụng là:

$$AGB = \exp(-2,23927 + 2,49596 * \log(DBH)) \quad (1.3)$$

$$\log(AGB) = -2,74348 + 0,693879 * \log(H * DBH^2) + 0,367445 * \log(WD * DBH^2) \quad (1.4)$$

$$\log(AGB) = -2,9766 + 0,535797 * \log(DBH) + 0,759321 * \log(H * DBH^2) \quad (1.5)$$

$$\log(AGB) = -2,05364 + 1,76966 * \log(DBH) + 0,376371 * \log(WD * DBH^2) \quad (1.6)$$

Tổng thể, chuỗi các nghiên cứu này khẳng định tầm quan trọng của việc chuẩn hóa mô hình sinh khối địa phương, kết hợp phương pháp chặt hạ mẫu hạn chế, phân

tích hồi quy phi tuyến và dữ liệu viễn thám, nhằm hướng tới ước lượng chính xác cho quản lý các-bon rừng trong bối cảnh REDD+ và phát thải ròng bằng không (Net Zero).

1.3. Ứng dụng GIS và viễn thám trong ước tính sinh khối cây rừng

Trong hơn ba thập kỷ qua, công nghệ viễn thám và hệ thống thông tin địa lý (GIS) đã trở thành công cụ quan trọng trong nghiên cứu và quản lý tài nguyên rừng. Các nguồn dữ liệu vệ tinh như Landsat, Sentinel, SPOT, MODIS, ASTER, hay các ảnh độ phân giải cao như IKONOS, QuickBird, WorldView cung cấp chuỗi quan sát dài hạn, diện bao phủ rộng và khả năng cập nhật liên tục, giúp giám sát hiện trạng và biến động rừng hiệu quả hơn nhiều so với các phương pháp truyền thống dựa hoàn toàn vào điều tra thực địa [90],[114].

Trong lĩnh vực lâm nghiệp, viễn thám được ứng dụng rộng rãi để phân loại thảm phủ, lập bản đồ hiện trạng rừng, theo dõi suy thoái, và đặc biệt là ước tính sinh khối – các-bon rừng. Độ chính xác của phân loại phụ thuộc vào độ phân giải ảnh, phương pháp phân loại và số lớp thông tin; ảnh độ phân giải cao thường cho kết quả tốt hơn nhưng chi phí lớn, trong khi các hệ thống miễn phí như Landsat và Sentinel có độ phân giải trung bình nhưng ổn định theo thời gian, phù hợp cho nghiên cứu dài hạn và quản lý rừng ở quy mô tỉnh, quốc gia [148].

Ước tính sinh khối từ viễn thám chủ yếu dựa trên việc phân tích mối quan hệ giữa đặc trưng phổ và cấu trúc được trích xuất từ ảnh vệ tinh, bao gồm các chỉ số thực vật như NDVI, EVI; phản xạ phổ ở các dải đỏ, cận hồng ngoại và hồng ngoại trung; đặc trưng cấu trúc tán; hoặc tín hiệu tán xạ của ra-đa. Nhiều nghiên cứu đã chứng minh mối tương quan chặt chẽ giữa các chỉ số phổ và biến số sinh khối như mật độ tán lá, trữ lượng và AGB [177],[184]. Các mô hình thống kê truyền thống và các kỹ thuật học máy, bao gồm hồi quy đa biến, k-nearest neighbour (kNN), cây quyết định, Random Forest (RF), Support Vector Machine (SVM) và các thuật toán học tập tổ hợp (ensemble), đã được ứng dụng rộng rãi trong dự đoán sinh khối, với độ chính xác thay đổi tùy theo loại ảnh, điều kiện tán rừng và mức độ phức tạp cấu trúc thảm thực

vật [134].

Viễn thám đóng vai trò như lớp “lấy mẫu đầu tiên”, hỗ trợ phân chia rừng thành các khu vực tương đối đồng nhất trước khi bố trí các ô điều tra thực địa, từ đó giảm chi phí và tăng hiệu quả kiểm kê [143]. Sự kết hợp giữa dữ liệu thực địa và ảnh vệ tinh là hướng tiếp cận được khuyến nghị trong các hướng dẫn của IPCC và chương trình REDD⁺, đặc biệt trong giám sát mất rừng, suy thoái rừng và kiểm kê các-bon phục vụ báo cáo khí nhà kính quốc gia.

Bên cạnh dữ liệu quang học, ảnh radar khẩu độ tổng hợp (SAR) như Sentinel-1 cho phép quan sát rừng trong điều kiện nhiều mây, vốn đặc trưng ở vùng nhiệt đới, và đã được chứng minh có hiệu quả trong ước tính sinh khối, đặc biệt tại các hệ sinh thái rừng rậm nơi tín hiệu radar nhạy với cấu trúc tầng tán, mật độ gỗ và độ thô ráp bề mặt [87], [150]. Radar ở các tần số như C-band, L-band hoặc P-band có khả năng xuyên qua tán ở các mức độ khác nhau, hỗ trợ mô tả cấu trúc rừng vượt trội hơn so với dữ liệu quang học trong điều kiện bão hòa phổ. Do đó, nhiều nghiên cứu gần đây hướng đến tích hợp dữ liệu quang học và radar nhằm tận dụng thế mạnh bổ sung của hai nguồn tín hiệu, qua đó nâng cao độ chính xác phân loại thảm rừng và mô hình hóa AGB [85],[162].

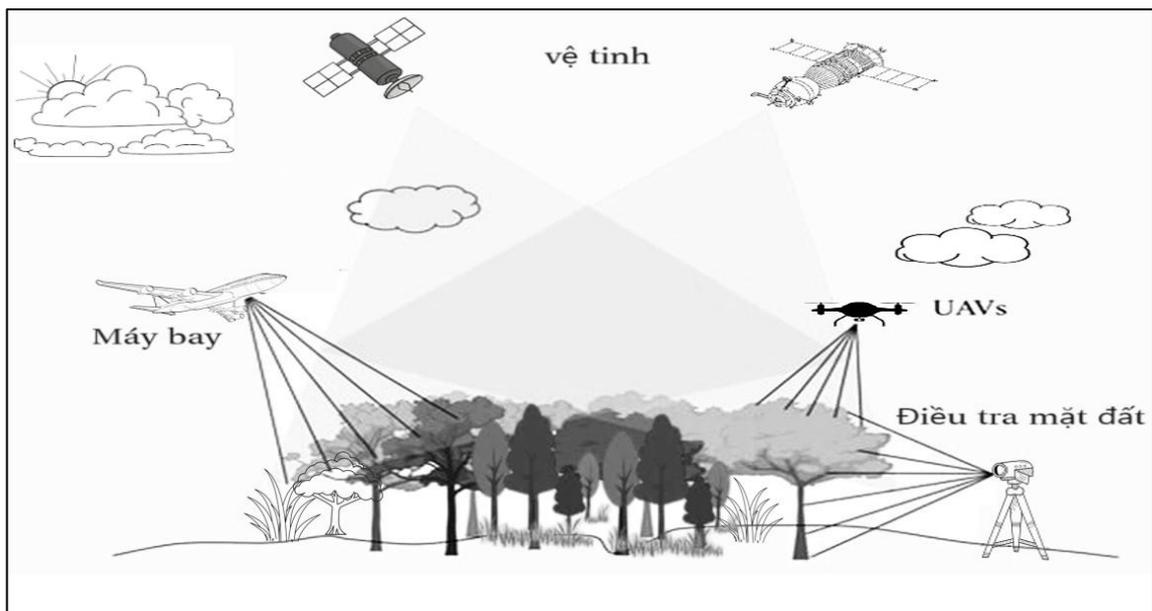
GIS giữ vai trò trung tâm trong việc tích hợp, xử lý và phân tích dữ liệu không gian, hỗ trợ xây dựng các bản đồ biến động rừng, bản đồ sinh khối AGB, bản đồ hấp thụ CO₂, cũng như đánh giá các yếu tố ảnh hưởng đến biến đổi rừng như địa hình, dân số, cơ sở hạ tầng hay khoảng cách đến đường giao thông. Các nghiên cứu đã cho thấy GIS có thể được sử dụng hiệu quả để dự báo rủi ro mất rừng, phân tích xu thế biến động sinh khối và xây dựng bản đồ các-bon ở nhiều quy mô khác nhau, từ quy mô lô rừng đến cấp vùng [81].

Sự kết hợp giữa viễn thám và GIS đang trở thành hướng tiếp cận tất yếu trong giám sát sinh khối và các-bon rừng ở quy mô lớn nhờ khả năng bao phủ rộng, độ tin cậy cao và chi phí hợp lý. Phương pháp tích hợp này đặc biệt quan trọng trong bối cảnh gia tăng nhu cầu kiểm kê các-bon, chi trả dịch vụ môi trường rừng và thực hiện

các cam kết giảm phát thải theo REDD⁺, chiến lược phát triển lâm nghiệp và mục tiêu quốc gia về phát thải ròng bằng không của Việt Nam. Nhờ ưu thế về tích hợp đa nguồn dữ liệu, GIS cung cấp nền tảng vững chắc để hỗ trợ ra quyết định, lập kế hoạch và theo dõi thực hiện các chính sách quản lý rừng và khí hậu [65].

1.3.1. Nguyên tắc Ước tính AGB sử dụng dữ liệu Viễn thám

Trái ngược với ước tính sinh khối rừng trực tiếp, các kỹ thuật viễn thám thường đánh giá AGB rừng thông qua việc xây dựng và sử dụng các thông số như sóng phản xạ bề mặt, các chỉ số thực vật (VIs), chỉ số diện tích tán lá (LAI), độ che phủ thực vật, chiều cao cây và đường kính tán cây nhằm thiết lập các mối quan hệ đóng vai trò là đại diện cho AGB [198]. Các kỹ thuật viễn thám được sử dụng để ước tính AGB của rừng được minh họa trong Hình 1.1.



Nguồn: Tian và cộng sự (2023)

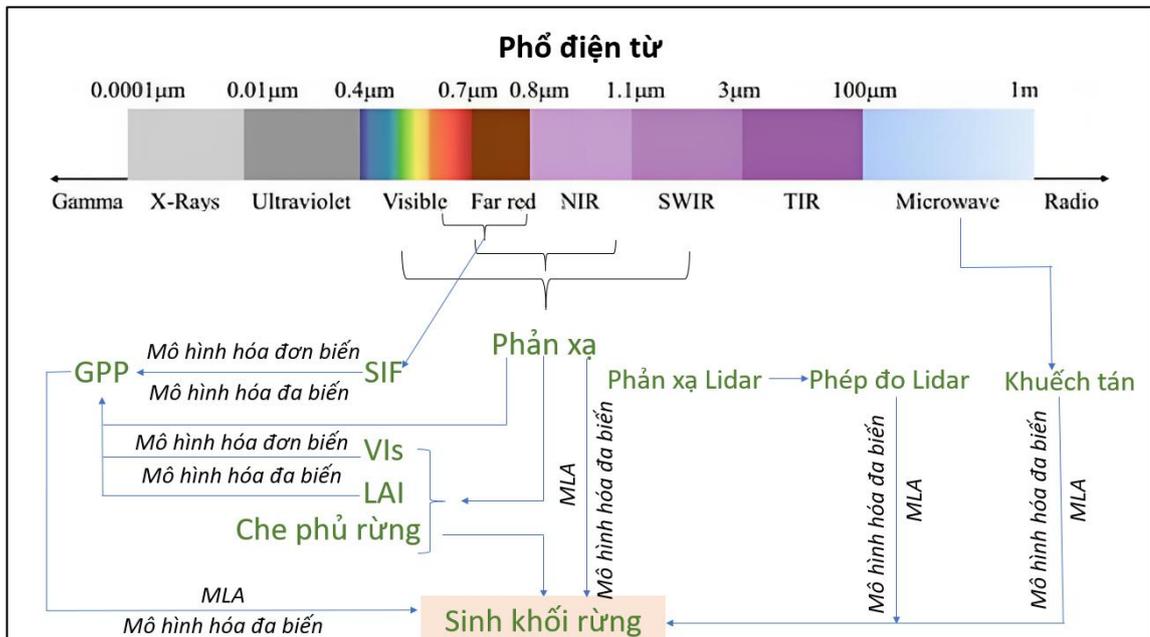
Hình 1.1. Minh họa ước tính AGB bằng kỹ thuật viễn thám

Ngoài thông tin các kênh ảnh đơn phổ của dữ liệu viễn thám quang học, AGB thường được ước tính thông qua VIs bao gồm chỉ số khác biệt thực vật chuẩn hóa (NDVI – Normal Differences Vegetation Index), chỉ số khác biệt thực vật (DVI - Differences Vegetation Index) và chỉ số thực vật tăng cường (EVI – Enhance Vegetation Index) [76],[186]. Tuy nhiên, khi thảm thực vật dày lên, sự hấp thụ mạnh

các bước sóng ở dải bước sóng màu đỏ dẫn đến hiệu ứng bão hòa, do đó làm giảm độ chính xác trong ước tính AGB [198]. Vì vậy, một số chỉ số thực vật khác như NDVI chuẩn hóa lại (RNDVI – Renormalized NDVI) và tỷ lệ đơn giản sửa đổi (MSR) đã được phát triển để cải thiện độ chính xác của ước tính sinh khối trong các khu vực có thảm thực vật dày [46],[63]. Đối với các khu vực có thảm thực vật thưa thớt, VI vuông góc dựa trên biến đổi trực giao VI (PVI – Perpendicular VI), VI hiệu chỉnh đất (SAVI – Soil-adjusted VI) và SAVI điều chỉnh (MSAVI – Modified SAVI) được sử dụng để giảm thiểu nhiễu từ khí quyển và nền đất [68],[129]. Hơn nữa, thông tin kết cấu ảnh viễn thám ngày càng được sử dụng nhiều hơn trong ước tính AGB của rừng [104],[166].

Các thông số bổ sung cần thiết cho ước tính AGB bao gồm các thông số mô tả cấu trúc rừng, chẳng hạn như chiều cao cây, đường kính ngang ngực. Chiều cao cây không chỉ phản ánh các đặc tính sinh học và khả năng sinh trưởng của cây, mà còn cho biết trữ lượng rừng [67]. Các nghiên cứu trước đây đã chứng minh AGB có mối liên hệ với chiều cao cây (thông qua một hệ số) ở các khu rừng rậm bằng cách sử dụng dữ liệu khảo sát về tuổi cây và chiều cao cây trung bình [64]. Tuy nhiên, rất khó để xác định chiều cao cây ở quy mô lớn, đặc biệt là ở các khu rừng kín có chiều cao cây lớn; do đó, thông thường chỉ xác định chiều cao cây của một số cây cá thể và sau đó ước tính chiều cao tổng thể của ô mẫu bằng cách thiết lập mối tương quan giữa chiều cao cây và DBH [67]. Hơn nữa, phương trình sinh học hàm mũ thể hiện tương quan giữa AGB và chiều cao cây được xây dựng ở quy mô ô mẫu vẫn được áp dụng trên quy mô lớn [161]. Đây là một lợi thế đáng kể của việc ước tính AGB bằng viễn thám kết hợp với các phép đo mặt đất [50],[89],[198]. Trong những năm gần đây, viễn thám siêu cao tần và viễn thám LiDAR (Light Detection and Ranging) đã được sử dụng rộng rãi để ước tính AGB. Chiều cao cây có thể được lấy chính xác và thuận tiện từ dữ liệu InSAR và LiDAR [34],[191]. Ngoài ra, chiều cao tán cây đã được chứng minh là cung cấp ước tính AGB chính xác [169]. Đáng chú ý, chiều cao tán không phải là chiều cao cây; Nó không chỉ phụ thuộc vào chiều cao của cây, mà còn phụ thuộc vào tán cây và mật độ cây gỗ [198].

Nhìn chung, AGB được ước tính bằng cách sử dụng dữ liệu viễn thám thu được trên một dải bước sóng điện từ rộng, từ ánh sáng khả kiến đến vi sóng (Hình 1.2).



Ghi chú: MLA: phương pháp học máy; NIR: cận hồng ngoại; SWIR: hồng ngoại sóng ngắn; TIR: hồng ngoại nhiệt; VIS; LAI; SIF: huỳnh quang diệp lục do năng lượng mặt trời gây ra; GPP: tổng tăng trưởng hàng năm của rừng. (Nguồn: Xiao và cộng sự, 2019)

Hình 1.2. Sử dụng phổ điện từ ước tính sinh khối rừng trên mặt đất

Ngoài các thông số quá trình sinh thái ở trên, các yếu tố môi trường (ví dụ: lượng mưa, nhiệt độ và áp suất khí quyển), địa hình và các yếu tố sinh học (ví dụ: đa dạng loài) cũng ảnh hưởng đến ước tính AGB của rừng. Cụ thể, các yếu tố như lượng mưa, nhiệt độ, độ cao và độ dốc thúc đẩy mô hình phân bố loài cây, trong khi tài nguyên đất và cường độ bức xạ xác định điều kiện tăng trưởng của thảm thực vật, tất cả đều ảnh hưởng đến AGB rừng [131].

1.3.2. Nguồn dữ liệu ảnh viễn thám sử dụng ước lượng AGB

AGB của rừng hiện được ước tính bằng nhiều dạng dữ liệu viễn thám khác nhau, bao gồm dữ liệu quang học thụ động (đa phổ, siêu phổ) và dữ liệu ra-đa chủ động; ngoài ra một số nghiên cứu cũng sử dụng dữ liệu LiDAR thu nhận từ mặt đất, trên không hoặc vệ tinh [30],[160],[167],[186],[198]. Mỗi nguồn dữ liệu cung cấp

thông tin đặc trưng về phổ phản xạ, tín hiệu tán xạ hoặc cấu trúc thẳng đứng của rừng, và có thể được kết hợp để nâng cao độ chính xác mô hình ước tính sinh khối.

1.3.2.1. Ảnh viễn thám quang học thụ động

Viễn thám thụ động là hệ thống viễn thám không tự mang nguồn bức xạ. Cụ thể, nó là một hệ thống viễn thám trong đó thiết bị thu nhận và ghi lại thông tin điện từ phát ra từ chính vật thể mục tiêu hoặc phản xạ từ nguồn bức xạ tự nhiên (mặt trời) trong quá trình viễn thám. Viễn thám quang học thụ động được sử dụng rộng rãi để ước tính AGB rừng vì nó rất nhạy cảm với các đặc tính của tán cây. Dữ liệu có độ phân giải thô (250–8000 m, ví dụ: MODIS, AVHRR) thường được sử dụng để ước tính AGB rừng ở quy mô khu vực hoặc toàn cầu [36],[49],[50],[197]. Hơn nữa, dữ liệu độ phân giải không gian trung bình (như ảnh Landsat, Sentinel-2 Multispectral Imager (MSI) và Terra/Aqua ASTER) được áp dụng cho các ước tính AGB rừng quy mô địa phương và khu vực cho các hệ sinh thái khác nhau [72],[125],[168]. Dữ liệu có độ phân giải không gian cao (<5 m, như IKONOS, QuickBird và WorldView-2) thường được sử dụng để tính toán AGB rừng quy mô lâm phần [60],[87]. Tuy nhiên, dữ liệu như vậy thường là các vệ tinh thương mại, điều này hạn chế ứng dụng rộng rãi hơn của chúng trong lĩnh vực ước tính AGB rừng.

Phản xạ quang phổ, VIs, kết cấu không gian và đặc tính tán rừng là các biến số chính có được thông qua viễn thám quang học thụ động để ước tính AGB. VIs đã được phát triển để góp phần đại diện cho điều kiện thực vật, chẳng hạn như nền đất, khí quyển và địa hình [37],[74],[194]. Ngoài các VIs thường được sử dụng (ví dụ: NDVI, EVI, SR và chỉ số chênh lệch điệp lục), biến đổi phổ (ST) và phân tích thành phần chính (PCA – Principal Component Analysis) cũng thường được sử dụng để ước tính AGB [186]). Kết cấu không gian ảnh mô tả các đặc điểm không gian của hình ảnh và có thể phản ánh lượng AGB của rừng ở một mức độ nào đó. Thông tin kết cấu này có thể được trích xuất bằng ma trận cấp độ xám, thường sử dụng kích thước cửa sổ 3×3 [155]. Ngoài ra, các nghiên cứu trước đây đã cải thiện độ chính xác của các ước tính AGB bằng cách bao gồm các chỉ số phản ánh các thuộc tính tán

rừng, chẳng hạn như LAI, mật độ tán rừng (FCD – Forest Canopy Density) và độ che phủ rừng [41],[113].

Viễn thám quang học là một trong những công cụ tốt nhất để ước tính AGB rừng vì độ phân giải không gian khác nhau từ phạm vi rộng như MODIS đến chi tiết như Wordview, đa thời gian, phạm vi phủ sóng toàn cầu và chi phí thấp. Tuy nhiên, khả năng xuyên qua đối tượng kém, chủ yếu ghi lại thông tin về cấu trúc nằm ngang, do đó khó ước lượng được cấu trúc đứng của rừng, đồng thời bị ảnh hưởng bởi những yếu tố khác như mây, mưa. Hơn nữa, tín hiệu viễn thám quang học có thể trở nên bão hòa trong các khu rừng rậm rạp, dẫn đến việc đánh giá thấp hoặc quá cao mật độ sinh khối. Do đó, để ước tính chính xác AGB rừng bằng cách sử dụng dữ liệu viễn thám quang học thụ động cũng là một thách thức [186].

1.3.2.2. Ảnh viễn thám siêu cao tần

Trái ngược với viễn thám quang học, công nghệ viễn thám ra-đa (viễn thám chủ động) có khả năng thu nhận dữ liệu cả ngày lẫn đêm, ít bị ảnh hưởng bởi mây che, điều kiện khí tượng và cường độ bức xạ Mặt Trời. Nhờ đặc tính sử dụng sóng vi ba có bước sóng dài, viễn thám ra-đa còn cho phép sóng điện từ xuyên qua tán rừng và tương tác với các thành phần cấu trúc bên trong như cành, thân và mặt đất. Do đó, công nghệ này mang lại nhiều ưu thế trong việc mô tả cấu trúc rừng và ước tính AGB, đặc biệt tại các khu vực nhiệt đới thường xuyên bị mây che phủ [117],[192].

Trong các ứng dụng lâm nghiệp, ảnh ra-đa khẩu độ tổng hợp được sử dụng rộng rãi để ước tính AGB thông qua phân tích hệ số tán xạ ngược. Mức độ và cơ chế tán xạ phụ thuộc chặt chẽ vào bước sóng (tần số) và trạng thái phân cực của tín hiệu SAR, do mỗi cấu phần của cây rừng tương tác khác nhau với sóng ra-đa. Hiện nay, các nghiên cứu ước tính AGB chủ yếu sử dụng dữ liệu SAR ở các băng tần X (9,6 GHz; 3,0 cm), C (5,6 GHz; 5,7 cm), S (3,0 GHz; 10 cm), L (1,27 GHz; 23,5 cm) và P (0,435 GHz; 70,0 cm), kết hợp với các trạng thái phân cực HH, HV, VH và VV.

Cụ thể, dải X chủ yếu tương tác với lá và lớp bề mặt của tán rừng, do đó phản

ánh thông tin sinh khối ở tầng trên. Dải C có khả năng xuyên qua lá và chịu ảnh hưởng mạnh từ các cành nhỏ và cấu trúc dưới tán. Trong khi đó, dải L có độ xuyên thấu cao hơn, cho phép tín hiệu tương tác với thân chính và các cành lớn, nên thường cho mối tương quan tốt hơn với AGB ở rừng có sinh khối trung bình đến cao. Băng tần P, với bước sóng dài nhất, có khả năng xuyên qua toàn bộ tán rừng và phần lớn tín hiệu tán xạ ngược xuất phát từ thân cây và tương tác giữa thân cây với mặt đất.

Bên cạnh đó, bốn tổ hợp phân cực SAR gồm HH, HV, VH và VV phản ánh các cơ chế tán xạ khác nhau của thảm rừng. Trong đó, phân cực chéo (HV và VH) thường nhạy cảm hơn với cấu trúc không đồng nhất của tán rừng, trong khi phân cực đồng hướng (HH và VV) phản ánh mạnh hơn các bề mặt và cấu trúc định hướng rõ rệt [78].

Các nghiên cứu trước đây đã phát hiện ra rằng dữ liệu đồng phân cực ngang và dọc (HH và VV) ở bước sóng dài hơn (ví dụ: băng tần P) rất nhạy cảm với điều kiện bề mặt thay đổi. Ngược lại, tín hiệu tán xạ ngược từ phân cực chéo (HV và VH) chủ yếu bao gồm nhiều tán xạ trong tán cây và ít bị ảnh hưởng bởi điều kiện bề mặt [182]. Đối với các khu vực sinh khối thấp, chẳng hạn như đồng cỏ, đầm lầy, rừng tái sinh, tín hiệu tán xạ ngược ở bước sóng dài hơn thấp hơn so với dải C; do đó, dải C được ưu tiên để ước tính sinh khối ở các khu vực sinh khối thực vật thấp hơn [78].

Dải C bị hạn chế bởi không có khả năng thâm nhập hiệu quả vào tán cây và mức độ bão hòa của nó (khoảng 60–70 Mg / ha). Tuy nhiên, những hạn chế này có thể được khắc phục bằng cách sử dụng các dải dài hơn có khả năng thâm nhập tán rừng cao hơn (ví dụ: băng L và băng P) [91]. Các nghiên cứu đã chỉ ra rằng băng L và dải P thường bão hòa ở mức 100 Mg/ha đối với các cấu trúc rừng nhiệt đới không đồng nhất phức tạp [164]; tuy nhiên, mức bão hòa này tăng lên khoảng 250 Mg/ha đối với các lâm phần có cấu trúc đơn giản và ít loài chiếm ưu thế [124].

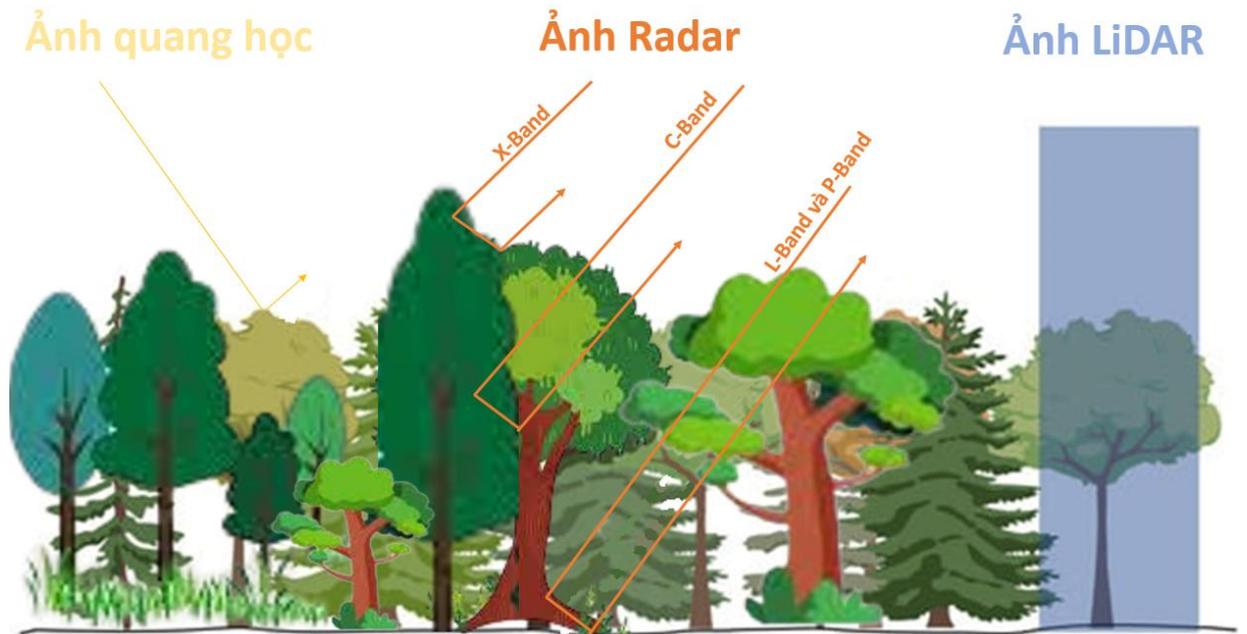
Mặc dù sóng ra-đa có khả năng trích xuất thông tin cấu trúc thẳng đứng của rừng và được sử dụng rộng rãi trong ước tính sinh khối rừng, nhưng vẫn tồn tại nhiều khó khăn trong ước tính sinh khối rừng dựa trên SAR. SAR phản ánh độ nhám của

bề mặt che phủ đất và do đó, không thể phân biệt giữa các loại thảm thực vật. Hơn nữa, tín hiệu SAR dễ bị nhiễu từ điều kiện tốc độ, độ ẩm và nhiệt độ, do đó làm phức tạp việc ước tính sinh khối [115]. Ngoài ra, độ bão hòa tín hiệu SAR cũng ảnh hưởng đến độ chính xác ước tính sinh khối rừng [184]. May thay, các nguồn dữ liệu đang được phát triển như PolInSAR và SAR chụp cắt lớp (TomoSAR), bao gồm BIOMASS, NISAR và TanDEM-L có thể khắc phục hiệu quả các vấn đề bão hòa trong ước tính AGB bằng cách đo trực tiếp các cấu trúc không gian rừng [186]. Do đó, việc sử dụng PolInSAR và TomoSAR cho thấy nhiều hứa hẹn trong tương lai gần.

Bên cạnh dữ liệu quang học và ra-đa, LiDAR cũng là nguồn dữ liệu quan trọng nhờ khả năng mô tả cấu trúc ba chiều của rừng. Tuy nhiên, chi phí cao và mức độ sẵn có còn hạn chế khiến LiDAR chưa được sử dụng rộng rãi trong lập bản đồ sinh khối ở quy mô lớn tại Việt Nam. Do đó, luận án này tập trung vào các nguồn dữ liệu miễn phí và dễ tiếp cận hơn như quang học và radar nhằm đảm bảo tính khả thi và ứng dụng thực tiễn.

1.2.3.3. Phối hợp dữ liệu viễn thám thụ động và chủ động

AGB rừng phụ thuộc vào bốn thông số: chiều cao cây, mật độ, DBH và chiều rộng tán cây. Tuy nhiên, việc đo DBH trực tiếp bằng cách sử dụng dữ liệu hàng không và vệ tinh là một thách thức. DBH là tham số có thể được sử dụng để ước tính ba tham số còn lại. Do đó, để tạo điều kiện cho các ước tính AGB rừng chính xác, việc kết hợp dữ liệu viễn thám thụ động và chủ động đã được sử dụng. Các thông số đo đếm của rừng liên quan đến các nguồn dữ liệu viễn thám khác nhau được thể hiện trong Hình 1.3.



Hình 1.3. Dữ liệu ảnh viễn thám khác nhau trong ước tính AGB (Nguồn: Tian và cộng sự, 2023)

Ước tính AGB rừng thông qua viễn thám quang học đa phân giải chủ yếu tập trung vào thông tin kết cấu ảnh được cung cấp bởi dữ liệu ảnh có độ phân giải cao được bổ sung cho ảnh có độ phân giải trung bình hoặc các chỉ số thực vật. Ước tính AGB bằng cách kết hợp thông tin kết cấu ảnh và chỉ số thực vật có độ chính xác cao hơn so với sử dụng từng biến đơn lẻ hoặc chỉ sử dụng các kênh quang phổ [39],[147].

Sự kết hợp giữa viễn thám chủ động và thụ động để ước tính AGB rừng chủ yếu dựa vào dữ liệu sinh khối thu được từ các điểm lấy mẫu hiện trường hoặc LiDAR làm điểm chuẩn và dữ liệu viễn thám thụ động hoặc SAR làm các biến độc lập. Sự kết hợp giữa SAR và dữ liệu quang học bổ sung cho nhau, do đó cải thiện độ chính xác ước tính và giảm vấn đề bão hòa, đặc biệt với chi phí thấp trên phạm vi rộng [110],[140].

Mặc dù viễn thám làm giảm đáng kể thời gian và chi phí ước tính AGB rừng, dữ liệu đo đếm hiện trường là không thể thiếu cho cả việc xây dựng mô hình và đánh giá kết quả ước tính AGB rừng. Ngoài ra, chi phí mua ảnh, phạm vi phủ sóng và sự

sẵn có hạn chế của các cảm biến khác nhau, việc lựa chọn một cảm biến phù hợp với tính sẵn có của dữ liệu cụ thể, ngân sách dự án và các yêu cầu kỹ năng kỹ thuật để phân tích dữ liệu vẫn là một thách thức thực tế.

1.3.3. Các phương pháp ước tính AGB

Với sự phát triển nhanh chóng của công nghệ cảm biến và xử lý dữ liệu, viễn thám ngày càng trở thành công cụ chủ lực trong ước tính sinh khối rừng, hỗ trợ hiệu quả cho các phương pháp truyền thống vốn đòi hỏi nhiều công sức và chi phí. Dựa trên sự khác biệt về cơ sở lý thuyết và cách biểu diễn mối quan hệ giữa tín hiệu viễn thám và sinh khối, các phương pháp ước tính AGB thường được chia thành bốn nhóm chính, gồm mô hình thực nghiệm, mô hình vật lý, mô hình cơ học và các mô hình tích hợp. Mỗi nhóm phương pháp có cơ sở lý thuyết, yêu cầu dữ liệu và mức độ phức tạp khác nhau, phù hợp cho các mục tiêu và quy mô nghiên cứu khác nhau.

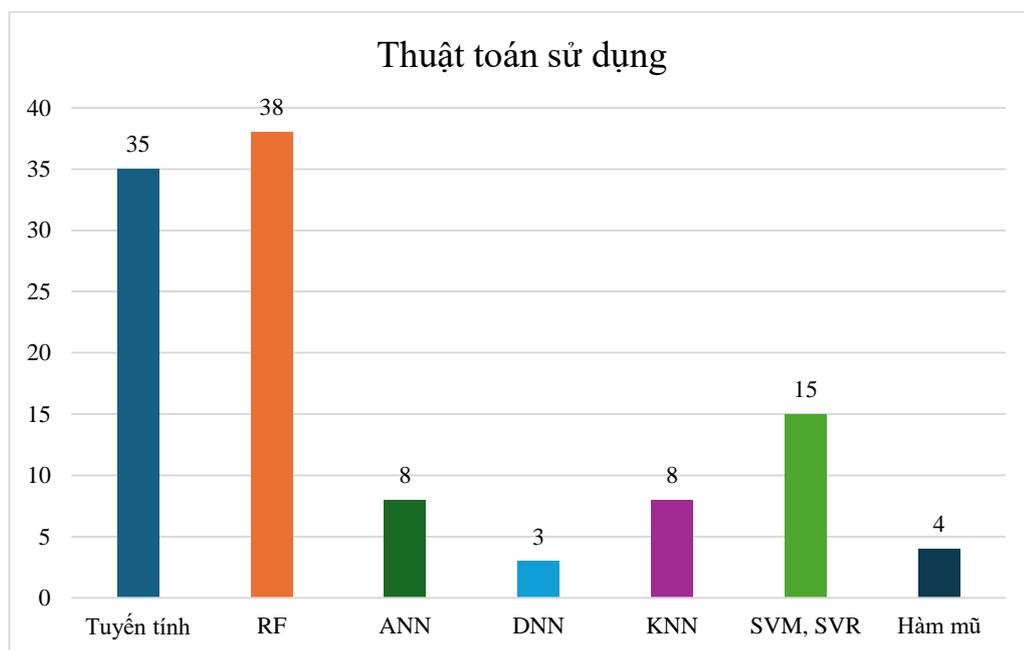
1.3.3.1. Mô hình thực nghiệm

Mô hình thực nghiệm là cách tiếp cận phổ biến nhất để ước tính AGB rừng. Cụ thể, một mô hình thống kê được xây dựng giữa các biến từ dữ liệu viễn thám và các dữ liệu ô mẫu, từ đó mô hình được sử dụng để ước tính AGB cho các vùng có diện tích lớn hơn [72],[121],[125]. Các mô hình thực nghiệm được phân loại làm mô hình tham số hoặc phi tham số. Các mô hình tham số chủ yếu đề cập đến hồi quy tuyến tính (LR), hồi quy đa biến (MR) và các phương pháp hồi quy phi tuyến tính. Theo đó, các phương trình và hàm tham số đóng vai trò là mô hình mô phỏng. Mặc dù các mô hình thực nghiệm rất đơn giản và dễ hiểu, tạo điều kiện cho việc hiểu và phân tích các phát hiện, độ chính xác ước tính của chúng thường không cao lắm [114],[117],[163].

Mô hình MR có thể tăng cường ước tính AGB bằng cách tích hợp độ phản xạ bề mặt, VIs và các yếu tố sinh lý [155]. Ngoài ra, giữa diện tích tán lá với trữ lượng các-bon ở rừng nhiệt đới cũng có mối quan hệ chặt chẽ [193]. Mô hình biến tối ưu phụ thuộc vào khu vực nghiên cứu [61],[186]. MR giả định rằng các biến dữ liệu viễn

thám ở các dải phổ khác nhau là không tương quan, điều này hiếm khi xảy ra trong viễn thám. Do đó, Lu và cộng sự (2012) đề xuất sử dụng hệ số tương quan và phân tích hồi quy từng bước để xác định các biến dữ liệu viễn thám có tương quan cao với sinh khối trong khi có tự tương quan yếu.

Trong giai đoạn trước đây, phần lớn các nghiên cứu trong nước về ước tính sinh khối rừng vẫn dựa trên hồi quy tuyến tính hoặc hàm hồi quy logarit để mô hình hóa mối quan hệ giữa chỉ số phổ và AGB. Tuy nhiên, cùng với sự phát triển mạnh mẽ của học máy, xu hướng gần đây cho thấy các thuật toán phi tuyến đang ngày càng được áp dụng rộng rãi hơn nhằm cải thiện độ chính xác mô hình. Tổng hợp thống kê (Hình 1.4) cho thấy các mô hình hồi quy truyền thống vẫn được sử dụng phổ biến trong khoảng 10 năm trở lại đây (35 nghiên cứu). Tuy nhiên các thuật toán phi tuyến vẫn được nghiên cứu rộng rãi đặc biệt là các thuật toán học máy được phát triển nhanh trong những năm gần đây.



Hình 1.4. Các thuật toán sử dụng trong các nghiên cứu gần đây

Trong bối cảnh nghiên cứu trong nước, các công trình giai đoạn 2014–2025 cho thấy mô hình thực nghiệm, đặc biệt là hồi quy tuyến tính và các thuật toán học máy, vẫn là nhóm phương pháp được sử dụng phổ biến nhất trong ước tính AGB

rừng. Các nghiên cứu tập trung chủ yếu tại các tỉnh có diện tích rừng lớn hoặc hệ sinh thái đặc thù như Đắk Lắk, Gia Lai, Quảng Bình, Bắc Kạn, Hòa Bình và Thừa Thiên Huế, với dữ liệu viễn thám chủ yếu là Landsat, Sentinel-1/2 và ALOS PALSAR. Trong những năm gần đây, các thuật toán học máy như RF, Cubist, SVM và ANN được áp dụng ngày càng rộng rãi và cho thấy khả năng cải thiện rõ rệt độ chính xác so với các mô hình hồi quy truyền thống [22],[27].

Trên phạm vi quốc tế, xu hướng chuyển dịch từ các mô hình hồi quy tuyến tính sang các thuật toán phi tuyến và học máy diễn ra mạnh mẽ, đặc biệt trong các hệ sinh thái rừng nhiệt đới và rừng hỗn loài phức tạp. Random Forest là thuật toán được sử dụng phổ biến nhất, tiếp theo là SVM, KNN, ANN và các mô hình học sâu, với nhiều nghiên cứu báo cáo hệ số xác định R^2 đạt từ 0,7 đến 0,95 khi kết hợp dữ liệu quang học, ra-đa và LiDAR [32],[98]. Tuy nhiên, việc mở rộng mô hình từ quy mô địa phương sang quy mô lớn vẫn chịu ảnh hưởng đáng kể của sai số dữ liệu thực địa và tính không đồng nhất không gian của rừng.

Các thuật toán phi tuyến truyền thống như hàm mũ hay GAM (mô hình cộng tính tổng quát) cũng được sử dụng nhưng không phổ biến. Mô hình GAM là một mở rộng linh hoạt của GLM (mô hình tuyến tính tổng quát), trong đó mỗi biến dự báo được mô tả bằng một hàm trơn phi tuyến thay vì quan hệ tuyến tính cố định. Theo Wood (2017), các hàm trơn như spline cho phép GAM mô hình hóa các mối quan hệ phức tạp mà không cần xác định trước dạng hàm, nhờ vậy đặc biệt phù hợp với các bài toán môi trường – lâm nghiệp có quan hệ sinh thái phi tuyến theo không gian và thời gian. Một ưu điểm quan trọng của GAM là khả năng điều chỉnh mức độ trơn nhằm cân bằng giữa tính linh hoạt và khả năng khái quát hóa giúp hạn chế hiện tượng quá khớp hoặc quá thấp. Do đó, trong các nghiên cứu ước tính sinh khối rừng – nơi mối quan hệ giữa chỉ số phổ, đặc trưng địa hình và AGB thường phi tuyến và không đồng nhất – việc lựa chọn GAM là hoàn toàn hợp lý.

Các mô hình tham số được xây dựng dựa trên các giả định phân phối dữ liệu lý tưởng, sao cho phân phối dữ liệu tuân theo phân bố chuẩn. Tuy nhiên, sự tương tác

giữa các biến thành phần viễn thám được sử dụng để ước tính AGB rừng rất phức tạp và việc phân phối dữ liệu rất khó đánh giá hoặc thiếu các đặc điểm phân biệt. Ngược lại, các mô hình phi tham số liên quan đến phân tích dữ liệu thống kê trực tiếp mà không dựa vào khái quát hóa toàn bộ phân phối mẫu.

Các mô hình phi tham số, thường được áp dụng trong học máy, là các mô hình được sử dụng nhiều nhất hiện nay trong các ước tính AGB rừng dựa trên viễn thám. Các mô hình được sử dụng nhiều như Người láng giềng gần nhất (kNN), Mạng thần kinh nhân tạo (ANN), Hỗ trợ máy véc-tơ (SVM), RF, Gradient boosting (GB), và Maximum entropy (ME) [71],[118],[148],[177]. Trong đó, phương pháp RF và ME đang ngày càng được áp dụng phổ biến trong lĩnh vực ước tính AGB rừng (Hình 1.4).

RF cải thiện độ chính xác dự đoán bằng cách xây dựng một "rừng" gồm nhiều cây quyết định phân loại thông qua một sơ đồ toàn diện, trong đó cả mẫu và biến được xử lý đồng thời thông qua phương pháp bootstrap và "thuật toán đóng gói" tương ứng [43],[118]. Ngoài ra, RF đạt được phân đoạn tối ưu tại mỗi nút bằng cách sử dụng cây phân loại và hồi quy. Mặc dù các cây riêng lẻ có thể yếu, sự kết hợp của tất cả các cây làm mạnh mẽ hơn các thuật toán khác và không bị giới hạn bởi sự xuất hiện của việc "học" quá mức [43].

Xu hướng nghiên cứu trong nước sử dụng các thuật toán hiện đại cũng bắt đầu hình thành. Các nghiên cứu của Nguyễn Thanh Tuấn và cộng sự (2022) cho thấy RF là thuật toán hiệu quả nhất khi ước tính AGB rừng thường xanh tại tỉnh Bình Phước. Gần đây, Đặng Thị Ngọc An và cộng sự (2019), Đỗ Thị Nhung và cộng sự (2024) cùng một số nhóm khác đã mở rộng áp dụng các thuật toán Cubist, ANN và RF, cho thấy tiềm năng lớn của học máy trong mô hình hóa mối quan hệ phi tuyến giữa đặc trưng phổ và sinh khối của các đối tượng rừng khác nhau.

Với các ưu và nhược điểm của các nghiên cứu trong và ngoài nước cho thấy việc thử nghiệm các thuật toán khác nhau từ tuyến tính, phi tuyến tính (GAM), và học máy (RF) cho nghiên cứu ước tính AGB rừng là phù hợp với xu thế nghiên cứu

hiện nay, đặc biệt là trên địa bàn tỉnh Đắk Lắk.

1.3.3.2. Mô hình hóa vật lý

Sinh khối có thể được ước tính bằng mô hình vật lý thông qua nội suy từ thông tin viễn thám bằng cách sử dụng mối quan hệ giữa các đặc điểm thực vật hai chiều và sinh khối. Các mô hình vật lý được sử dụng để ước tính AGB rừng chủ yếu bao gồm truyền bức xạ và mô hình quang hình học. Sử dụng LiDAR và dữ liệu viễn thám quang học để ước tính AGB thông qua các thông số cấu trúc của rừng như độ che phủ thực vật, LAI và chiều cao cây cho thấy có sự cải thiện đáng kể độ chính xác trong ước tính AGB của rừng [109]. AGB cũng được ước tính thông qua dữ liệu độ che phủ của tán cây và chiều cao trung bình của tán được xác định từ nguồn dữ liệu phản xạ đa góc làm mô hình quang hình học [48]. Mặc dù mô hình vật lý có ý nghĩa vật lý rõ ràng và tính ổn định cũng như khả năng ứng dụng tốt, việc tính toán rất phức tạp và hiện chỉ áp dụng cho các ước tính AGB quy mô nhỏ [121].

Các nghiên cứu quốc tế cho thấy mô hình vật lý và quang hình học được ứng dụng hiệu quả trong các nghiên cứu có dữ liệu cấu trúc rừng chi tiết, đặc biệt khi kết hợp ảnh đa góc, LiDAR hoặc dữ liệu không gian ba chiều [48],[109]. Tuy nhiên, tại Việt Nam, việc ứng dụng nhóm mô hình này còn rất hạn chế do thiếu dữ liệu LiDAR diện rộng, cũng như sự phức tạp trong thu thập và chuẩn hóa các tham số cấu trúc rừng. Phần lớn các nghiên cứu trong nước vẫn ưu tiên các mô hình thực nghiệm nhờ tính khả thi cao hơn về dữ liệu và nguồn lực.

1.3.3.3. Mô hình hóa theo thuyết cơ học

Mô hình cơ học (hay mô hình quá trình) mô phỏng sự tăng trưởng sinh khối hàng năm (NPP) của rừng dựa trên các nguyên lý sinh lý và sinh thái thực vật (Yan và Shugart, 2005). Nhóm mô hình này mô tả các quá trình như quang hợp, thoát hơi nước, chuyển đổi năng lượng mặt trời thành năng lượng sinh học và trao đổi nước – dinh dưỡng trong cây. Các dạng mô hình phổ biến gồm mô hình dựa trên khí hậu, mô hình sinh học, sinh thái và mô hình hiệu quả sử dụng ánh sáng [120],[174]. Một số

mô hình sử dụng thông tin từ viễn thám như độ che phủ thực vật, độ ẩm đất hoặc bức xạ quang hợp hấp thụ [154].

So với mô hình thực nghiệm, mô hình cơ học có ưu điểm là phản ánh rõ ràng các quá trình sinh thái và thường cho kết quả ổn định hơn. Tuy nhiên, chúng đòi hỏi lượng lớn tham số đầu vào khó thu thập (đất, sinh lý cây, khí tượng, bức xạ...), khiến khả năng ứng dụng còn hạn chế trong nhiều điều kiện nghiên cứu [187].

Trên thế giới, các mô hình cơ học được phát triển và ứng dụng chủ yếu trong các nghiên cứu quy mô lớn, gắn với các chương trình giám sát sinh quyển và mô hình hóa chu trình các-bon toàn cầu. Tuy nhiên, ngay cả trong các nghiên cứu quốc tế, nhóm mô hình này vẫn gặp khó khăn khi thiếu dữ liệu sinh lý, sinh thái chi tiết và đồng bộ. Trong bối cảnh nghiên cứu tại Việt Nam, các hạn chế về dữ liệu khí tượng, sinh lý cây và đặc điểm đất đai khiến việc áp dụng mô hình cơ học cho ước tính AGB chưa thực sự khả thi.

1.3.3.4. Mô hình hóa toàn diện

Mô hình toàn diện kết hợp các nguyên lý sinh thái với dữ liệu viễn thám để mô phỏng động thái rừng và thay đổi sinh khối theo thời gian. Các mô hình như FAREAST, LANDIS/LANDIS-II, FVS hay SORTIE-ND được sử dụng để mô phỏng kế thừa sinh thái, tái sinh, cạnh tranh và tác động xáo trộn đến sinh khối [189]. Nhóm mô hình này có cấu trúc linh hoạt và cho phép mô phỏng biến động AGB dưới ảnh hưởng của khai thác và biến đổi khí hậu [134],[185].

Tuy nhiên, độ chính xác mô phỏng phụ thuộc mạnh vào chất lượng tham số sinh thái của từng loài và độ chi tiết của dữ liệu đầu vào. Khi độ phân giải ảnh viễn thám thấp hoặc thiếu thông tin sinh học – sinh thái, mô hình khó phản ánh đúng trạng thái rừng [120]. Vì vậy, mặc dù có tiềm năng lớn, các mô hình toàn diện chưa phổ biến trong bối cảnh hạn chế dữ liệu chi tiết ở nhiều khu vực, trong đó có Việt Nam.

Các mô hình toàn diện được ứng dụng rộng rãi trong các nghiên cứu quốc tế nhằm mô phỏng động thái rừng và biến động sinh khối dưới tác động của khai thác

và biến đổi khí hậu [134],[185]. Tuy nhiên, việc triển khai các mô hình này đòi hỏi bộ dữ liệu sinh thái – sinh học chi tiết theo loài và theo không gian, điều kiện vẫn còn thiếu ở nhiều quốc gia đang phát triển. Thực tế này cũng phản ánh rõ trong các nghiên cứu tại Việt Nam, nơi mà dữ liệu đầu vào chưa đủ chi tiết để vận hành ổn định các mô hình toàn diện ở quy mô lớn.

Tóm lại, mặc dù các nhóm mô hình vật lý, cơ học và mô hình toàn diện đều có tiềm năng mô phỏng sinh khối ở mức độ chi tiết và mang tính sinh thái cao, việc vận hành các mô hình này đòi hỏi tập dữ liệu lớn, tham số sinh học – sinh thái chuyên sâu và cấu trúc dữ liệu đầu vào phức tạp, vốn chưa đầy đủ trong bối cảnh nghiên cứu tại Việt Nam. Do đó, luận án tập trung vào nhóm mô hình thực nghiệm, đặc biệt là hồi quy và thuật toán RF, như một cách tiếp cận khả thi và phù hợp nhất với điều kiện dữ liệu viễn thám và thực địa hiện có, đồng thời đáp ứng mục tiêu xây dựng mô hình ước tính AGB và phân tích biến động sinh khối theo chuỗi thời gian.

1.4. Thách thức và triển vọng khi sử dụng ảnh viễn thám để ước tính AGB

Trong những thập kỷ gần đây, ước tính sinh AGB ở các quy mô không gian khác nhau đã có nhiều tiến bộ nhờ sự phát triển nhanh chóng của dữ liệu và kỹ thuật viễn thám. Các nghiên cứu cho thấy AGB là một biến số then chốt để hiểu rõ chu trình các-bon toàn cầu và đánh giá vai trò của rừng trong bối cảnh biến đổi khí hậu [89][186]. Tuy nhiên, mặc dù tiềm năng ứng dụng rất lớn, việc ước tính AGB bằng ảnh viễn thám vẫn đối mặt với nhiều thách thức liên quan đến dữ liệu, phương pháp và mô hình hóa [138].

Một nguồn sai số quan trọng bắt nguồn từ dữ liệu viễn thám. Tín hiệu phản xạ phổ, VIs, SIF, hay LAI chịu ảnh hưởng mạnh của điều kiện khí quyển, độ ẩm bề mặt, góc chiếu sáng, suy giảm cảm biến và các sai số trong quá trình tiền xử lý ảnh [180]. Ngoài ra, sự không đồng bộ về thời gian giữa ảnh viễn thám và điều tra thực địa cũng làm gia tăng độ không chắc chắn trong xây dựng mô hình AGB. Đối với dữ liệu ra-đa, hiện tượng bão hòa tín hiệu ở các lâm phần có sinh khối cao, ảnh hưởng của địa hình và sự không tương xứng giữa kích thước pixel với diện tích ô mẫu thực

địa là những hạn chế phổ biến, đặc biệt ở rừng nhiệt đới có cấu trúc phức tạp. Bên cạnh đó dữ liệu ảnh viễn thám độ phân giải cao thường tốn kém chi phí và phạm vi phủ hạn chế làm giảm khả năng tiếp cận, đặc biệt là các quốc gia đang phát triển.

Sai số từ dữ liệu thực địa cũng là một thách thức lớn. Các lỗi đo đạc DBH, chiều cao cây, sai lệch trong lựa chọn phương trình sinh khối, cũng như sai số định vị ô mẫu so với pixel ảnh đều có thể gây sai lệch đáng kể cho kết quả ước tính [105]. Nhiều nghiên cứu cho thấy sai số lấy mẫu có thể ảnh hưởng lớn hơn cả sai số mô hình hóa [73]. Bên cạnh đó, việc thu thập dữ liệu thực địa thường đòi hỏi nguồn lực lớn về thời gian, nhân lực và kinh phí, đặc biệt ở các khu vực rừng nhiệt đới rậm rạp, địa hình phức tạp hoặc khó tiếp cận. Số lượng ô mẫu hạn chế làm giảm khả năng đại diện cho cấu trúc và mức độ dị hợp không gian của rừng, từ đó ảnh hưởng trực tiếp đến độ tin cậy của các mô hình viễn thám [18].

Từ góc độ mô hình hóa, độ chính xác của ước tính AGB chịu tác động đồng thời của sai số dữ liệu đầu vào, giả định mô hình và quá trình ước lượng tham số. Khả năng mô tả khoảng biến thiên AGB giữa các mô hình khác nhau là rất khác nhau, khiến việc lựa chọn mô hình và biến đầu vào trở thành yếu tố then chốt [186]. Ngoài ra, cấu trúc lâm phần nhiều tầng, sự hiện diện của thảm cây bụi và thảm tươi dưới tán cũng gây nhiễu tín hiệu viễn thám, đặc biệt đối với dữ liệu quang học và ra-đa băng tần ngắn.

Song song với các thách thức, triển vọng ứng dụng ảnh viễn thám trong ước tính AGB là rất lớn. Viễn thám là phương tiện gần như duy nhất cho phép thu thập dữ liệu liên tục về không gian và thời gian ở quy mô lớn, từ khu vực đến toàn cầu [57]. Các nguồn dữ liệu quang học miễn phí như Landsat và Sentinel, đặc biệt là bộ dữ liệu kết hợp Harmonized Landsat–Sentinel (HLS), cho phép xây dựng chuỗi thời gian dày với độ phân giải không gian trung bình, tạo điều kiện thuận lợi cho việc theo dõi biến động AGB một cách nhất quán. Bên cạnh đó, dữ liệu độ phân giải cao từ UAV, ảnh vệ tinh thương mại và LiDAR hàng không có thể cung cấp các mẫu AGB chi tiết, phục vụ hiệu chỉnh và huấn luyện mô hình.

Trong lĩnh vực radar, dữ liệu SAR băng tần L và P, cùng với các kỹ thuật tiên tiến như InSAR, PolInSAR và TomoSAR, mở ra khả năng khai thác thông tin cấu trúc tán rừng ba chiều, góp phần giảm ảnh hưởng của hiện tượng bão hòa và cải thiện độ chính xác ước tính sinh khối. Đồng thời, sự phát triển nhanh của công nghệ LiDAR không gian (như GEDI và ICESat-2) đã cung cấp các thông tin trực tiếp về chiều cao và cấu trúc rừng với độ tin cậy cao, đóng vai trò quan trọng trong xây dựng các mô hình AGB ở quy mô khu vực và toàn cầu.

Xu hướng chung hiện nay là tích hợp đa nguồn dữ liệu viễn thám, kết hợp quang học, SAR và LiDAR cùng với dữ liệu thực địa. Cách tiếp cận này cho phép khai thác ưu điểm của từng loại dữ liệu, giảm thiểu hạn chế riêng lẻ và nâng cao độ ổn định của bản đồ AGB. Đồng thời, các phương pháp phi tham số và học máy ngày càng cho thấy ưu thế trong xử lý dữ liệu lớn và mô hình hóa mối quan hệ phi tuyến phức tạp giữa AGB và các biến viễn thám. Tuy nhiên, việc lựa chọn phương pháp vẫn cần cân nhắc mục tiêu nghiên cứu, phạm vi không gian, đặc điểm hệ sinh thái rừng và tính sẵn có của dữ liệu để đảm bảo tính khả thi và độ tin cậy của kết quả.

1.5. Thảo luận

Tổng hợp và phân tích hơn 120 công trình khoa học trong và ngoài nước giai đoạn 2014–2025 cho thấy, ước tính sinh khối rừng và đánh giá khả năng hấp thụ CO₂ bằng công nghệ viễn thám và GIS là một hướng nghiên cứu phát triển nhanh, song tồn tại sự khác biệt rõ rệt giữa nghiên cứu quốc tế và nghiên cứu trong nước về quy mô không gian, mức độ tích hợp dữ liệu, phương pháp mô hình hóa và khả năng ứng dụng thực tiễn.

Xét về phạm vi không gian và đối tượng nghiên cứu, các nghiên cứu quốc tế đã tiếp cận ở quy mô lưu vực, quốc gia đến toàn cầu, bao phủ nhiều kiểu hệ sinh thái rừng khác nhau và cho phép theo dõi sinh khối động theo chuỗi thời gian dài. Trong khi đó, phần lớn các nghiên cứu trong nước mới dừng ở quy mô địa phương (tỉnh, khu bảo tồn hoặc lâm phần), tập trung vào một số kiểu rừng phổ biến như rừng thường xanh, rừng ngập mặn và rừng trồng. Điều này phản ánh thực tế rằng Việt Nam

vẫn đang trong giai đoạn xây dựng dữ liệu nền và kiểm chứng mô hình, chưa hình thành được hệ thống giám sát sinh khối liên tục và đồng bộ ở quy mô vùng hoặc quốc gia.

Về nguồn dữ liệu viễn thám, xu hướng chung trên thế giới là tích hợp đa nguồn, kết hợp ảnh quang học, radar và LiDAR (GEDI, ICESat-2), thậm chí UAV/TLS, nhằm khắc phục hiện tượng bão hòa tín hiệu, mây che phủ và nâng cao độ chính xác ước tính AGB. Ngược lại, các nghiên cứu trong nước chủ yếu vẫn dựa vào ảnh quang học miễn phí như Landsat và Sentinel-2; dữ liệu ra-đa Sentinel-1 mới được khai thác trong một số nghiên cứu gần đây, còn dữ liệu LiDAR chưa được sử dụng rộng rãi do hạn chế về chi phí và khả năng tiếp cận. Việc tích hợp dữ liệu quang học – ra-đa đa thời gian vẫn còn ở mức thử nghiệm, chưa hình thành quy trình ổn định cho giám sát dài hạn.

Về phương pháp mô hình hóa, các nghiên cứu quốc tế đã chuyển mạnh sang các thuật toán học máy và trí tuệ nhân tạo (RF, SVM, ANN, DNN, XGBoost), cho phép mô tả tốt hơn mối quan hệ phi tuyến giữa tín hiệu viễn thám và sinh khối rừng, đồng thời nâng cao độ chính xác dự báo. Trong khi đó, tại Việt Nam, các mô hình hồi quy tuyến tính và lô-ga-rít vẫn được sử dụng phổ biến; dù gần đây đã xuất hiện các nghiên cứu áp dụng RF, Cubist, SVM hoặc ANN, nhưng quy mô mẫu còn hạn chế và chưa đánh giá đầy đủ hiệu quả mô hình trên chuỗi thời gian dài.

Xét về kết quả và khả năng ứng dụng, nhiều nghiên cứu quốc tế đã xây dựng được bản đồ sinh khối và các-bon có độ chính xác cao ($R^2 > 0,7-0,9$), phục vụ hiệu quả cho REDD+, PFES và hệ thống MRV. Ngược lại, các nghiên cứu trong nước phần lớn mới tập trung vào ước tính AGB tại một thời điểm, chưa đánh giá được xu thế biến động dài hạn của sinh khối và khả năng hấp thụ CO₂, cũng như chưa tích hợp đầy đủ kết quả vào hệ thống quản lý rừng quốc gia.

Từ tổng hợp kinh nghiệm quốc tế có thể rút ra ba xu hướng chủ đạo: (i) tích hợp dữ liệu viễn thám đa nguồn và đa thời gian, (ii) ứng dụng các mô hình học máy để xử lý mối quan hệ phi tuyến phức tạp, và (iii) giám sát biến động sinh khối và các-

bon theo chuỗi thời gian phục vụ hoạch định chính sách giảm phát thải. So sánh với thực tiễn nghiên cứu trong nước cho thấy, Việt Nam vẫn thiếu các nghiên cứu tích hợp đồng thời cả ba xu hướng này, đặc biệt đối với rừng tự nhiên có cấu trúc phức tạp như rừng thường xanh.

Trong bối cảnh đó, việc nghiên cứu biến động sinh khối và khả năng hấp thụ CO₂ của rừng lá rộng thường xanh tại tỉnh Đắk Lắk thông qua tích hợp dữ liệu viễn thám đa nguồn, mô hình học máy và phân tích thống kê chuỗi thời gian là hướng tiếp cận phù hợp, mang tính kế thừa và phát triển các thành tựu nghiên cứu trong và ngoài nước.

1.6. Khoảng trống nghiên cứu và lý do lựa chọn đề tài

1.6.1. Khoảng trống nghiên cứu

Tổng quan các công trình giai đoạn 2014–2025 cho thấy, mặc dù nghiên cứu ước tính sinh khối rừng và khả năng hấp thụ CO₂ bằng viễn thám – GIS đã đạt được nhiều tiến bộ, song vẫn tồn tại những khoảng trống khoa học và kỹ thuật đáng kể trong bối cảnh Việt Nam.

Thứ nhất, khoảng trống về không gian và quy mô thời gian: Các nghiên cứu trong nước về rừng thường xanh chủ yếu mang tính cắt ngang, tập trung vào hiện trạng tại một thời điểm hoặc giai đoạn ngắn, trong khi thiếu các nghiên cứu xây dựng chuỗi giám sát dài hạn và liên tục. Điều này hạn chế khả năng đánh giá xu thế biến động sinh khối và vai trò của rừng thường xanh trong cân bằng các-bon ở quy mô tỉnh và vùng.

Thứ hai, khoảng trống về nguồn dữ liệu: Việc phụ thuộc chủ yếu vào ảnh quang học trong khi chưa khai thác đầy đủ dữ liệu ra-đa đa thời gian dẫn đến hạn chế trong điều kiện mây che phủ và thiếu tính liên tục dữ liệu. Sự thiếu đồng bộ giữa dữ liệu quá khứ và hiện tại làm giảm khả năng phân tích xu thế dài hạn của sinh khối rừng.

Thứ ba, khoảng trống về phương pháp mô hình hóa: Các mô hình hồi quy

truyền thống vẫn chiếm ưu thế, trong khi các mô hình học máy mới được áp dụng thử nghiệm, chưa có nghiên cứu so sánh hệ thống giữa mô hình tuyến tính, phi tuyến và học máy trên cùng bộ dữ liệu để lựa chọn mô hình tối ưu cho rừng thường xanh.

Thứ tư, khoảng trống về phân tích xu thế sinh khối: Các phương pháp thống kê mạnh như kiểm định Mann–Kendall và Sen’s slope, vốn được sử dụng rộng rãi trong nghiên cứu quốc tế để phát hiện xu thế biến động AGB, hầu như chưa được áp dụng một cách đầy đủ trong nghiên cứu rừng thường xanh tại Việt Nam.

1.6.2. Lý do lựa chọn đề tài

Xuất phát từ các khoảng trống nêu trên, đề tài “Ước lượng khả năng hấp thụ CO₂ của kiểu rừng lá rộng thường xanh trên địa bàn tỉnh Đắk Lắk dựa vào dữ liệu viễn thám và kỹ thuật GIS” được lựa chọn dựa trên các cơ sở sau:

Về tính cấp thiết khoa học, Đắk Lắk là địa phương có diện tích rừng tự nhiên lớn của Tây Nguyên, trong đó rừng thường xanh giữ vai trò quan trọng trong hấp thụ các-bon và điều hòa khí hậu. Tuy nhiên, rừng của tỉnh đang chịu áp lực suy thoái và chuyển đổi đất mạnh mẽ, làm biến động đáng kể sinh khối và khả năng hấp thụ CO₂. Trong bối cảnh Việt Nam cam kết đạt phát thải ròng bằng không vào năm 2050, việc định lượng và theo dõi biến động sinh khối rừng thường xanh ở quy mô tỉnh có ý nghĩa khoa học và thực tiễn rõ rệt.

Về tính mới trong cách tiếp cận, đề tài xây dựng chuỗi giám sát dài hạn (2015–2025) dựa trên tích hợp dữ liệu Sentinel-1 và Landsat 8, cho phép phân tích biến động sinh khối theo không gian và thời gian. Việc áp dụng mô hình học máy RF và so sánh với các mô hình hồi quy tuyến tính – phi tuyến, kết hợp các kiểm định thống kê Mann–Kendall và Sen’s slope, giúp nhận diện xu thế biến động AGB một cách định lượng và khách quan, là hướng tiếp cận còn ít được triển khai trong các nghiên cứu trước tại Việt Nam.

Về tính thực tiễn và ứng dụng, kết quả nghiên cứu cung cấp bản đồ sinh khối và khả năng hấp thụ CO₂ của rừng thường xanh, tạo lập cơ sở dữ liệu GIS phục vụ

quản lý rừng cấp tỉnh, hỗ trợ các chương trình PFES, REDD+ và tiến trình hoàn thiện hệ thống MRV trong lĩnh vực lâm nghiệp.

Về tính kế thừa và phát triển, đề tài kế thừa các kết quả nghiên cứu đã có tại khu vực Tây Nguyên, đồng thời phát triển theo hướng chuẩn hóa dữ liệu, tích hợp đa nguồn viễn thám, áp dụng mô hình học máy và phân tích xu thế dài hạn. Việc lựa chọn rừng thường xanh tại Đắk Lắk – đại diện cho hệ sinh thái rừng nhiệt đới vùng Tây Nguyên – tạo tiền đề cho khả năng mở rộng và nhân rộng mô hình nghiên cứu trong tương lai.

Tóm lại, luận án hướng tới giải quyết các khoảng trống hiện hữu trong nghiên cứu sinh khối rừng ở Việt Nam, đồng thời đóng góp cơ sở khoa học cho quản lý rừng bền vững và thực hiện các cam kết giảm phát thải khí nhà kính trong bối cảnh biến đổi khí hậu.

CHƯƠNG 2. ĐỐI TƯỢNG, NỘI DUNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

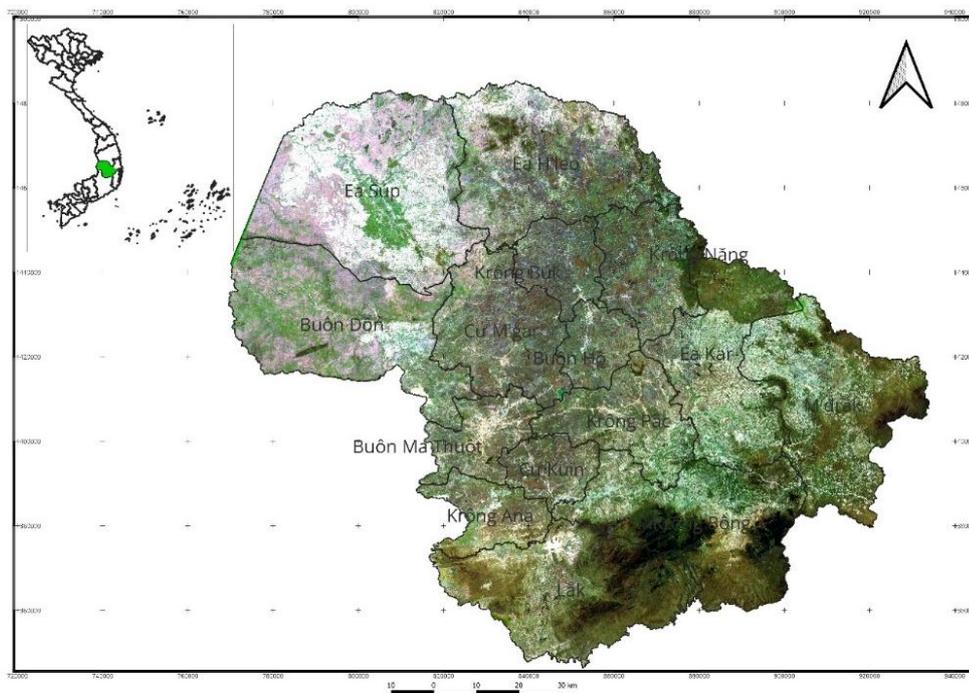
2.1. Đối tượng và địa điểm nghiên cứu

2.1.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài là kiểu rừng thường xanh (RTX) và AGB của kiểu rừng này. Các phân tích tập trung vào mối quan hệ giữa đặc trưng ảnh viễn thám và AGB, nhằm ước tính lượng CO₂ hấp thụ thông qua mô hình tích hợp dữ liệu viễn thám, GIS và số liệu thực địa.

2.1.2. Địa điểm nghiên cứu

Nghiên cứu được thực hiện tại tỉnh Đắk Lắk (Hình 2.1), thuộc khu vực Tây Nguyên, nơi phân bố diện tích lớn rừng thường xanh với tính đa dạng sinh học cao. Địa hình chủ yếu là đồi núi, độ dốc lớn, phân hóa mạnh theo vùng sinh thái; khí hậu nhiệt đới gió mùa, mùa khô kéo dài. Đây cũng là khu vực chịu tác động rõ rệt của suy thoái rừng, chuyển đổi mục đích sử dụng đất và hạn hán kéo dài, tạo điều kiện phù hợp để nghiên cứu biến động sinh khối – các-bon bằng công nghệ viễn thám.



Hình 2.1. Vị trí khu vực nghiên cứu

Tỉnh Đắk Lắk nằm ở vị trí trung tâm của địa bàn Tây Nguyên, trong khoảng toạ độ địa lý từ 107⁰28'57" - 108⁰59'37" độ kinh Đông và từ 12⁰09'45" – 13⁰25'06" độ vĩ Bắc. Phía Bắc giáp tỉnh Gia Lai, phía Nam giáp tỉnh Lâm Đồng, phía Đông giáp tỉnh Phú Yên và tỉnh Khánh Hoà và phía Tây giáp Vương quốc Campuchia và tỉnh Đắk Nông.

2.2. Nội dung nghiên cứu

Nghiên cứu được triển khai xoay quanh bốn nhóm nội dung chính, tương ứng với các kết quả mà luận án cần đạt được:

(1) Đánh giá thay đổi diện tích rừng thường xanh (RTX) tỉnh Đắk Lắk giai đoạn 2015–2025:

- Thu thập và tiền xử lý dữ liệu viễn thám đa thời gian (Landsat, Sentinel).
- Xây dựng bộ mẫu phân loại, lựa chọn chỉ số và biên giải đoán.
- Thực hiện phân loại RTX cho các mốc thời gian.
- Đánh giá độ chính xác phân loại và lập bản đồ phân bố RTX.
- Phân tích biến động diện tích RTX theo không gian và thời gian giai đoạn 2015–2025.

(2) Phân tích mối quan hệ giữa đặc trưng ảnh viễn thám và AGB:

- Chuẩn hóa và xử lý dữ liệu thực địa AGB.
- Trích xuất các biến viễn thám gồm chỉ số thực vật, băng phổ, cấu trúc tán che, địa hình, ra-đa.
- Phân tích tương quan, kiểm định ý nghĩa và lựa chọn tập biến đầu vào phù hợp cho mô hình AGB.
- Đánh giá mức độ ảnh hưởng của từng biến đến sự thay đổi AGB.

(3) Xây dựng mô hình ước tính AGB từ dữ liệu viễn thám:

- Xây dựng mô hình AGB dựa trên các thuật toán Hồi quy tuyến tính, Hồi

quy phi tuyến tính và Random Forest (RF).

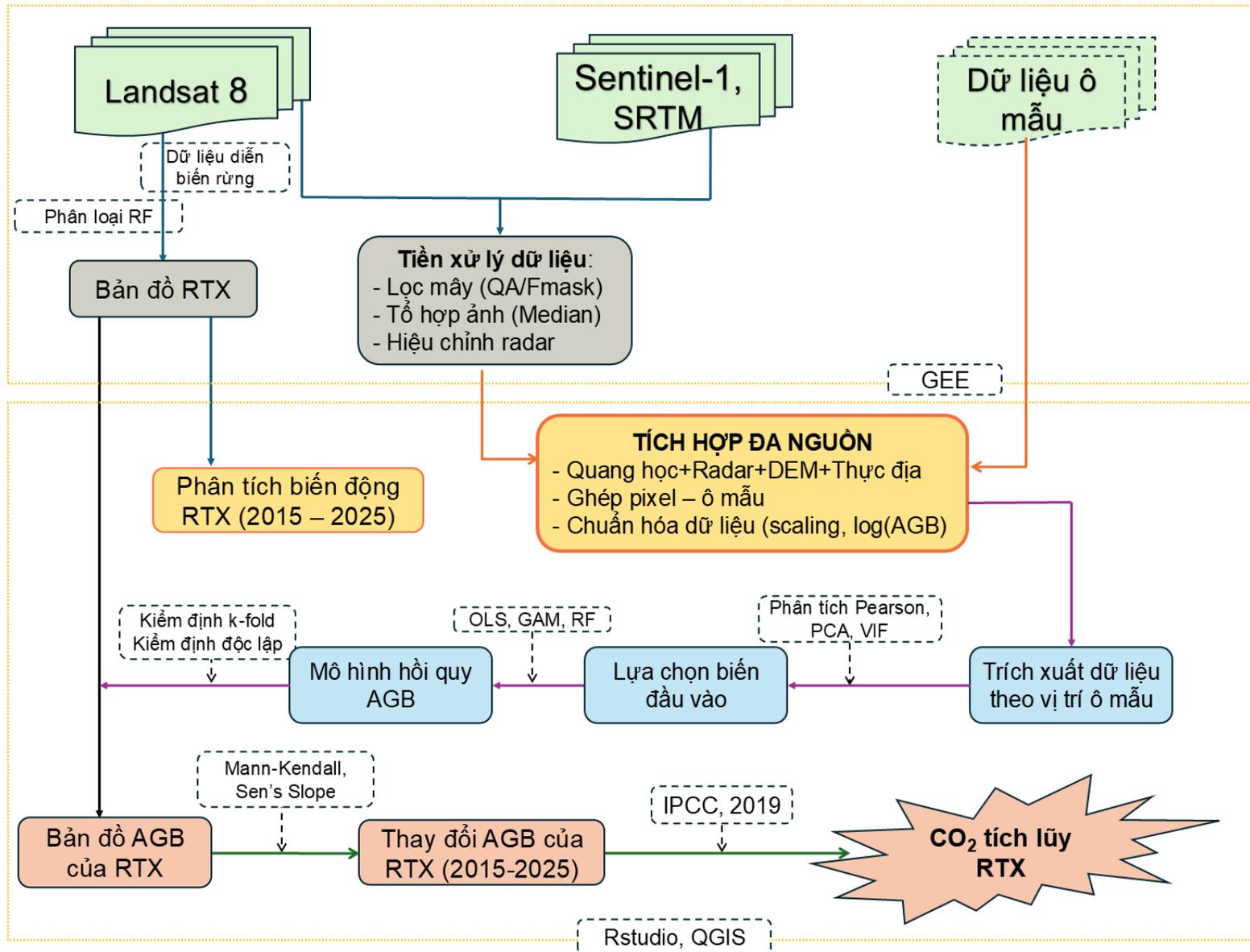
- Đánh giá hiệu quả mô hình bằng các chỉ tiêu MAE, RMSE, R^2 và kiểm định bằng dữ liệu độc lập.
- Lựa chọn mô hình tối ưu cho khu vực nghiên cứu.

(4) Lập bản đồ AGB và lượng CO₂ hấp thụ, đánh giá biến động theo thời gian:

- Áp dụng mô hình tối ưu để ước tính AGB cho toàn tỉnh Đắk Lắk các mốc thời gian.
- Lập bản đồ phân bố AGB và CO₂ hấp thụ.
- Phân tích biến động sinh khối – các-bon theo không gian và thời gian giai đoạn nghiên cứu.
- Chuyển đổi AGB sang lượng CO₂ hấp thụ theo hệ số mặc định.
- Đánh giá xu hướng biến động hấp thụ CO₂ để hỗ trợ quản lý tài nguyên rừng và chính sách các-bon.

2.3. Phương pháp nghiên cứu

Nghiên cứu này được tiếp cận theo hướng tích hợp đa nguồn dữ liệu, kết hợp dữ liệu thực địa với chuỗi ảnh viễn thám quang học đa thời gian, bao gồm cả ảnh quang học ở mức phản xạ bề mặt (Surface Reflectance – SR) và ảnh ra-đa Sentinel 1A để ước tính, lập bản đồ và phân tích biến động AGB, đồng thời suy tính khả năng hấp thụ CO₂ của rừng theo không gian và thời gian. Tiếp cận chung của nghiên cứu được tóm tắt trong sơ đồ Hình 2.2.



Hình 2.2. Sơ đồ tiếp cận nghiên cứu

Trên thực địa, AGB được tính từ các ô mẫu tiêu chuẩn thông qua đo đạc DBH và H, áp dụng phương trình sinh trắc đã được kiểm định cho điều kiện địa phương; ô mẫu được bố trí theo phân khối trạng thái nhằm bảo đảm tính đại diện không gian và phù hợp với độ phân giải ảnh. Về viễn thám, sử dụng chuỗi ảnh SR nhất quán theo mùa giữa các năm; từ đó trích xuất các dải phổ, địa hình (độ cao và độ dốc) và các chỉ số phản ánh đặc tính sinh lý, sinh thái của tán rừng gồm (NDVI, EVI, SAVI, ARVI, SIPI, và NDWI). Dữ liệu viễn thám được ghép đồng thời với số đo AGB (cùng năm để tránh sai lệch) để đồng nhất điều kiện giữa ảnh và thực địa, thời vụ. Trước khi mô hình hóa, các biến độc lập được sàng lọc để loại bỏ đa cộng tuyến và giảm nhiễu thông tin, sau đó sàng lọc biến nhằm hạn chế đa cộng tuyến trước khi mô hình hóa.

Mô hình ước tính AGB dựa trên viễn thám được xây dựng theo ba cách tiếp cận: i) mô hình tham số (hồi quy tuyến tính); ii) mô hình hồi quy phi tham số; và iii) mô hình học máy (RF). Hiệu suất của mô hình được đánh giá dựa trên các chỉ số định lượng (R^2 , RMSE, MAE). Mô hình tối ưu được lựa chọn theo bằng chứng thực nghiệm, ưu tiên nguyên tắc đơn giản hóa hợp lý khi các mô hình đạt hiệu suất tương đương. Mô hình tối ưu sau đó được áp dụng để ước tính AGB cho từng pixel trên toàn vùng, đồng thời định lượng bất định dự báo (theo sai tiêu chuẩn/độ rộng khoảng dự báo đối với mô hình tham số, hoặc theo phân bố/quantile dự báo đối với mô hình phi tham số) nhằm làm cơ sở cho diễn giải kết quả và phát hiện thay đổi: một pixel/khu vực chỉ được kết luận tăng/giảm có ý nghĩa khi chênh lệch ước lượng vượt ngưỡng thay đổi tối thiểu (MDC) tính từ bất định của hai thời điểm. Biến động dài hạn của AGB tiếp tục được kiểm định bằng Mann–Kendall để xác định ý nghĩa thống kê của xu thế và được định lượng bằng hệ số dốc Sen (tấn/ha/năm). Kết luận xu thế chỉ được chấp nhận khi $p < 0,05$, khoảng tin cậy của dốc không bao trùm 0 và độ dốc vượt MDC suy từ bất định ước lượng. Cuối cùng, AGB được chuyển đổi sang các-bon và CO₂ theo hệ số khuyến nghị của IPCC (2019), qua đó ước tính và lập bản đồ khả năng hấp thụ CO₂ theo không gian - thời gian, hỗ trợ quản lý rừng bền vững, chi

trả dịch vụ môi trường rừng và hoạch định chính sách ở quy mô tỉnh và vùng.

2.3.1. Thu thập và xử lý dữ liệu thực địa

2.3.1.1. Thu thập dữ liệu thứ cấp

Một số thông tin, dữ liệu thứ cấp được thu thập từ các đơn vị quản lý hành chính nhà nước như Sở Nông nghiệp và Môi trường, Chi cục Kiểm Lâm, các Hạt kiểm lâm, ... Các dữ liệu này được sử dụng để đối chiếu, bổ trợ cho phân tích biến động rừng và chuẩn hóa bản đồ nền GIS. Các thông tin thu thập bao gồm:

- Thông tin, số liệu diễn biến rừng trong giai đoạn nghiên cứu
- Bản đồ diễn biến và kiểm kê rừng trong giai đoạn nghiên cứu
- Dữ liệu GIS liên quan đến ranh giới hành chính, địa hình tỉnh Đắk Lắk.

2.3.1.2. Thu thập dữ liệu hiện trường

Trong nghiên cứu 2 loại dữ liệu được thu thập với các mục đích khác nhau: i) dữ liệu mẫu để phân loại và đánh giá kết quả phân loại thảm phủ; ii) dữ liệu ô mẫu dùng để xây dựng mô hình sinh khối.

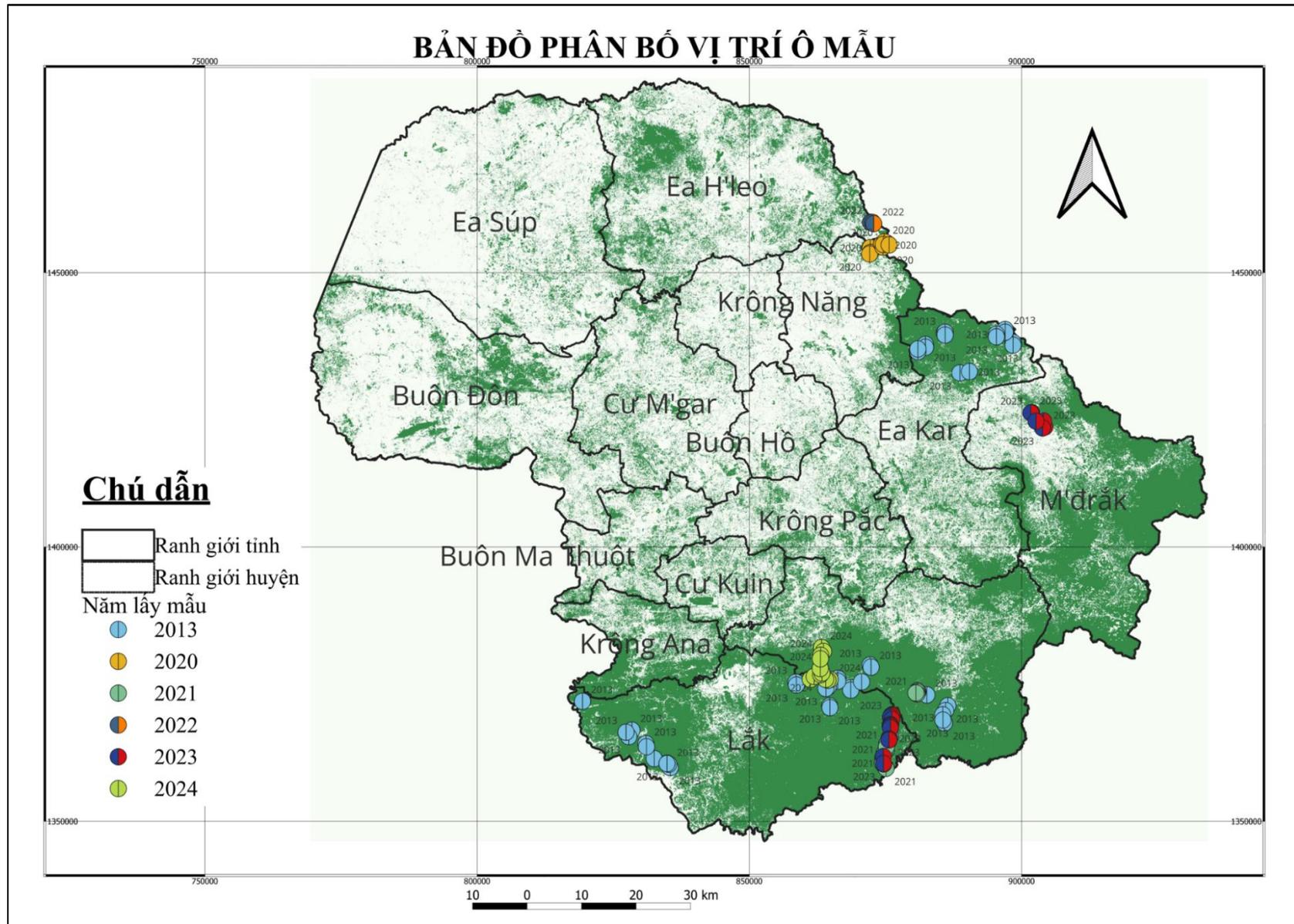
- Dữ liệu mẫu phục vụ phân loại và kiểm định độ chính xác được bố trí ngẫu nhiên dựa trên dữ liệu theo dõi diễn biến rừng. mẫu được lấy theo 2 loại thảm phủ: rừng thường xanh và khác.

- Dữ liệu ô mẫu dùng để xây dựng mô hình sinh khối được bố trí trong các khu vực rừng tự nhiên nhằm đảm bảo tính đại diện cho các trạng thái rừng thường xanh trên địa bàn tỉnh Đắk Lắk. Việc bố trí được thực hiện theo phương pháp phân tầng dựa trên trạng thái và cấu trúc rừng, bao gồm mức độ khép tán, mật độ cây, đường kính và chiều cao tầng ưu thế, cũng như điều kiện địa hình (độ dốc, hướng dốc, độ cao). Các tiêu chí này được nhận diện thông qua quan sát và đo đạc trực tiếp trên thực địa kết hợp với thông tin hỗ trợ từ ảnh vệ tinh, kết quả phân loại và bản đồ hiện trạng rừng gần nhất để xác định các khối trạng thái trước khi bố trí ô mẫu. Tuy nhiên, với địa hình các khu vực RTX thường khó tiếp cận do địa hình núi cao, độ dốc

lớn và bị chia cắt trong khi đó diện tích ô mẫu lớn và nguồn lực có hạn nên số lượng ô mẫu bố trí trên hiện trường có giới hạn và không thể phủ đều trên toàn bộ diện tích RTX của tỉnh.

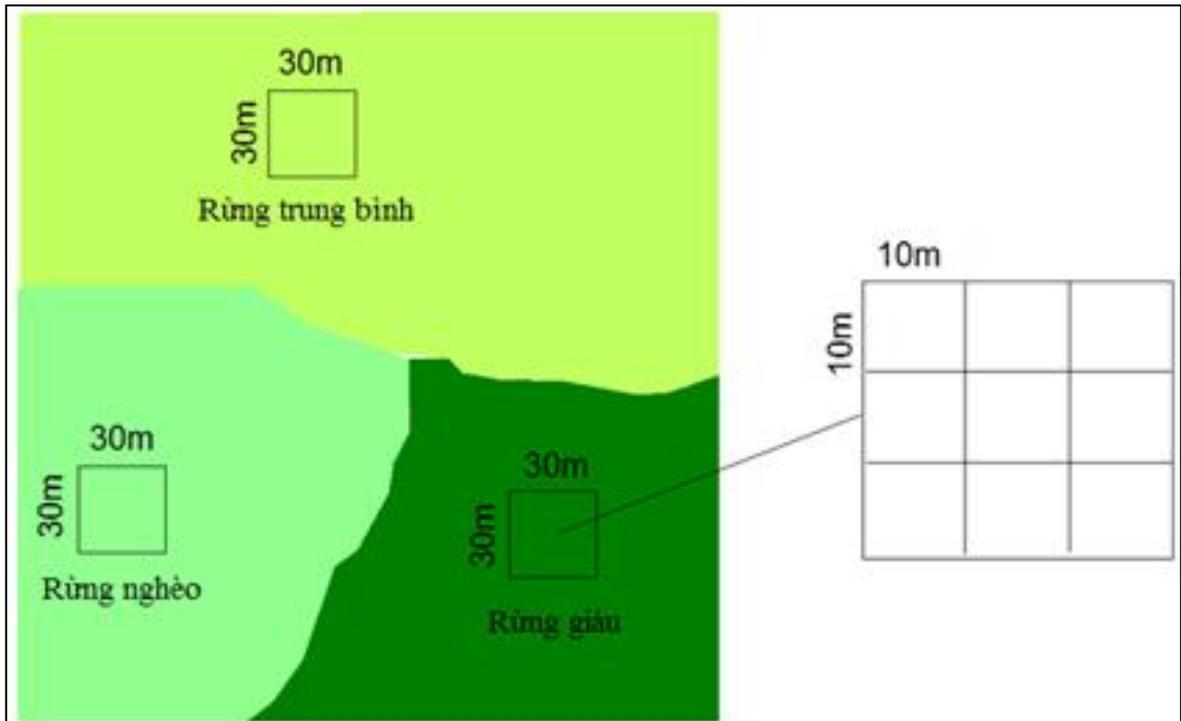
Mỗi ô mẫu được thiết kế hình vuông, diện tích 900 m^2 ($30 \times 30 \text{ m}$), tương ứng với độ phân giải không gian của ảnh Landsat. Cấu trúc này cho phép đối sánh trực tiếp giữa dữ liệu thực địa và giá trị pixel viễn thám, góp phần giảm sai số không gian trong quá trình trích xuất thông tin và xây dựng mô hình ước tính sinh khối.

Tổng cộng có 70 ô mẫu mới được bố trí tại các huyện M'Đrăk, Ea Kar, Krông Bông và Lắk là những khu vực có diện tích RTX điển hình và thể hiện sự biến động về địa hình và trạng thái rừng. Việc thu thập dữ liệu được tiến hành trong giai đoạn từ 2020 đến năm 2024. Ngoài ra, nghiên cứu kế thừa 47 ô mẫu từ dự án “Lập bản đồ phân bố thực vật quý hiếm” (2013) do nhóm nghiên cứu FREM – Trường Đại học Tây Nguyên thực hiện. Các ô kế thừa này đã được kiểm tra, cập nhật hiện trạng và chuẩn hóa tọa độ, thời gian nhằm đồng bộ với bộ dữ liệu mới. Bản đồ vị trí các ô mẫu được thể hiện ở Hình 2.3.



Hình 2.3. Bản đồ vị trí các ô mẫu

Kết hợp dữ liệu mới và kế thừa, tổng số 117 ô mẫu được sử dụng trong nghiên cứu, phân bố trên các khu vực có RTX nhiều của tỉnh Đắk Lắk. Bộ dữ liệu này được bố trí đại diện về không gian, đồng thời cung cấp nguồn thông tin thực tế cho việc ước tính sinh khối, hiệu chỉnh và kiểm định mô hình viễn thám. Hình dạng ô mẫu được minh họa như ở Hình 2.4.



Nguồn: Nguyễn Thị Thanh Hương, 2021

Hình 2.4. Phương pháp lập ô mẫu đo đếm

Các ô mẫu được bố trí theo hướng Đông – Tây, Nam – Bắc; trong ô mẫu được chia thành 9 ô phụ và được đánh số thứ tự từ 1 – 9 theo đường chéo từ Tây sang Đông và từ Bắc xuống Nam. Đối với địa hình dốc, ô mẫu được cải bằng tùy theo độ dốc. Trong các ô thứ cấp xác định loài cây và đo đếm đường kính tại vị trí 1,3m ($D_{1,3}$ -cm), chiều cao (H-m) và đường kính tán (CD-m) của các cây gỗ có đường kính ngang ngực từ 5cm trở lên. Đường kính cây rừng được đo bằng thước đo đường kính với độ chính xác đến 0,1cm; chiều cao cây được đo bằng máy đo cây đa năng Laser và máy đo cao Sunnto với độ chính xác 0,1m; đường kính tán được đo theo 02 hướng Đông – Tây và Nam – Bắc bằng thước đo dài với độ chính xác 0,1m.

Tính sinh khối rừng trên ô mẫu

Việc ước tính AGB tại các ô mẫu là bước quan trọng nhằm tạo lập bộ dữ liệu chuẩn cho xây dựng mô hình viễn thám. Để bảo đảm tính nhất quán giữa các nguồn dữ liệu thực địa khác nhau theo thời gian, nghiên cứu thực hiện đồng bộ hóa dữ liệu thông qua việc: chuẩn hóa hệ thống đo đếm (DBH tại 1,3 m; chiều cao cây H đo bằng máy đo Lazer) và thống nhất công thức tính AGB cho tất cả các năm đo đếm thực địa nhằm tránh sai lệch do dùng nhiều phương trình khác nhau.

Đối với lựa chọn phương trình sinh trắc, luận án sử dụng mô hình của Bảo Huy và cộng sự (2016a) – phương trình đã được xây dựng trên tập dữ liệu lớn gồm 968 cây giải tích thuộc các trạng thái rừng thường xanh ở Tây Nguyên, trong đó có nhiều mẫu thu thập tại tỉnh Đắk Lắk. Do đó, đây là mô hình có *tính đại diện cao* cho cấu trúc sinh thái – lâm học của rừng RTX khu vực nghiên cứu. Phương trình có dạng:

$$AGB = 263,9977 \times (DBH^2 \times H)^{0,93645} \quad (2.1)$$

Trong đó:

- DBH là đường kính ngang ngực (cm),
- H là chiều cao cây (m),
- AGB tính theo kg/cây.

Mô hình sử dụng hai biến dễ đo đếm và có độ ổn định cao qua các năm (DBH, H), phù hợp với điều kiện đo đạc tại hiện trường ở rừng tự nhiên Tây Nguyên. Bảo Huy và cộng sự (2016a) cũng đã chứng minh rằng phương trình này có sai số thấp (R^2 hiệu chỉnh = 0,896) và phù hợp hơn so với các mô hình sinh trắc quốc tế khi áp dụng trong điều kiện sinh thái địa phương.

Trong nghiên cứu này, phương trình (2.1) được áp dụng thống nhất cho toàn bộ dữ liệu thực địa các năm, nhằm đảm bảo:

- Tính đồng bộ về phương pháp tính AGB;

- Hạn chế sai số hệ thống giữa các năm;
- Tăng độ tin cậy cho mô hình hóa mối quan hệ giữa AGB và dữ liệu viễn thám.

Nhờ đó, bộ dữ liệu AGB đầu vào đạt tính nhất quán cao, phù hợp để tích hợp với dữ liệu ảnh vệ tinh đa thời gian trong các bước phân tích tiếp theo của luận án.

Để đối chiếu, nghiên cứu cũng tham khảo mô hình toàn cầu của Chave và cộng sự (2014), được xây dựng trên dữ liệu rừng nhiệt đới ở nhiều khu vực

$$AGB = 0,0673 \times (\rho \times DBH^2 \times H)^{0,976} \quad (2.2)$$

Trong đó: ρ là mật độ gỗ.

So sánh giữa hai cách tiếp cận cho thấy, mô hình toàn cầu của Chave và cộng sự (2014) có ưu điểm là được kiểm chứng trên nhiều kiểu rừng nhiệt đới khác nhau, tuy nhiên việc yêu cầu thông tin mật độ gỗ (ρ) cho từng loài cây khiến việc áp dụng trong thực tế gặp nhiều hạn chế, nhất là tại các hệ sinh thái nhiệt đới có tính đa dạng loài rất cao và thông tin về ρ thường không đầy đủ hoặc không đồng nhất. Ngược lại, mô hình của Bảo Huy và cộng sự (2016a) được xây dựng dựa trên dữ liệu thực địa tại Tây Nguyên, phản ánh đặc trưng sinh thái – cấu trúc của RTX khu vực nghiên cứu. Vì vậy, việc nghiên cứu sử dụng mô hình này giúp tăng độ tin cậy và phù hợp hơn cho mục tiêu ước tính AGB tại Đắk Lắk.

2.3.2. Dữ liệu viễn thám

2.3.2.1. Thu thập dữ liệu ảnh viễn thám

Dữ liệu ảnh viễn thám Landsat được thu thập cho giai đoạn 2013–2025, phục vụ đồng thời hai mục tiêu chính của nghiên cứu là phân loại lớp phủ rừng và ước tính sinh khối trên mặt đất (AGB). Toàn bộ dữ liệu được khai thác miễn phí từ nền tảng Google Earth Engine (GEE), cho phép truy cập trực tiếp và xử lý hiệu quả các sản phẩm Landsat Collection 2 do USGS cung cấp.

Trong nghiên cứu này, dữ liệu phản xạ bề mặt (Surface Reflectance – SR) từ

ảnh vệ tinh Landsat 8 OLI/TIRS được sử dụng cho giai đoạn 2013–2025. Các sản phẩm này thuộc Collection 2, đã được hiệu chỉnh hình học (orthorectified) và hiệu chỉnh khí quyển theo chuẩn của USGS. Đối với Landsat 8, phản xạ bề mặt được xử lý bằng thuật toán LaSRC (Land Surface Reflectance Code), đảm bảo tính nhất quán và độ tin cậy của dữ liệu đầu vào. Ảnh Landsat 8 OLI/TIRS cung cấp các dải phổ quan trọng cho nghiên cứu rừng, bao gồm 5 dải VNIR, 2 dải SWIR và 1 dải TIR (theo *Earth Engine Data Catalog*).

Do đặc điểm khí hậu nhiệt đới gió mùa của khu vực nghiên cứu, hiện tượng mây và bóng mây xuất hiện với tần suất cao, đặc biệt trong mùa mưa, gây ảnh hưởng đáng kể đến quá trình phân loại và mô hình hóa sinh khối. Vì vậy, loại bỏ mây và bóng mây được xem là bước tiền xử lý bắt buộc. Trong GEE, các cảnh ảnh Landsat 8 được xử lý loại bỏ mây bằng mặt nạ chất lượng (QA band) tích hợp sẵn trong sản phẩm SR, cho phép loại bỏ các pixel bị ảnh hưởng bởi mây, bóng mây và sương mù.

Để xây dựng ảnh không mây đại diện cho từng năm, nghiên cứu tiến hành tổng hợp toàn bộ các ảnh Landsat 8 hợp lệ trong khoảng thời gian từ ngày 01/01 đến 31/12 của từng năm trong khu vực nghiên cứu. Trên tập hợp ảnh đã được lọc mây này, giá trị trung vị (median) của từng pixel được tính toán để tạo ra ảnh tổ hợp theo năm (annual median composite). Phương pháp tổ hợp trung vị giúp hạn chế ảnh hưởng của các giá trị ngoại lai, giảm nhiễu do mây sót, đồng thời phản ánh ổn định đặc trưng phổ trung bình của bề mặt rừng trong năm nghiên cứu.

Sau khi tạo ảnh tổ hợp không mây, các ảnh được cắt theo ranh giới khu vực nghiên cứu nhằm giảm dung lượng dữ liệu và tối ưu thời gian xử lý. Toàn bộ quy trình tiền xử lý, từ lọc mây, tổng hợp ảnh theo năm đến cắt theo vùng nghiên cứu, được xây dựng và thực hiện thống nhất bằng mã lệnh trên nền tảng Google Earth Engine (Hình 2.5), đảm bảo tính tự động, lặp lại và nhất quán cho toàn bộ chuỗi thời gian nghiên cứu.

```

PhD_AGB_Dec
Get Link Save Run Reset
25 function makeStack(year) {
26   // ----- ẢNH QUANG HỌC (Landsat 8 làm ví dụ) -----
27   var landsat = ee.ImageCollection("LANDSAT/LC08/C02/T1_L2")
28     .filterBounds(roi)
29     .filter(ee.Filter.calendarRange(year, year, 'year'))
30   .map(function(img){
31     var qa = img.select('QA_PIXEL');
32     var cloudMask = qa.bitwiseAnd(1 << 3).eq(0).and(qa.bitwiseAnd(1 << 4).eq(0));
33     var img_sr = img.updateMask(cloudMask)
34       .multiply(0.0000275).add(-0.2) // scale Landsat SR
35       .select(['SR_B2', 'SR_B3', 'SR_B4', 'SR_B5', 'SR_B6', 'SR_B7'],
36         ['B2', 'B3', 'B4', 'B5', 'B6', 'B7']);
37     return img_sr;
38   })
39   .median();
40
41   // ----- CHỈ SỐ THỰC VẬT -----
42   var ndvi = landsat.normalizedDifference(['B5', 'B4']).rename('NDVI');
43   var evi = landsat.expression(
44     '2.5 * ((B5 - B4) / (B5 + 6*B4 - 7.5*B2 + 1))',
45     {'B2': landsat.select('B2'),
46      'B4': landsat.select('B4'),
47      'B5': landsat.select('B5')}).rename('EVI');
48   var savi = landsat.expression(
49     '((B5 - B4) / (B5 + B4 + 0.5)) * 1.5',
50     {'B4': landsat.select('B4'),
51      'B5': landsat.select('B5')}).rename('SAVI');
52   var arvi = landsat.expression(

```

Hình 2.5. Minh họa đoạn mã code thu thập ảnh trên GEE

Dữ liệu Sentinel-1 được thu từ vệ tinh ra-đa khẩu độ tổng hợp (SAR) băng tần C với tần số 5,405 GHz, cung cấp sản phẩm ảnh Ground Range Detected (GRD) đã được xử lý và hiệu chỉnh bằng Sentinel-1 Toolbox. Bộ dữ liệu được cập nhật hằng ngày và bao gồm các cảnh có độ phân giải 10 m, 25 m hoặc 40 m, với các chế độ phân cực khác nhau (VV, HH, VV+VH, HH+HV) tùy thuộc cấu hình thu nhận. Mỗi cảnh còn kèm theo dải dữ liệu gốc tới, được nội suy từ lưới geolocation để phục vụ hiệu chỉnh địa hình. Quy trình tiền xử lý ảnh Sentinel-1 bao gồm khử nhiễu nhiệt, hiệu chỉnh bức xạ, và hiệu chỉnh địa hình bằng SRTM-30. Sau cùng, dữ liệu được chuyển đổi sang giá trị phản xạ chuẩn hóa ở đơn vị decibel (dB) thông qua phép biến đổi lô-ga-rít ($10 \cdot \log_{10}(x)$), tạo cơ sở cho các phân tích viễn thám ra-đa và tích hợp với dữ liệu quang học trong nghiên cứu (Theo *Earth Engine Data Catalog*).

Dữ liệu địa hình: Trong nghiên cứu này, hai loại dữ liệu địa hình gồm độ cao (*elevation*) và độ dốc (*slope*) được sử dụng. Các dữ liệu này được trích xuất từ Mô hình số độ cao SRTM (Shuttle Radar Topography Mission) với độ phân giải không gian 30 m, do NASA phát triển, cung cấp thông tin địa hình toàn cầu đáng tin cậy cho

các phân tích không gian và môi trường [66]. Sau khi thu thập, dữ liệu được chuẩn hóa về cùng hệ tọa độ và độ phân giải với ảnh vệ tinh quang học và ra-đa nhằm đảm bảo tính đồng nhất và khả năng tích hợp trong mô hình hóa sinh khối rừng.

2.3.2.2. Tiền xử lý ảnh viễn thám

Tiền xử lý ảnh viễn thám được thực hiện nhằm đảm bảo tính đồng nhất và độ tin cậy của chuỗi dữ liệu phục vụ phân loại và mô hình hóa sinh khối. Nguồn ảnh Landsat 8 được sử dụng ở mức phản xạ bề mặt (Surface Reflectance – SR) do USGS cung cấp sẵn, giúp loại bỏ ảnh hưởng khí quyển và đảm bảo tính nhất quán phổ theo thời gian. Đối với Sentinel-1, dữ liệu quét tầm xa (Ground Range Detected-GRD) được hiệu chỉnh sơ cấp theo quy trình chuẩn gồm hiệu chỉnh nhiễu (thermal noise removal), hiệu chỉnh địa hình (terrain correction) và chuyển đổi sang hệ số tán xạ ngược sigma-naught, bảo đảm sự ổn định tín hiệu radar khi so sánh giữa các thời điểm.

Sau khi hiệu chỉnh, các pixel bị mây và bóng mây trên ảnh Landsat được loại bỏ bằng cách kết hợp dải đánh giá chất lượng (QA band) với thuật toán Fmask. Các ảnh hợp lệ sau lọc mây được tổ hợp theo mùa sinh trưởng của rừng (mùa khô và mùa mưa). Ảnh đại diện mỗi mùa được tạo bằng phương pháp hợp nhất trung vị (median composite) nhằm giảm nhiễu thời vụ và tăng tính ổn định của phổ phản xạ. Chuỗi ảnh SR theo mùa sau khi tổ hợp được sử dụng làm nền tảng để tính toán các biến độc lập cho mô hình AGB.

Trên cơ sở dữ liệu Landsat 8 SR đã được xử lý, nghiên cứu tính toán một loạt chỉ số phản ánh đặc tính quang học, sinh lý, sinh thái và tình trạng ẩm của thảm thực vật, bao gồm NDVI, EVI, SAVI, ARVI, SIPI, NDWI, MSI, GNDVI và tỷ số phổ SR (Bảng 2.1). Các chỉ số này giúp mô tả cấu trúc, sinh lý của tán rừng theo nhiều khía cạnh, giảm hiện tượng bão hòa tín hiệu và hỗ trợ tăng độ chính xác của mô hình ước tính sinh khối. Tín hiệu tán xạ ngược VH/VV từ Sentinel-1 sau khi chuẩn hóa góp phần bổ sung thông tin cấu trúc tán và độ ẩm, giúp khắc phục hạn chế do mây che phủ của ảnh quang học. Quy trình này tạo ra bộ dữ liệu viễn thám đa nguồn, nhất

quán và ổn định theo thời gian, phù hợp cho các bước phân loại thảm phủ rừng và xây dựng mô hình ước tính AGB trong khu vực nghiên cứu. Tóm tắt đặc điểm ứng dụng của các chỉ số và công thức tính toán được trình bày trong Bảng 2.1:

Bảng 2.1. Các chỉ số ảnh vệ tinh Landsat

Chỉ số	Công thức	Ý nghĩa chính	Ứng dụng
ARVI (Atmospherically Resistant Vegetation Index)	$ARVI = (NIR - (2 * Red) + Blue) / (NIR + (2 * Red) + Blue)$	Hiệu chỉnh tác động aerosol	Giám sát ở vùng nhiều bụi, khói, ô nhiễm không khí
EVI (Enhanced Vegetation Index)	$EVI = 2,5 * ((NIR - Red) / ((NIR) + (C1 * Red) - (C2 * Blue) + L))$	Khắc phục bão hòa NDVI, hiệu chỉnh đất và khí quyển	Phân tích rừng mưa nhiệt đới tán dày
GNDVI (Green Normalized Difference Vegetation Index)	$GNDVI = (NIR - Green) / (NIR + Green)$	Nhạy với hàm lượng diệp lục và trạng thái dinh dưỡng	Giám sát hàm lượng diệp lục, đánh giá tình trạng dinh dưỡng và stress
NDVI (Normalized Difference Vegetation Index)	$NDVI = (NIR - RED) / (NIR + RED)$	Phản ánh độ che phủ xanh, cường độ quang hợp	Giám sát thảm thực vật tổng quát, phân biệt khu vực có/không có thảm xanh
NDWI (Normalized Difference Water Index)	$NDWI = (Green - NIR) / (Green + NIR)$	Phản ánh hàm lượng nước trong thảm thực vật	Giám sát trạng thái ẩm và stress do khô hạn
MSI (Moisture Stress Index)	$MSI = SWIR / NIR$	Phản ánh hàm lượng nước của tán lá, tỷ lệ nghịch với độ ẩm	Theo dõi tình trạng khô hạn, stress do thiếu nước
SAVI (Soil Adjusted Vegetation Index)	$SAVI = ((NIR - Red) / (NIR + Red + L)) * (1 + L)$	Giảm ảnh hưởng nền đất (L : hệ số hiệu chỉnh 0–1)	Hữu ích cho rừng thưa, cây non, đất trống
SIPI (Structure Insensitive Pigment Index)	$SIPI = (NIR - Blue) / (NIR - Red)$	Phản ánh tỷ lệ carotenoids/diệp lục, chỉ thị stress sinh lý	Theo dõi sức khỏe thực vật, phát hiện bệnh hoặc suy thoái sinh lý
SR (Simple Ratio)	$SR = NIR / Red$	Phản ánh năng lượng phản xạ cận hồng ngoại so với vùng đỏ, tỷ lệ thuận với sinh khối và diện tích lá	Ước lượng sinh khối, năng suất, độ che phủ tán lá

2.3.3. Phân tích thay đổi thảm phủ rừng giai đoạn 2015-2025

2.3.3.1. Phân loại thảm phủ RTX

Trong nghiên cứu này, việc phân loại thảm phủ sử dụng ảnh vệ tinh Landsat 8 OLI/TIRS nhằm mục tiêu xây dựng bản đồ phân bố không gian của kiểu RTX. Quá trình phân loại được thực hiện thông qua ba bước chính: i) thu thập điểm mẫu giải đoán, ii) lựa chọn và huấn luyện các thuật toán phân loại, và iii) đánh giá kết quả phân loại.

- Thu thập điểm mẫu giải đoán

Dựa trên lớp dữ liệu diễn biến rừng và ảnh vệ tinh đa nguồn các năm nghiên cứu, tổng cộng 1.110 điểm mẫu đã được xây dựng, bao gồm 510 điểm cho lớp rừng thường xanh (RTX) và 600 điểm cho lớp “Khác” (gồm đất trống, cây trồng, khu dân cư, mặt nước, v.v.). Việc lựa chọn điểm mẫu được thực hiện theo phương pháp lấy mẫu ngẫu nhiên phân tầng (stratified random sampling) nhằm đảm bảo tính đại diện không gian cho từng lớp phủ và hạn chế sai lệch trong quá trình huấn luyện mô hình phân loại.

Theo khuyến nghị của Congalton và Green (2009), số lượng mẫu tối thiểu để đánh giá và giải đoán một lớp phủ là 30 điểm; do đó, quy mô mẫu sử dụng trong nghiên cứu này được xem là đủ lớn và đảm bảo độ tin cậy thống kê cho cả quá trình huấn luyện và đánh giá độ chính xác phân loại. Tập dữ liệu mẫu sau đó được chia ngẫu nhiên thành 80% cho huấn luyện mô hình và 20% cho kiểm định, bảo đảm các điểm kiểm định hoàn toàn độc lập với tập huấn luyện, từ đó phản ánh khách quan năng lực khái quát hóa của mô hình phân loại [107],[157].

Việc xác định lớp phủ rừng ở các thời điểm quá khứ được thực hiện theo phương pháp “lấy không gian thay thế thời gian” (space-for-time substitution). Ảnh vệ tinh hiện tại (Google Earth, Sentinel, Landsat) kết hợp với khảo sát thực địa được dùng để nhận dạng đặc trưng phổ và hình thái của RTX. Kiểu rừng này có đặc điểm phổ ổn định, dễ phân biệt với các lớp phủ khác, và phân loại chỉ ở cấp “kiểu rừng”,

nên giảm sai số. Từ các vùng mẫu hiện tại, đối chiếu với ảnh và bản đồ rừng các năm trước (Landsat) để suy luận lớp phủ quá khứ, qua đó xây dựng bộ dữ liệu mẫu suy diễn cho các năm không có điều tra, đảm bảo tính nhất quán chuỗi thời gian. Dữ liệu mẫu dùng để đánh giá kết quả phân loại suy diễn được kiểm tra bằng cách đối chiếu với ảnh vệ tinh đa thời gian (Landsat, Sentinel) và các bản đồ hiện trạng rừng hoặc dữ liệu kiểm kê cùng kỳ. Một số điểm mẫu được kiểm tra trực quan trên ảnh GE lịch sử để xác nhận xu thế lớp phủ. Việc so sánh đặc trưng phổ ổn định và logic biến động lớp phủ cho phép đánh giá và hiệu chỉnh tập mẫu quá khứ, đảm bảo tính nhất quán và hợp lý theo chuỗi thời gian.

- *Phân loại RTX bằng thuật toán Random Forest*

Thuật toán RF được lựa chọn để phân loại thảm phủ cho khu vực tỉnh Đắk Lắk do mang lại độ chính xác cao trong phân loại thảm phủ tại khu vực nghiên cứu [3],[14],[20],[24]. Thuật toán RF là phương pháp học máy tổ hợp dựa trên cây quyết định (decision trees) được huấn luyện ngẫu nhiên. Trong nghiên cứu này, số lượng cây (*n_{tree}*) được thiết lập cho phân loại là 100, đây là số cây phân loại tối ưu đối với thảm phủ chính trên địa bàn nghiên cứu [3]. Thuật toán RF có ưu điểm nổi bật là giảm phương sai mô hình, chống hiện tượng quá khớp (overfitting), đồng thời nâng cao độ ổn định và độ chính xác dự báo so với các mô hình cây đơn lẻ. Phân loại được triển khai trên nền tảng GEE thông qua công cụ *Classifier*.

- *Đánh giá kết quả phân loại:*

Đánh giá độ chính xác của kết quả phân loại được xem là bước quan trọng trước khi kết quả phân loại được sử dụng cho các phân tích tiếp theo [51],[143]. Sử dụng các chỉ tiêu đánh giá theo hướng dẫn của Congalton và Green (2009) như: sử dụng bảng ma trận để đánh giá độ chính xác dựa trên tiêu chí độ chính xác chung (Overall accuracy - OA), độ chính xác người sản xuất (Producer accuracy - PA) và độ chính xác người sử dụng (User accuracy - UA). Ngoài ra, việc đánh giá mức độ quan hệ giữa thực tế và kết quả phân loại còn dựa vào chỉ số Kappa. Hệ số Kappa dao động trong phạm vi từ 0 đến 1 và biểu thị sự giảm theo tỷ lệ về sai số được thực

hiện bằng một yếu tố phân loại hoàn toàn ngẫu nhiên.

Việc đánh giá độ chính xác cho từng lớp phủ và của toàn bộ ảnh phân loại được thực hiện dựa trên bảng ma trận sai số theo các tiêu chí đề nghị của Congalton và Green (2009) như sau:

$$\text{Độ chính xác tổng thể: } OA = \frac{\sum_{i=1}^r x_{ii}}{N} \times 100 \quad (2.3)$$

$$\text{Độ chính xác người sử dụng: } UA = \frac{x_{ii}}{x_{i+}} \times 100 \quad (2.4)$$

$$\text{Độ chính xác người sản xuất: } PA = \frac{x_{ii}}{x_{+i}} \times 100 \quad (2.5)$$

Hệ số Kappa (K):

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})} \quad (2.6)$$

Trong đó: N là tổng số pixel lấy mẫu; r là số lớp đối tượng phân loại; x_{ii} là số pixel đúng trong lớp thứ i ($i= 1,2,\dots,r$); x_{i+} là tổng pixel lớp thứ i của mẫu (loại thực tế/tổng giá trị theo hàng); x_{+i} là tổng pixel của lớp thứ i sau phân loại (loại giải đoán/tổng giá trị theo cột).

Chỉ số Kappa với các mức độ phân biệt được căn cứ vào bảng giá trị mức độ chặt chẽ theo chỉ số Kappa (Bảng 2.2) cụ thể như sau:

Bảng 2.2. Giá trị và mức độ chặt chẽ theo chỉ số Kappa (Nguồn: Navulur, 2007)

TT	Giá trị K	Mức độ phân biệt
1	<0,00	Thấp
2	0,00 – 0,20	Nhẹ
3	0,21 – 0,40	Vừa
4	0,41 – 0,60	Tương đối chặt
5	0,61 – 0,80	Chặt
6	0,81 – 1,00	Rất chặt

2.3.3.2. Phân tích thay đổi thảm phủ RTX giai đoạn 2015-2025

Dựa trên các bản đồ lớp phủ chuẩn hóa cho các mốc 2015, 2020, 2025 (cùng hệ phân loại, độ phân giải và phạm vi), nghiên cứu tiến hành ghép cặp theo giai đoạn (2015–2020; 2020–2025) và chồng lớp (overlay) từng cặp để xác định chuyển đổi from → to giữa các lớp. Từ lớp chuyển đổi này, ma trận biến động (change matrix) được lập bằng cách tổng hợp diện tích và tỷ lệ của từng hướng chuyển đổi [97], qua

đó tính toán biến động cho mỗi giai đoạn. Kết quả được trình bày dưới dạng bản đồ chuyển đổi và bảng ma trận theo giai đoạn, cung cấp số liệu định lượng và trực quan không gian về mức độ và hướng thay đổi thảm phủ trong toàn bộ giai đoạn 2015–2025.

2.3.4. Phân tích mối quan hệ giữa đặc trưng ảnh viễn thám và AGB

Hai bộ dữ liệu được sử dụng để ước tính AGB bao gồm: (1) Dữ liệu điều tra thực địa về sinh khối rừng; và (2) Dữ liệu viễn thám (ảnh vệ tinh và các chỉ số thực vật; dữ liệu địa hình).

Trước khi phân tích mối quan hệ giữa hai nguồn dữ liệu này, cần tiến hành chuẩn hóa và tiền xử lý để đảm bảo dữ liệu đầu vào đáp ứng các giả định thống kê của mô hình dự báo.

2.3.4.1. Chuẩn hóa dữ liệu AGB

Bộ dữ liệu gồm 117 ô mẫu thực địa, trong đó 47 ô được kế thừa từ nghiên cứu năm 2013 được sử dụng để đánh giá độc lập; 70 ô thu thập mới trong giai đoạn 2020–2024, được sử dụng để xây dựng mô hình ước lượng AGB.

Dữ liệu AGB được tiến hành kiểm tra nhằm đảm bảo các giả định thống kê cho quá trình mô hình hóa. Trước hết, AGB được kiểm tra phân bố thông qua các thống kê mô tả (giá trị trung bình, trung vị, độ lệch chuẩn, độ lệch (Skewness), độ nhọn (Kurtosis) và biểu đồ histogram đồng thời kiểm tra phân phối chuẩn của dữ liệu AGB. Kết quả cho thấy dữ liệu AGB có xu hướng lệch phải (right skewed), điều này có thể ảnh hưởng bất lợi đến độ ổn định của mô hình hồi quy, vốn giả định dữ liệu có phân phối gần chuẩn và phương sai đồng nhất. Để khắc phục, phép biến đổi lô-ga-rít ($\log AGB = \log(AGB + 1)$) được áp dụng cho toàn bộ bộ dữ liệu. Nguyên lý của phép biến đổi này là thu nhỏ khoảng cách giữa các giá trị lớn, đồng thời làm rõ sự khác biệt giữa các giá trị nhỏ, từ đó giảm độ lệch phải của phân bố. Phép biến đổi lô-ga-rít này giúp nén khoảng cách giữa các giá trị lớn, đồng thời làm nổi bật sự khác biệt giữa các giá trị nhỏ, từ đó giảm hiện tượng lệch phải và ổn định phương sai. Nhờ đó, bộ

dữ liệu AGB sau biến đổi phù hợp hơn cho việc xây dựng mô hình hồi quy và các phân tích thống kê tiếp theo, đảm bảo giả định về phân phối chuẩn và tính đồng nhất phương sai của phần dư.

Trong tập dữ liệu AGB, một số giá trị có mức cao bất thường nhưng vẫn nằm trong phạm vi hợp lý của sinh khối rừng nhiệt đới [112]. Song do số lượng ô mẫu hạn chế, các giá trị này được giữ lại nhằm bảo đảm phản ánh đầy đủ phổ biến thiên thực tế của sinh khối. Ảnh hưởng của chúng được kiểm soát thông qua phép biến đổi lô-ga-rít và áp dụng mô hình RF, vốn ít nhạy cảm với giá trị ngoại lai [42],[53]. Cách tiếp cận này cho phép duy trì toàn vẹn thông tin trong khi vẫn đảm bảo tính ổn định của mô hình.

2.3.4.2. Chuẩn hóa dữ liệu ảnh vệ tinh

Mối quan hệ giữa AGB rừng và ảnh vệ tinh được thực hiện dựa trên phân tích tương quan giữa AGB rừng và các biến số của ảnh vệ tinh cùng với biến số địa hình (bao gồm độ dốc và độ cao. Trong đó, biến số ảnh được sử dụng bao gồm các band ảnh gốc và các band chỉ số. Các bước thực hiện bao gồm:

Tọa độ trung tâm của từng ô mẫu được sử dụng để trích xuất giá trị pixel tương ứng từ các lớp dữ liệu viễn thám, bao gồm ảnh Landsat (band 1–7) và các chỉ số thực vật như NDVI, SAVI, EVI, ARVI, SIPI, NDWI..., cùng với dữ liệu ra-đa (VV, VH) và dữ liệu địa hình (*elevation, slope*) theo thời gian điều tra. Quá trình trích xuất được thực hiện bằng công cụ *Sample Raster Value* trong phần mềm QGIS, đảm bảo tính đồng nhất về tọa độ và độ phân giải giữa các lớp dữ liệu. Trong trường hợp ranh giới ô mẫu không trùng khớp hoàn toàn với lưới pixel, giá trị tại tọa độ trung tâm ô mẫu được sử dụng như một đại diện không gian (*spatial representative value*), theo cách tiếp cận được khuyến nghị bởi Foody (2002) và Lu & Weng (2007) nhằm giảm sai số do hiệu ứng trộn pixel và đảm bảo tính đại diện cho đơn vị mẫu.

Để đảm bảo sự đồng nhất giữa các biến viễn thám có đơn vị và thang đo khác nhau, các giá trị băng phổ và chỉ số thực vật được chuẩn hóa về cùng khoảng [0–1]

bằng phương pháp Min–Max Scaling [75]. Phương pháp này giúp giảm ảnh hưởng của các giá trị lớn bất thường, tăng độ ổn định cho quá trình huấn luyện mô hình và phù hợp với các thuật toán học máy phi tuyến như Random Forest. Việc chuẩn hóa được thực hiện theo biểu thức:

$$X_{scaled} = \frac{x_i - X_{min}}{X_{max} - X_{min}} \quad (2.7)$$

Trong đó: x_i là giá trị gốc của biến, X_{min} và X_{max} lần lượt là giá trị nhỏ nhất và lớn nhất của biến trong tập dữ liệu. Đây là phương pháp chuẩn hóa được khuyến nghị rộng rãi trong học máy, đặc biệt trong xử lý dữ liệu số có biên độ lớn [75]. Bên cạnh đó ảnh vệ tinh Sentinel-1 cũng được đưa về cùng độ phân giải với ảnh vệ tinh Landsat 8 để đồng nhất dữ liệu đầu vào.

2.3.4.3. Phân tích tương quan và chọn biến đầu vào cho mô hình AGB

Các giá trị phổ phản xạ ảnh và các chỉ số sinh học được kết hợp với dữ liệu AGB đo từ thực địa, tạo thành bộ cơ sở dữ liệu phục vụ mô hình hóa. Để hạn chế hiện tượng đa cộng tuyến, phân tích tương quan Pearson được áp dụng nhằm lựa chọn những biến có hệ số tương quan cao ($|r| \geq 0,8$) trước khi đưa vào mô hình.

Tiếp theo, phân tích thành phần chính (PCA – Principal Component Analysis) được áp dụng nhằm giảm chiều dữ liệu và nhận diện các biến đại diện cho từng nhóm nhân tố viễn thám. Phương pháp PCA giúp chuyển đổi tập biến ban đầu có tương quan cao sang hệ trục tọa độ mới gồm các thành phần chính không tương quan, qua đó tối ưu hóa khả năng giải thích phương sai của dữ liệu và hạn chế hiện tượng đa cộng tuyến khi đưa vào mô hình ước lượng sinh khối [75]. Đồng thời, PCA cung cấp cơ sở tham chiếu để đánh giá mức độ đóng góp tương đối của các nhóm biến (chỉ số thực vật, băng phổ, radar, địa hình) đối với sự biến thiên của AGB, hỗ trợ lựa chọn tập biến đầu vào tối ưu cho mô hình dự đoán.

Để kiểm tra và loại trừ hiện tượng đa cộng tuyến đa chiều còn lại giữa các biến độc lập, nghiên cứu sử dụng hệ số phóng đại phương sai (*Variance Inflation Factor – VIF*). VIF phản ánh mức độ mà phương sai của ước lượng hệ số hồi quy bị

phóng đại do sự tương quan giữa các biến độc lập [108]. Theo quy tắc thông thường, giá trị VIF lớn hơn 10 được xem là dấu hiệu của đa cộng tuyến nghiêm trọng và cần được xem xét loại bỏ hoặc đổi biến số [59]. Tuy nhiên, O'Brien (2007) cảnh báo rằng việc áp dụng cứng nhắc ngưỡng “ $VIF > 10$ ” có thể dẫn đến loại bỏ những biến có ý nghĩa thực tiễn, do đó cần đánh giá trong bối cảnh cụ thể của mô hình và dữ liệu. Một số nghiên cứu thực hành, đặc biệt trong mô hình hồi quy sinh thái và viễn thám, chấp nhận các biến có VIF dưới 20 để đảm bảo không loại bỏ các biến có tiềm năng giải thích cao [31],[173]. Trong trường hợp của nghiên cứu này, các biến có $VIF > 10$ được loại bỏ khỏi mô hình. Cách tiếp cận này giúp cân bằng giữa việc giảm thiểu đa cộng tuyến và duy trì khả năng giải thích của mô hình [103]. Các biến được chọn lọc sau khi phân tích được sử dụng để làm định hướng cho phân tích lập mô hình ước tính AGB và lượng CO₂ hấp thụ của RTX trên địa bàn tỉnh Đắk Lắk.

2.3.5. Xây dựng mô hình ước tính AGB từ dữ liệu viễn thám

Bộ dữ liệu đã chuẩn hóa tương ứng với 70 ô mẫu thu thập mới trong giai đoạn 2020–2024, được sử dụng để phân tích mối quan hệ giữa AGB với các nhân tố ảnh hưởng viễn thám. Mỗi ô mẫu đã được ghép với các biến viễn thám tương ứng theo năm, bao gồm dữ liệu địa hình, các kênh phổ Landsat và các chỉ số thực vật (NDVI, EVI, SAVI, ARVI, NDWI...). Các biến ra-đa Sentinel-1 chỉ được sử dụng cho giai đoạn 2020–2024 do hạn chế dữ liệu trước năm 2015; do đó, hai mô hình phân tích được xây dựng song song: (i) mô hình có tích hợp dữ liệu Sentinel-1, và (ii) mô hình chỉ sử dụng các biến quang học để đảm bảo khả năng áp dụng cho chuỗi ảnh từ 2013 đến 2025.

2.3.5.1. Xây dựng mô hình hồi quy AGB

Để tìm ra mô hình tối ưu mô tả mối quan hệ giữa AGB và các nhân tố ảnh hưởng viễn thám với số mẫu nhỏ, ba hướng tiếp cận chính được thử nghiệm: (i) mô hình hồi quy tuyến tính (Linear Regression – LR), (ii) mô hình hồi quy phi tuyến và (iii) mô hình học máy - RF. Toàn bộ quy trình được thực hiện trên phần mềm RStudio, bao gồm các bước: chuẩn hóa dữ liệu, xây dựng mô hình và đánh giá hiệu suất mô hình.

- *Mô hình hồi quy tuyến tính*

Mô hình hồi quy tuyến tính được thực hiện ở cả hai dạng: đơn biến và đa biến. Mô hình hồi quy đơn biến nhằm xác định nhanh mối quan hệ riêng lẻ giữa các biến viễn thám và sinh khối rừng, hỗ trợ đánh giá biến có khả năng dự báo tốt nhất trong điều kiện dữ liệu hạn chế hoặc ứng dụng thăm dò. Ngược lại, mô hình hồi quy tuyến tính đa biến được xây dựng để mô tả đồng thời ảnh hưởng của nhiều biến viễn thám đến sinh khối rừng, qua đó tăng khả năng giải thích biến động AGB và đánh giá hiệu quả kết hợp giữa dữ liệu quang học và ra-đa.

Mối quan hệ tuyến tính giữa AGB (biến phụ thuộc) và các biến viễn thám (biến độc lập) được thiết lập bằng hồi quy tuyến tính từng bước (stepwise regression) trên phần mềm RStudio. Công thức tổng quát của mô hình có dạng [108]:

$$AGB = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (2.8)$$

Trong đó:

- ✓ AGB là sinh khối trên mặt đất (biến phụ thuộc).
- ✓ β_0 là hằng số (intercept).
- ✓ $\beta_1, \beta_2, \dots, \beta_n$ là các hệ số hồi quy của các biến dự đoán.
- ✓ X_1, X_2, \dots, X_n là các biến dự đoán trích xuất từ dữ liệu viễn thám (chẳng hạn như chỉ số thực vật, kênh ảnh, v.v.).
- ✓ ε là sai số ngẫu nhiên.

- *Mô hình hồi quy phi tuyến đa biến*

Để mô tả tốt hơn mối quan hệ phi tuyến giữa sinh khối rừng và các biến viễn thám, nghiên cứu áp dụng Mô hình cộng tuyến tổng quát (Generalized Additive Model – GAM). Phương pháp này được xem là mở rộng của hồi quy tuyến tính tổng quát (GLM), trong đó mối quan hệ giữa biến phụ thuộc và các biến độc lập được mô tả thông qua các hàm trơn (smooth functions) thay vì các hệ số tuyến tính cố định.

Mô hình GAM được biểu diễn tổng quát như sau:

$$g(E(Y)) = \beta_0 + s_1(X_1) + s_2(X_2) + \dots + s_n(X_n) \quad (2.9)$$

Trong đó:

- Y: biến phụ thuộc (logAGB – sinh khối trên mặt đất theo lô-ga-rít tự nhiên),
- $s_i(X_i)$: hàm trơn mô tả mối quan hệ phi tuyến giữa biến độc lập X_i (các kênh phổ hoặc chỉ số thực vật) và logAGB,
- $g(E(Y))$: hàm liên kết (link function), trong nghiên cứu này sử dụng dạng tuyến tính đơn giản ($g(y) = y$).
- β_0 : là hằng số

Các hàm trơn $s_i(X_i)$ được xác định thông qua spline hồi quy có phạt (penalized regression splines), cho phép linh hoạt trong việc mô tả dạng cong của dữ liệu mà vẫn tránh hiện tượng quá khớp (overfitting).

Trong nghiên cứu này, GAM được triển khai trong môi trường R (gói “mgcv”), với hai mô hình:

- Mô hình GAM toàn bộ biến (gam1) – bao gồm tất cả các biến quang học, chỉ số phổ và ra-đa được lựa chọn từ phân tích tương quan.
- Mô hình GAM rút gọn (gam2) – sử dụng tập biến sau khi loại bỏ đa cộng tuyến dựa trên kết quả PCA và VIF.

Hiệu suất mô hình được đánh giá thông qua các chỉ tiêu:

- R^2 hiệu chỉnh: thể hiện mức độ giải thích biến thiên của logAGB,
- Deviance explained (%): tỷ lệ phương sai được giải thích,
- p -value của các hàm trơn $s(X_i)$: kiểm tra ý nghĩa thống kê của từng biến.

Việc áp dụng GAM cho phép phát hiện và mô hình hóa linh hoạt các quan hệ phi tuyến giữa phản xạ phổ – chỉ số thực vật – ra-đa và sinh khối rừng, giúp tăng độ

chính xác ước lượng so với hồi quy tuyến tính. Phương pháp này được chứng minh hiệu quả trong nhiều nghiên cứu gần đây về ước lượng sinh khối và các-bon rừng nhiệt đới [32].

- *Thuật toán RF:*

Thuật toán hồi quy RF được sử dụng nhằm ước lượng AGB. Mô hình được thiết lập với các biến đầu vào bao gồm các chỉ số phổ, kênh ảnh Landsat và dữ liệu ra-đa. Trong quá trình huấn luyện, RF tự động tạo các tập mẫu bootstrap (~70% dữ liệu) cho từng cây và sử dụng phần còn lại (~30%) làm mẫu “ngoài túi” (OOB) để đánh giá hiệu năng nội bộ mô hình. Tập dữ liệu được chia thành 70% cho huấn luyện (bagging) và 30% “ngoài túi” để kiểm định nội bộ. Trong quá trình huấn luyện, các siêu tham số chính của mô hình RF bao gồm số lượng cây quyết định (*ntree*) và số biến được chọn ngẫu nhiên tại mỗi lần phân chia (*mtry*) được hiệu chỉnh tự động thông qua đánh giá sai số nội bộ OOB. Giá trị tối ưu của các siêu tham số được xác định tại điểm mà sai số OOB đạt giá trị nhỏ nhất và độ ổn định của mô hình cao nhất. Mô hình được chạy thử với nhiều tổ hợp giá trị khác nhau của *ntree* (từ 500 đến 3000) và *mtry* (từ 3 đến 6), sau đó lựa chọn cấu hình có sai số OOB nhỏ nhất và độ ổn định cao nhất làm bộ tham số tối ưu. Việc lựa chọn tham số theo cách này giúp đảm bảo mô hình đạt hiệu năng tối ưu, tránh hiện tượng quá khớp (overfitting), đồng thời phản ánh khách quan mối quan hệ phi tuyến giữa sinh khối (AGB) và các đặc trưng viễn thám [40],[44].

Công thức tổng quát của mô hình hồi quy phi tuyến tính có thể được biểu diễn như sau [42]:

$$\gamma = \frac{1}{T} \sum_{t=1}^T f_t(X) \quad (2.10)$$

Trong đó:

- γ là giá trị dự đoán trung bình của mô hình RF.
- T là số lượng cây quyết định trong mô hình RF.
- $f_t(X)$ là giá trị dự đoán của cây quyết định thứ t dựa trên biến đầu vào X .

2.3.5.2. **Đánh giá độ chính xác của mô hình**

Để đánh giá độ chính xác và khả năng khái quát hóa của các mô hình ước lượng AGB từ dữ liệu viễn thám, nghiên cứu áp dụng hai phương pháp kiểm định gồm: kiểm định chéo k-fold (k-fold cross-validation) và kiểm định độc lập theo thời gian (temporal validation).

- *Kiểm định chéo k-fold*

Phương pháp kiểm định chéo k-fold là một kỹ thuật thống kê phổ biến trong học máy, cho phép đánh giá hiệu suất mô hình dự đoán trên các tập dữ liệu độc lập được tạo ra từ chính tập dữ liệu ban đầu, giúp giảm sai lệch ngẫu nhiên trong quá trình chia tách dữ liệu và hạn chế hiện tượng quá khớp (overfitting) [135]. Phương pháp này được áp dụng thống nhất cho cả ba nhóm mô hình gồm tuyến tính đa biến-OLS, mô hình phi tuyến tổng quát-GAM và mô hình học máy-RF. Các mô hình được huấn luyện và kiểm định trên cùng một cơ chế chia mẫu nhằm đảm bảo tính khách quan và khả năng so sánh trực tiếp giữa các thuật toán [158].

Toàn bộ 70 ô mẫu thực địa được chia ngẫu nhiên thành các nhóm (folds) có kích thước tương đương, với các giá trị k được thử nghiệm lần lượt là 5 và 10. Trong mỗi vòng lặp, k-1 nhóm (tương đương (k-1)/k dữ liệu) được dùng để huấn luyện mô hình và nhóm còn lại để kiểm định. Quá trình được lặp lại k lần, bảo đảm rằng mỗi ô mẫu được sử dụng đúng một lần trong kiểm định. Cấu hình 5-fold, được nhiều nghiên cứu khuyến nghị cho bộ dữ liệu cỡ nhỏ [58],[196], được lựa chọn làm phương án kiểm định chính trong nghiên cứu này.

Để đánh giá hiệu suất mô hình, nghiên cứu sử dụng bộ chỉ tiêu gồm R^2 , R^2 hiệu chỉnh, RMSE và MAE. Các chỉ tiêu này được lựa chọn vì phản ánh trực tiếp năng lực dự đoán của mô hình theo đơn vị thực tế (tấn/ha), dễ diễn giải và đặc biệt phù hợp khi so sánh các mô hình có bản chất khác nhau. Trong bối cảnh nghiên cứu kết hợp đồng thời mô hình tham số (OLS, GAM) với mô hình phi tham số (RF), việc sử dụng một hệ tiêu chí chung, không phụ thuộc vào giả định phân phối hay dạng hàm, là yêu cầu quan trọng để đảm bảo tính công bằng khi so sánh.

Mặc dù tồn tại các chỉ tiêu truyền thống như AIC, BIC, Mallows' Cp hoặc

các chỉ số dựa trên giả định phân phối, nhưng các tiêu chí này chủ yếu áp dụng cho mô hình tham số và phụ thuộc vào số lượng tham số, cấu trúc hàm và giả định sai số. Điều này khiến những tiêu chí này không phù hợp khi đánh giá đồng thời các mô hình phi tham số hoặc mô hình học máy vốn không dựa trên cấu trúc hàm cố định. Do đó, việc sử dụng R^2 , R^2 hiệu chỉnh, RMSE và MAE là phù hợp hơn trong nghiên cứu này vì cho phép đánh giá trực tiếp và nhất quán hiệu quả dự báo của các mô hình ước tính sinh khối từ dữ liệu viễn thám.

Các chỉ số đánh giá được tính riêng cho từng vòng lặp và giá trị trung bình của toàn bộ quá trình được sử dụng để phản ánh hiệu năng tổng thể của từng mô hình [58],[135],[196]. Mô hình có giá trị R^2 cao và sai số (MAE, RMSE) thấp hơn được xem là mô hình có hiệu suất tốt và được sử dụng để ước lượng sinh khối cho chuỗi thời gian 2015–2025.

Chỉ số MAE phản ánh mức sai lệch trung bình giữa giá trị dự đoán và quan sát, được tính theo công thức (2.11). Vì ít nhạy cảm với giá trị ngoại lai hơn RMSE, MAE thường được sử dụng song song để phản ánh mức sai số trung bình thực tế của mô hình. Sai số trung phương RMSE được tính theo công thức (2.12), thể hiện mức độ phân tán của sai số xung quanh giá trị trung bình.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2.11)$$

Sai số trung phương - RMSE được tính theo công thức:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2.12)$$

Trong đó: \hat{y}_i giá trị được ước lượng, y_i là giá trị thực được đo tính trên ô mẫu độc lập tức là ô mẫu không tham gia trong quá trình ước lượng.

n : số ước lượng tham gia đánh giá

Căn cứ vào kết quả đánh giá và so sánh hiệu suất, mô hình có độ chính xác cao nhất được lựa chọn để ước lượng và xây dựng bản đồ sinh khối cho toàn khu vực nghiên cứu.

- *Kiểm định độc lập theo thời gian*

Nhằm kiểm tra khả năng ổn định và khái quát của mô hình theo thời gian, tập dữ liệu gồm 70 ô mẫu thu thập trong giai đoạn 2020–2024 được sử dụng để xây dựng mô hình, trong khi 47 ô mẫu độc lập thu năm 2013 được dùng để kiểm định độ chính xác ước lượng. Do ảnh vệ tinh năm 2013 không có dữ liệu ra-đa, mô hình tối ưu được huấn luyện lại chỉ với các biến quang học và địa hình trước khi đánh giá độc lập để đảm bảo tính nhất quán của biến đầu vào giữa các thời kỳ.

2.3.6. Lập bản đồ AGB và lượng CO₂ hấp thụ, đánh giá biến động AGB giai đoạn 2015-2025

2.3.6.1. Áp dụng mô hình tối ưu để ước tính AGB tỉnh Đắk Lắk

Sau khi lựa chọn được mô hình dự báo AGB tối ưu (dựa trên các chỉ tiêu RMSE, MAE, R²), mô hình được áp dụng cho toàn bộ không gian tỉnh Đắk Lắk bằng cách:

Chuẩn bị bộ dữ liệu ảnh viễn thám đa thời gian (Landsat 8 2015–2025) đã được tiền xử lý (hiệu chỉnh khí quyển, lọc mây, trích xuất chỉ số phổ).

Chuẩn hóa các biến đầu vào theo cùng cấu trúc với mô hình huấn luyện (đặc biệt là quy tắc chuẩn hóa Min–Max [0–1] và chuyển đổi SAR sang dB).

Tính toán giá trị AGB cho toàn tỉnh thông qua mô hình tối ưu, theo từng mốc thời gian: 2015-2025.

Phương pháp này đảm bảo tính nhất quán trong ước lượng không gian, phù hợp với khuyến nghị của GFOI (2020) và IPCC (2019) cho theo dõi rừng dựa trên ảnh vệ tinh.

2.3.6.2. Phân tích biến động AGB tỉnh Đắk Lắk giai đoạn 2015-2025

Chuỗi bản đồ AGB theo năm được sử dụng để phân tích sự biến đổi sinh khối rừng trong giai đoạn 2015–2025. Sau khi tổng hợp và tính toán các thống kê mô tả cho từng năm, ảnh raster AGB được xuất ra và áp dụng phương pháp phát hiện thay

đòi để đánh giá sự biến động theo thời gian. Để nâng cao độ tin cậy của ước lượng, nghiên cứu đã kết hợp ba bước để hỗ trợ nhau bao gồm: i) phân tích độ bất định nhằm đánh giá mức độ tin cậy của giá trị ước lượng và xác định ngưỡng thay đổi tối thiểu để loại bỏ biến động giả do sai số mô hình; ii) kiểm định Mann–Kendall để xác định xu thế tăng, giảm hay ổn định của AGB; và iii) hệ số Sen's slope để định lượng mức độ biến đổi qua các năm. Cách tiếp cận này giúp diễn giải biến động sinh khối một cách chặt chẽ, nhất quán và ít bị ảnh hưởng bởi nhiễu của ảnh viễn thám.

Phân tích độ bất định và ngưỡng thay đổi tối thiểu (MDC): Để đảm bảo việc diễn giải biến động AGB có ý nghĩa thống kê, tiến hành phân tích độ bất định và xác định ngưỡng thay đổi tối thiểu (Minimum Detectable Change – MDC) tức là mức thay đổi nhỏ nhất của AGB giữa hai thời điểm có thể khẳng định là thực sự có sự khác biệt chứ không phải do sai số ngẫu nhiên.

Ngưỡng MDC được xác định dựa trên sai số chuẩn (Standard Error – SE) của mô hình ước lượng theo công thức (2.13):

$$MDC = z_{\alpha/2} \times \sqrt{SE_{t1}^2 + SE_{t2}^2} \quad (2.13)$$

Trong đó:

- $z_{\alpha/2}$ là giá trị tới hạn của phân phối chuẩn ứng với mức ý nghĩa (chọn $\alpha=0,05 \Rightarrow z=1,96$);
- SE_{t1}, SE_{t2} là sai số chuẩn của ước lượng AGB tại hai thời điểm so sánh.

Việc áp dụng ngưỡng MDC trong nghiên cứu này giúp loại bỏ các biến động, tăng độ tin cậy của kết quả phân tích và cung cấp cơ sở khoa học vững chắc khi đánh giá mức độ thay đổi AGB và khả năng hấp thụ CO₂ theo không gian – thời gian.

Phân tích xu thế sử dụng kiểm định Mann-Kendall: Để đánh giá xu thế thay đổi của AGB theo thời gian, kiểm định Mann-Kendall (MK) đã được áp dụng [84],[102]. Kiểm định MK là một phương pháp thống kê phi tham số, được sử dụng rộng rãi để phát hiện xu thế đơn điệu trong dữ liệu mà không yêu cầu dữ liệu tuân theo phân phối cụ thể nào. Phương pháp này đặc biệt phù hợp để xác định xu hướng dài hạn trong dữ liệu nghiên cứu.

Gọi x_1, x_2, \dots, x_n là các giá trị AGB theo thời gian. Thống kê kiểm định Mann-Kendall (S) được tính bằng phương trình sau:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(x_j - x_i) \quad (2.14)$$

Trong đó:

$$\text{Sign}(x_j - x_i) = \begin{cases} +1, & \text{nếu } x_j - x_i > 0 \\ 0, & \text{nếu } x_j - x_i = 0 \\ -1, & \text{nếu } x_j - x_i < 0 \end{cases}$$

Phương sai của S được tính theo công thức:

$$\text{Var}(S) = \frac{n(n-1)(2n+5)}{18} \quad (2.15)$$

Thống kê kiểm định Z được tính như sau:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}}, & S > 0 \\ 0, & S = 0 \\ \frac{S+1}{\sqrt{\text{Var}(S)}}, & S < 0 \end{cases} \quad (2.16)$$

Mức ý nghĩa của xu thế được xác định bằng giá trị P tương ứng với thống kê kiểm định Z. Ngưỡng ý nghĩa 0,05 được sử dụng để đánh giá mức độ có ý nghĩa thống kê của xu thế. Nếu $P < 0,05$, xu thế được coi là có ý nghĩa thống kê.

Ước lượng mức độ thay đổi AGB bằng ước lượng Sen's Slope (Sen's Slope Estimator): Để định lượng cường độ thay đổi của xu thế AGB, phương pháp ước lượng dốc Sen's Slope được áp dụng [165]. Phương pháp này tính toán độ dốc trung vị giữa tất cả các cặp điểm dữ liệu, cung cấp ước lượng về tốc độ thay đổi theo thời gian. Độ dốc cho mỗi cặp điểm dữ liệu (x_i, x_j) được tính theo công thức:

$$Q_i = \frac{x_j - x_i}{j - i} \text{ với } i < j \quad (2.17)$$

Trong đó: i và j là các chỉ số thời gian. Ước lượng Sen's slope là giá trị trung vị của tất cả các độ dốc đã tính:

$$Q = \text{median}(Q_1, Q_2, \dots, Q_m) \quad (2.18)$$

Trong đó m là số lượng độ dốc được tính. Dấu của dốc Sen's slope biểu thị

chiều hướng xu thế:

$$Q > 0 \rightarrow \text{xu thế tăng}$$

$$Q < 0 \rightarrow \text{xu thế giảm}$$

Khoảng tin cậy của độ dốc được tính bằng phân phối chuẩn. Nếu kiểm định Mann-Kendall cho kết quả có ý nghĩa thống kê và giá trị dốc Sen's slope dương, điều đó cho thấy xu thế tăng. Ngược lại, dốc Sen's slope âm cho thấy xu thế giảm.

Tất cả việc xử lý dữ liệu và phân tích thống kê được thực hiện bằng phần mềm R với gói "trend". Dữ liệu AGB được tổ chức theo loại hình che phủ đất và theo năm. Các phân tích kiểm định Mann-Kendall và ước lượng dốc Sen được thực hiện riêng biệt cho từng loại che phủ đất. Kết quả được diễn giải dựa trên ý nghĩa thống kê và mức độ của xu thế được phát hiện.

2.3.6.3. Khả năng hấp thụ và biến động CO₂ của RTX

Khả năng hấp thụ CO₂ của rừng thường xanh được ước tính gián tiếp thông qua AGB đã được mô hình hóa từ dữ liệu viễn thám. Trên cơ sở bản đồ AGB cho từng thời điểm nghiên cứu, lượng các-bon tích lũy và lượng CO₂ hấp thụ được tính toán theo hệ số mặc định của IPCC (2019). Phương pháp chuyển đổi được thực hiện như sau:

$$\text{Các-bon} = \text{AGB} \times 0,47 \quad (2.19)$$

$$\text{CO}_2 = \text{Các-bon} \times 3,67 \quad (2.20)$$

Trong đó:

- 0,47 là hệ số chuyển đổi sinh khối khô sang lượng các-bon theo hướng dẫn IPCC (2019), phù hợp cho rừng nhiệt đới trong điều kiện không có hệ số đặc thù địa phương.
- 3,67 là hệ số chuyển đổi các-bon thành CO₂ (dựa trên tỷ lệ khối lượng phân tử CO₂/C).

Sau khi chuyển đổi, các bản đồ phân bố CO₂ hấp thụ được thành lập cho từng

mốc thời gian (2015-2025) nhằm phản ánh sự khác biệt theo không gian của khả năng tích lũy các-bon. Các bản đồ này tiếp tục được chồng xếp và phân tích biến động theo chuỗi thời gian nhằm:

- Xác định các khu vực tăng hoặc giảm khả năng hấp thụ CO₂;
- Đánh giá tốc độ biến động CO₂ giữa các giai đoạn;
- Xác định những vùng có xu hướng suy giảm hấp thụ CO₂ do mất rừng, suy thoái rừng hoặc giảm sinh trưởng;
- Xác định các vùng có khả năng phục hồi các-bon cao, từ đó hỗ trợ lập luận cho quản lý rừng, chi trả dịch vụ môi trường rừng và xây dựng đường cơ sở các-bon trong REDD⁺.

Cách tiếp cận này đảm bảo tính nhất quán giữa mô hình sinh khối, bản đồ AGB và đánh giá các-bon, đồng thời tuân thủ khuyến nghị của IPCC và GFOI về tích hợp dữ liệu viễn thám trong theo dõi các-bon rừng.

CHƯƠNG 3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Đánh giá thay đổi RTX giai đoạn 2015 – 2025

3.1.1. Phân loại ảnh xác định kiểu RTX

Sử dụng mô hình phân loại RF để chiết tách thảm phủ RTX từ ảnh vệ tinh. Kết quả phân loại được đánh giá dựa vào bộ mẫu độc lập tức là điểm không tham gia vào quá trình huấn luyện. Kết quả được tổng hợp trong Bảng 3.1 và trực quan hóa thông qua ma trận nhầm lẫn (Confusion Matrix) của từng năm. Việc sử dụng tập dữ liệu độc lập giúp đảm bảo tính khách quan và độ tin cậy, đồng thời phản ánh năng lực khái quát hóa thực tế của mô hình RF khi áp dụng cho khu vực nghiên cứu.

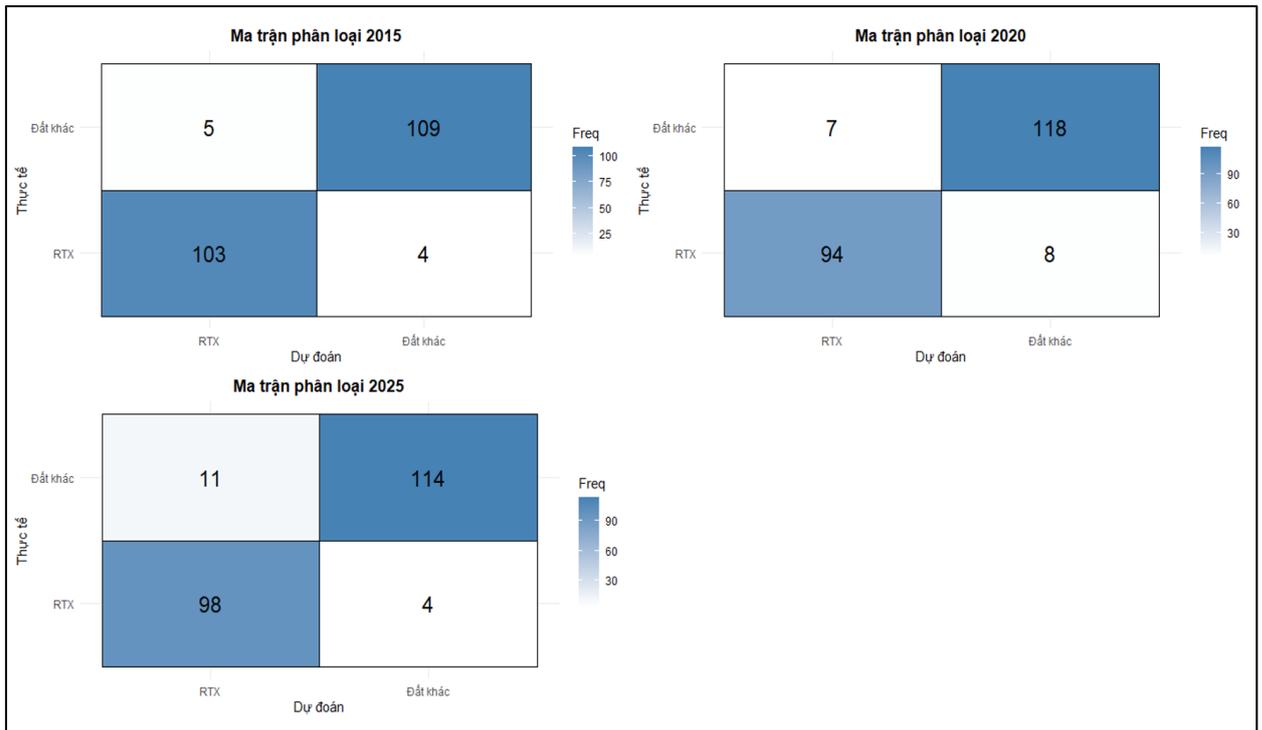
Bảng 3.1. Kết quả tổng hợp độ chính xác phân loại các năm

Năm	OA (%)	Kappa	PA RTX (%)	PA Đất khác (%)	UA RTX (%)	UA Đất khác (%)
2015	95,93	0,918	96,26	95,61	95,37	96,46
2020	93,39	0,866	92,16	94,40	93,07	93,65
2025	93,39	0,867	96,08	91,20	89,91	96,61

Chú thích: OA = Overall Accuracy (Độ chính xác tổng thể), PA = Producer's Accuracy (Độ chính xác của nhà sản xuất), UA = User's Accuracy (Độ chính xác của người sử dụng)

Kết quả cho thấy: Lớp RTX được nhận diện rất tốt qua các năm, với $PA \geq 92\%$ và $UA \geq 90\%$, cho thấy mô hình RF phân loại chính xác lớp rừng mục tiêu. Lớp “Đất khác” cũng duy trì độ chính xác cao, đặc biệt năm 2015 đạt $PA = 96,26\%$ và $UA = 95,37\%$, phản ánh khả năng mô hình phân biệt rừng và các loại đất phi rừng hiệu quả.

So với các nghiên cứu trước của Hồ Đình Bảo và cộng sự (2025) phân loại 4 loại thảm phủ tại Đắc Lắc đạt $OA = 0,91$, lớp rừng $PA/UA = 0,95/0,92$. Nguyễn Thị Thanh Hương và cộng sự (2020) sử dụng Sentinel-2 cũng cho kết quả tương tự, với lớp RTX ở Đắc Nông được nhận diện chính xác hơn các loại rừng khác. Khi tăng số lớp giải đoán hoặc chi tiết hóa lớp phủ, độ chính xác giảm do khả năng nhầm lẫn cao hơn [1],[20],[21],[23]. Ma trận nhầm lẫn phân loại được thể hiện ở Hình 3.1.



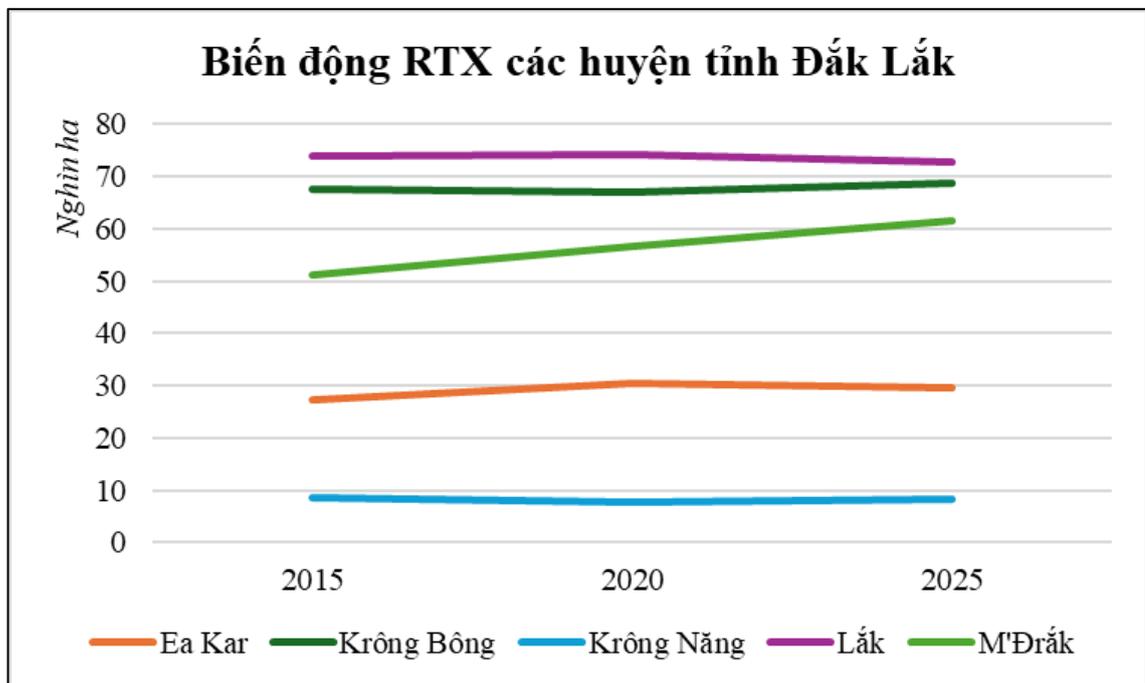
Hình 3.1. Ma trận nhầm lẫn phân loại

Độ chính xác toàn cục của nghiên cứu này cũng có sự cải thiện hơn so với khi sử dụng ảnh Landsat 5 và thuật toán RF để phân loại cho 4 loại thảm phủ chính là Rừng, Đất nông nghiệp, Dân cư và Mặt nước trên địa bàn tỉnh Đắk Lắk của Hồ Đình Bảo và cộng sự (2025) khi độ chính xác tổng thể cho 4 loại thảm phủ chỉ đạt 0,91 trong đó lớp rừng có độ chính xác của người sản xuất và người sử dụng đạt lần lượt là 0,95 và 0,92; lớp phủ nông nghiệp chỉ đạt lần lượt là 0,83 và 0,81. Nguyễn Thị Thanh Hương và cộng sự (2020) khi sử dụng ảnh vệ tinh Sentinel-2 để phân loại thảm phủ trên địa bàn tỉnh Đắk Nông cũng cho kết quả tương đồng khi RTX cho kết quả phân loại có độ chính xác cao hơn so với các loại rừng khác, đặc biệt là đối với cây công nghiệp. Khi tăng số lớp giải đoán, loại thảm phủ chi tiết hơn thì kết quả phân loại thường giảm độ chính xác do khả năng nhầm lẫn giữa các lớp cao hơn. Ngô Đức Anh và cộng sự (2022) đã phân loại cho 10 loại thảm phủ khác nhau ở tỉnh Lâm Đồng độ chính xác tổng thể chỉ đạt 0,76; hay nghiên cứu của Nguyễn Thị Thanh Hương và Đoàn Minh Trung (2018), Mai Thị Huyền và cộng sự (2021) cho 5 loại thảm phủ rừng khác nhau cũng chỉ đạt 0,90; nghiên cứu khác cho 7 loại thảm phủ khác nhau

trên địa bàn tỉnh Đắk Lắk của Nguyễn Thị Ngọc Quyên và cộng sự (2016) cũng chỉ đạt độ chính xác toàn cục là 0,74. Như vậy, việc chiết tách RTX ra khỏi các loại thảm phủ khác đã cho độ chính xác tốt hơn và phù hợp với mục tiêu nghiên cứu là ước lượng sinh khối và khả năng hấp thụ CO₂ chung cho kiểu rừng này.

Kết quả này cho thấy việc sử dụng mô hình RF để phân loại ảnh vệ tinh đa phổ như Landsat là lựa chọn phù hợp cho việc lập bản đồ thảm phủ rừng RTX tại Đắk Lắk. Đồng thời, góp phần mở rộng cơ sở dữ liệu chuỗi thời gian dài (2015–2025), tạo tiền đề quan trọng cho các bước phân tích biến động sinh khối và hấp thụ CO₂ của kiểu rừng này.

Kết quả phân loại và thống kê diện tích RTX từ ảnh viễn thám cho thấy sự phân bố của kiểu rừng này tại tỉnh Đắk Lắk có sự chênh lệch rõ rệt giữa các khu vực địa lý (Hình 3.2).



Hình 3.2. Diện tích rừng thường xanh theo huyện giai đoạn 2015-2025

Kết quả thống kê diện tích RTX tại các huyện Ea Kar, Krông Bông, Krông Năng, Lắk và M'Đrắk trong giai đoạn 2015–2025 cho thấy sự biến động không đồng đều theo không gian và thời gian, phản ánh rõ các xu thế khác biệt về mức độ phục

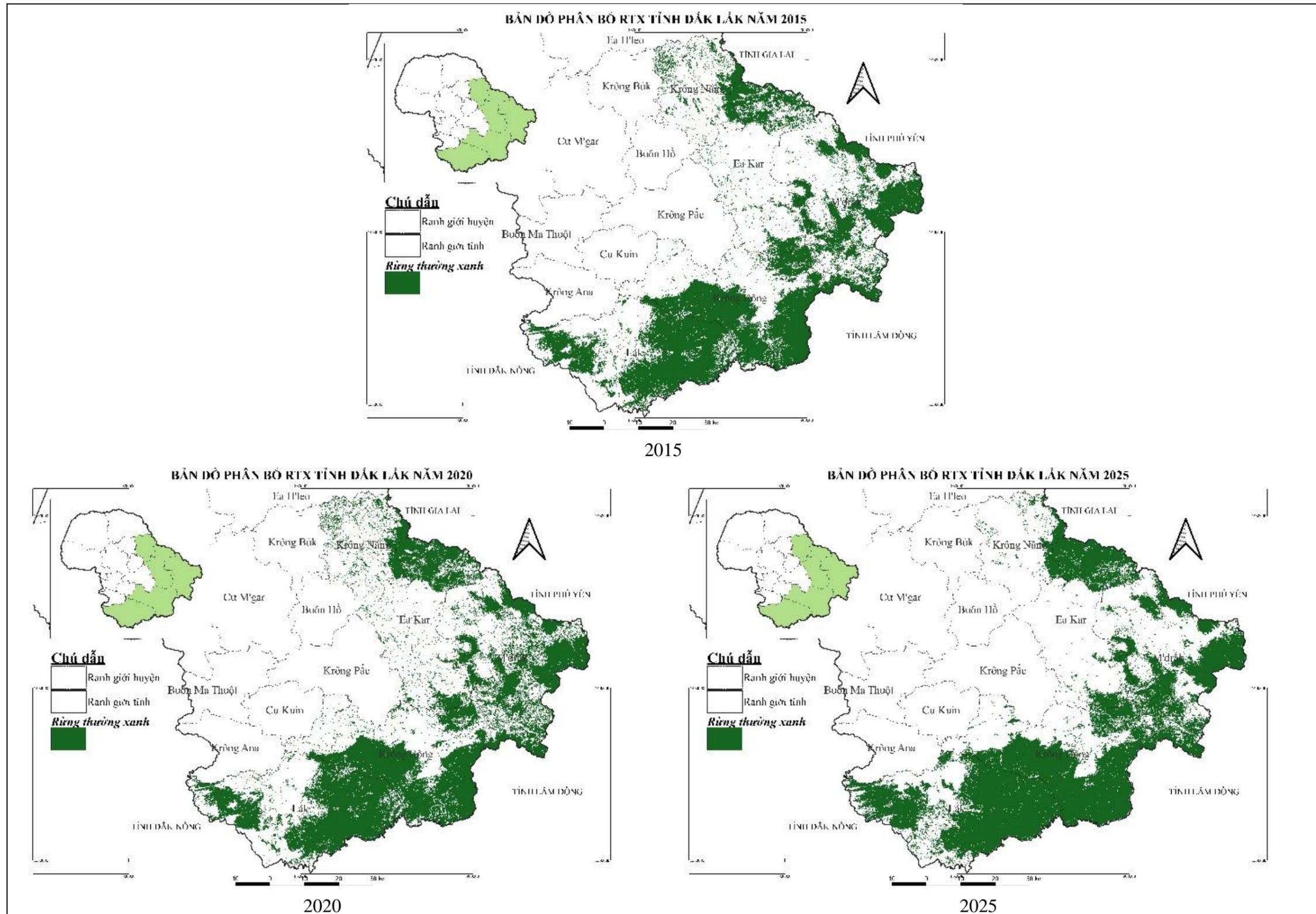
hồi và suy giảm rừng giữa các khu vực trong tỉnh.

Trong giai đoạn 2015–2025, tổng diện tích RTX có xu hướng tăng nhẹ, chủ yếu nhờ các hoạt động phục hồi rừng và quy định đóng cửa rừng tự nhiên trong giai đoạn này. Tuy nhiên, mức độ biến động khác biệt đáng kể giữa các huyện, cho thấy tác động tổng hợp của điều kiện tự nhiên, áp lực kinh tế - xã hội và công tác quản lý rừng.

Về tổng thể, diện tích rừng thường xanh (RTX) tại các huyện có xu hướng tăng nhẹ trong giai đoạn 2015–2025, mặc dù mức độ biến động khác nhau giữa các địa phương. M'Đrăk là huyện ghi nhận mức tăng mạnh nhất (tăng 10.110,42 ha), tiếp đến là Ea Kar (tăng 2.361,87 ha) và Krông Bông (tăng 1.254,24 ha). Ngược lại, diện tích RTX giảm nhẹ tại Lắk (giảm 1.104,93 ha) và Krông Năng (giảm 273,87 ha).

Trong giai đoạn 2020–2025, diện tích RTX tăng rõ rệt tại Krông Bông và đặc biệt là M'Đrăk, trong khi Ea Kar và Lắk có xu hướng giảm nhẹ; Krông Năng phục hồi nhưng chưa đạt mức năm 2015. Tổng hợp toàn tỉnh cho thấy diện tích RTX tăng từ 228.474,72 ha năm 2015 lên 240.822,45 ha năm 2025, tương ứng mức tăng 12.347,73 ha. Kết quả này cho thấy xu hướng phục hồi rừng thường xanh đã xuất hiện, góp phần bù đắp diện tích suy giảm cục bộ ở một số địa phương.

Phân bố diện tích rừng theo không gian cho thấy RTX của tỉnh tập trung phân bố chủ yếu ở phía Đông – Nam gồm các huyện Krông Bông, Lắk, M'Đrăk, Ea Kar – nơi địa hình dốc, ít tác động sản xuất và là vùng lõi của các Vườn quốc gia Chư Yang Sin và các khu bảo tồn thiên nhiên Nam Ka và Ea Sô (Hình 3.3).



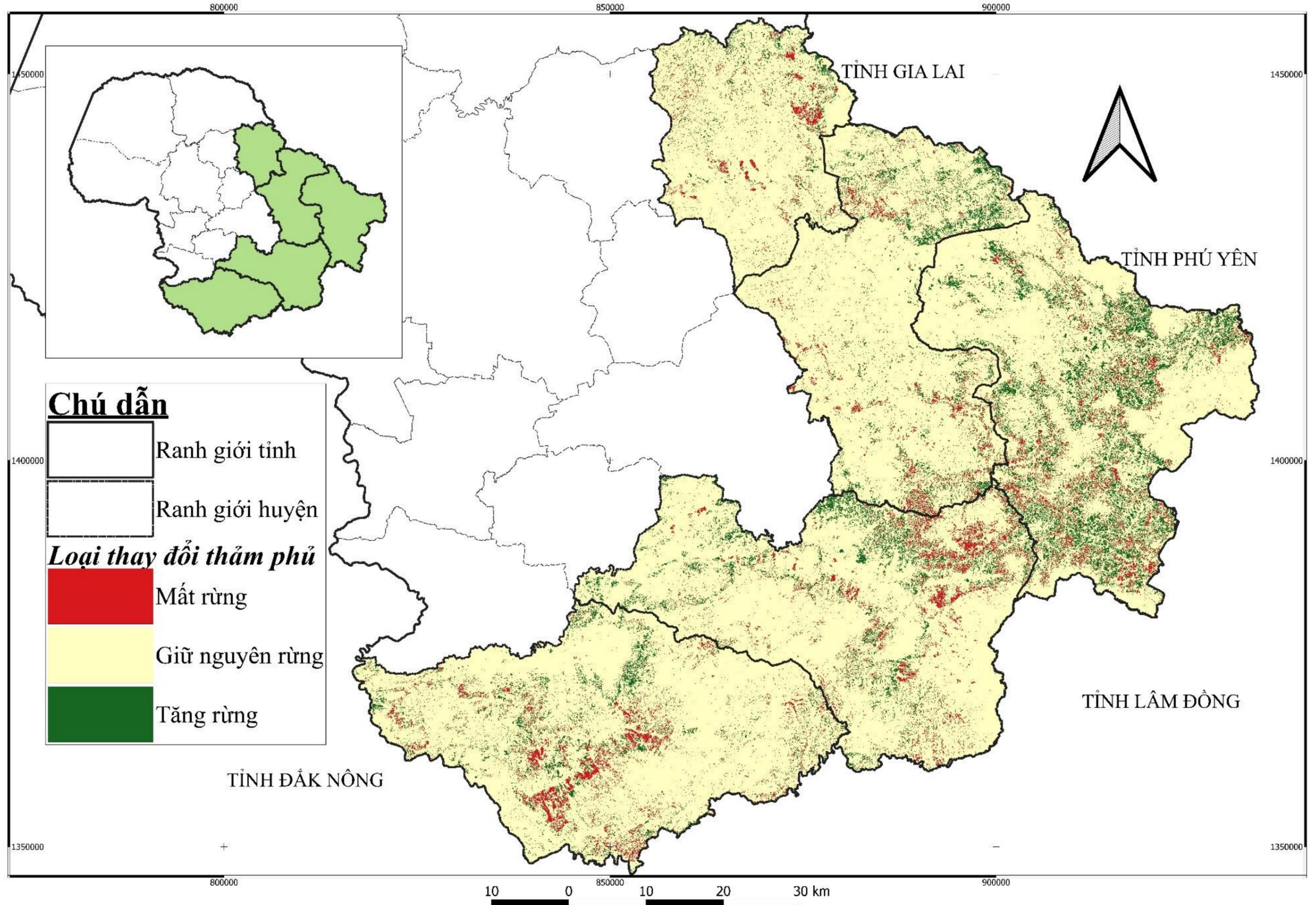
Hình 3.3. Phân bố thảm phủ RTX tỉnh Đắk Lắk giai đoạn 2015 - 2025

Nhìn chung, kết quả phân tích cho thấy rừng RTX tại Đắk Lắk phân bố không đồng đều, tập trung mạnh ở vùng đồi núi phía Đông và Nam, nơi điều kiện tự nhiên thuận lợi và mức độ bảo vệ cao hơn. Các khu vực trung tâm và phía Bắc có độ che phủ thấp hơn, chủ yếu là rừng nghèo hoặc rừng phục hồi. Sự khác biệt này phản ánh mối liên hệ chặt chẽ giữa yếu tố địa hình, khí hậu, mức độ bảo vệ rừng và áp lực sử dụng đất, đồng thời khẳng định vai trò đặc biệt của các huyện Krông Bông, Lắk và M'Đrăk trong việc duy trì quỹ rừng RTX có mức đa dạng sinh học cao của tỉnh.

Rừng thường xanh có cấu trúc sinh thái phức tạp với nhiều tầng cây gỗ, độ che phủ lớn và mức độ đa dạng loài cao. Các đặc điểm này giúp duy trì chỉ số thực vật (NDVI, EVI) ổn định quanh năm, tạo nên tín hiệu quang phổ đặc trưng và khác biệt rõ rệt so với các lớp “Khác” như nương rẫy, đất trống hay cây công nghiệp. Điều này lý giải tại sao mô hình RF đạt hiệu quả cao khi nhận dạng lớp RTX. Phát hiện này cũng tương đồng với nghiên cứu của Nguyễn Thị Thanh Hương và Đoàn Minh Trung (2018), Mallinis và cộng sự (2020), Nguyễn Thị Thanh Hương và cộng sự (2020), trong đó RF cho kết quả phân loại thảm phủ có độ chính xác cao hơn so với các thuật toán phân loại khác như kNN (k-nearest neighbor) hay Maximum likelihood,... đặc biệt là đối với lớp phủ là RTX.

3.1.2. Phân tích thay đổi diện tích RTX giai đoạn 2015 – 2025

Sử dụng các bản đồ phân loại RTX các năm 2015, 2020 và 2025 để phân tích xu hướng thay đổi RTX trên địa bàn tỉnh giai đoạn 2015 – 2025. Kết quả cho thấy xu hướng ổn định và tăng nhẹ ở một số địa phương trong 10 năm qua (Hình 3.4). Phân tích này cung cấp cơ sở tin cậy để phân tích AGB và khả năng hấp thụ CO₂ trong nghiên cứu tiếp theo.



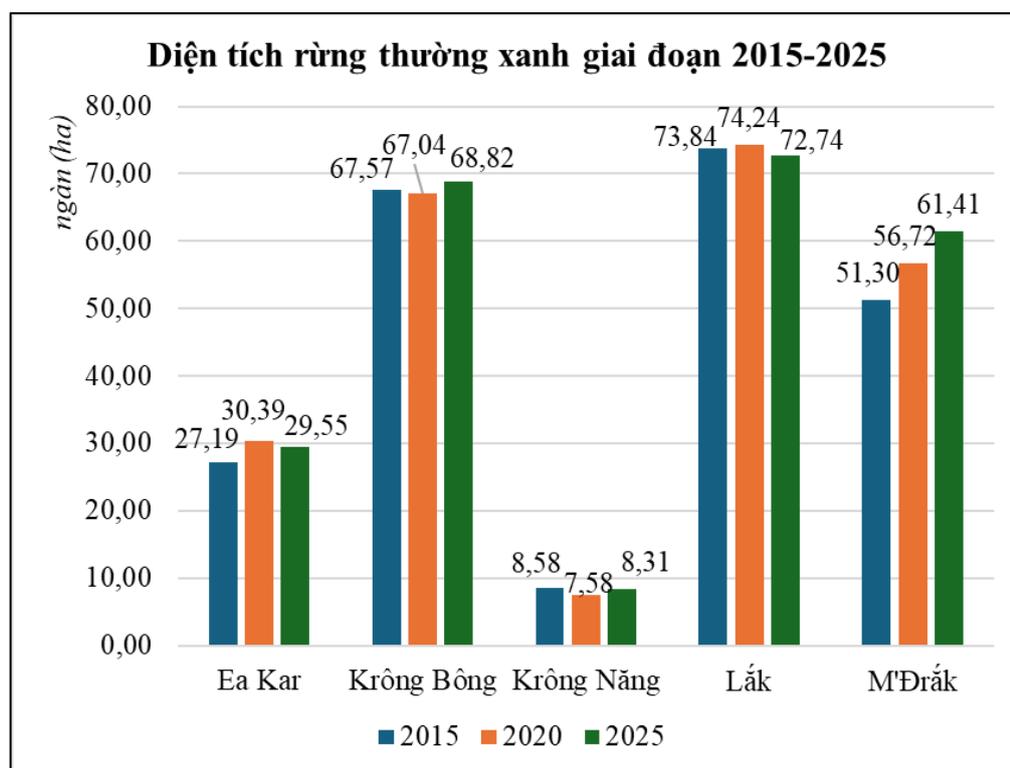
Hình 3.4. Bản đồ thay đổi diện tích rừng thường xanh giai đoạn 2015-2025

Kết quả đánh giá thay đổi diện tích RTX ở các huyện (địa giới hành chính cũ) có diện tích rừng thường xanh lớn trong tỉnh là Lắk, Krông Bông, M'Đrăk, Ea Kar và Krông Năng cho thấy:

- Phần lớn các huyện có diện tích RTX ổn định trong giai đoạn nghiên cứu, một phần do quy định đóng cửa rừng tự nhiên giai đoạn này và kết quả quản lý bảo vệ rừng của các chủ rừng lớn. Đặc biệt là các vườn quốc gia, khu bảo tồn và rừng phòng hộ có diện tích RTX lớn như Chư Yang Sin, Nam Ka, Núi Vọng Phu, Lắk cũng như một số công ty Lâm nghiệp.
- Diện tích tăng rừng thường xanh rải rác và chủ yếu là các khu vực phục hồi xung quanh các rừng đặc dụng đặc biệt là các huyện M'Đrăk và Krông Bông.
- Khu vực giảm RTX nhiều nằm trên địa bàn huyện Krông Năng, Krông Bông và Lắk.

Nhìn chung, diện tích rừng thường xanh tại khu vực nghiên cứu ổn định và có xu hướng tăng nhẹ ở quy mô toàn tỉnh, phản ánh tác động tích cực của chính sách đóng cửa rừng tự nhiên, công tác quản lý bảo vệ rừng và các hoạt động phục hồi rừng, đặc biệt tại các khu rừng đặc dụng và rừng phòng hộ.

Phân tích theo từng huyện có diện tích rừng thường xanh lớn trong tỉnh cũng có sự khác biệt đáng kể về mức độ thay đổi ở các địa phương khác nhau (Hình 3.5). Trong giai đoạn 2015–2025, diện tích rừng thường xanh nhìn chung có xu hướng ổn định, song mức độ biến động khác nhau giữa các địa phương. Huyện Lắk và Krông Bông là hai khu vực có diện tích RTX lớn nhất và tương đối ổn định. Tại huyện Lắk, diện tích RTX tăng từ khoảng 73,84 nghìn ha (2015) lên 74,24 nghìn ha (2020), sau đó giảm nhẹ còn 72,74 nghìn ha (2025), cho thấy xu thế ổn định với biến động nhỏ. Huyện Krông Bông ghi nhận xu hướng tăng nhẹ và liên tục, từ 67,57 nghìn ha (2015) lên 68,82 nghìn ha (2025).



Hình 3.5. Thay đổi diện tích RTX của các huyện giai đoạn 2015-2025

Ngược lại, huyện M'Đrắk thể hiện xu hướng tăng rõ rệt nhất trong giai đoạn nghiên cứu, với diện tích RTX tăng từ 51,30 nghìn ha (2015) lên 61,41 nghìn ha (2025). Điều này phản ánh khả năng phục hồi rừng và/hoặc hiệu quả của các biện pháp quản lý, bảo vệ rừng tại địa phương. Huyện Ea Kar cũng có xu hướng tăng diện tích RTX từ 27,19 nghìn ha lên 29,55 nghìn ha, tuy mức tăng không lớn.

Riêng huyện Krông Năng có diện tích RTX nhỏ nhất trong số các huyện phân tích và biến động không đáng kể, giảm nhẹ giai đoạn 2015–2020 (từ 8,58 xuống 7,58 nghìn ha), sau đó tăng trở lại đạt 8,31 nghìn ha vào năm 2025.

Tổng hợp số liệu cho thấy, trong giai đoạn 2015–2025, tổng diện tích rừng thường xanh của tỉnh Đắk Lắk tăng khoảng 12,35 nghìn ha. Xu thế tăng diện tích tập trung chủ yếu tại huyện M'Đrắk, trong khi các huyện Lắk và Krông Năng ghi nhận suy giảm nhẹ. Điều này cho thấy, mặc dù xu hướng mất rừng đã được kiểm chế ở quy mô toàn tỉnh, song hiện tượng suy giảm rừng vẫn còn diễn ra cục bộ tại một số địa phương. Kết quả này là cơ sở quan trọng cho các phân tích tiếp theo về biến động

sinh khối (AGB) và khả năng hấp thụ CO₂ của rừng thường xanh trong nghiên cứu. Chi tiết thay đổi diện tích RTX các giai đoạn thể hiện ở Bảng 3.2.

Bảng 3.2. Diện tích RTX các huyện giai đoạn 2015-2025 (ha)

Tên huyện	2015	2020	2025	Thay đổi
Ea Kar	27.187,02	30.390,48	29.548,89	2.361,87
Krông Bông	67.570,11	67.041,45	68.824,35	1.254,24
Krông Năng	8.580,24	7.578,00	8.306,37	-273,87
Lắk	73.841,85	74.239,29	72.736,92	-1.104,93
M'Đrắk	51.295,50	56.719,89	61.405,92	10.110,42
Tổng	228.474,72	235.969,11	240.822,45	12.347,73

Nhìn chung, trong giai đoạn này tổng diện tích rừng thường xanh trên địa bàn tỉnh có biến động nhưng không nhiều, xu thế mất rừng chậm lại và xu thế phục hồi nhờ vào công tác phục hồi rừng và quản lý bảo vệ rừng tốt hơn đặc biệt là tại các khu rừng đặc dụng trên địa bàn tỉnh.

3.2. Phân tích mối quan hệ giữa đặc trưng ảnh viễn thám và AGB

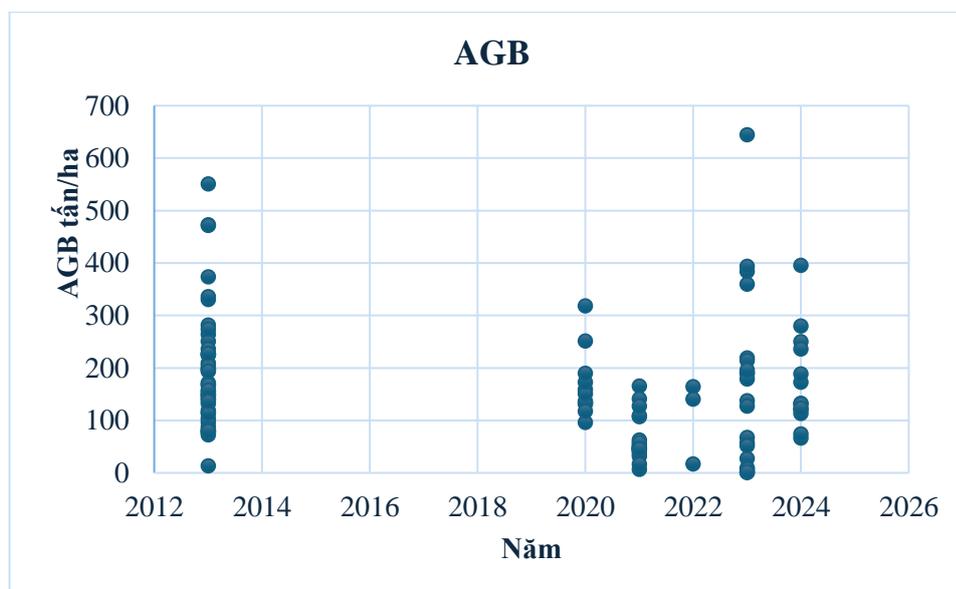
3.2.1. Chuẩn hóa dữ liệu sinh khối ô mẫu

Biểu đồ Hình 3.6 thể hiện phân bố giá trị AGB (đơn vị: tấn/ha) của các ô mẫu điều tra trong giai đoạn 2013 – 2024. Dữ liệu được thu thập tại các thời điểm khảo sát thực địa khác nhau, bao gồm năm 2013, 2020, 2021, 2022, 2023 và 2024.

- Năm 2013 có số lượng ô mẫu nhiều nhất, với AGB trung bình dao động rộng từ khoảng 5 tấn/ha đến trên 500 tấn/ha, trong đó phần lớn tập trung trong khoảng 100 – 300 tấn/ha.
- Năm 2020 có phạm vi biến động AGB hẹp hơn, chủ yếu trong khoảng 50 – 200 tấn/ha, chỉ một số ít ô mẫu đạt trên 300 tấn/ha.
- Giai đoạn 2021 – 2022 có nhiều ô mẫu với AGB thấp (<50 tấn/ha), phản ánh đặc trưng của rừng phục hồi hoặc các lâm phần nghèo kiệt.
- Năm 2023 – 2024 có một số ô mẫu có AGB rất cao (>600 tấn/ha).

Phân bố tổng thể cho thấy các giá trị AGB trong toàn bộ giai đoạn nghiên cứu

chủ yếu nằm trong khoảng 100–250 tấn/ha, với một số ít điểm có giá trị cao hoặc thấp hơn đáng kể. Các kết quả này phản ánh đặc trưng phân bố sinh khối của các ô mẫu đo đạc trong từng thời điểm khảo sát, qua đó cũng phản ánh sự đa dạng về trạng thái rừng, tuổi rừng và điều kiện sinh thái tại khu vực nghiên cứu.



Hình 3.6. Phân bố giá trị AGB của ô mẫu theo năm

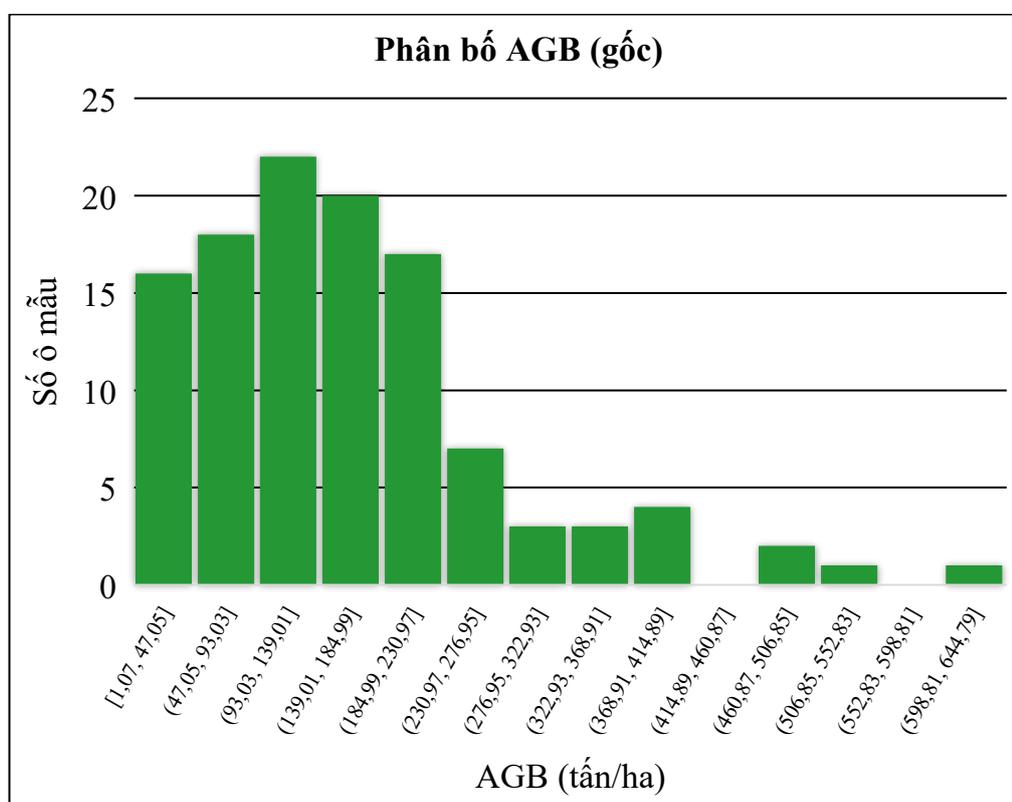
Kết quả thống kê mô tả (Bảng 3.3) cho thấy: giá trị trung bình AGB đạt 159,81 tấn/ha, độ lệch chuẩn 117,28 tấn/ha, phản ánh mức biến động lớn giữa các ô mẫu mẫu do sự khác biệt về trạng thái rừng, độ tán che và cấu trúc lâm phần trong khu vực nghiên cứu. Hệ số lệch (Skewness = 1,43) và độ nhọn (Kurtosis = 2,93) cho thấy phân bố dữ liệu lệch phải rõ rệt, tồn tại một số giá trị ngoại lai cao. Giá trị trung bình 159,8 tấn/ha nằm trong khoảng điển hình của các hệ sinh thái rừng nhiệt đới ẩm, tương tự các kết quả đã được công bố ở Tây Nguyên [2]; Bình Phước [27]; Bắc Kạn [22].

Bảng 3.3. Thống kê mô tả dữ liệu AGB

Chỉ số thống kê	Giá trị
Trung bình Mean	159,8054
Sai tiêu chuẩn (Standard Error)	10,98392
Trung vị (Median)	140,955
Mode	191
Độ lệch chuẩn (SD)	117,2761
Phương sai mẫu (Sample Variance)	13753,69
Độ nhọn (Kurtosis)	2,926741

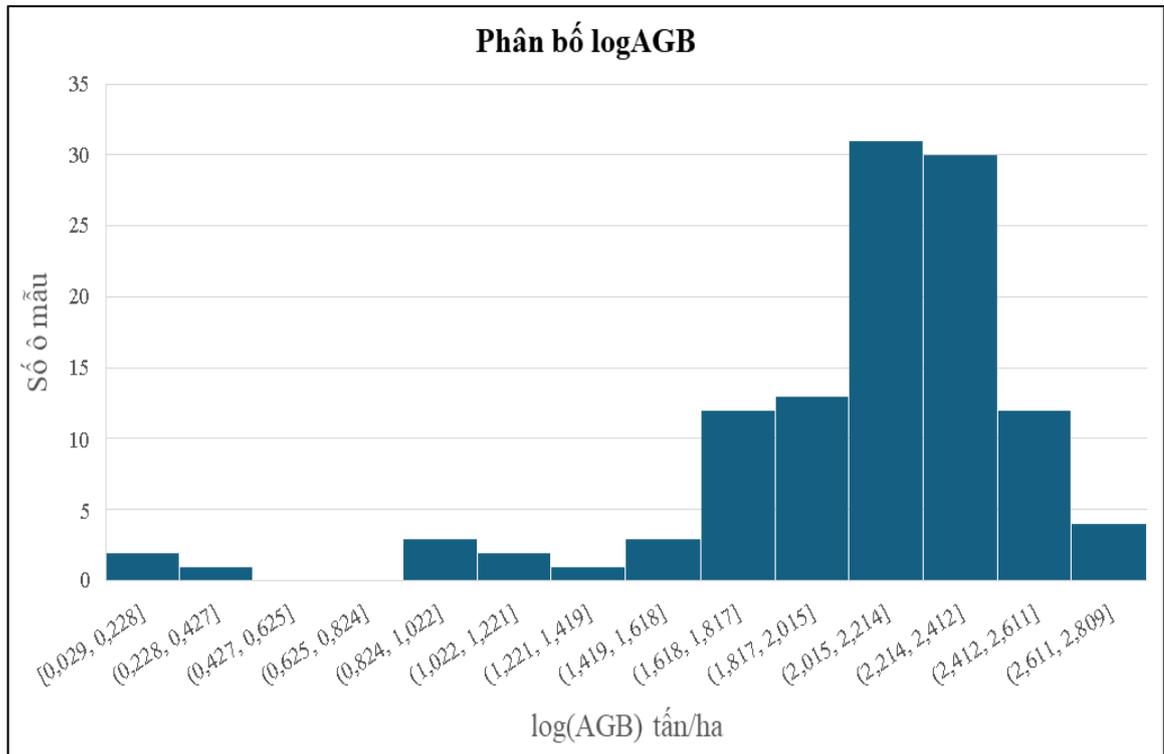
Chỉ số thống kê	Giá trị
Độ lệch (Skewness)	1,425821
Khoảng biến thiên (Range)	643,72
Giá trị nhỏ nhất (Min)	1,07
Giá trị lớn nhất (Max)	644,79
Số mẫu (n)	117
Mức tin cậy 95%	21,76112

Biểu đồ histogram (Hình 3.7) cho thấy phân bố AGB gốc không tuân theo dạng chuẩn, phần đuôi kéo dài về phía giá trị cao, đặc biệt có một số điểm >400 tấn/ha. Điều này có thể ảnh hưởng đến độ ổn định của các mô hình hồi quy tuyến tính.



Hình 3.7. Biểu đồ phân bố AGB ô mẫu (dữ liệu gốc)

Để khắc phục hạn chế trên, phép biến đổi lô-ga-rít được áp dụng cho dữ liệu AGB. Phân bố logAGB (Hình 3.8) cho thấy phần lớn các ô mẫu tập trung trong khoảng log(AGB) từ 1,8 đến 2,6, tương ứng với 63,1 – 398,1 tấn/ha. Việc lô-ga-rít hóa làm giảm biến động dữ liệu và cải thiện sự phù hợp với các giả định thống kê trong mô hình hồi quy.



Hình 3.8. Biểu đồ phân bố AGB ô mẫu sau biến đổi lô-ga-rít

Phép biến đổi lô-ga-rít được áp dụng để khắc phục hiện tượng lệch phải và ổn định phương sai, điều này phù hợp với khuyến nghị của nhiều nghiên cứu mô hình hóa sinh khối rừng bằng hồi quy tuyến tính [15],[95],[130],[172]. Việc lô-ga-rít hóa giúp dữ liệu trở nên gần phân phối chuẩn hơn, từ đó đảm bảo các giả định thống kê cơ bản trong phân tích hồi quy được đáp ứng.

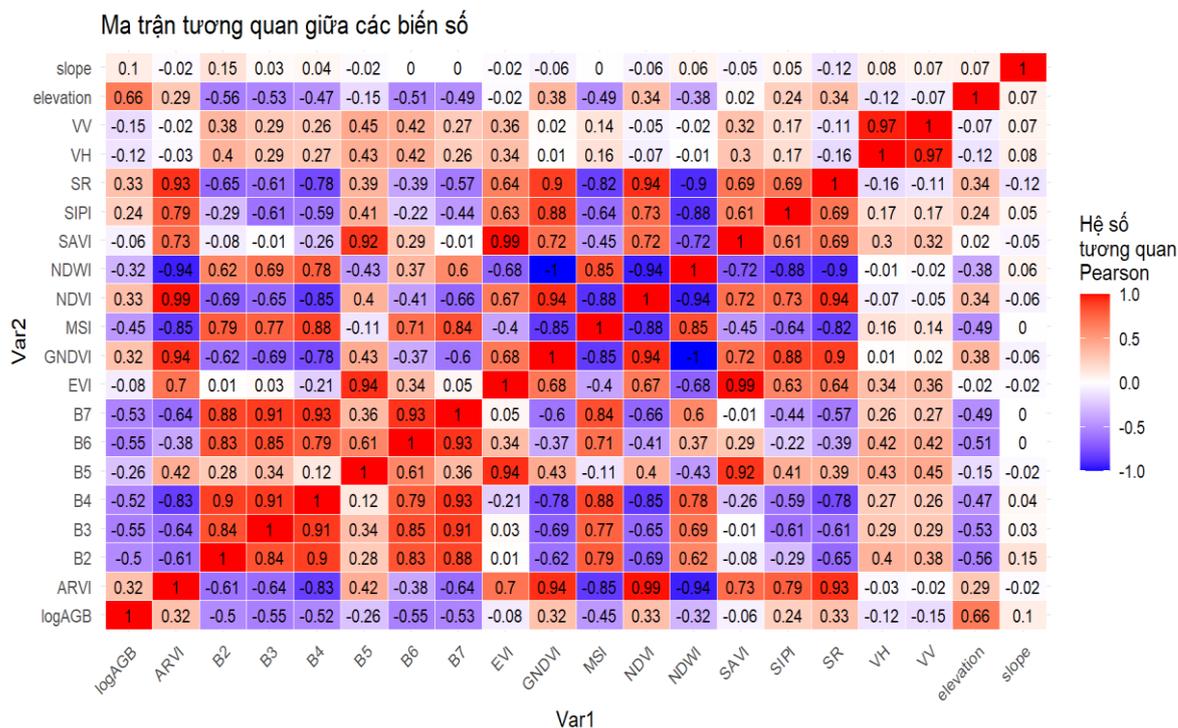
Trong quá trình xây dựng mô hình, logAGB được sử dụng làm biến phụ thuộc. Sau khi dự đoán, kết quả được chuyển ngược về đơn vị gốc (tán/ha) bằng hàm mũ.

Tóm lại, phân tích phân bố cho thấy:

- Dữ liệu AGB gốc biến động lớn và lệch phải mạnh.
- Biến đổi lô-ga-rít giúp giảm ảnh hưởng của ngoại lai, cải thiện tính đối xứng và sự ổn định phương sai.
- Bộ dữ liệu sau biến đổi lô-ga-rít phù hợp hơn để áp dụng các mô hình ước lượng sinh khối từ dữ liệu viễn thám.

3.2.2. Phân tích tương quan Pearson

Ma trận tương quan Pearson (Hình 3.9) phản ánh mối quan hệ giữa sinh khối rừng đã lô-ga-rít hóa (logAGB) và các biến độc lập bao gồm nhóm chỉ số thực vật, chỉ số độ ẩm, các băng phổ Landsat, yếu tố địa hình (độ cao – *elevation*, độ dốc – *slope*) và dữ liệu ra-đa Sentinel 1A (VV, VH).



Hình 3.9. Ma trận tương quan giữa các biến viễn thám và logAGB

Nhìn chung, các biến quang học và chỉ số phổ thể hiện mối tương quan mạnh với nhau, trong khi các biến ra-đa và địa hình có mối tương quan thấp đến trung bình. Giá trị hệ số tương quan (r) dao động trong khoảng từ $-0,94$ đến $+1,00$, phản ánh sự tồn tại của hiện tượng đa cộng tuyến đáng kể, đặc biệt trong nhóm các chỉ số thực vật và các kênh phổ Landsat.

Mối quan hệ giữa logAGB và các biến độc lập thể hiện sự khác biệt rõ rệt giữa các nhóm nhân tố (Bảng 3.4). Trong nhóm biến địa hình, *elevation* có mối tương quan dương rõ rệt với logAGB ($r = 0,66$), phản ánh xu thế sinh khối tăng theo độ cao, đặc biệt ở các khu vực rừng tự nhiên lâu năm có độ che phủ lớn. Ngược lại, *slope* có

tương quan rất yếu ($r = 0,10$), cho thấy độ dốc không phải là yếu tố chi phối đáng kể sự biến động sinh khối ở quy mô toàn tỉnh.

Bảng 3.4. Phân loại nhóm biến theo mối tương quan với logAGB

Nhóm biến	Biến đại diện	Hệ số tương quan (r)	Đặc điểm nổi bật
Địa hình	Elevation, Slope	0,66; 0,10	Elevation có tương quan dương mạnh; slope không đáng kể
Chỉ số thực vật	NDVI, SAVI, GNDVI, SR, ARVI	0,24 – 0,33	Tương quan dương yếu – trung bình; có hiện tượng bão hòa ở vùng rừng dày
Chỉ số độ ẩm	NDWI, MSI	-0,32; -0,45	Tương quan âm vừa; phản ánh ảnh hưởng độ ẩm tán và đất
Băng phổ Landsat	B3, B4, B5, B6, B7	-0,26 : -0,55	Quan hệ nghịch mạnh; B3 và B6 nhạy với biến động sinh khối
Ra-đa Sentinel-1	VH, VV	-0,12; -0,15	Tương quan rất yếu; tín hiệu bị bão hòa ở vùng rừng dày
Địa hình tổng hợp	Elevation, Slope	0,32; 0,10	Ảnh hưởng tích cực đến sinh khối do phân bố rừng theo độ cao

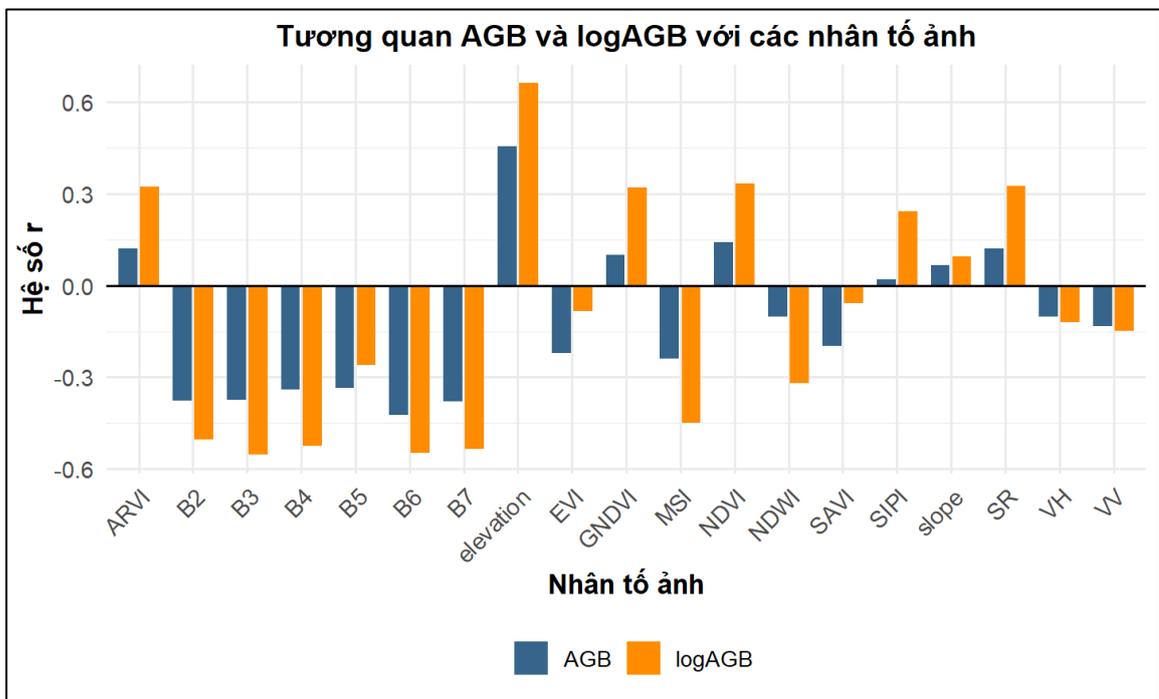
Đối với dữ liệu quang học Landsat, các kênh phổ *B3 (Green)*, *B4 (Red)*, *B5 (NIR)* và *B6 (SWIR1)* thể hiện mối tương quan âm khá mạnh với logAGB (r từ -0,26 đến -0,55). Trong đó, *B3* và *B6* có hệ số tương quan âm lớn nhất ($r = -0,55$), phản ánh sự nghịch biến giữa cường độ phản xạ phổ vùng đỏ – hồng ngoại sóng ngắn và mật độ tán rừng: khi sinh khối tăng, tán rừng hấp thụ nhiều năng lượng hơn, làm giảm phản xạ phổ ghi nhận trên ảnh vệ tinh.

Nhóm chỉ số thực vật gồm *NDVI*, *SAVI*, *GNDVI* và *SR* có mối tương quan dương chặt chẽ giữa chúng ($r > 0,85$), phản ánh tính trùng lặp cao về thông tin quang phổ. Tuy nhiên, mối tương quan của các chỉ số này với logAGB chỉ ở mức yếu đến trung bình ($r = 0,24 - 0,33$), cho thấy hiện tượng bão hòa phổ (spectral saturation) tại các vùng rừng có sinh khối cao, khi giá trị chỉ số không còn tăng tỷ lệ thuận với mật độ tán rừng.

Ngược lại, các chỉ số độ ẩm như *NDWI* và *MSI* thể hiện tương quan âm đáng kể với logAGB ($r = -0,32$ và $-0,45$), phản ánh rằng các khu vực có độ ẩm cao thường không tương ứng với sinh khối lớn, do ảnh hưởng của đất trống hoặc thảm thực vật thấp tầng.

Đối với dữ liệu ra-đa Sentinel-1A, hai kênh *VH* và *VV* có tương quan rất yếu với *logAGB* ($r = -0,12$ và $-0,15$). Điều này cho thấy tín hiệu tán xạ ra-đa bị bão hòa ở rừng dày, khiến khả năng phản ánh biến động sinh khối bị giới hạn trong vùng giá trị cao.

Để làm rõ hơn mối tương quan giữa sinh khối rừng và các biến ảnh, biểu đồ so sánh hệ số tương quan giữa *AGB* và *logAGB* (Hình 3.10) cho thấy xu thế biến động tương đồng, tuy nhiên hệ số r của *logAGB* có độ lớn và tính ổn định cao hơn, phản ánh mối quan hệ tuyến tính được cải thiện sau khi biến đổi lô-ga-rít dữ liệu sinh khối.



Hình 3.10. So sánh tương quan AGB và logAGB với các nhân tố ảnh

Việc biến đổi lô-ga-rít không chỉ giúp tăng cường mức độ tương quan với các biến quang học và chỉ số phổ (đặc biệt là NDVI, GNDVI, SR, ARVI), mà còn làm nổi bật hơn mối quan hệ nghịch với các băng phổ vùng SWIR (B6, B7). Kết quả này cho thấy việc sử dụng *logAGB* giúp nâng cao độ tin cậy và tính ổn định của mô hình ước lượng sinh khối từ dữ liệu viễn thám.

Các phát hiện này phù hợp với nhiều nghiên cứu trước đây, Morel và cộng sự (2012) và Singh và cộng sự (2014) đã chỉ ra rằng các chỉ số thực vật như NDVI và EVI

thường bị bão hòa ở vùng rừng có độ che phủ cao, dẫn đến mối tương quan yếu với AGB. Tương tự, Lu và cộng sự (2012) và Næsset và cộng sự (2013) nhận thấy việc áp dụng lô-ga-rít hóa giúp cải thiện độ phù hợp của các mô hình hồi quy tuyến tính khi ước lượng sinh khối rừng.

Đối với dữ liệu ra-đa, hệ số tương quan thấp của VH và VV phù hợp với kết quả nghiên cứu của Omar và cộng sự (2017) tại rừng mưa Malaysia, nơi tín hiệu tán xạ C-band bị giới hạn trong vùng sinh khối cao. Tuy nhiên, các nghiên cứu gần đây đều cho rằng việc kết hợp dữ liệu ra-đa và quang học có thể cải thiện đáng kể độ chính xác của mô hình ước lượng sinh khối [25],[55],[179],[188].

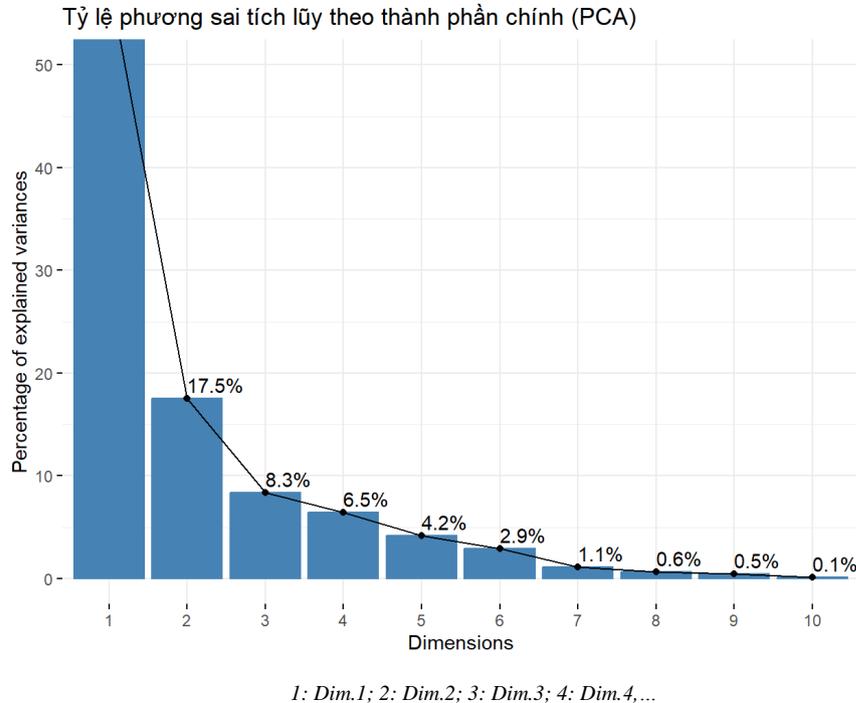
Tổng hợp lại, phân tích tương quan Pearson cho thấy dữ liệu AGB gốc có mối quan hệ yếu với các chỉ số viễn thám do phân bố lệch và hiệu ứng bão hòa tín hiệu, trong khi phép biến đổi lô-ga-rít giúp cải thiện đáng kể tương quan, đặc biệt với các chỉ số NDVI, GNDVI và SR. Kết quả này khẳng định tính cần thiết của việc chuẩn hóa logAGB trong quá trình xây dựng mô hình hồi quy ước lượng sinh khối rừng từ dữ liệu viễn thám đa nguồn. Để giảm bớt số lượng biến tham gia vào mô hình và giảm sai số thì PCA được sử dụng để lựa chọn các biến chính có khả năng giải thích tốt dữ liệu phân bố của logAGB, phục vụ cho các phân tích tương quan và hồi quy.

3.2.3. Phân tích thành phần chính (PCA)

Để giảm bớt các biến không hoặc ít đóng góp vào tương quan với AGB, sử dụng phân tích PCA để lựa chọn các biến đóng góp chính vào tương quan với AGB. Phân tích PCA được sử dụng nhằm giảm số lượng biến đầu vào, loại bỏ hiện tượng đa cộng tuyến và nhận diện các biến có mức đóng góp cao nhất trong việc giải thích biến thiên AGB. Kết quả phân tích PCA (Hình 3.11) cho thấy:

- Thành phần chính thứ nhất (Dim.1) giải thích 58,3% phương sai.
- Thành phần chính thứ hai (Dim.2) giải thích thêm 17,5% phương sai.
- Tổng cộng hai thành phần đầu tiên giải thích hơn 75% thông tin của toàn bộ tập dữ liệu.

- Khi bổ sung thành phần chính thứ ba (Dim.3), tổng phương sai giải thích đạt 84,1%, cho thấy PCA phản ánh tốt cấu trúc dữ liệu.



Hình 3.11. Tỷ lệ phương sai giải thích của các thành phần chính

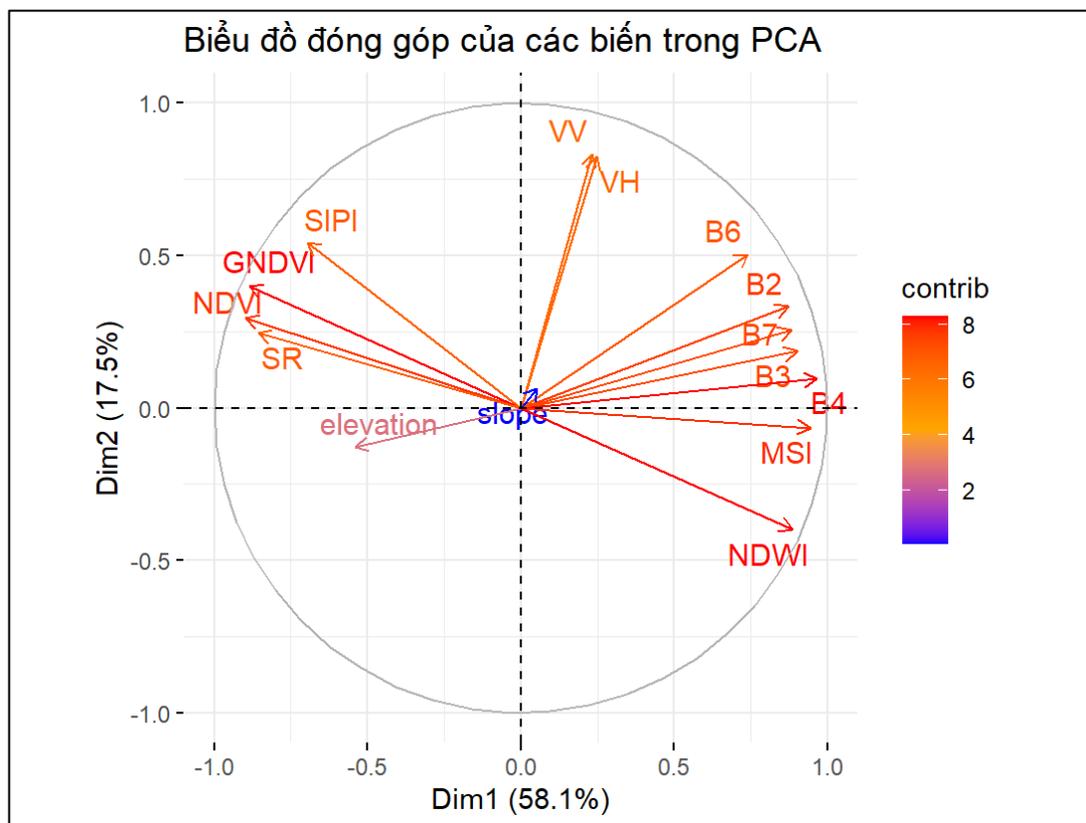
Phân tích biểu đồ biplot (Hình 3.12) cho thấy các biến được chia thành ba nhóm rõ rệt:

- Nhóm 1 – Các chỉ số thực vật (*NDVI*, *GNDVI*, *SR*, *SIP1*): Các vector của nhóm này có hướng và chiều tương đồng, thể hiện mối tương quan dương chặt chẽ, phản ánh khả năng mô tả mức độ xanh và mật độ tán rừng. Đây là nhóm có ý nghĩa sinh thái cao trong phản ánh hàm lượng diệp lục và cường độ quang hợp. Tuy nhiên, do tương quan nội tại mạnh ($r > 0,85$), chỉ nên chọn một hoặc hai chỉ số đại diện như *NDVI* và *GNDVI* để giảm đa cộng tuyến trong mô hình hồi quy [121].
- Nhóm 2 – Các kênh phổ quang học (*B2*, *B3*, *B4*, *B5*, *B6*, *B7*) và chỉ số độ ẩm (*NDWI*, *MSI*): Các biến này tập trung chủ yếu dọc theo trục Dim.1, phản ánh sự khác biệt về phản xạ phổ giữa các vùng đất trống, cây trồng thấp tầng và rừng có mật độ cao. Trong nhóm này, các kênh SWIR (*B6*, *B7*) cùng

NDWI và *MSI* có độ đóng góp cao nhất (màu đỏ, *contrib* > 6), cho thấy vai trò nổi bật trong phản ánh đặc tính độ ẩm và cấu trúc tán lá. Kết quả phù hợp với nhận định của Foody và cộng sự (2003), theo đó vùng phổ SWIR có độ nhạy cao đối với hàm lượng nước và vật chất khô của thực vật.

- Nhóm 3 – Các biến ra-đa (*VV*, *VH*): Hai biến này định hướng gần như vuông góc với các nhóm quang học, phản ánh nguồn thông tin độc lập. Ra-đa nhạy với đặc tính cấu trúc 3D của tán rừng và độ gồ ghề bề mặt [122],[132]. Sự tách biệt rõ ràng trên đồ thị PCA cho thấy việc kết hợp dữ liệu ra-đa và quang học là cần thiết để nâng cao khả năng mô hình hóa sinh khối rừng.

Ngoài ra, hai biến *elevation* và *slope* có vector ngắn và định hướng khác biệt, cho thấy ảnh hưởng địa hình tuy không mạnh nhưng mang tính bổ sung thông tin quan trọng, đặc biệt ở các vùng rừng tự nhiên phân bố theo đai cao.



Hình 3.12. Biểu đồ phân tích mức đóng góp của các biến

Từ kết quả phân tích trên, để vừa giảm thiểu hiện tượng đa cộng tuyến vừa đảm bảo giữ lại thông tin quan trọng, các biến được lựa chọn đưa vào phân tích tiếp theo gồm:

- Nhóm chỉ số thực vật: *NDVI*, *GNDVI*
- Chỉ số độ ẩm và phổ SWIR: *NDWI*, *MSI*, *B6*, *B7*
- Yếu tố địa hình: *elevation*, *slope*
- Ra-đa: *VV*, *VH*

Các biến này được xem là có mức đóng góp cao nhất vào tổng phương sai và phản ánh đa chiều các đặc tính quang học – cấu trúc – địa hình của thảm phủ rừng.

Đặc biệt, sự tách biệt của nhóm biến ra-đa (*VV*, *VH*) và địa hình (*elevation*, *slope*) trong không gian PCA chứng tỏ thông tin từ sóng ra-đa và địa hình phản ánh cấu trúc tán và độ gồ ghề bề mặt – những đặc trưng không thể quan sát được bằng cảm biến quang học [122],[132],[149]. Kết quả này cũng có hướng tiếp cận tích hợp quang học – ra-đa nhằm nâng cao độ chính xác trong ước tính AGB.

Việc áp dụng PCA không chỉ giúp loại bỏ hiện tượng đa cộng tuyến giữa các chỉ số quang học vốn có tính tương quan cao (ví dụ *NDVI*, *SAVI*, *SR*) mà còn nâng cao tính ổn định của mô hình học máy thông qua việc giảm số chiều dữ liệu đầu vào [80],[99]. Hơn nữa, PCA cho phép tích hợp thông tin từ nhiều nguồn dữ liệu viễn thám (quang học – ra-đa – địa hình) trong một không gian đặc trưng tối ưu, qua đó cải thiện khả năng giải thích biến động sinh khối rừng [181].

Việc PCA xác định được các nhóm biến có tính đại diện cao đồng thời cung cấp nền tảng cho các mô hình hồi quy và thuật toán học máy (như RF, SVM). Theo Han và Joe (2024) việc giảm số chiều đầu vào thông qua PCA không chỉ cải thiện tốc độ xử lý mà còn giúp mô hình học máy ổn định hơn, tránh hiện tượng quá khớp khi dữ liệu có tương quan cao giữa các biến. Như vậy, PCA không chỉ đóng vai trò tiền xử lý dữ liệu mà còn góp phần quan trọng trong việc tối ưu hóa mô hình ước tính sinh khối rừng, đặc biệt trong bối cảnh kết hợp đa nguồn dữ liệu (quang học – ra-đa).

Nhìn chung, kết quả phân tích dữ liệu mẫu, kiểm tra phân bố, biến đổi lô-garit, kiểm định tương quan và PCA đã thiết lập được cơ sở khoa học quan trọng cho việc xây dựng các mô hình ước tính sinh khối rừng từ dữ liệu viễn thám. Đây là bước chuẩn bị cần thiết để tiến hành các phân tích hồi quy và ứng dụng các thuật toán học máy Random Forest trong các phân tiếp theo của nghiên cứu.

3.2.4. Phân tích đa cộng tuyến (VIF - Variance Inflation Factor)

Sau khi xác định các nhóm biến có ý nghĩa thông qua phân tích PCA, bước tiếp theo nhằm đảm bảo tính ổn định và hiệu quả của mô hình là kiểm tra hiện tượng đa cộng tuyến (multicollinearity) giữa các biến độc lập. Hiện tượng đa cộng tuyến xảy ra khi hai hoặc nhiều biến giải thích có mối tương quan tuyến tính mạnh, dẫn đến ước lượng hệ số hồi quy không ổn định, làm tăng phương sai của sai số chuẩn và giảm độ tin cậy của mô hình [59].

Kiểm tra hệ số VIF cho các biến được lựa chọn sau phân tích PCA, bao gồm nhóm biến quang học (NDVI, GNDVI, NDWI, MSI, B6, B7), ra-đa (VV, VH) và địa hình (elevation, slope).

Kết quả kiểm tra và lựa chọn VIF (Bảng 3.5) cho thấy hầu hết các biến có giá trị VIF thấp, nằm trong giới hạn chấp nhận được ($VIF = 1,01 - 9,68$), chứng tỏ mức độ phụ thuộc giữa các biến độc lập không cao. Tuy nhiên, biến MSI có $VIF = 9,68$, tiến sát ngưỡng cảnh báo, cho thấy biến này có thể chia sẻ thông tin tương đối lớn với NDWI và B7 - các chỉ số đều phản ánh đặc tính độ ẩm và cấu trúc tán rừng ở vùng phổ SWIR.

Bảng 3.5. Các biến lựa chọn có VIF thấp

Biến	B7	MSI	NDWI	VH	elevation	slope
VIF	4,36	9,68	4,47	1,14	1,37	1,01

Ngược lại, các biến còn lại như B7 ($VIF = 4,36$) và NDWI ($VIF = 4,47$) đều có mức tương quan trung bình và mang thông tin bổ sung về đặc tính quang học – độ ẩm tán lá. Hai biến địa hình *elevation* ($VIF = 1,37$) và *slope* ($VIF = 1,01$) có giá trị rất thấp, thể hiện tính độc lập cao, phản ánh điều kiện sinh thái của rừng theo độ cao

và địa hình. Biến ra-đa *VH* ($VIF = 1,14$) cũng cho thấy tính độc lập rõ rệt, minh chứng rằng thông tin phản xạ từ ra-đa không bị trùng lặp với các biến quang học, phù hợp với đặc tính phản ánh cấu trúc tán rừng và độ gồ ghề bề mặt [122],[132]. Sử dụng các biến này để thực hiện phân tích hồi quy tuyến tính đa biến, phi tuyến tính đa biến và RF cho dữ liệu sinh khối rừng để lựa chọn mô hình hồi quy phù hợp với dữ liệu khu vực nghiên cứu.

Tóm lại, kết quả phân tích thống kê, tương quan, PCA và kiểm tra VIF cho thấy bộ dữ liệu viễn thám và địa hình có mối quan hệ chặt chẽ với AGB và có thể sử dụng hiệu quả trong mô hình hóa. Việc biến đổi $\log AGB$ giúp ổn định phương sai và cải thiện tính tuyến tính của dữ liệu, đáp ứng giả định của mô hình hồi quy. Các biến đầu vào được lựa chọn cuối cùng bao gồm các đặc trưng quang học, chỉ số thực vật, thông tin địa hình và kênh ra-đa, đảm bảo tính đa dạng thông tin phổ – sinh thái và hạn chế hiện tượng đa cộng tuyến.

Bộ biến này được sử dụng thống nhất cho ba mô hình ước lượng gồm hồi quy tuyến tính đa biến (OLS), mô hình phi tuyến tổng quát (GAM) và mô hình học máy RF. Cách tiếp cận đồng bộ này giúp so sánh hiệu năng giữa các thuật toán trên cùng cơ sở dữ liệu, đồng thời kiểm chứng khả năng mô hình hóa các mối quan hệ tuyến tính, phi tuyến và phi tham số trong ước lượng sinh khối rừng.

3.3. Xây dựng mô hình ước tính AGB từ dữ liệu viễn thám

3.3.1. Ước tính AGB sử dụng mô hình hồi quy tuyến tính

3.3.1.1. Hồi quy tuyến tính đơn biến

Từ các biến đầu vào đã lựa chọn ở trên, để xác định mối quan hệ giữa các chỉ số viễn thám và sinh khối rừng, nghiên cứu tiến hành phân tích hồi quy đơn biến với biến phụ thuộc đã được chuyển đổi $\log AGB$ (tấn/ha). Các biến độc lập bao gồm chỉ số thực vật (*MSI*, *NDWI*), kênh phổ Landsat (*B7 – SWIR2*), cùng dữ liệu ra-đa Sentinel-1 (*VH*) và dữ liệu địa hình (*elevation*, *slope*).

Kết quả phân tích hồi quy tuyến tính đơn biến (Bảng 3.6) cho thấy có sự khác

hầu hết các biến qua phân tích chọn lọc biến đều có mối tương quan có ý nghĩa thống kê với logAGB.

Bảng 3.6. Kết quả hồi quy đơn biến giữa logAGB và các biến viễn thám

Biến	R ²	P-value
B7	0,2852	1,922e-06
MSI	0,2016	9,671e-05
NDWI	0,1025	0,0069
VH	0,0145	-1,845e-05
Elevation	0,4386	4,271e-10
Slope	0,0092	0,4297

Kết quả cho thấy biến độ cao địa hình (*elevation*) có hệ số xác định cao nhất trong ($R^2 = 0,4386$), thể hiện mối quan hệ chặt và có ý nghĩa thống kê rất cao ($p < 0,001$) với logAGB. Kết quả này phản ánh xu hướng sinh khối rừng tăng dần theo độ cao, phù hợp với đặc điểm phân bố rừng tự nhiên ở khu vực Đắc Lắc, nơi các khu vực địa hình núi cao, có độ che phủ rừng lớn hơn và ít chịu tác động của hoạt động nhân sinh.

Trong nhóm biến quang học, kênh phổ hồng ngoại sóng ngắn B7 thể hiện mối quan hệ mạnh và có ý nghĩa cao với logAGB ($R^2 = 0,2852$; $p < 0,001$). Kết quả này khẳng định vai trò của vùng phổ SWIR trong phản ánh cấu trúc tán rừng, độ ẩm thực vật và mật độ sinh khối. Chỉ số độ ẩm MSI cũng thể hiện mối tương quan đáng kể ($R^2 = 0,2016$; $p < 0,001$), cho thấy khả năng phản ánh sự thay đổi hàm lượng nước trong tán lá và đất – yếu tố có quan hệ chặt chẽ với trữ lượng sinh khối rừng.

Chỉ số NDWI có mối quan hệ yếu hơn nhưng vẫn đạt ý nghĩa thống kê ($R^2 = 0,1025$; $p < 0,01$), phản ánh rằng độ ẩm bề mặt và mật độ tán lá vẫn góp phần nhất định vào việc giải thích sự biến thiên sinh khối. Biến ra-đa (VH) có tương quan thấp với logAGB tuy nhiên vẫn có ý nghĩa thống kê. Ngược lại, độ dốc (*slope*) cho giá trị R^2 rất thấp (0,0092) và không có ý nghĩa thống kê ($p = 0,4297 > 0,05$), cho thấy tác động của chúng đến sinh khối rừng là không đáng kể khi xét riêng biệt. Đặc biệt, tín hiệu ra-đa VH (band C) bị ảnh hưởng bởi hiện tượng bão hòa tán xạ ở các khu vực rừng dày, làm giảm khả năng phản ánh chính xác sự thay đổi sinh khối [146].

Tổng hợp lại, các biến có ảnh hưởng đáng kể và có ý nghĩa thống kê cao gồm *elevation*, *B7*, *MSI* và *NDWI*, là những biến có tiềm năng cao để đưa vào mô hình hồi quy đa biến hoặc thuật toán học máy trong các bước tiếp theo. Kết quả này cũng phù hợp với nhận định của Lu và cộng sự (2012) và Zhang và cộng sự (2019), khi các tác giả nhấn mạnh rằng việc kết hợp thông tin độ cao, độ ẩm và vùng phổ SWIR có thể cải thiện đáng kể khả năng ước lượng sinh khối rừng từ dữ liệu viễn thám đa nguồn. Ngoài ra, theo tiêu chuẩn phân loại mức độ tương quan của Cohen (1988), các mô hình logAGB có $|r|$ tương ứng với R^2 từ 0,04 đến 0,12, tương đương mức tương quan yếu đến trung bình, song đủ để phản ánh xu hướng biến động của sinh khối theo biến quang học trong điều kiện rừng nhiệt đới phức tạp.

Trong số các biến viễn thám được kiểm tra, *elevation* và *B7* (SWIR2) cho thấy mối tương quan chặt chẽ và có ý nghĩa thống kê với logAGB. Điều này là do kênh phổ *B7* là kênh phổ hồng ngoại sóng ngắn, nhạy cảm với hàm lượng nước trong thực vật, độ ẩm đất và cấu trúc tán lá (Foody và cộng sự, 2003; Lu và cộng sự, 2014). Rừng có sinh khối cao thường có mật độ tán dày, chứa nhiều nước và vật chất hữu cơ, dẫn đến sự thay đổi rõ rệt trong phản xạ SWIR2 [54],[128]. Do đó, *B7* thể hiện khả năng phân biệt tốt hơn giữa các trạng thái sinh khối khác nhau.

Kết quả phân tích hồi quy tuyến tính đơn biến cho thấy rằng việc kết hợp các kênh phổ vùng đặc biệt là *B7*, độ cao địa hình và chỉ số *NDWI* có ý nghĩa quan trọng trong xây dựng mô hình ước tính sinh khối rừng, và có tiềm năng cao hơn so với các chỉ số phổ cận hồng ngoại hoặc chỉ số thực vật truyền thống. Đây là những biến tiềm năng có thể được sử dụng trong các mô hình hồi quy đa biến hoặc mô hình học máy để ước tính sinh khối rừng chính xác hơn.

3.3.1.2. Hồi quy tuyến tính đa biến

Sau khi thực hiện kiểm tra tương quan Pearson, phân tích PCA và loại bỏ các biến gây đa cộng tuyến (mục 3.2), mô hình hồi quy tuyến tính đa biến được xây dựng với biến phụ thuộc là logAGB và sáu biến độc lập gồm *elevation*, *slope*, *B7*, *MSI*, *NDWI*, và *VH*.

Kết quả phân tích cho thấy mô hình hồi quy cho kết quả cải thiện hơn so với phân tích hồi quy tuyến tính đơn biến. Mô hình cho hệ số xác định $R^2 = 0,506$, R^2 hiệu chỉnh = 0,459. Với các thông số thống kê được trình bày ở Bảng 3.7.

Bảng 3.7. Kết quả phân tích hồi quy tuyến tính đa nhân tố với logAGB

	Estimate	Std. Error	t_value	Pr(> t)
(Intercept)	2,577	4,365	0,590	0,557
B7	-23,683	10,538	-2,247	0,028*
NDWI	-0,702	4,443	-0,158	0,875
VH	0,101	0,567	0,179	0,859
Elevation	2,652	0,512	5,183	2,45e-06***
Slope	0,289	0,465	0,623	0,536
MSI	2,318	3,367	0,689	0,494

Ghi chú: *: $p < 0,05$ **: $p < 0,01$ ***: $p < 0,001$

Kết quả mô hình hồi quy đa nhân tố với logAGB cho thấy mức độ giải thích của mô hình được cải thiện đáng kể so với mô hình sử dụng đơn biến (R^2 tăng từ 0,44 (*elevation*) lên 0,51). Điều này chứng tỏ các biến khác ngoài *elevation* đã giúp tăng cường tính tuyến tính và giảm phương sai của sai số, từ đó mô hình trở nên ổn định và phù hợp hơn.

Trong các biến độc lập, hai biến có ý nghĩa thống kê cao là:

- *elevation* (Estimate = 2,65; $p < 0,001$) – thể hiện ảnh hưởng dương mạnh đến logAGB. Điều này phù hợp với thực tế khi các khu vực có độ cao lớn hơn (đặc biệt vùng núi phía Nam và Đông Nam tỉnh) thường có rừng tự nhiên lâu năm, tán dày và trữ lượng sinh khối cao hơn.
- *B7* (Estimate = -23,68; $p < 0,05$) – có ảnh hưởng nghịch đến logAGB. *B7* phản ánh phản xạ vùng hồng ngoại sóng ngắn, nhạy với độ ẩm và cấu trúc tán lá; khi sinh khối tăng, năng lượng phản xạ giảm mạnh do tán hấp thụ nhiều bức xạ, dẫn đến mối quan hệ âm rõ rệt.

Các biến còn lại như *NDWI*, *MSI*, *VH* và *slope* không thể hiện ý nghĩa thống kê rõ ràng ($p > 0,1$), cho thấy ảnh hưởng riêng lẻ của chúng đến sinh khối rừng không mạnh, hoặc đã bị bao hàm trong các biến quang học và địa hình khác.

Nghiên cứu tiếp tục lựa chọn 2 biến độc lập có tương quan cao với logAGB

là *elevation* và *B7* để thử nghiệm phân tích hồi quy với mô hình đơn giản (2 biến). Kết quả cho thấy mô hình tuyến tính đa biến đơn giản có dạng:

$$\log AGB = 3,9214 - 15,6052 * B7 + 2,6016 * elevation \quad (3.1)$$

Mô hình có hệ số xác định $R^2 = 0,4954$, R^2 hiệu chỉnh = 0,4803, $F(2,67) = 32,89$ và $p < 0,001$; Sai số chuẩn của phần dư (Residual SE) = 0,8506. Kết quả này cho thấy mô hình giải thích được khoảng 49,5% biến thiên của $\log AGB$, một mức độ phù hợp trung bình thấp đối với mô hình tuyến tính trong nghiên cứu viễn thám sinh khối. Cả hai biến độc lập đều có ý nghĩa thống kê cao ($p < 0,01$), chứng minh rằng chúng đóng vai trò quan trọng trong việc giải thích sự biến động của sinh khối rừng.

So với mô hình đơn biến ($R^2 \approx 0,18$), mô hình đa biến đã cải thiện đáng kể khả năng ước lượng và giảm sai số phần dư ($R^2 = 0,51$ đối với mô hình 6 biến và $R^2 = 0,4954$ với mô hình đơn giản với 2 biến) cho thấy hiệu quả khi kết hợp dữ liệu quang học và ra-đa trong ước tính sinh khối bằng mô hình tuyến tính đa biến cũng như mô hình đa biến đơn giản cho hiệu quả ước lượng tốt hơn hẳn so với đơn biến. Mặc dù hệ số xác định chưa cao, kết quả này được xem là mức chấp nhận được trong các nghiên cứu viễn thám sinh khối rừng nhiệt đới, vốn chịu ảnh hưởng bởi đặc tính phi tuyến và độ phức tạp của cấu trúc tán [137],[178].

Như vậy, mô hình rút gọn đã loại bỏ hiện tượng đa cộng tuyến, đồng thời sử dụng các biến có ý nghĩa sinh học và thống kê trong ước lượng sinh khối rừng. Kết quả này phù hợp với xu hướng chung của các nghiên cứu gần đây, cho rằng việc kết hợp dữ liệu quang học (*NDWI*, *SWIR*, *MSI*) và ra-đa (*VH*) giúp cải thiện đáng kể độ chính xác trong ước lượng sinh khối [27],[156].

Với mong đợi nâng cao độ chính xác ước lượng, nghiên cứu tiếp tục áp dụng mô hình hồi quy phi tuyến và thuật toán học máy RF. Các phương pháp này cho phép mô phỏng các mối quan hệ phi tuyến phức tạp, tự động nhận diện tương tác giữa các biến độc lập và hạn chế ảnh hưởng của đa cộng tuyến. Việc so sánh hiệu suất giữa hồi quy tuyến tính, hồi quy phi tuyến và RF giúp đánh giá toàn diện hơn khả năng

ước tính sinh khối rừng từ dữ liệu viễn thám đa nguồn, từ đó xác định mô hình tối ưu cho giai đoạn ước lượng và lập bản đồ AGB.

3.3.2. Ước tính AGB rừng sử dụng mô hình hồi quy phi tuyến

Để khắc phục những hạn chế của mô hình hồi quy tuyến tính (R^2 hiệu chỉnh thấp và hiện tượng đa cộng tuyến giữa các biến quang học), nghiên cứu tiếp tục áp dụng mô hình hồi quy phi tuyến bằng phương pháp Mô hình cộng tuyến tổng quát - GAM. Ưu điểm của GAM là cho phép mô tả linh hoạt mối quan hệ phi tuyến giữa sinh khối rừng ($\log\text{AGB}$) và các biến độc lập từ ảnh viễn thám thông qua các hàm tron. Do mô hình hồi quy GAM phản ánh cộng tuyến tính của các biến riêng lẻ và không bị ảnh hưởng nghiêm trọng bởi đa cộng tuyến nên nghiên cứu thử nghiệm xây dựng mô hình hồi quy phi tuyến giữa $\log\text{AGB}$ với toàn bộ các biến đầu vào (gam1) và với các biến chọn lọc (gam2).

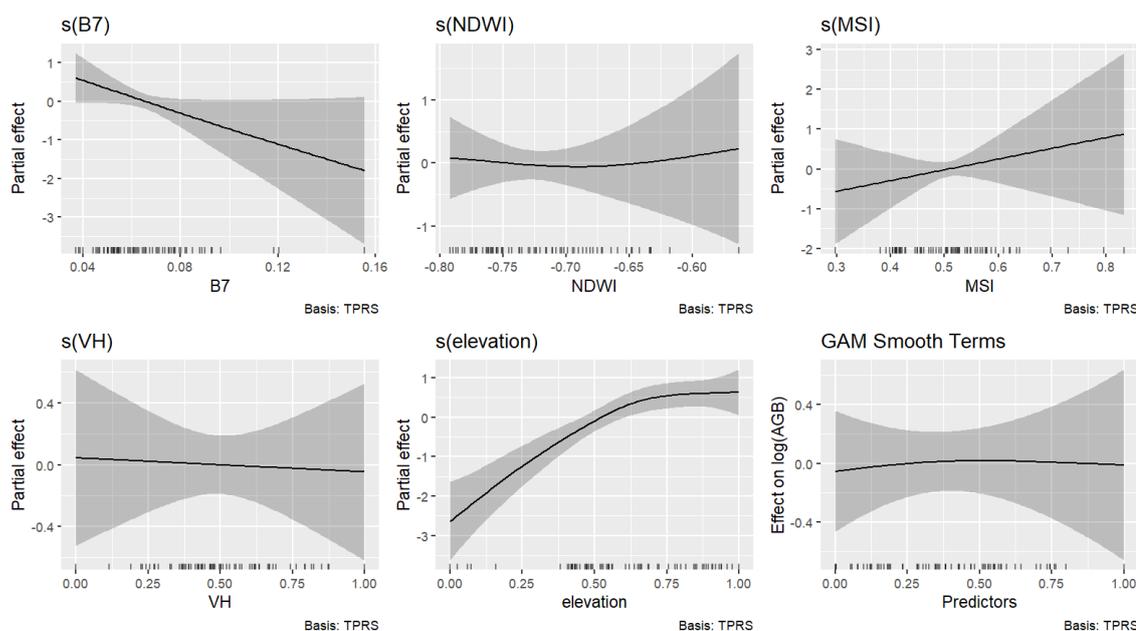
Kết quả mô hình GAM toàn bộ biến (gam1) cho thấy R^2 hiệu chỉnh đạt 0,619 và tỉ lệ giải thích phương sai (Deviance explained) là 72,5%. Một số biến có ý nghĩa thống kê rõ rệt, gồm: $ARVI$ ($p = 0,018$), VH ($p = 0,005$), VV ($p = 0,007$) và $elevation$ ($p = 1,24e-06$); bên cạnh đó, $B5$ ($p = 0,079$) có ý nghĩa ở mức cận biên. Điều này chứng tỏ rằng một số chỉ số thực vật, ra-đa và địa hình có quan hệ phi tuyến đáng kể với $\log\text{AGB}$. Kết quả này cũng tương đồng với một số nhận định của các nghiên cứu trước như Wai và cộng sự, 2022; Đinh Bá Duy, 2024; Đỗ Thị Nhung và cộng sự, 2024, Song và cộng sự, 2024.

Sử dụng các biến được lựa chọn sau phân tích PCA, mô hình GAM rút gọn (gam2) được xây dựng với $B7$, $NDWI$, MSI , VH , $elevation$ và $slope$. Kết quả cho thấy R^2 hiệu chỉnh = 0,535 và tỉ lệ giải thích phương sai = 59%. Mặc dù thấp hơn so với mô hình đầy đủ, mức độ phù hợp vẫn được duy trì đồng thời giúp đơn giản hóa cấu trúc mô hình. Đáng chú ý, biến $elevation$ thể hiện ảnh hưởng mạnh mẽ và có ý nghĩa thống kê cao ($p < 0,001$), trong khi các biến còn lại chỉ có xu hướng ảnh hưởng nhưng chưa đạt mức ý nghĩa. Kết quả mô hình GAM được trình bày trong Bảng 3.8.

Bảng 3.8. Kết quả mô hình GAM

Mô hình	Biến quan trọng ($p < 0,05$)	R^2 hiệu chỉnh	Tỉ lệ giải thích phương sai
GAM đầy đủ (gam1)	<i>ARVI, VH, VV, B5</i> (cận biên)	0,619	72,5%
GAM rút gọn (gam2)	<i>elevation</i>	0,535	59,0%

Do các biến trong GAM được ước lượng bằng hàm trơn thay vì hệ số cố định, nên mô hình không thể viết thành biểu thức tuyến tính “đóng” như hồi quy truyền thống. Kết quả phân tích được minh họa thông qua các đồ thị hàm trơn (Hình 3.13). Các đường cong (nét liền) thể hiện quan hệ phi tuyến giữa biến độc lập (chỉ số viễn thám, kênh ra-đa) và biến phụ thuộc $\log\text{AGB}$, trong khi các phần màu xám đen biểu thị khoảng tin cậy 95%.

**Hình 3.13. Đồ thị hàm trơn từ mô hình GAM**

Kết quả đồ thị hàm trơn cho thấy, đường cong hồi quy thể hiện mối quan hệ nghịch rõ rệt giữa $B7$ và $\log\text{AGB}$. Khi giá trị phản xạ ở kênh SWIR2 tăng, $\log\text{AGB}$ giảm mạnh, đặc biệt trong khoảng giá trị $B7 < 0,10$, phản ánh khả năng hấp thụ mạnh năng lượng ở vùng SWIR của các tán rừng dày có sinh khối lớn. Sau ngưỡng này, đường cong dần ổn định, cho thấy hiệu ứng bão hòa khi phản xạ SWIR2 không còn tăng tỷ lệ nghịch rõ rệt với sinh khối. Kết quả này khẳng định tính nhạy cảm cao của

kênh SWIR2 đối với độ ẩm, cấu trúc tán và mật độ rừng. Xu hướng này phù hợp với các nghiên cứu của Foody và cộng sự (2003) và Lu và cộng sự (2012).

NDWI có xu hướng tác động yếu và không tuyến tính rõ ràng đến $\log\text{AGB}$. Trong hầu hết khoảng giá trị, đường cong gần như phẳng, cho thấy *NDWI* không phải là biến chi phối trong mô hình, mặc dù có sự tăng nhẹ của $\log\text{AGB}$ ở các vùng có *NDWI* cao (thảm thực vật ẩm hơn). Điều này cho thấy *NDWI* chỉ phản ánh một phần độ ẩm bề mặt và lá cây, không đủ mạnh để giải thích biến động sinh khối trong khu vực RTX có cấu trúc tán phức tạp như Đắc Lắc.

Biến *MSI* thể hiện xu hướng tăng dần và phi tuyến yếu với $\log\text{AGB}$. Khi *MSI* tăng (nghĩa là mức căng thẳng ẩm giảm), $\log\text{AGB}$ có xu hướng tăng, phản ánh mối liên hệ thuận chiều giữa độ ẩm và sinh khối. Tuy nhiên, khoảng tin cậy 95% khá rộng, cho thấy biến này có mức đóng góp vừa phải trong mô hình.

Biến *VH* cho thấy mối quan hệ rất yếu với $\log\text{AGB}$, thể hiện qua đường cong gần như nằm ngang và khoảng tin cậy rộng. Điều này phản ánh hiện tượng bão hòa tín hiệu ra-đar thường gặp ở rừng nhiệt đới, khi mật độ tán cao khiến sóng ra-đa tán xạ mạnh và không còn tỷ lệ thuận với sinh khối. Kết quả tương đồng với các nghiên cứu của Joshi và cộng sự (2017) khẳng định rằng dữ liệu ra-đa tần số C (Sentinel-1) có giới hạn trong việc ước tính sinh khối cao.

Biến *elevation* thể hiện quan hệ phi tuyến thuận mạnh nhất với $\log\text{AGB}$. Sinh khối tăng rõ rệt từ vùng thấp lên vùng trung bình, sau đó xu hướng ổn định ở khu vực cao hơn. Điều này phù hợp với quy luật phân bố rừng theo độ cao, nơi các khu vực cao thường có độ che phủ lớn, ẩm độ cao và ít chịu tác động của hoạt động nhân sinh. Đây là biến đóng góp mạnh nhất trong mô hình GAM, được thể hiện qua đường cong có độ dốc lớn và khoảng tin cậy hẹp.

Mô hình GAM đã cải thiện đáng kể khả năng mô phỏng phi tuyến của các biến, thể hiện rõ hơn bản chất phi tuyến trong quan hệ giữa sinh khối rừng và các yếu tố phản xạ phổ - địa hình. Trong các biến, *elevation* và *B7* thể hiện ảnh hưởng mạnh

và ổn định nhất, trong khi *VH*, *NDWI* và *MSI* chỉ đóng vai trò bổ trợ.

Sự khác biệt hình dạng giữa các hàm hồi quy riêng phần ($s(x)$) chứng tỏ quan hệ giữa $\log AGB$ và các nhân tố ảnh hưởng không tuân theo mô hình tuyến tính đơn giản, do đó việc sử dụng GAM là phù hợp trong bối cảnh sinh thái phức tạp của RTX ở Đắk Lắk.

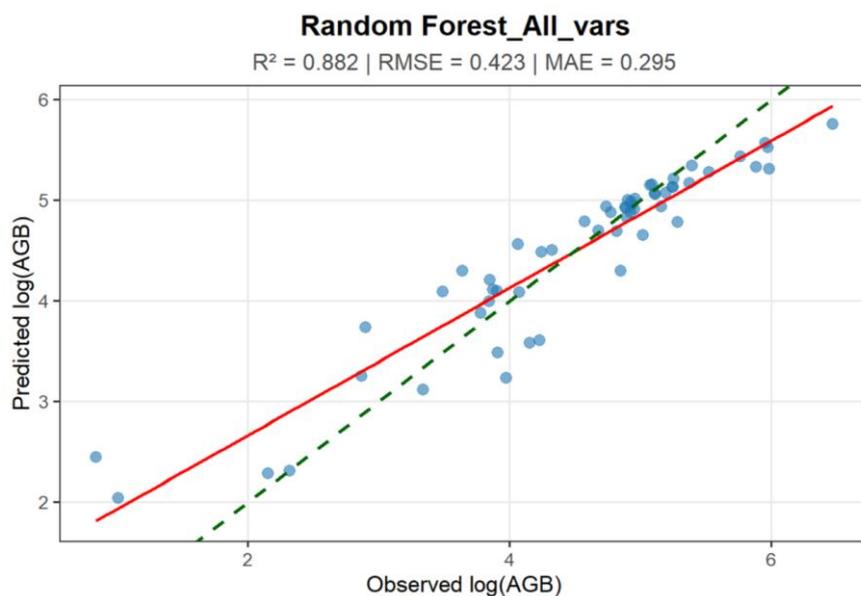
Các kết quả cho thấy quan hệ giữa tín hiệu viễn thám và sinh khối không hoàn toàn tuyến tính, mà có dạng phi tuyến yếu. Điều này phù hợp với nhận định của Lu và cộng sự (2014) rằng mối quan hệ giữa AGB và chỉ số phổ thường là phi tuyến, do sự thay đổi trong cấu trúc tán, chiều cao cây và độ che phủ. Việc sử dụng mô hình GAM cho phép xác định xu hướng phi tuyến này mà mô hình hồi quy tuyến tính không thể nắm bắt. Theo Levine và cộng sự (2020), GAM giúp mô tả linh hoạt các quan hệ phi tuyến mà không cần giả định trước dạng hàm, do đó phù hợp với các bài toán viễn thám rừng có tín hiệu đa chiều và nhiễu cao. Việc mô hình GAM phát hiện được quan hệ phi tuyến yếu cho thấy tiềm năng phát triển các mô hình phi tuyến nâng cao như RF nhằm nâng cao độ chính xác ước tính AGB [54]. So với hồi quy tuyến tính, mô hình GAM cho thấy khả năng nắm bắt mối quan hệ phi tuyến giữa $\log AGB$ và các biến phổ viễn thám, đồng thời cải thiện mức giải thích phương sai. Tuy nhiên, giá trị R^2 hiệu chỉnh (0,535) và tỉ lệ giải thích phương sai (59%) vẫn chưa cao, phản ánh những hạn chế khi dữ liệu có sự tương tác phức tạp. Do đó, nghiên cứu tiếp tục áp dụng phương pháp RF nhằm khắc phục các nhược điểm trên.

3.3.3. Ước tính AGB rừng sử dụng mô hình RF

Đối với mô hình hồi quy phi tham số, đặc biệt là mô hình học máy như RF thì phân phối chuẩn của dữ liệu và đa cộng tuyến không ảnh hưởng đến kết quả hồi quy cũng như nhằm xác định phương án tối ưu trong việc lựa chọn biến đầu vào và đánh giá ảnh hưởng của các nhóm biến đến hiệu quả mô hình, nghiên cứu tiến hành thử nghiệm hai cấu hình của mô hình RF: (i) RF–All_var.: sử dụng toàn bộ biến đầu vào và (ii) RF–Selected (VH): sử dụng nhóm biến quan trọng đã được lựa chọn (*B7*, *NDWI*, *MSI*, *VH*, *elevation*, *slope*).

3.3.3.1. Kết quả phân tích RF cho toàn bộ các biến (RF-All_var.)

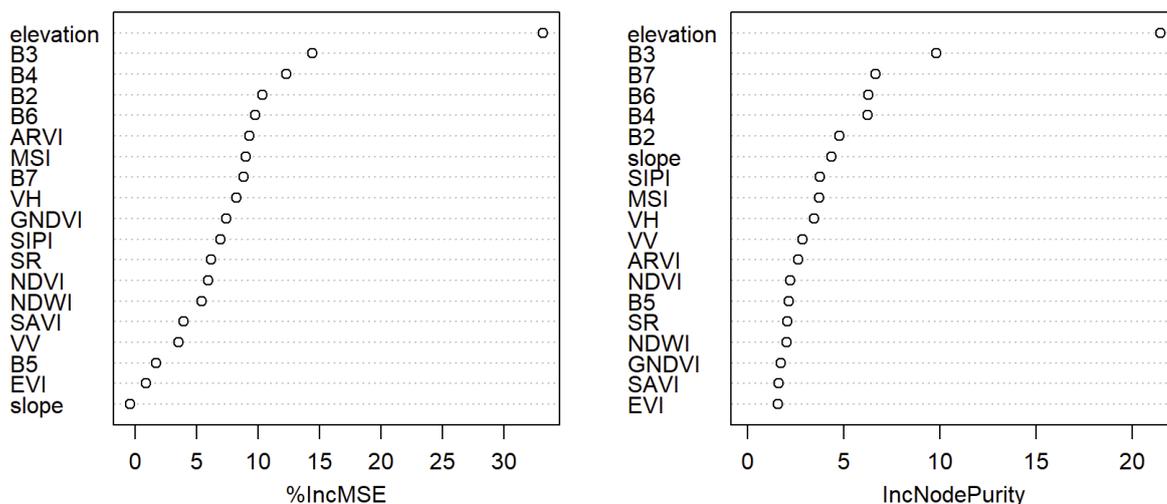
Mô hình RF được huấn luyện với số cây quyết định (n_{tree}) = 2.000 và số biến được chọn tại mỗi lần chia nhánh (m_{try}) = 6 trên toàn bộ 18 biến đầu vào (bao gồm các chỉ số quang học, ra-đa và địa hình). Kết quả cho thấy mô hình đạt độ phù hợp cao với $R^2 = 0,882$, $RMSE = 0,423$ và $MAE = 0,295$; Các giá trị này chứng tỏ mô hình RF có độ phù hợp rất tốt giữa giá trị ước lượng và giá trị thực tế của logAGB, thể hiện khả năng ước lượng của thuật toán trong việc mô hình hóa các mối quan hệ phi tuyến và tương tác phức tạp giữa các biến viễn thám. Mức R^2 cao cho thấy mô hình giải thích được tới 88,2% phương sai của log(AGB), vượt trội so với OLS ($R^2 \approx 0,49$) và GAM ($\sim 0,61$). Tuy nhiên, khi đánh giá bằng dữ liệu kiểm định nội tại (OOB), hiệu suất suy giảm rõ rệt ($R^2_{test} = 0,395$, $RMSE_{test} = 1,073$, $MAE_{test} = 0,745$), cho thấy mô hình có xu hướng quá khớp (overfitting). Tổng thể, RF-All_var. đạt độ phù hợp cao trên tập huấn luyện, nhưng khả năng khái quát hóa còn hạn chế do ảnh hưởng của đa cộng tuyến và sự dư thừa thông tin giữa các biến (Hình 3.14). Ngoài ra đồ thị phân bố tương quan giữa dữ liệu thực đo và dữ liệu ước tính qua mô hình (đường màu đỏ) cũng cho thấy mô hình vẫn còn hiện tượng ước lượng cao đối với vùng giá trị thấp và ước lượng thấp cho vùng giá trị cao của AGB.



Hình 3.14. Tương quan logAGB thực tế và ước tính

Trong cấu trúc của mô hình RF, các biến có độ quan trọng (variable importance) cao nhất bao gồm *elevation*, *B7*, *MSI*, và *NDWI*, cho thấy sự đóng góp đồng thời của các yếu tố địa hình, phản xạ phổ hồng ngoại trung, và độ ẩm trong tán rừng đối với sự biến thiên sinh khối. Sự hiện diện của các biến ra-đa (*VH*, *VV*) tuy có mức đóng góp thấp hơn, nhưng giúp tăng tính ổn định của mô hình nhờ khả năng khắc phục ảnh hưởng của mây che phủ và độ ẩm đất, đặc biệt trong khu vực rừng nhiệt đới ẩm như Đắc Lắc.

Kết quả phân tích độ quan trọng của biến (Hình 3.15) trong mô hình RF (với 2000 cây và tất cả 18 biến đầu vào) cho thấy sự khác biệt rõ rệt giữa mức đóng góp của các nhóm biến quang học, địa hình và ra-đa đối với khả năng ước lượng sinh khối rừng ($\log AGB$).



Hình 3.15. Mức độ quan trọng của các biến trong mô hình RF

Cả hai thước đo %IncMSE IncMSE (mức tăng sai số trung bình khi loại biến) và IncNodePurity IncNodePurity (mức giảm độ nhiễu trung bình của các nút chia trong cây quyết định) đều chỉ ra rằng *elevation* là biến có mức độ quan trọng cao nhất, vượt trội so với các biến còn lại. Điều này khẳng định vai trò chi phối của yếu tố địa hình trong phân bố và biến thiên sinh khối, phản ánh thực tế sinh thái – rừng tự nhiên tại Đắc Lắc thường tập trung ở khu vực có độ cao trung bình, nơi điều kiện ẩm độ và thổ nhưỡng thuận lợi cho sinh trưởng cây gỗ lớn.

Tiếp theo, các băng phổ vùng nhìn thấy và cận hồng ngoại (*B2, B3, B4, B6, B7*) đóng góp đáng kể, đặc biệt *B3* và *B7* đạt giá trị %IncMSE cao. Điều này cho thấy khả năng phản ánh cấu trúc tán rừng và hàm lượng nước của các kênh phổ này. *B7* thể hiện quan hệ nghịch với logAGB — phù hợp với bản chất quang học khi sinh khối tăng thì phản xạ vùng hồng ngoại trung giảm. Các kết quả này tương đồng với phân tích hồi quy và PCA, trong đó nhóm kênh hồng ngoại sóng ngắn và kênh ánh sáng đỏ có tương quan mạnh với logAGB.

Nhóm chỉ số phổ (*ARVI, MSI, SIPI*) thể hiện đóng góp trung bình, cho thấy các chỉ số này giúp bổ sung thông tin phổ tổng hợp về tình trạng thực vật và căng thẳng ẩm. Trong đó, *MSI* thể hiện vai trò quan trọng tương đối, khẳng định ảnh hưởng của độ ẩm tán và cấu trúc tán lá đến sinh khối.

Các biến ra-đa (*VH, VV*) có mức quan trọng thấp hơn so với biến quang học và địa hình, tuy nhiên vẫn đóng vai trò bổ sung, phản ánh cấu trúc tán và độ gồ ghề bề mặt - các yếu tố khó nắm bắt bằng dữ liệu quang học. Sự kết hợp ra-đa và quang học được chứng minh là giúp tăng độ chính xác ước lượng, như nhiều nghiên cứu trước đây [35],[110].

Cuối cùng, biến *slope* có mức đóng góp rất thấp trong cả hai chỉ số, cho thấy độ dốc không phải là yếu tố có ảnh hưởng đáng kể đến sự phân bố sinh khối ở quy mô tỉnh.

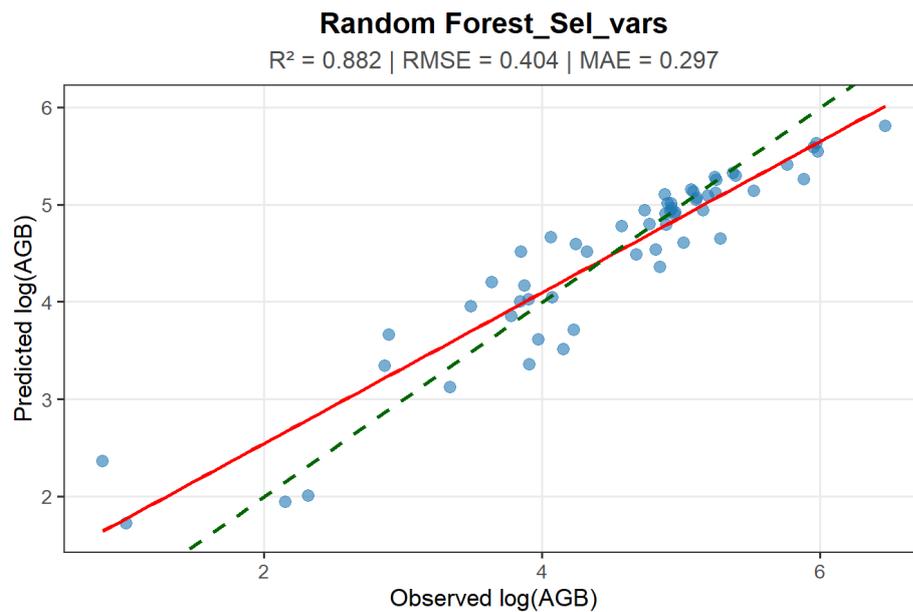
Tổng thể, kết quả độ quan trọng biến từ mô hình RF khẳng định:

- *elevation* là yếu tố địa hình chi phối chính,
- Các kênh phổ SWIR và Green (*B6, B7* và *B3*,) cung cấp thông tin quang học then chốt,
- Chỉ số *MSI* và *ARVI* có vai trò hỗ trợ về độ ẩm và trạng thái sinh lý thực vật,
- Kênh ra-đa *VV* và *VH* mang tính thông tin cấu trúc, bổ sung cho dữ liệu quang học.

Như vậy, việc kết hợp hài hòa giữa biến địa hình – quang học – ra-đa là hướng tiếp cận hiệu quả nhằm nâng cao độ chính xác của mô hình ước lượng sinh khối rừng bằng viễn thám đa nguồn. Bên cạnh đó, để giảm nhiễu và tối ưu hóa mô hình hồi quy, mô hình hồi quy RF cho các biến chọn lọc đã được xác định thông qua phân tích PCA được thử nghiệm để đối chiếu, so sánh hiệu quả ước lượng của mô hình RF.

3.3.3.2. Mô hình RF chọn lọc (RF_Selected)

Với cùng cấu hình tham số ($n_{tree} = 2.000$; $m_{try} = 6$), mô hình RF được huấn luyện lại chỉ với 6 biến chọn lọc. Kết quả đạt $R^2_{train} = 0,882$, $RMSE_{train} = 0,404$, $MAE_{train} = 0,297$ (tương đương $RMSE$ và MAE của AGB lần lượt là 2,54 tấn/ha và 1,98 tấn/ha), phản ánh mức độ phù hợp cao tương đương mô hình toàn biến Hình 3.16.

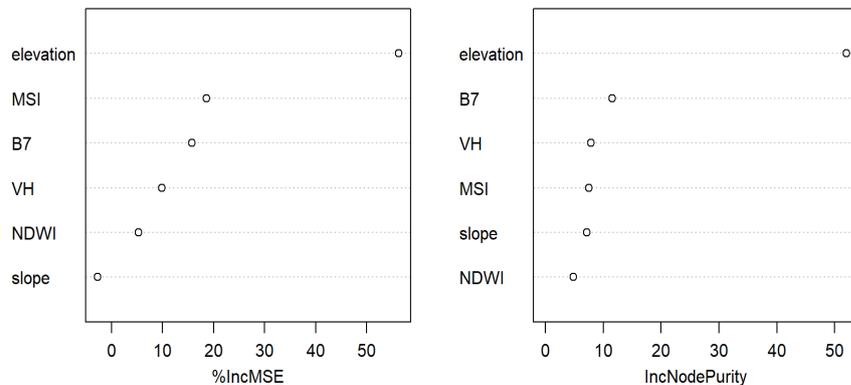


Hình 3.16. Tương quan logAGB thực tế và ước tính (chọn lọc biến)

Đáng chú ý, khi kiểm định bằng dữ liệu OOB, R^2_{test} tăng lên 0,606 (cao hơn 0,211 so với mô hình toàn biến), trong khi $RMSE_{test} = 0,894$ và $MAE_{test} = 0,614$, giảm tương ứng 16,7 và 17,6% sai số so với mô hình gồm tất cả các biến. Kết quả này cho thấy việc loại bỏ các biến dư thừa giúp mô hình giảm hiện tượng overfitting và tăng tính ổn định.

Biểu đồ phân tán giữa logAGB thực tế và logAGB ước tính cho thấy các điểm dữ liệu phân bố gần đường hồi quy lý tưởng (đường nét đứt), chứng tỏ mô hình RF có khả năng dự báo tốt trên toàn bộ dải giá trị sinh khối. Sai số xuất hiện chủ yếu ở các vùng có sinh khối rất thấp hoặc rất cao, do ảnh hưởng của sự biến thiên phổ và độ bão hòa tín hiệu quang học trong vùng rừng dày.

Phân tích độ quan trọng biến dựa trên hai chỉ tiêu %IncMSE và IncNodePurity (Hình 3.17) chỉ ra *elevation*, *B7 (SWIR2)* và *MSI* là ba yếu tố có đóng góp lớn nhất trong giải thích logAGB, các biến ra-đa (*VH*) và chỉ số *NDWI* có vai trò hỗ trợ trong phân biệt vùng rừng – không rừng và phản ánh tình trạng ẩm của thảm thực vật.



Hình 3.17. Tương quan logAGB thực tế và ước tính (chọn lọc biến)

Biến *elevation* có giá trị %IncMSE cao nhất (56,26%) và IncNodePurity lớn nhất (52,03), cho thấy yếu tố địa hình đóng vai trò quyết định trong phân bố và tích lũy sinh khối rừng. Sinh khối thường cao ở các khu vực rừng tự nhiên có độ cao trung bình – nơi có độ ẩm khí hậu và cấu trúc tán phát triển mạnh.

Biến *B7* thể hiện mức độ quan trọng thứ hai, phản ánh mối quan hệ nghịch giữa phản xạ phổ vùng hồng ngoại sóng ngắn và mật độ tán rừng; độ phản xạ giảm khi sinh khối tăng do tán cây hấp thụ mạnh năng lượng quang học.

Biến *MSI* có giá trị %IncMSE = 18,55%, cho thấy chỉ số độ ẩm có ảnh hưởng đáng kể đến sinh khối rừng, đặc biệt trong việc phản ánh mức khô hạn của tán lá và đất nền.

Biến *VH* đóng góp ở mức trung bình, phản ánh cấu trúc tán rừng nhưng bị giới hạn do hiện tượng bão hòa tín hiệu ra-đa ở rừng có mật độ cao.

Biến *NDWI* có vai trò hỗ trợ trong phản ánh vùng đất trống hoặc ẩm thấp, góp phần cải thiện phân biệt giữa rừng và không rừng.

Biến *slope* có tầm quan trọng thấp nhất, phù hợp với kết quả hồi quy tuyến tính trước đó khi độ dốc không thể hiện mối quan hệ rõ ràng với sinh khối ở quy mô tỉnh.

Để xác định các thông số của mô hình RF phù hợp với dữ liệu biến chọn lọc, nghiên cứu tiến hành thử nghiệm mô hình RF với các giá trị *ntree* và *mtry* khác nhau để lựa chọn mô hình phù hợp nhất với dữ liệu chọn lọc biến. Kết quả thử nghiệm các cấu hình RF với ba mức số cây (*ntree* = 1.000, 1.500, 2.000) và bốn giá trị *mtry* (3–6) cho thấy hiệu năng mô hình nhìn chung ổn định như được trình bày ở Bảng 3.9, với R^2_{train} cao (0,882–0,888) và sai số huấn luyện thấp. Tuy nhiên, tiêu chí quan trọng hơn để lựa chọn mô hình phù hợp là khả năng khái quát hóa, thể hiện qua R^2_{test} và các chỉ số sai số $RMSE_{\text{test}}$, MAE_{test} .

Bảng 3.9. Hiệu suất mô hình RF theo các cấu hình khác nhau

<i>ntree</i>	<i>mtry</i>	Train			Test		
		R^2	RMSE	MAE	R^2	RMSE	MAE
1000	3	0,884	0,416	0,303	0,629	0,892	0,617
1000	4	0,883	0,408	0,301	0,621	0,887	0,609
1000	5	0,887	0,401	0,293	0,630	0,877	0,603
1000	6	0,886	0,398	0,292	0,606	0,893	0,604
1500	3	0,883	0,417	0,299	0,626	0,890	0,600
1500	4	0,886	0,405	0,296	0,623	0,885	0,607
1500	5	0,888	0,399	0,292	0,631	0,877	0,607
1500	6	0,884	0,401	0,295	0,632	0,871	0,596
2000	3	0,888	0,409	0,296	0,616	0,900	0,613
2000	4	0,887	0,401	0,296	0,627	0,880	0,602
2000	5	0,884	0,403	0,295	0,626	0,881	0,607
2000	6	0,882	0,404	0,297	0,606	0,894	0,614

Thứ nhất, xét theo số lượng cây quyết định, khi *ntree* tăng từ 1000 lên 2000 không làm cải thiện rõ rệt hiệu quả mô hình. R^2_{test} dao động quanh 0,616–0,632, khác biệt rất nhỏ giữa ba mức *ntree*. Điều này cho thấy mô hình RF đã đạt trạng thái

hội tụ từ khoảng 1000–1500 cây, và việc tăng thêm số cây chỉ làm tăng thời gian tính toán mà không mang lại cải thiện đáng kể.

Thứ hai, *mtry* có ảnh hưởng rõ rệt hơn đến hiệu năng dự báo. Trong cả ba mức *n tree*, giá trị *mtry* = 5 hoặc 6 thường cho R^2_{test} cao nhất và RMSE thấp nhất. Cụ thể, cấu hình 1.500 cây với *mtry* = 6 đạt $R^2_{\text{test}} = 0,632$ (cao nhất trong toàn bộ các tổ hợp), đồng thời $\text{RMSE}_{\text{test}} = 0,871$ và $\text{MAE}_{\text{test}} = 0,596$ – đây cũng là một trong những trường hợp có sai số thấp nhất. Cấu hình 1.000 cây với *mtry* = 5 và 2.000 cây với *mtry* = 6 cũng có hiệu năng tương đương, nhưng mức cải thiện không vượt trội hơn so với *n tree* = 1500.

Tổng hợp các chỉ số đánh giá, cấu hình của mô hình RF với *n tree* = 1.500 và *mtry* = 6 được xem là mô hình tối ưu cho dữ liệu các biến được chọn lọc, với R^2_{test} cao nhất (0,632), sai số thấp và mô hình ổn định. Cấu hình này được lựa chọn cho các bước phân tích tiếp theo nhằm đảm bảo độ chính xác và tính khái quát hóa của mô hình ước lượng AGB cho RTX của khu vực nghiên cứu.

3.3.3.3. So sánh hiệu suất mô hình RF toàn bộ biến và biến chọn lọc

Kết quả chỉ số thống kê của các mô hình RF cho toàn bộ biến và các biến chọn lọc (Bảng 3.10) cho thấy, mặc dù giá trị R^2_{train} của hai phương án tương đương nhau (0,882), mô hình chọn lọc cho R^2_{test} cao hơn +0,211, đồng thời giảm sai số $\text{RMSE}_{\text{test}}$ và MAE_{test} khoảng 16,7-17,6% so với mô hình toàn biến. Cụ thể, R^2_{train} đạt 0,882 và R^2_{test} (OOB) đạt 0,606 – cao hơn rõ rệt so với mô hình toàn biến ($R^2_{\text{test}} = 0,395$). Các chỉ tiêu sai số cũng cải thiện tương ứng ($\text{RMSE}_{\text{test}} = 0,894$; $\text{MAE}_{\text{test}} = 0,614$) so với mô hình toàn biến ($\text{RMSE}_{\text{test}} = 1,073$; $\text{MAE}_{\text{test}} = 0,745$). Kết quả cho thấy việc giảm số lượng biến đầu vào đã cải thiện đáng kể khả năng khái quát hóa của mô hình RF.

Bảng 3.10. So sánh hiệu suất của mô hình RF dựa trên 2 bộ dữ liệu khác nhau

Mô hình	R^2_{train}	R^2_{test}	$\text{RMSE}_{\text{train}}$	$\text{RMSE}_{\text{test}}$	$\text{MAE}_{\text{train}}$	MAE_{test}
All_var.	0,882	0,395	0,423	1,073	0,295	0,745
Selected (VH)	0,884	0,632	0,401	0,871	0,295	0,596

Kết quả phân tích cho thấy, mô hình RF–Selected (VH) với số lượng $n_{tree} = 1.500$ và $m_{try} = 6$ không chỉ duy trì được độ phù hợp cao trên tập huấn luyện mà còn cải thiện rõ rệt khả năng tổng quát hóa so với mô hình toàn biến, nhờ loại bỏ các biến dư thừa và giảm hiện tượng quá khớp. Mô hình này đạt cân bằng tốt giữa độ chính xác, tính ổn định và cấu trúc đơn giản, do đó được lựa chọn là mô hình tối ưu để ước lượng và lập bản đồ AGB rừng thường xanh của tỉnh Đắk Lắk.

So với mô hình RF toàn bộ biến, mô hình chỉ sử dụng các biến chọn lọc đạt hiệu quả tương đương thậm chí tốt hơn. Điều này khẳng định việc lựa chọn biến qua các bước kiểm định tương quan, PCA và VIF là hợp lý, giúp mô hình giảm độ phức tạp nhưng vẫn giữ được độ chính xác cao hơn.

Kết quả này phù hợp với xu hướng chung của các nghiên cứu trước đây khi sử dụng RF để ước tính sinh khối rừng. Chẳng hạn, Đặng Thị Ngọc An và cộng sự (2019) khi áp dụng Sentinel-2 và RF tại Vườn quốc gia Yok Đôn, Việt Nam, đã đạt $R^2 = 0,81$ và $RMSE = 36,67$ tấn/ha, khẳng định khả năng của RF trong việc tận dụng cả biến phổ và biến kết cấu để mô hình hóa AGB hiệu quả. Tương tự, Han và cộng sự (2022) khi kết hợp Gaofen-1 và Sentinel-1 cũng ghi nhận mô hình RF cho kết quả ước lượng tốt nhất với $R^2 = 0,7$ và $RMSE = 16,26$ tấn/ha, thấp hơn đáng kể so với nghiên cứu này, cho thấy dữ liệu quang học độ phân giải cao kết hợp tiền xử lý phù hợp có thể cải thiện kết quả ước tính.

Trong nghiên cứu của Talebiesfandarani & Shamsoddini (2022), việc áp dụng RF và SVR cho ước tính sinh khối toàn cầu cũng cho thấy RF duy trì hiệu năng tốt và ổn định hơn so với các thuật toán phi tuyến khác, dù có xu hướng ước lượng thấp ở khu vực sinh khối cao và ước lượng cao ở khu vực sinh khối thấp. Hiện tượng này ít xuất hiện trong kết quả hiện tại nhờ áp dụng phép biến đổi lô-ga-rít cho AGB, giúp giảm phương sai và ổn định mô hình.

Các nghiên cứu ở Iran [77] và Nepal [101] cũng cho thấy RF thường vượt trội hơn các mô hình tuyến tính (MLR) và SVM trong ước tính sinh khối rừng, với R^2 dao động từ 0,69 đến 0,87, tương tự mức đạt được trong nghiên cứu này. RF đặc biệt hiệu

quả khi làm việc với bộ dữ liệu đa nguồn (quang học + ra-đa + địa hình) và trong các hệ sinh thái có tính dị thể cao, như rừng nhiệt đới lá rộng thường xanh hoặc rừng phục hồi sau suy thoái.

Việc rút gọn biến trong mô hình RF mà không làm giảm đáng kể độ chính xác cũng được ghi nhận trong các nghiên cứu gần đây, chẳng hạn Đặng Thị Ngọc An và cộng sự (2019) chỉ sử dụng 11 biến quan trọng thay vì 132 biến ban đầu mà vẫn giữ $R^2 > 0,8$. Điều này cho thấy lựa chọn biến tối ưu (thông qua đánh giá tầm quan trọng biến) có vai trò quan trọng trong việc đơn giản hóa mô hình và giảm hiện tượng đa cộng tuyến, đồng thời tăng khả năng ứng dụng thực tiễn khi triển khai trên diện rộng.

Như vậy, kết quả của nghiên cứu này không chỉ củng cố hiệu quả của thuật toán RF trong ước tính sinh khối rừng bằng dữ liệu viễn thám, mà còn chỉ ra rằng một tập hợp biến nhỏ, được chọn lọc kỹ lưỡng (như *elevation*, B7, MSI, NDWI, VH, *slope*) có thể mang lại hiệu suất tương đương với mô hình sử dụng toàn bộ biến. Đây là cơ sở khoa học quan trọng để phát triển các mô hình RF gọn nhẹ, hiệu quả và có tính tổng quát cao, phục vụ cho giám sát sinh khối và các-bon rừng ở quy mô cảnh quan.

3.3.4. Đánh giá độ chính xác của các mô hình ước lượng

3.3.4.1. Kiểm định bằng các thông số nội bộ của mô hình

Đánh giá nội bộ mô hình được thực hiện dựa trên ba chỉ số thống kê chính gồm R^2 , RMSE và MAE, nhằm phản ánh đồng thời mức độ giải thích phương sai, sai số bình phương trung bình và sai số tuyệt đối trung bình trong dự báo AGB. Kết quả tổng hợp cho ba mô hình OLS, GAM và Random Forest được trình bày tại Bảng 3.11.

Bảng 3.11. Hiệu suất mô hình trên tập huấn luyện

Mô hình	RMSE	MAE	R^2
OLS	0,824	0,6194	0,510
GAM	0,750	0,5340	0,590
RF	0,401	0,2950	0,884

Mô hình OLS cho kết quả $RMSE = 0,824$, $MAE = 0,6194$ và $R^2 = 0,51$, phản

ánh mức độ mô phỏng chỉ ở mức trung bình. Sai số RMSE và MAE đều khá cao, cho thấy mô hình gặp hạn chế trong việc nắm bắt biến động thực của AGB, đặc biệt do giả định tuyến tính giữa các biến viễn thám và sinh khối rừng. Điều này phù hợp với bản chất phi tuyến đặc trưng của tín hiệu quang học và ra-đa đối với cấu trúc tán rừng, khiến OLS khó mô phỏng chính xác các vùng sinh khối thấp–cao.

Mô hình GAM cải thiện hiệu suất với $RMSE = 0,750$, $MAE = 0,534$ và $R^2 = 0,59$, cho thấy khả năng mô phỏng tốt hơn so với OLS. Điều này xuất phát từ việc GAM sử dụng các hàm trơn (spline) cho phép mô tả các quan hệ phi tuyến một biến giữa AGB và các chỉ số phổ – ra-đa – địa hình. Tuy nhiên, mức cải thiện vẫn còn hạn chế vì GAM khó nắm bắt các tương tác đa biến phức hợp, vốn rất đặc trưng trong mối quan hệ giữa sinh khối và đặc trưng ảnh viễn thám.

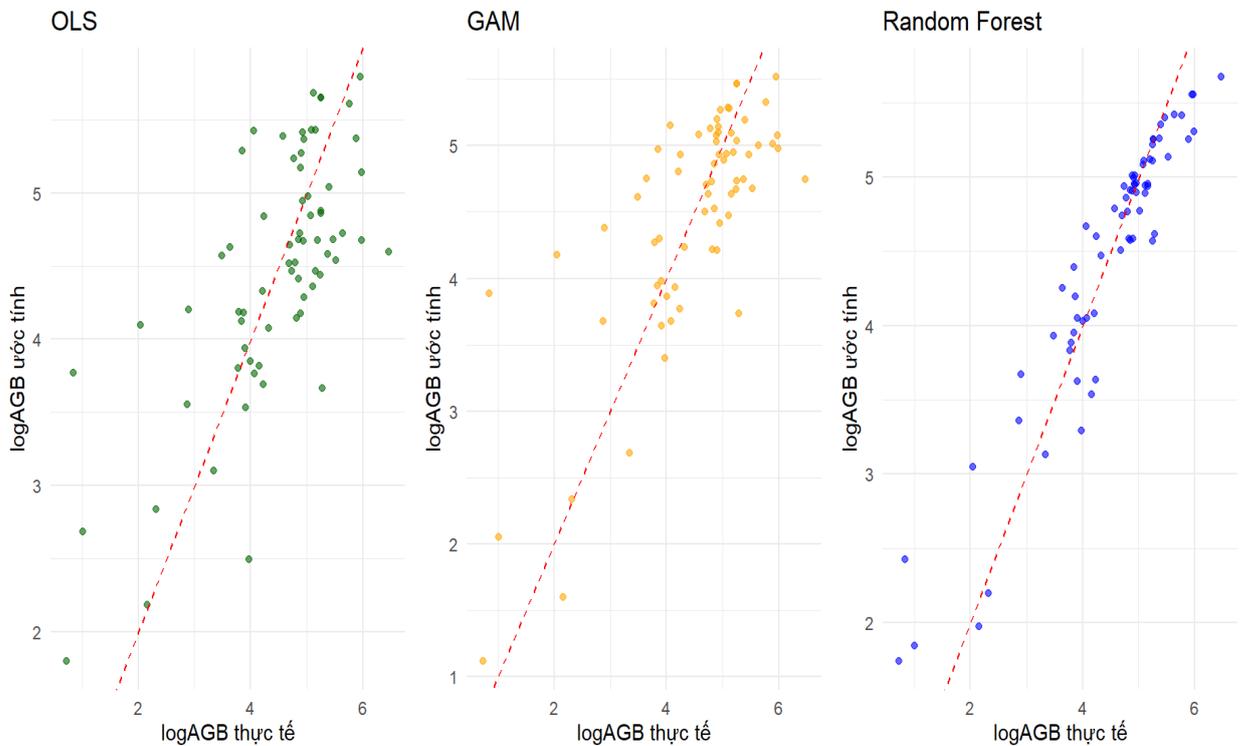
Mô hình RF đạt hiệu suất vượt trội với $RMSE = 0,401$, $MAE = 0,295$ và $R^2 = 0,884$, cao nhất trong ba phương pháp. Điều này chứng tỏ RF có khả năng xử lý tốt dữ liệu dị thể và cấu trúc phi tuyến cao, nhờ cơ chế kết hợp nhiều cây quyết định và đánh giá tương tác biến trên nhiều mức độ. Mô hình mô phỏng tốt cả hai vùng sinh khối thấp và cao, hạn chế đáng kể sai số hệ thống và giảm hiện tượng bão hòa tín hiệu – vốn là hạn chế thường gặp trong dữ liệu quang học ở rừng nhiệt đới.

So sánh ba mô hình cho thấy:

- OLS có độ chính xác thấp nhất do giả định tuyến tính và không xử lý được quan hệ phi tuyến giữa AGB và các biến giải đoán.
- GAM cải thiện độ chính xác nhờ mô tả phi tuyến đơn biến nhưng vẫn thiếu khả năng mô phỏng tương tác đa chiều.
- RF cho kết quả tốt nhất với RMSE và MAE thấp nhất, R^2 cao nhất, phản ánh khả năng mô phỏng tối ưu trong điều kiện dữ liệu phức tạp của rừng thường xanh Đắk Lắk.

Đồ thị so sánh logAGB thực tế và logAGB ước tính (Hình 3.18) cho thấy mức bám sát đường chéo của RF cao hơn rõ rệt so với hai mô hình còn lại, minh chứng

trực quan cho hiệu suất vượt trội được thể hiện ở các chỉ số thống kê. Điều này là cơ sở để lựa chọn RF làm mô hình tối ưu cho bước ước tính không gian AGB và CO₂ trong các phần tiếp theo.



Hình 3.18. Mức độ phù hợp giữa dữ liệu thực tế và ước tính từ các mô hình

Những kết quả trên là cơ sở để tiếp tục đánh giá chi tiết hơn bằng phương pháp kiểm định chéo phân tầng Stratified k-fold và dữ liệu độc lập nhằm kiểm tra tính ổn định, khả năng khái quát hóa và mức độ tin cậy của từng mô hình.

3.3.4.2. Kiểm định chéo Stratified k-fold

Để đánh giá độ chính xác và khả năng khái quát hóa của các mô hình ước lượng sinh khối rừng, nghiên cứu áp dụng phương pháp kiểm định chéo phân tầng k-fold cho cả ba mô hình OLS, GAM và RF. Toàn bộ 70 ô mẫu được chia ngẫu nhiên thành các nhóm (folds) có kích thước tương đương, với hai kịch bản kiểm định phổ biến là $k = 5$ và $k = 10$. Chỉ tiêu được sử dụng để đánh giá gồm R^2 , RMSE và MAE (Bảng 3.12).

Bảng 3.12. Kết quả kiểm định các mô hình hồi quy bằng cross validation

Mô hình	k	R ²	RMSE	MAE
OLS	5	0,475	0,896	0,669
OLS	10	0,480	0,878	0,709
GAM	5	0,491	0,845	0,646
GAM	10	0,515	0,817	0,639
RF	5	0,515	0,881	0,651
RF	10	0,550	0,818	0,637

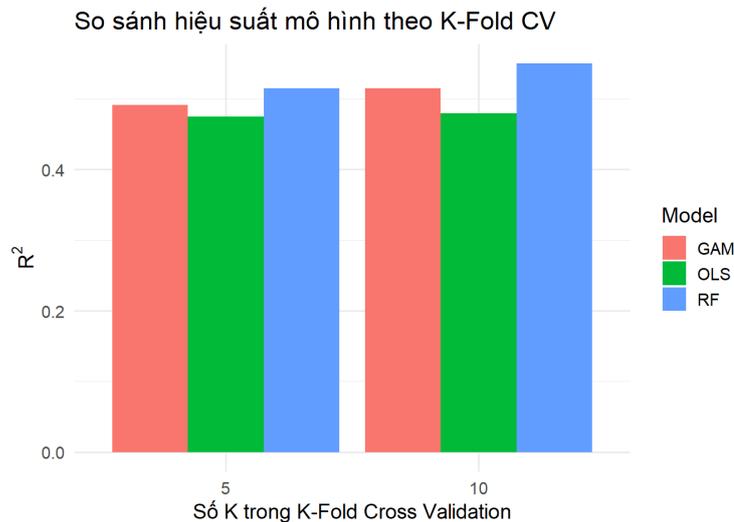
Kết quả cho thấy rằng khi tăng số lần chia dữ liệu từ $k = 5$ lên $k = 10$, tất cả các mô hình đều có cải thiện nhẹ về độ chính xác – thể hiện qua việc giảm RMSE và tăng R^2 . Tuy nhiên, mức cải thiện rõ rệt nhất được ghi nhận ở mô hình RF, trong khi OLS và GAM chỉ có cải thiện nhỏ hoặc không đáng kể. Biểu đồ so sánh hiệu suất mô hình kiểm định theo k-fold (Hình 3.19) cho thấy:

Mô hình OLS: OLS cho kết quả trung bình với R^2 tăng rất ít từ $R^2 = 0,475$ với $k=5$ lên $R^2 = 0,480$ với $k=10$; RMSE giảm từ 0,896 với $k=5$ xuống còn 0,878 với $k=10$; tuy nhiên sai số trung bình tuyệt đối MAE tăng từ 0,669, lên 0,709. Điều này phản ánh giới hạn của mô hình tuyến tính cổ điển trong việc mô phỏng mối quan hệ phi tuyến giữa sinh khối rừng và các biến viễn thám (như $B7$, $NDWI$, MSI), vốn có tính chất phi tuyến hoặc bão hòa tín hiệu quang học ở sinh khối cao.

Mô hình GAM: GAM thể hiện hiệu suất tốt hơn OLS, với R^2 cao hơn mô hình OLS và R^2 cũng tăng khi tăng số k tương ứng với $R^2=0,491$ và 0,515; RMSE giảm so với mô hình OLS và có xu hướng giảm khi tăng số k với giá trị RMSE là 0,845 và 0,817; sai số MAE cũng giảm đáng kể so với OLS với giá trị tương ứng với k lần lượt là 0,646 và 0,639. Mô hình này cho phép mô phỏng các quan hệ phi tuyến mềm giữa logAGB và các biến giải thích thông qua hàm spline, giúp phản ánh chính xác hơn xu hướng biến thiên sinh khối. Tuy nhiên, độ chính xác vẫn chưa đạt mức cao do GAM vẫn chịu ảnh hưởng của hiện tượng bão hòa phổ và nhiễu không gian trong dữ liệu viễn thám.

Mô hình RF: RF thể hiện hiệu suất vượt trội so với hai mô hình còn lại. đặc biệt ở $k = 10$ ($R^2 = 0,551$), đồng thời có RMSE và MAE thấp nhất lần lượt là 0,818 và 0,637. Điều này khẳng định ưu thế của RF khi xử lý dữ liệu có tính dị thể cao và

mối quan hệ phi tuyến đa chiều. Kết quả này chứng minh năng lực mô hình hóa phi tuyến và tương tác đa biến mạnh mẽ của RF, nhờ việc sử dụng tập hợp nhiều cây quyết định và cơ chế lấy mẫu bootstrap. Như vậy, RF tiếp tục là mô hình có hiệu suất cao nhất khi đánh giá bằng kiểm định chéo, phù hợp với kết quả đánh giá nội bộ.



Hình 3.19. So sánh hiệu suất mô hình RF bằng kiểm định chéo k-fold

Khi tăng số fold từ $k = 5$ lên $k = 10$ trong thiết lập chia mẫu ngẫu nhiên phân tầng, hiệu suất mô hình có xu hướng được cải thiện: RMSE và MAE giảm ở cả GAM và RF (ngoại trừ MAE của OLS tăng nhẹ). Giá trị R^2 tăng rõ rệt, đặc biệt đối với RF, từ 0,515 ($k = 5$) lên 0,551 ($k = 10$). Điều này là do khi k lớn hơn, tỷ lệ dữ liệu dùng cho huấn luyện tăng (từ 80% lên 90%), và nhờ cơ chế phân tầng theo logAGB, mỗi tập huấn luyện – kiểm định vẫn giữ được sự đại diện của các mức sinh khối. Vì vậy mô hình, đặc biệt là RF, khai thác tốt hơn các quan hệ phi tuyến và giảm sai lệch do chia mẫu.

Đồ thị (Hình 3.19) cho thấy: OLS đạt hiệu quả thấp nhất trong ba mô hình. Mức R^2 thấp (~0,48) và giá trị RMSE cao (~0,88–0,89) cho thấy mô hình tuyến tính không phù hợp để mô tả cấu trúc phức tạp của sinh khối rừng nhiệt đới, vốn chịu chi phối mạnh bởi các quan hệ phi tuyến giữa đặc trưng phổ – cấu trúc tán – địa hình. Mô hình GAM thể hiện mức cải thiện đáng kể so với OLS, đặc biệt ở $k = 10$ với $R^2 = 0,515$. Điều này cho thấy các hàm trơn đã mô phỏng tốt hơn sự biến đổi phi tuyến

của AGB theo các biến giải đoán, nhưng vẫn chưa đủ mạnh để mô phỏng tương tác đa biến. RF vượt trội nhất trong cả hai cấu hình k , với: R^2 cao nhất (0,55), RMSE thấp nhất ($\sim 0,818$), MAE thấp nhất ($\sim 0,637$). Điều này chứng minh ưu thế của RF khi mô hình hóa mối quan hệ phi tuyến phức tạp và tương tác giữa nhiều biến viển thám đồng thời. Đặc biệt, sự ổn định giữa hai cấu hình k cho thấy RF có tính nhất quán cao.

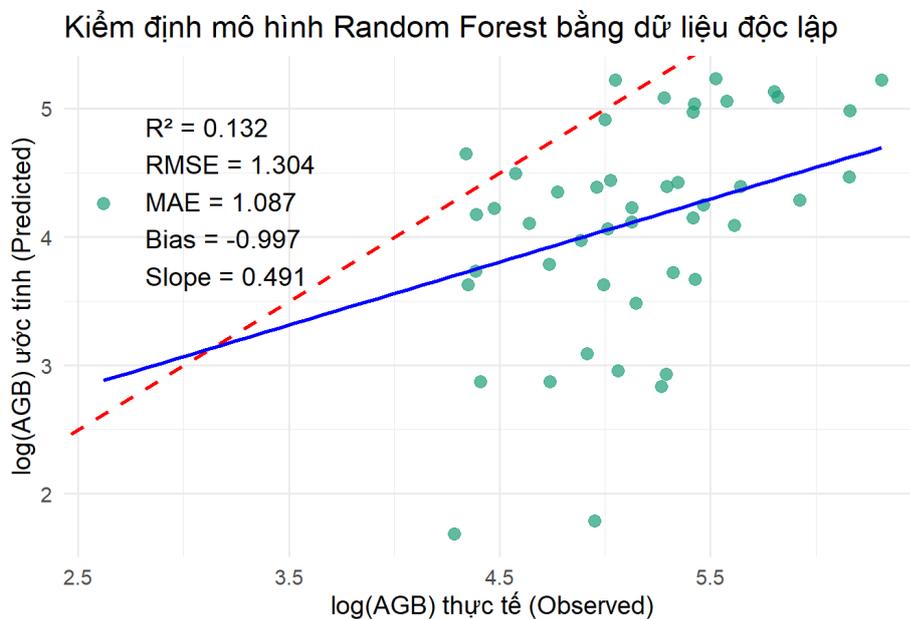
Kết quả xếp hạng hiệu quả mô hình (RF > GAM > OLS) phản ánh xu hướng thường thấy trong nghiên cứu viển thám sinh khối: các mô hình dựa vào tổ hợp cây quyết định, đặc biệt RF, thường vượt trội về độ chính xác và tính ổn định khi xử lý dữ liệu viển thám đa nguồn có tính phi tuyến và dị hướng cao. RF có khả năng chịu được dữ liệu có chiều cao, đa cộng tuyến giữa biến, nhiễu phổ và hiệu ứng địa hình nhờ cơ chế lấy mẫu bootstrap và chọn ngẫu nhiên biến đầu vào cho từng cây con, từ đó giảm quá khớp và mô phỏng tốt các tương tác phi tuyến phức tạp giữa biến quang học và ra-đa [40]. Nhiều nghiên cứu so sánh thuật toán cũng cho thấy RF thường đạt R^2 và độ chính xác cao hơn các mô hình tham số như OLS và đôi khi vượt hoặc tương đương với các phương pháp bán/phi tham số như GAM, đặc biệt khi dữ liệu có tính dị hướng không tuyến tính [88]. Các phân tích so sánh giữa ba nhóm mô hình - tham số (parametric), bán tham số (semiparametric) và phi tham số (nonparametric) - trong ước tính sinh khối trên mặt đất cho thấy rằng các phương pháp phi tham số, đặc biệt là các mô hình dựa trên cây quyết định như RF, thường cho kết quả sai số thấp hơn so với mô hình hồi quy tuyến tính thông thường (OLS) trong các khu rừng có cấu trúc phức tạp và đa dạng. Ngoài ra, RF còn có ưu thế trong việc khai thác và mô tả mối quan hệ tương tác phi tuyến giữa nhiều nguồn dữ liệu khác nhau [171]. Do đó, lựa chọn RF làm mô hình chính cho ước lượng sinh khối trong khu vực nghiên cứu là phù hợp về mặt thực nghiệm và lý thuyết.

3.3.4.3. Kiểm định mô hình RF bằng dữ liệu độc lập

Để đánh giá khả năng tổng quát hóa theo thời gian (temporal transferability) của mô hình, nghiên cứu áp dụng mô hình RF đã huấn luyện giai đoạn 2020–2024 cho bộ dữ liệu độc lập của năm 2013. Tập dữ liệu này khác biệt về điều kiện cảm biến

(chỉ có dữ liệu Landsat, thiếu dữ liệu Sentinel-1), đặc tính phổ và trạng thái rừng, qua đó phản ánh mức độ ổn định của mô hình khi áp dụng ngoài thời kỳ huấn luyện.

Kết quả cho thấy mô hình có giá trị R^2 chỉ đạt 13,2%, phản ánh mô hình chỉ giải thích được một phần nhỏ biến thiên của $\log(\text{AGB})$ trong năm 2013 (Hình 3.20). Điều này cho thấy mức suy giảm lớn về hiệu suất so với giai đoạn huấn luyện ($R^2_{\text{train}} = 0,884$). Đồng thời, các giá trị $\text{RMSE} = 1,304$ và $\text{MAE} = 1,087$ cao hơn mô hình huấn luyện cho thấy sai số dự báo tăng đáng kể khi áp dụng mô hình cho dữ liệu ngoài mẫu.



Hình 3.20. Đồ thị tương quan AGB ước lượng và thực tế (độc lập)

Hệ số $\text{Slope} = 0,491$ cho thấy mô hình bị nén giá trị ước lượng - hiện tượng ước lượng thấp tại vùng sinh khối cao và ước lượng quá cao ở vùng sinh khối thấp, phản ánh ảnh hưởng của bão hòa tín hiệu quang học khi không có dữ liệu ra-đa hỗ trợ. Kết quả này phù hợp với các nghiên cứu trước, vốn chỉ ra rằng mô hình được huấn luyện tại một thời kỳ thường suy giảm độ chính xác khi áp dụng cho giai đoạn khác, do khác biệt về đặc điểm phổ, cảm biến và cấu trúc rừng [35],[47],[123]. Dù vậy, mô hình RF vẫn thể hiện giá trị khoa học và thực tiễn, khi duy trì được xu hướng tổng thể của biến thiên sinh khối rừng, qua đó chứng minh tính khả dụng cho ước

lượng không gian – thời gian ở quy mô cấp tỉnh.

3.4. Lập bản đồ AGB rừng và khả năng hấp thụ CO₂ giai đoạn 2015 – 2025

3.4.1. Xây dựng bản đồ AGB giai đoạn 2015 - 2025

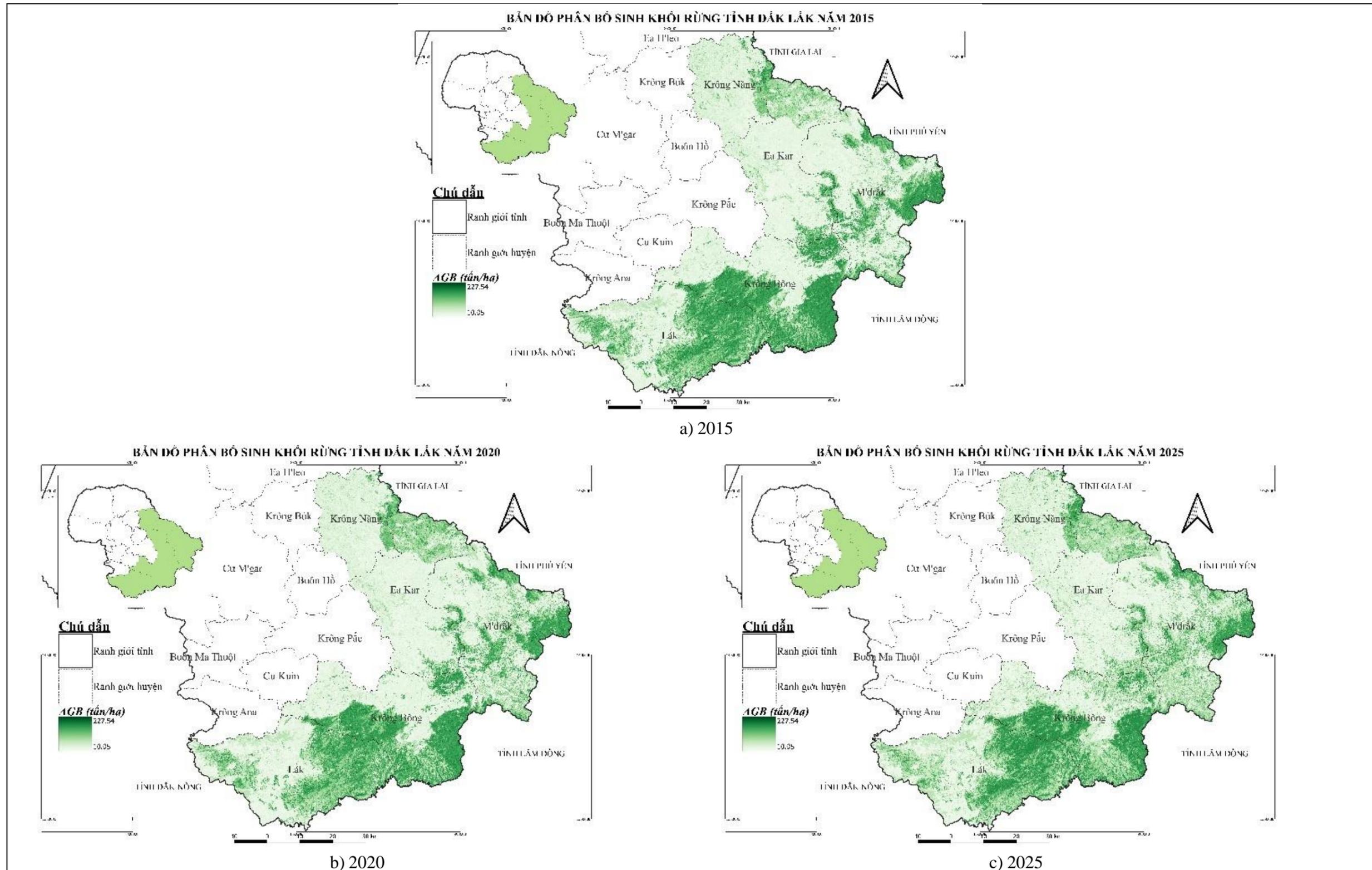
Dựa trên mô hình RF đã được huấn luyện và kiểm định bằng dữ liệu sử dụng nhóm biến quan trọng đã được lựa chọn (*B7, NDWI, MSI, VH, elevation, slope*) như đã phân tích ở phần trên, nghiên cứu tiến hành ước lượng và lập bản đồ phân bố không gian sinh khối trên mặt đất của RTX tỉnh Đắk Lắk giai đoạn 2015 - 2025.

Việc chỉ giới hạn phân tích cho giai đoạn này nhằm đảm bảo tính nhất quán của biến đầu vào, do dữ liệu ra-đa thu thập từ ảnh Sentinel-1 là một trong các biến quan trọng phản ánh cấu trúc tán rừng chỉ có sẵn từ năm 2015 trở lại đây.

Để nâng cao độ tin cậy, việc ước lượng tập trung vào các huyện có diện tích RTX lớn như Lắk, Krông Bông, M'Đrắk, Ea Kar và Krông Năng để tránh nhiễu dữ liệu từ các kiểu rừng và che phủ thực vật khác trong tỉnh.

Cách tiếp cận này không chỉ góp phần làm rõ động thái sinh khối rừng dưới tác động của các hoạt động quản lý, bảo vệ và phục hồi rừng, mà còn cung cấp cơ sở khoa học cho giám sát phát thải và báo cáo REDD⁺ ở cấp tỉnh.

Bản đồ AGB của RTX các năm 2015, 2020 và 2025 (Hình 3.21) cho thấy phân bố sinh khối trên mặt đất tại tỉnh Đắk Lắk có xu hướng tương đối ổn định về không gian. Nhìn chung, các khu vực có AGB cao vẫn tập trung chủ yếu ở phía Nam và Đông Nam tỉnh (huyện Lắk và Krông Bông), nơi có diện tích rừng tự nhiên liên tục và điều kiện sinh thái thuận lợi. Các dãy núi và khu vực có địa hình cao cũng duy trì giá trị sinh khối lớn, phản ánh sự tồn tại của các quần thể rừng ít bị tác động mạnh. Ngược lại, các khu vực phía Bắc và Đông Bắc (Ea Kar, Krông Năng, M'đrắk) thể hiện AGB thấp hơn, với phân bố loang lổ và rời rạc, chủ yếu do chịu ảnh hưởng của quá trình khai thác gỗ trước đây, nương rẫy và chuyển đổi mục đích sử dụng đất. Các vùng ven đô thị và khu dân cư nhìn chung cũng có giá trị sinh khối thấp, phù hợp với thực tế suy giảm diện tích rừng tự nhiên.



Hình 3.21. Phân bố sinh khối rừng qua các giai đoạn

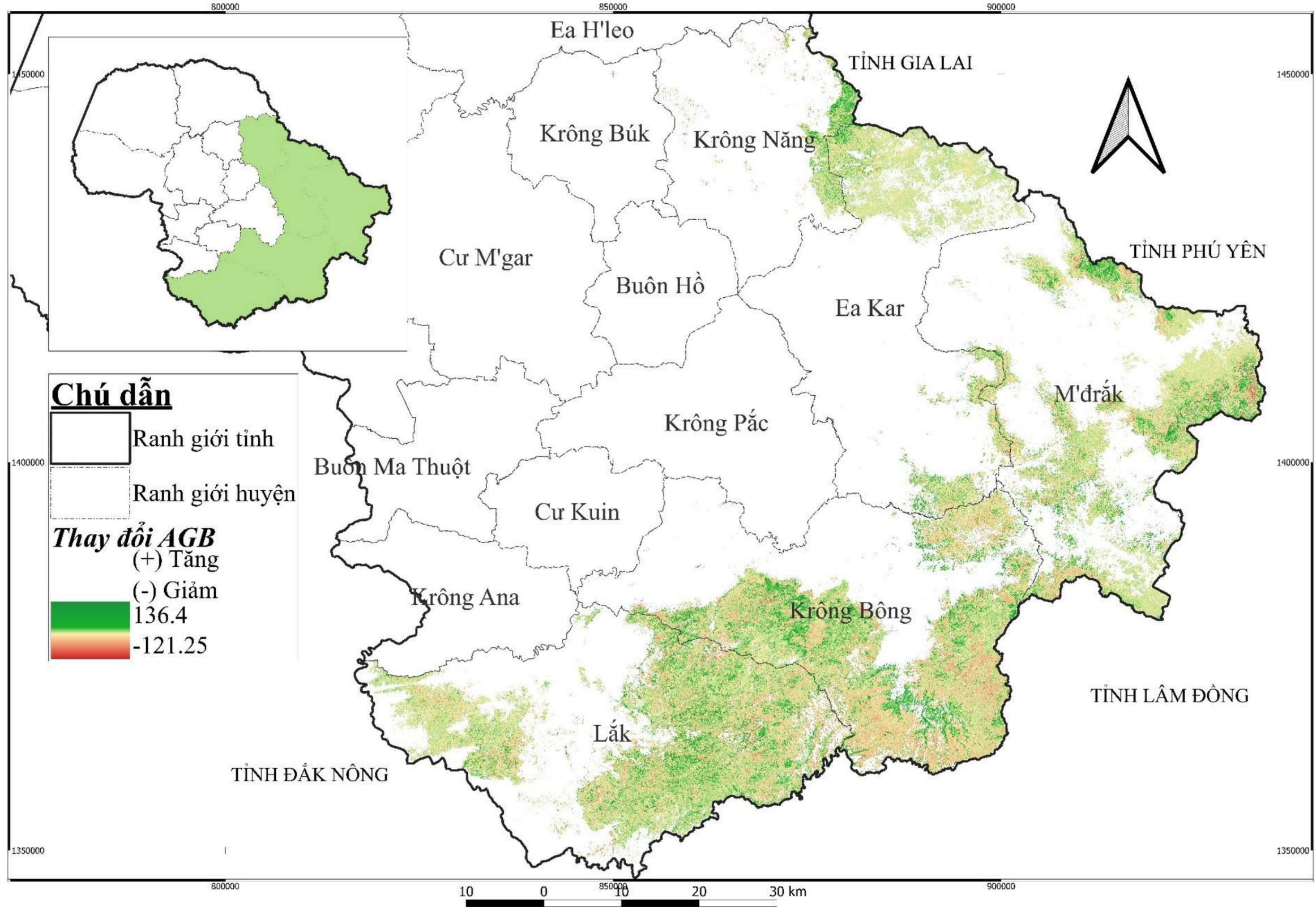
Tóm lại, trong suốt giai đoạn 2015–2025, bản đồ AGB cho thấy phân bố AGB tương đối đồng nhất giữa các năm, với sự phân hóa rõ rệt giữa các khu vực rừng giàu ở phía Nam và rừng suy giảm ở phía Bắc – Đông Bắc. Điều này cho phép khẳng định đặc trưng không gian của sinh khối RTX tại Đắk Lắk ít thay đổi, mặc dù có sự biến động về giá trị cụ thể qua các giai đoạn. Những biến động này phản ánh tác động tổng hợp của khai thác rừng trái phép, mở rộng diện tích canh tác nông nghiệp, quá trình phục hồi rừng tự nhiên, cũng như các hoạt động quản lý và bảo vệ rừng trong từng giai đoạn. Do vậy, việc phân tích sự thay đổi AGB theo không gian và thời gian sẽ cung cấp những bằng chứng khoa học quan trọng, góp phần đánh giá hiệu quả quản lý rừng và định hướng chính sách phát triển lâm nghiệp bền vững.

3.4.2. Phân tích biến động AGB giai đoạn 2015 - 2025

Để phân tích biến động AGB, ngưỡng thay đổi có thể phát hiện tối thiểu (MDC) được xác định dựa trên sai số chuẩn của các mô hình ước tính sinh khối năm 2013 và 2024. Với RMSE lần lượt là 1,679 và 0,414, số lượng điểm kiểm định $n = 70$, và mức ý nghĩa $\alpha = 0,05$ ($z = 1,96$), giá trị MDC được tính là 0,405. Điều này có nghĩa là chỉ những thay đổi logAGB lớn hơn 0,405 (tương ứng 2,5 tấn AGB/ha) mới có thể được coi là có ý nghĩa thống kê, vượt quá mức sai số của mô hình ước lượng.

Kết quả phân tích biến động AGB (Hình 3.22) cho thấy xu hướng thay đổi rõ rệt về không gian và thời gian trong giai đoạn 2015–2025. Nhìn chung, AGB có xu hướng suy giảm ở nhiều khu vực, đặc biệt tập trung ở các huyện M'Đrăk, Ea Kar và một phần Krông Bông, nơi chịu tác động mạnh của khai thác gỗ trái phép và chuyển đổi mục đích sử dụng đất. Các vùng màu đỏ và cam trên bản đồ thể hiện sự suy giảm mạnh về sinh khối, quá trình mất rừng và suy thoái do mở rộng nông nghiệp. Ngược lại, các khu vực có màu xanh lá cây đậm cho thấy sự gia tăng sinh khối, chủ yếu phân bố tại một số vùng thuộc huyện Lắk và Krông Bông, nơi có điều kiện thuận lợi cho tái sinh rừng tự nhiên và trồng rừng, nhờ các chương trình khoanh nuôi phục hồi và chi trả dịch vụ môi trường rừng.

BẢN ĐỒ BIẾN ĐỘNG AGB TỈNH ĐẮK LẮK GIAI ĐOẠN 2015 - 2025



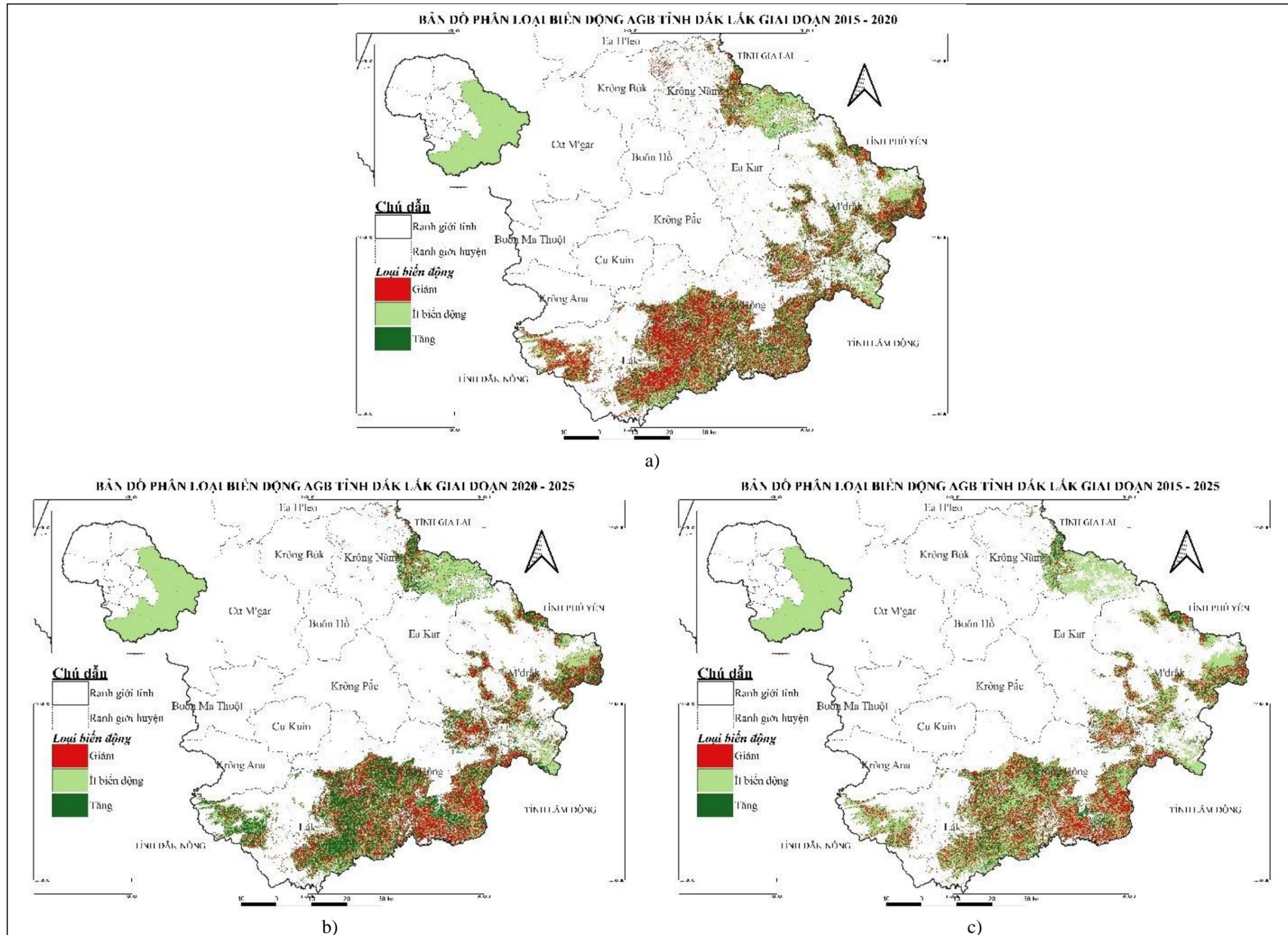
Hình 3.22. Bản đồ biến động AGB giai đoạn 2015 – 2025

Tổng thể, các bản đồ phân loại biến động AGB cho thấy sinh khối rừng tỉnh Đắk Lắk trong giai đoạn 2015–2025 có sự đan xen giữa các khu vực suy giảm và phục hồi, song xu thế suy giảm vẫn chiếm ưu thế về không gian. Các vùng giảm AGB phân bố rộng và liên tục hơn so với các vùng tăng AGB, phản ánh áp lực kéo dài của khai thác cây gỗ trái phép, chuyển đổi mục đích sử dụng đất và suy thoái rừng, đồng thời cho thấy các nỗ lực phục hồi rừng chưa đủ mạnh để bù đắp mức suy giảm tổng thể.

Xét theo từng giai đoạn, giai đoạn 2015–2020 (Hình 3.23a) cho thấy các khu vực giảm AGB chiếm tỷ lệ lớn, tập trung rõ rệt ở khu vực phía Đông và Nam tỉnh, đặc biệt dọc theo ranh giới các huyện Lắk và Krông Bông. Đây là những khu vực chịu tác động mạnh của hoạt động canh tác nông nghiệp và khai thác tài nguyên rừng. Ngược lại, các điểm tăng AGB xuất hiện rải rác ở các khu vực rừng thuộc huyện Krông Bông và M'Đrắk, nơi xa khu dân cư, khó tiếp cận.

Trong giai đoạn 2020–2025 (Hình 3.23b), phân bố biến động AGB có xu hướng phân tán hơn. Diện tích tăng AGB gia tăng tại một số khu vực rừng sâu, ít chịu tác động của con người, đặc biệt ở phía Nam tỉnh như Krông Bông và Lắk. Tuy nhiên, các vùng giảm AGB vẫn xuất hiện với mật độ cao tại phía Đông huyện Krông Bông và các huyện Ea Kar và M'Đrắk, cho thấy áp lực suy thoái rừng tại các khu vực này chưa được kiểm soát hiệu quả dù đã có sự cải thiện cục bộ ở một số nơi.

Tổng hợp toàn bộ giai đoạn 2015–2025 (Hình 3.23c), bản đồ biến động AGB cho thấy xu thế suy giảm sinh khối rừng vẫn là chủ đạo, tập trung nhiều ở các huyện phía Đông và phía Nam tỉnh (Ea Kar, M'Đrắk, Krông Bông và Lắk). Trong khi đó, các khu vực có xu hướng tăng AGB phân bố rải rác ở các khu vực thuộc rừng đặc dụng ở các huyện Lắk, Krông Bông. Kết quả này phản ánh rằng trong vòng một thập kỷ qua, quá trình suy thoái rừng diễn ra liên tục, còn các hoạt động phục hồi và tái sinh rừng mới chỉ mang tính cục bộ, chưa tạo ra sự chuyển biến rõ rệt về sinh khối rừng ở quy mô toàn tỉnh.



Hình 3.23. Bản đồ phân loại biến động AGB rừng tỉnh Đắk Lắk các giai đoạn
 a) 2015–2020; b) 2020–2025; c) 2015–2025.

Kết quả phân tích biến động AGB kiểu rừng thường xanh tỉnh Đắk Lắk giai đoạn 2015–2025 cho thấy xu thế suy giảm sinh khối chiếm ưu thế, diễn ra liên tục và với cường độ khác nhau ở các giai đoạn. Các điểm tăng AGB chỉ xuất hiện cục bộ tại những khu vực rừng sâu, xa dân cư hoặc nơi có hoạt động trồng rừng và tái sinh tự nhiên, trong khi phần lớn diện tích rừng, đặc biệt ở các huyện Ea Kar, M’Đrăk, Krông Bông và Lắk, vẫn ghi nhận suy giảm về sinh khối. Điều này phản ánh thực trạng áp lực lớn từ chuyển đổi mục đích sử dụng đất, khai thác gỗ và nông nghiệp. Nếu không có các biện pháp quản lý và phục hồi kịp thời, nguồn dự trữ các-bon của rừng sẽ tiếp tục suy giảm, làm gia tăng rủi ro về biến đổi khí hậu và mất cân bằng sinh thái.

Kết quả phân tích định lượng sự thay đổi sinh khối RTX giai đoạn 2015–2025 (Bảng 3.13) cho thấy xu thế vừa tích lũy vừa suy giảm, song tổng thể toàn tỉnh vẫn bị thâm hụt nhẹ. Cụ thể, tổng lượng sinh khối bị suy giảm trong giai đoạn nghiên cứu ước tính 1,45 triệu tấn, trong khi lượng tăng chỉ đạt 1,18 triệu tấn, dẫn đến mức suy giảm ròng 265,04 nghìn tấn AGB. Các khu vực ghi nhận suy giảm mạnh tập trung ở huyện M’Đrăk, Krông Bông, Lắk và Ea Kar, phản ánh quá trình khai thác gỗ trái phép, chuyển đổi sử dụng đất và suy thoái rừng tự nhiên. Các điểm tăng sinh khối rõ rệt tập trung tại huyện Krông Bông, Lắk và Ea Kar, nơi các hoạt động quản lý bảo vệ rừng và phục hồi rừng diễn ra tích cực ở các khu rừng đặc dụng và các chủ rừng. Kết quả này phản ánh sự mất cân bằng giữa quá trình suy thoái và phục hồi rừng, cho thấy mặc dù một số huyện (như Lắk và Krông Bông có xu hướng tích cực, song chưa đủ để bù đắp tình trạng suy thoái mạnh tại các huyện khác, đặc biệt là M’Đrăk và Ea Kar.

Bảng 3.13. Kết quả phân tích định lượng thay đổi sinh khối RTX tỉnh Đắk Lắk giai đoạn 2015–2025

Chỉ tiêu	Giảm AGB (tấn)	Tăng AGB (tấn)
Giá trị nhỏ nhất (tấn/ha)	0,00	1,00
Giá trị lớn nhất (tấn/ha)	-175,99	136,81
Khoảng biến động (tấn/ha)	175,98	135,81
Tổng lượng (tấn)	1.448.379,41	1.183.339,32
Giá trị trung bình (tấn/ha)	7,42	14,33
Độ lệch chuẩn	12,86	14,93
Cân đối (tấn)	-	-265.040,09

Kết quả phân tích thay đổi sinh khối giai đoạn 2015–2025 theo từng huyện (Bảng 3.14) cho thấy sự khác biệt rõ rệt về mức độ suy giảm và phục hồi sinh khối rừng trong tỉnh Đắk Lắk. Huyện Krông Bông là khu vực có mức suy giảm sinh khối lớn nhất, với tổng lượng AGB giảm đạt 609,38 nghìn tấn, trong khi lượng AGB tăng chỉ đạt 398,16 nghìn tấn, dẫn đến mức thâm hụt ròng cao nhất toàn tỉnh (–211,21 nghìn tấn). Tương tự, huyện Lắk cũng ghi nhận xu thế suy giảm chiếm ưu thế, với lượng AGB giảm 489,45 nghìn tấn, vượt quá mức tăng 410,67 nghìn tấn, làm cho cân đối sinh khối rừng âm 78,78 nghìn tấn.

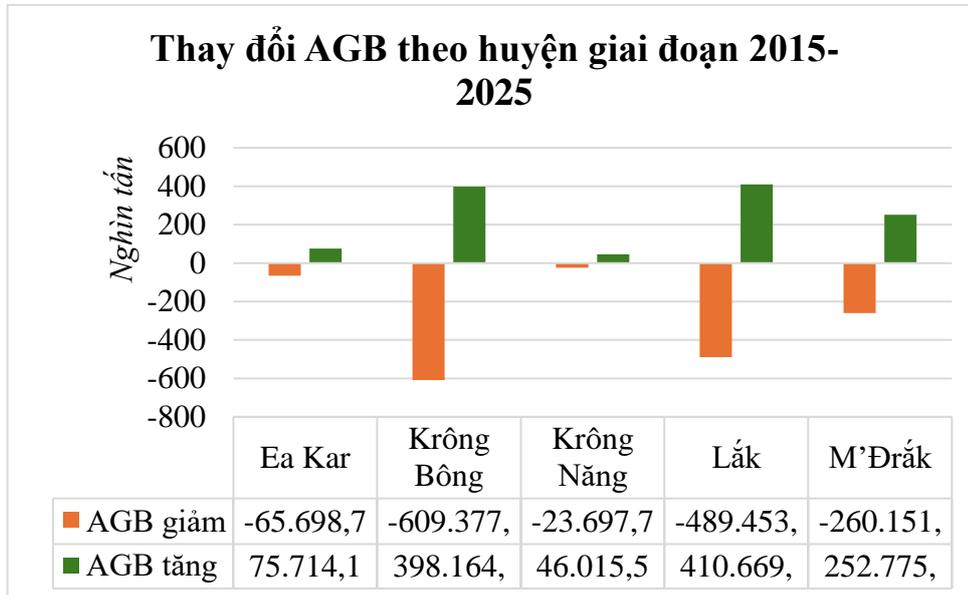
Huyện M'Đrăk có mức suy giảm sinh khối tương đối lớn (260,15 nghìn tấn), song đồng thời cũng ghi nhận lượng AGB tăng đáng kể (252,78 nghìn tấn), do đó mức suy giảm ròng chỉ còn 7,38 nghìn tấn, cho thấy sự bù đắp tương đối giữa mất rừng và phục hồi rừng trong giai đoạn nghiên cứu. Ngược lại, hai huyện Krông Năng và Ea Kar thể hiện xu thế tích cực, với lượng AGB tăng vượt quá lượng giảm, tương ứng đạt cân đối sinh khối dương lần lượt là 22,32 nghìn tấn và 10,02 nghìn tấn. Kết quả này phản ánh quá trình phục hồi sinh khối rừng diễn ra rõ rệt hơn tại hai địa phương này so với các khu vực còn lại trong tỉnh..

Bảng 3.14. Kết quả phân tích định lượng thay đổi sinh khối RTX theo huyện, giai đoạn 2015–2025 (tấn AGB)

Tên huyện	AGB giảm (tấn)	AGB tăng (tấn)	Cân đối
Ea Kar	-65.698,75	75.714,18	10.015,43
Krông Bông	-609.377,52	398.164,82	-211.212,70
Krông Năng	-23.697,75	46.015,53	22.317,78
Lắk	-489.453,63	410.669,59	-78.784,04
M'Đrăk	-260.151,75	252.775,20	-7.376,55
Tổng	-1.448.379,40	1.183.339,32	-265.040,08

Từ kết quả bảng trên, để trực quan hóa sự khác biệt về mức tăng – giảm sinh khối giữa các huyện. Hình 3.24 trình bày biểu đồ thay đổi AGB giai đoạn 2015–2025. Biểu đồ cho thấy rõ ràng sự đối lập: trong khi Ea Kar và Krông Năng có xu hướng tích lũy sinh khối, thì M'Đrăk, Lắk và Krông Bông lại có xu hướng suy giảm đáng kể. Điều này phản ánh sự khác biệt trong quản lý, áp lực sử dụng đất và khả năng

phục hồi rừng giữa các địa phương trong tỉnh.



Hình 3.24. Biểu đồ thay đổi AGB theo huyện giai đoạn 2015-2025

Nhìn chung, kết quả phân tích biến động sinh khối RTX giai đoạn 2015–2025 cho thấy xu hướng suy giảm chiếm ưu thế so với tích lũy, với tổng lượng sinh khối giảm nhiều hơn 265,04 nghìn tấn so với lượng tăng. Tuy nhiên, sự thay đổi này không đồng đều giữa các huyện: Ea Kar và Krông Năng thể hiện khả năng phục hồi và tích lũy sinh khối, trong khi M'Đrắk, Lắk và Krông Bông lại mất mát đáng kể. Những khác biệt này phản ánh sự phức tạp trong tác động của hoạt động quản lý rừng, chuyển đổi sử dụng đất và áp lực khai thác tài nguyên tại từng địa bàn.

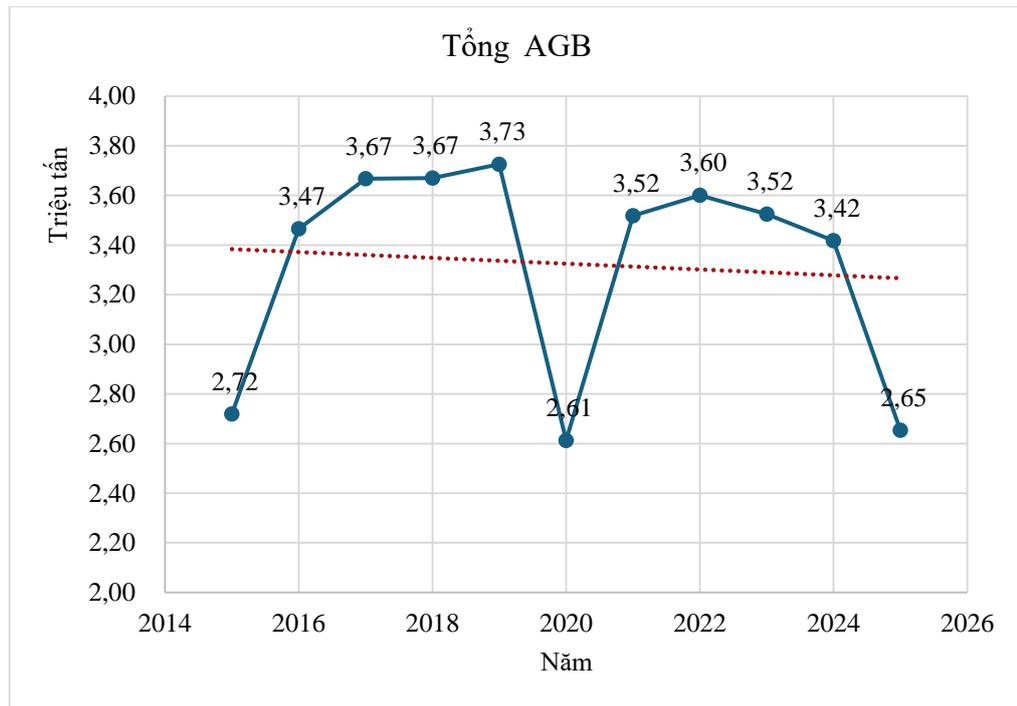
Tổng sinh khối rừng thường xanh của tỉnh trong giai đoạn 2015–2025 có sự dao động đáng kể, phản ánh biến động không ổn định trong trữ lượng sinh khối và khả năng hấp thụ các-bon của hệ sinh thái rừng. Đồ thị thay đổi AGB (Hình 3.25) cho thấy giá trị AGB tăng nhanh trong giai đoạn 2015–2019, từ 2,72 triệu tấn lên 3,73 triệu tấn, tương ứng mức tăng khoảng 37,13%, cho thấy giai đoạn này rừng có xu hướng phục hồi mạnh mẽ. Sự gia tăng này nhiều khả năng gắn liền với các chương trình trồng rừng thay thế và chính sách chi trả dịch vụ môi trường rừng (PFES) được triển khai mạnh sau năm 2015, đặc biệt ở các huyện có độ che phủ lớn như Krông Bông, Lắk và M'Đrắk.

Tuy nhiên, năm 2020, tổng sinh khối giảm đột ngột xuống còn 2,61 triệu tấn, giảm khoảng 30,03% so với năm 2019. Sự suy giảm này có thể phản ánh tác động tổng hợp của nhiều yếu tố, bao gồm áp lực suy thoái rừng, chuyển đổi mục đích sử dụng đất, cũng như ảnh hưởng của các điều kiện khí hậu bất lợi trong giai đoạn này đặc biệt là các cơn bão lớn ảnh hưởng trực tiếp đến khu vực phía Đông và Đông Nam của tỉnh. Ngoài ra, chất lượng đầu vào của ảnh viễn thám đa thời gian không đồng bộ giữa các năm cũng có thể góp phần làm gia tăng mức độ biến động ước tính AGB giữa các năm. Kết quả này cho thấy sinh khối rừng tại khu vực nghiên cứu có độ nhạy cao với các tác động bên ngoài, đặc biệt trong các giai đoạn có biến động mạnh về môi trường và quản lý tài nguyên rừng.

Sau giai đoạn suy giảm mạnh, tổng AGB phục hồi trở lại trong các năm 2021–2023, dao động quanh mức 3,52–3,60 triệu tấn, cho thấy khả năng phục hồi tự nhiên hoặc tác động tích cực từ các hoạt động quản lý và bảo vệ rừng. Tuy nhiên, mức phục hồi này vẫn chưa đạt lại ngưỡng cao nhất trước năm 2020, phản ánh rằng quá trình phục hồi sinh khối diễn ra chậm và chưa bền vững.

Đến giai đoạn 2024–2025, tổng AGB tiếp tục giảm xuống còn 2,65 triệu tấn, tương đương mức của các năm 2015 và 2020. Xu hướng này cho thấy biến động sinh khối không ổn định theo thời gian, với các chu kỳ suy giảm – phục hồi xen kẽ, nhưng xu thế chung vẫn là giảm nhẹ trong toàn giai đoạn (khoảng 0,07 triệu tấn). Điều này nhấn mạnh sự cần thiết của việc tăng cường giám sát rừng theo chuỗi thời gian dài và áp dụng các giải pháp quản lý bền vững nhằm hạn chế suy thoái và nâng cao khả năng tích lũy các-bon của hệ sinh thái rừng.

Về mặt xu thế, nếu xét trung bình, tổng AGB giai đoạn 2015–2025 duy trì ở mức khoảng 3,40 triệu tấn, song thể hiện chu kỳ tăng – giảm rõ rệt, đặc trưng cho biến động rừng nhiệt đới do chịu tác động kết hợp giữa tự nhiên và nhân sinh. Diễn biến này cho thấy, mặc dù có sự phục hồi cục bộ, tổng trữ lượng sinh khối chưa ổn định và có xu hướng suy giảm trong giai đoạn gần đây.



Hình 3.25. Tổng sinh khối rừng giai đoạn 2015 – 2025

Chuỗi biến động này cho thấy:

- Giai đoạn 2015–2020: sinh khối suy giảm, gắn liền với khai thác bất hợp pháp, chuyển đổi mục đích sử dụng đất và suy thoái rừng;
- Giai đoạn 2020–2025: sinh khối có dấu hiệu phục hồi, nhưng chưa đạt mức trước suy giảm.

Nhìn chung, mặc dù vẫn tồn tại các khu vực có khả năng phục hồi sinh khối, song tổng thể toàn tỉnh Đắk Lắk vẫn đang đối diện xu thế suy giảm nhẹ sinh khối rừng, đòi hỏi các giải pháp quản lý bền vững, bảo vệ diện tích rừng hiện có và tăng cường phục hồi rừng ở những khu vực trọng điểm suy thoái.

Do vậy, để xác định liệu các biến động này mang xu thế dài hạn hay chỉ là dao động ngắn hạn, cần phân tích xu thế tổng thể bằng kiểm định thống kê Mann-Kendall kết hợp với ước lượng độ dốc Theil–Sen nhằm đánh giá xu hướng tổng thể của AGB trong giai đoạn 2015–2025.

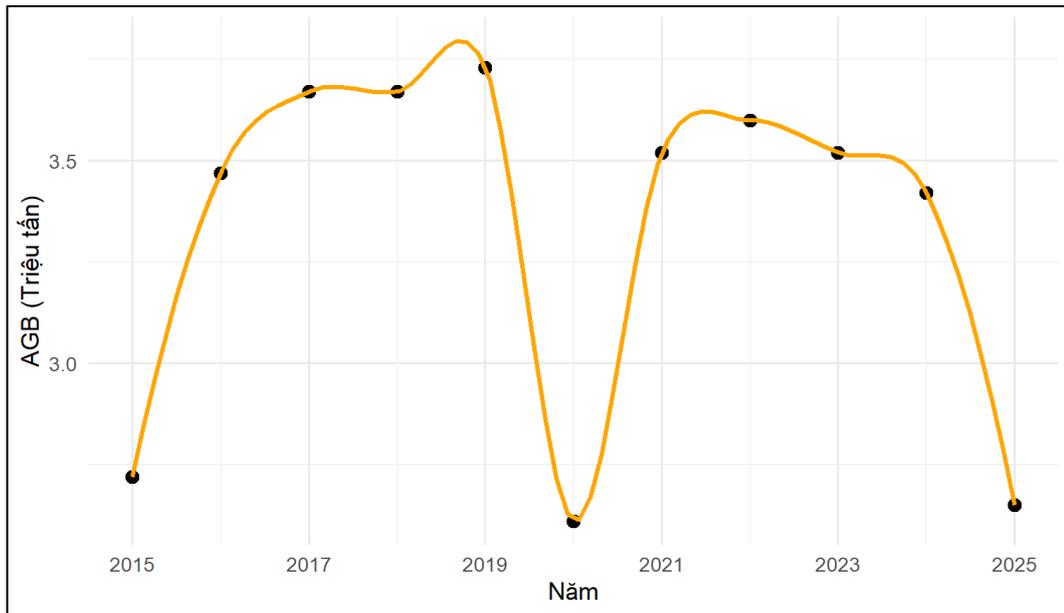
3.4.3. Kiểm định xu thế AGB giai đoạn 2015-2025

Để đánh giá sự biến động dài hạn của sinh khối RTX trong giai đoạn 2015–2025, ngoài việc phân tích bản đồ phân bố và biến động theo từng thời kỳ, nghiên cứu sử dụng các kiểm định thống kê nhằm xác định xu thế thay đổi tổng thể. Phương pháp kiểm định phi tham số Mann–Kendall kết hợp với ước lượng độ dốc Theil–Sen được áp dụng để phát hiện xu hướng tăng hoặc giảm có ý nghĩa thống kê trong chuỗi sinh khối. Đây là phương pháp phổ biến trong nghiên cứu thủy văn, khí hậu và tài nguyên rừng [102],[165], với ưu điểm không yêu cầu phân phối chuẩn và ít bị ảnh hưởng bởi ngoại lai. Các nghiên cứu tiếp theo cũng cho thấy MK test đặc biệt phù hợp với chuỗi dữ liệu ngắn và có nhiễu – đặc trưng điển hình của sinh khối ước tính từ ảnh viễn thám [84],[190].

Song song với kiểm định thống kê, nghiên cứu sử dụng đường làm mượt LOESS để mô tả xu thế phi tuyến của tổng sinh khối theo thời gian. LOESS cho phép nhận diện tốt các pha tăng – giảm ngắn hạn và thể hiện rõ dạng biến động không đơn điệu (non-monotonic), vốn là đặc trưng của rừng nhiệt đới chịu tác động của khai thác chọn, cháy rừng, suy thoái cục bộ hoặc phục hồi sau nhiễu loạn [45],[132]. Việc kết hợp LOESS và MK tạo cơ sở trực quan và thống kê để đánh giá toàn diện xu thế tổng sinh khối RTX. Kết quả tổng hợp thể hiện trong Bảng 3.15 và Hình 3.26.

Bảng 3.15. Tổng sinh khối RTX giai đoạn 2015-2025

Năm	AGB_total (triệu tấn)	Năm	AGB_total (triệu tấn)
2015	2,72	2021	3,52
2016	3,47	2022	3,60
2017	3,67	2023	3,52
2018	3,67	2024	3,42
2019	3,73	2025	2,65
2020	2,61		



Hình 3.26. Đường LOESS mô tả xu thế phi tuyến của tổng sinh khối RTX giai đoạn 2015–2025

Bảng 3.15 cho thấy tổng sinh khối AGB dao động trong khoảng 29,0–41,4 triệu tấn. Sinh khối tăng rõ rệt giai đoạn 2015–2019, giảm mạnh vào năm 2020, phục hồi trong giai đoạn 2021–2023 và tiếp tục suy giảm vào năm 2024–2025.

Đường LOESS thể hiện rõ xu thế phi tuyến này: tăng mạnh (2015–2019), giảm sâu (2020), phục hồi (2021–2023), sau đó giảm trở lại (2024–2025). Mẫu hình tăng–giảm xen kẽ cho thấy sinh khối chủ yếu chịu tác động của các yếu tố ngắn hạn như mất rừng cục bộ, khai thác chọn hoặc biến động khí hậu theo mùa.

Kết quả kiểm định MK và Sen's slope: Kết quả kiểm định MK đối với tổng sinh khối RTX giai đoạn 2015–2025 (Bảng 3.16) cho thấy hệ số $\tau = -0,1667$ với $p = 0,53 (> 0,05)$, nghĩa là không tồn tại xu thế dài hạn có ý nghĩa thống kê. Giá trị Sen's slope $-0,0175$ triệu tấn/năm phản ánh xu hướng giảm nhẹ theo thời gian nhưng cũng không có ý nghĩa thống kê. Điều này chứng tỏ sự giảm sinh khối không mang tính hệ thống mà chủ yếu xuất phát từ các biến động ngắn hạn.

Kết quả này hoàn toàn phù hợp với dạng xu thế phi tuyến từ đường LOESS: chuỗi AGB không tăng hoặc giảm đều mà biến động theo nhiều pha đối lập, do đó MK không phát hiện được xu thế đơn điệu.

Bảng 3.16. Kết quả kiểm định xu thế AGB giai đoạn 2015–2025

Chỉ số thống kê	Giá trị	Ý nghĩa thống kê
Hệ số <i>tau</i>	-0,1667	Xu thế giảm nhẹ, không có ý nghĩa thống kê
<i>Sen's slope</i>	-0,0175	$p = 0,53 (>0,05)$, không có xu thế rõ rệt.
Số quan sát (năm)	11	2015 - 2025

Như vậy, mặc dù tổng sinh khối có sự dao động đáng kể qua các năm, nhưng kiểm định thống kê cho thấy xu thế biến đổi AGB trong giai đoạn 2015–2025 là không đáng kể, chủ yếu chịu ảnh hưởng bởi các yếu tố ngắn hạn như mất rừng cục bộ hoặc phục hồi sinh khối ở những khu vực nhất định. Trên cơ sở đó, nghiên cứu tiếp tục sử dụng các khu vực có xu hướng tăng AGB để phân tích khả năng hấp thụ và lưu giữ các-bon của RTX, phục vụ định hướng quản lý rừng bền vững và chiến lược giảm phát thải tại Đắk Lắk.

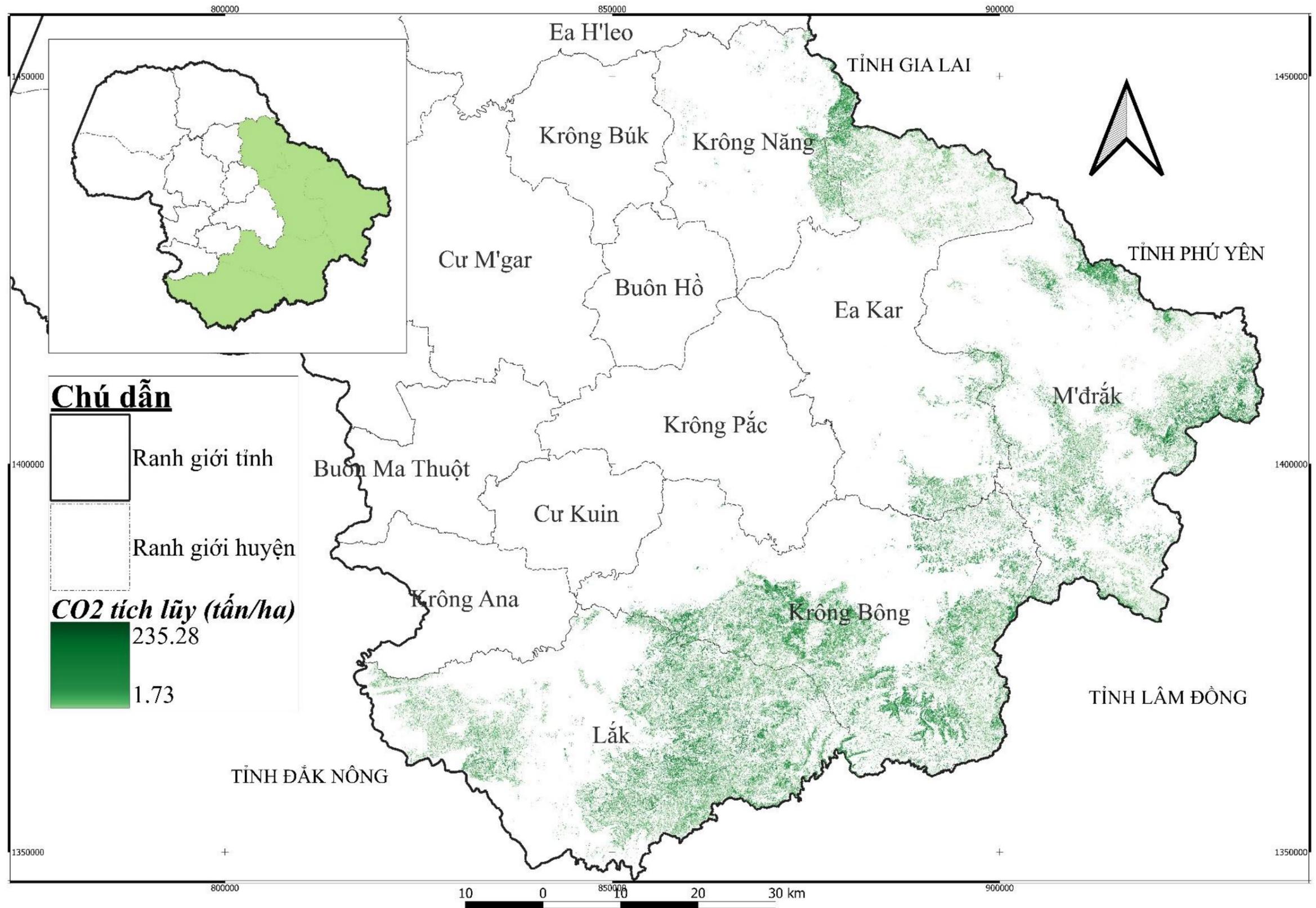
3.4.4. Ước tính khả năng hấp thụ CO₂ của rừng

Sinh khối trên mặt đất của rừng không chỉ phản ánh tiềm năng sản xuất và tình trạng sinh thái của hệ sinh thái rừng mà còn là cơ sở quan trọng để ước tính lượng các-bon được tích lũy và khả năng hấp thụ khí CO₂ từ khí quyển. Với đặc tính quang hợp, rừng hấp thụ CO₂, chuyển hóa thành chất hữu cơ và lưu trữ lâu dài trong sinh khối cây, qua đó góp phần giảm phát thải khí nhà kính và điều hòa khí hậu toàn cầu [47],[133].

Trong bối cảnh biến đổi khí hậu ngày càng phức tạp, việc định lượng khả năng hấp thụ CO₂ của RTX tại tỉnh Đắk Lắk có ý nghĩa đặc biệt quan trọng. Đây không chỉ là cơ sở khoa học để đánh giá vai trò của rừng trong cân bằng các-bon, mà còn cung cấp thông tin thiết yếu cho việc xây dựng các chính sách quản lý rừng bền vững, tham gia các cơ chế chi trả dịch vụ môi trường rừng (PFES), giảm phát thải từ mất rừng và suy thoái rừng (REDD⁺) hay trung hòa các-bon.

Hình 3.27 thể hiện phân bố tích lũy CO₂ của RTX tỉnh Đắk Lắk trong giai đoạn 2015–2025. Nhìn chung, giá trị CO₂ tích lũy có sự khác biệt rõ rệt giữa các khu

vực, phản ánh sự phân hóa về phân bố sinh khối rừng. Các khu vực có giá trị tích lũy CO₂ cao (từ 150–235 tấn/ha) tập trung chủ yếu tại huyện Lắk, phía nam Krông Bông và một phần huyện M'Đrăk, nơi còn duy trì diện tích rừng tự nhiên tương đối lớn, địa hình đồi núi, độ ẩm cao và mật độ dân cư thấp. Những điều kiện này hạn chế tác động suy thoái và tạo thuận lợi cho việc tích lũy sinh khối – các-bon lâu dài. Ngoài ra các khu vực phía bắc Krông Năng và rìa phía đông M'Đrăk có nhiều mảng rừng có khả năng tích lũy CO₂ cao, chủ yếu do địa hình đồi núi, điều kiện ẩm thuận lợi và còn duy trì được các quần thể rừng tự nhiên thứ sinh ít bị khai thác. Đặc điểm này giúp thảm rừng tại đây duy trì mật độ tán lá dày, khả năng quang hợp và tốc độ tăng trưởng sinh khối cao hơn so với các vùng dễ tiếp cận hoặc khu vực chịu tác động sản xuất nông nghiệp. Kết quả này phù hợp với xu hướng chung đã được ghi nhận ở các nghiên cứu quy mô nhiệt đới, cho thấy rừng ẩm thứ sinh hoặc rừng tái sinh trưởng thành tại Đông Nam Á có khả năng tích lũy CO₂ cao tương đương hoặc thậm chí vượt một số vùng rừng nguyên sinh nhờ tốc độ phục hồi mạnh mẽ sau suy thoái [35],[45],[159].



Hình 3.27. Bản đồ phân bố CO₂ tích lũy giai đoạn 2015-2025

Ngược lại, các khu vực có giá trị tích lũy thấp (dưới 50 tấn/ha) phân bố rải rác ở những vùng đã chịu tác động mạnh của khai thác gỗ trong quá khứ, chuyển đổi mục đích sử dụng đất hoặc chịu áp lực từ sản xuất nông nghiệp lâu dài, điển hình là vùng trung tâm Ea Kar và một số xã phía đông Krông Bông. Sự khác biệt này cho thấy tiềm năng hấp thụ CO₂ của rừng không đồng đều, phụ thuộc vào cấu trúc và tình trạng rừng. Xu hướng này tương tự với nhận định của Avitabile và cộng sự (2016), Saatchi và cộng sự (2011) đều chỉ ra rằng rừng bị phân mảnh hoặc suy thoái do canh tác nông nghiệp thường có mật độ sinh khối thấp hơn đáng kể so với rừng tự nhiên nguyên vẹn, dẫn đến khả năng lưu giữ và hấp thụ các-bon giảm rõ rệt.

Về mặt định lượng, tổng lượng CO₂ tích lũy của RTX trên toàn tỉnh Đắk Lắk giai đoạn 2015–2025 đạt khoảng 2,04 triệu tấn. Giá trị tích lũy CO₂ dao động trong khoảng 1,73 – 235,98 tấn/ha, với giá trị trung bình đạt 24,72 tấn/ha. Độ lệch chuẩn tương đối lớn (25,75 tấn/ha) cho thấy mức độ biến động cao về khả năng hấp thụ CO₂ giữa các khu vực rừng khác nhau. Điều này phản ánh sự khác biệt trong cấu trúc và chất lượng rừng: những khu vực rừng tự nhiên giàu trữ lượng có khả năng tích lũy CO₂ cao, trong khi các khu vực rừng suy thoái hoặc bị tác động mạnh chỉ tích lũy ở mức thấp.

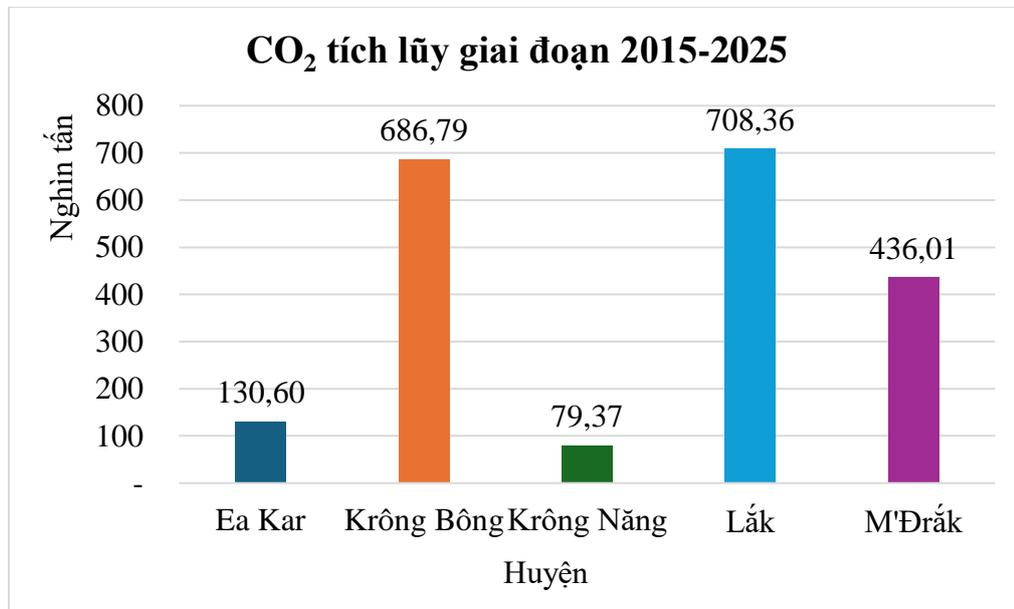
Bảng 3.17 dưới đây thể hiện thống kê mô tả về lượng CO₂ tích lũy của rừng thường xanh giai đoạn 2015–2025:

Bảng 3.17. Thống kê lượng CO₂ tích lũy RTX tỉnh Đắk Lắk giai đoạn 2015–2025

Chỉ tiêu	Giá trị
Minimum value (tấn/ha)	1,73
Maximum value (tấn/ha)	235,98
Range (tấn/ha)	234,26
Sum (tấn CO ₂)	2.041.141,99
Mean value (tấn/ha)	24,72
Standard deviation (tấn/ha)	25,75

Kết quả thống kê CO₂ tích lũy cho thấy có sự khác biệt rõ rệt giữa các huyện trong tỉnh Đắk Lắk giai đoạn 2015–2025 (Hình 3.28). Trong đó, huyện Lắk là khu vực có lượng CO₂ tích lũy lớn nhất, đạt 707,36 nghìn tấn, chiếm tỷ lệ cao nhất trong

toàn tỉnh. Tiếp theo là huyện Krông Bông với 686,79 nghìn tấn, đóng vai trò là vùng rừng trọng điểm về khả năng hấp thụ và lưu giữ CO₂. Các huyện M'Đrăk và Ea Kar lần lượt đạt 436,01 nghìn tấn và 130,60 nghìn tấn, thể hiện mức tích lũy trung bình. Trong khi đó, huyện Krông Năng ghi nhận lượng CO₂ tích lũy thấp nhất, chỉ khoảng 79,37 nghìn tấn, phản ánh diện tích RTX hạn chế và mức độ suy thoái cao hơn so với các khu vực khác. Xu hướng phân bố này phản ánh quy luật chung về mối quan hệ giữa cấu trúc và chất lượng rừng với khả năng hấp thụ CO₂. Các nghiên cứu tại vùng nhiệt đới cũng ghi nhận rằng rừng tự nhiên và rừng đầu nguồn có mật độ sinh khối cao thường tích lũy CO₂ vượt trội so với các khu vực rừng suy thoái hoặc chuyển đổi mục đích sử dụng đất [35],[159].



Hình 3.28. Biểu đồ tổng tích lũy CO₂ theo huyện

Như vậy, có thể nhận định rằng sự phân bố không gian của CO₂ tích lũy trên địa bàn tỉnh Đắk Lắk chịu ảnh hưởng mạnh bởi diện tích, trạng thái rừng và mức độ tác động nhân sinh. Những huyện có diện tích rừng thường xanh lớn, ít bị chuyển đổi mục đích sử dụng đất (như Lắk, Krông Bông) có khả năng hấp thụ và lưu giữ CO₂ cao hơn rõ rệt so với những huyện có áp lực khai thác, chuyển đổi rừng lớn. Kết quả này góp phần khẳng định vai trò quan trọng của RTX trong điều hòa khí hậu và giảm phát thải khí nhà kính.

KẾT LUẬN VÀ KIẾN NGHỊ

Kết luận

Luận án “Ước lượng khả năng hấp thụ CO₂ của kiểu RTX trên địa bàn tỉnh Đắk Lắk dựa vào dữ liệu viễn thám và kỹ thuật GIS” đã đạt được các mục tiêu nghiên cứu đề ra và có những đóng góp quan trọng về mặt khoa học và thực tiễn trong lĩnh vực quản lý tài nguyên rừng, giám sát các-bon và ứng phó biến đổi khí hậu.

Luận án đã hệ thống hóa và cập nhật các tiến bộ trong ứng dụng công nghệ viễn thám và GIS cho ước tính sinh khối rừng, đồng thời lựa chọn và xây dựng quy trình phân tích phù hợp với điều kiện sinh thái dữ liệu của tỉnh Đắk Lắk; Phương pháp nghiên cứu tích hợp gồm: (i) xử lý ảnh Landsat và Sentinel; (ii) tính toán và lựa chọn các chỉ số phổ – ra-đa có ý nghĩa; (iii) kiểm định thống kê (Pearson, PCA, hồi quy đơn và đa biến); và (iv) mô hình học máy Random Forest (RF) kết hợp kiểm định chéo phân tầng k-fold. Kết hợp giữa dữ liệu quang học (B7, MSI, NDWI), dữ liệu địa hình (elevation, slope) và dữ liệu ra-đa (VH) đã chứng minh hiệu quả trong mô phỏng biến động sinh khối rừng. Phép biến đổi lô-ga-rít của AGB được áp dụng giúp giảm sai số và nâng cao độ ổn định thống kê của các mô hình ước lượng.

Kết quả phân loại thảm phủ bằng thuật toán RF đạt độ chính xác tổng thể trung bình 94,24% và chỉ số Kappa 0,88, thể hiện khả năng nhận dạng tốt RTX. Phân bố diện tích rừng theo không gian cho thấy RTX của tỉnh tập trung phân bố chủ yếu ở phía Đông – Nam, nơi có địa hình cao và dốc, khó tiếp cận.

Dữ liệu AGB ô mẫu sau khi biến đổi lô-ga-rít cho thấy phân bố dữ liệu trở nên chuẩn hóa hơn, tạo điều kiện thuận lợi cho các mô hình hồi quy và học máy. Kết quả phân tích tương quan và hồi quy cho thấy các biến B7, MSI, NDWI, VH, elevation và slope có mối quan hệ có ý nghĩa thống kê với logAGB ($p < 0,05$). Kết hợp ảnh quang học, địa hình và ra-đa với mô hình RF có cải thiện độ chính xác của mô hình hồi quy giữa AGB với các nhân tố viễn thám. Mô hình hồi quy RF cho kết quả ước lượng có độ tin cậy cao hơn so với mô hình hồi quy tuyến tính đa biến và mô

hình hồi quy phi tuyến GAM với hệ số R^2 lần lượt là 0,51, 0,59 và 0,88. Kiểm định chéo phân tầng k-fold RF vượt trội nhất trong cả hai cấu hình k, với: R^2 cao nhất (0,55), RMSE thấp nhất ($\sim 0,818$), MAE thấp nhất ($\sim 0,637$). Điều này chứng minh ưu thế của RF khi mô hình hóa mối quan hệ phi tuyến phức tạp và tương tác giữa nhiều biến viễn thám đồng thời. Sử dụng dữ liệu độc lập để đánh giá sai số ước lượng cho thấy mô hình RF áp dụng cho ảnh vệ tinh quang học có hệ số $R^2 = 0,132$ và các sai số $RMSE = 1,304$, $MAE = 1,087$ cho thấy ảnh vệ tinh quang học đơn lẻ không giải thích tốt mối quan hệ giữa sinh khối và nhân tố ảnh viễn thám so với kết hợp ảnh radar và ảnh quang học.

Tổng lượng CO_2 tích lũy của rừng tự nhiên toàn tỉnh giai đoạn 2015-2025 ước đạt khoảng 2,04 triệu tấn CO_2 . Tính trung bình, lượng CO_2 hấp thụ của RTX trong khu vực nghiên cứu đạt trung bình 24,72 tấn CO_2/ha . Các địa phương có tích lũy CO_2 cao trong giai đoạn này là huyện Krông Bông và huyện Lắk với lượng CO_2 tích lũy lần lượt là 686,79 và 708,36 nghìn tấn CO_2 .

Những kết quả này không chỉ góp phần khẳng định vai trò của rừng trong chu trình các-bon, mà còn là cơ sở khoa học quan trọng để hoạch định chính sách quản lý, bảo vệ và phát triển rừng, cũng như tham gia vào các cơ chế chi trả dịch vụ môi trường rừng, giảm phát thải từ mất rừng và suy thoái rừng (REDD+) và trung hòa các-bon.

Tồn tại

Mặc dù luận án đã đạt được các mục tiêu nghiên cứu đề ra và xây dựng được mô hình ước tính sinh khối và khả năng hấp thụ CO_2 cho rừng thường xanh tỉnh Đắk Lắk với độ tin cậy khá tốt, nghiên cứu vẫn tồn tại một số hạn chế nhất định.

Trước hết, dữ liệu thực địa còn hạn chế cả về số lượng và mức độ đồng bộ theo thời gian so với quy mô lớn và tính phân mảnh cao của rừng trong toàn tỉnh. Việc thiếu đồng nhất về thời điểm thu thập giữa dữ liệu thực địa và ảnh viễn thám, cùng với việc chưa thu thập đầy đủ các thông tin cấu trúc rừng chi tiết, có thể làm gia

tăng sai số và hạn chế khả năng diễn giải sinh thái của mô hình.

Thứ hai, hạn chế từ dữ liệu viễn thám chủ yếu xuất phát từ độ phân giải không gian của ảnh Landsat và Sentinel (10–30 m), chưa phản ánh đầy đủ cấu trúc rừng hỗn loài, nhiều tầng và sinh khối cao. Bên cạnh đó, ảnh quang học chịu ảnh hưởng của mây che và điều kiện khí quyển, trong khi dữ liệu radar C-band có xu hướng bão hòa ở rừng giàu sinh khối. Việc chưa tích hợp dữ liệu LiDAR mặt đất hoặc không gian cũng là một hạn chế trong việc mô tả cấu trúc thẳng đứng của rừng.

Thứ ba, về mô hình hóa, luận án mới tập trung thử nghiệm một số nhóm mô hình đại diện (OLS, GAM và RF). Mặc dù RF cho kết quả tối ưu, nghiên cứu chưa mở rộng so sánh với các mô hình học máy nâng cao hơn, chưa đánh giá đầy đủ khả năng chuyển tiếp mô hình theo thời gian và chưa phân tích sâu sự khác biệt sinh thái theo các điều kiện địa hình – môi trường khác nhau. Do đó, khả năng khái quát và mở rộng mô hình sang các kiểu rừng hoặc vùng sinh thái khác vẫn còn hạn chế.

Cuối cùng, về tính ứng dụng quản lý, các kết quả ước tính sinh khối và CO₂ mới dừng ở mức cung cấp cơ sở khoa học và dữ liệu tham khảo, chưa được tích hợp đầy đủ vào các hệ thống MRV phục vụ REDD⁺ hoặc PFES.

Kiến nghị

Trên cơ sở những kết quả đạt được và các hạn chế còn tồn tại của luận án, tác giả đề xuất một số kiến nghị sau nhằm nâng cao hiệu quả nghiên cứu và ứng dụng trong thời gian tới:

Thứ nhất, về dữ liệu thực địa, cần tiếp tục mở rộng và chuẩn hóa hệ thống ô mẫu rừng theo hướng tăng số lượng, cải thiện tính đại diện và đồng bộ hóa thời gian thu thập với dữ liệu viễn thám. Việc bổ sung các thông tin cấu trúc rừng như chiều cao tán, mật độ cây, tái sinh và đa dạng loài sẽ góp phần nâng cao độ chính xác của mô hình ước tính sinh khối, đồng thời tăng khả năng diễn giải sinh thái của kết quả nghiên cứu.

Thứ hai, về dữ liệu viễn thám, cần đẩy mạnh tích hợp các nguồn dữ liệu có

khả năng mô tả cấu trúc thẳng đứng của rừng, đặc biệt là dữ liệu LiDAR mặt đất và LiDAR không gian (GEDI, ICESat-2), kết hợp với ảnh quang học và ra-đa đa thời gian. Cách tiếp cận đa nguồn này sẽ giúp khắc phục hiện tượng bão hòa tín hiệu ở rừng sinh khối cao, nâng cao độ tin cậy của bản đồ sinh khối và CO₂ hấp thụ.

Thứ ba, về phương pháp và mô hình hóa, các nghiên cứu tiếp theo nên mở rộng so sánh các thuật toán học máy nâng cao như Gradient Boosting, XGBoost, hoặc các thuật toán ANN, Deep learning. Việc phân tích độ nhạy của sinh khối theo các điều kiện sinh thái – địa hình khác nhau cũng là hướng nghiên cứu cần thiết nhằm làm rõ cơ chế biến động sinh khối rừng.

Thứ tư, về tính ứng dụng quản lý, cần từng bước tích hợp kết quả ước tính sinh khối và CO₂ vào các hệ thống MRV phục vụ REDD+ và chi trả dịch vụ môi trường rừng (PFES). Đồng thời, việc xây dựng các quy trình kỹ thuật và bộ công cụ đơn giản hóa trên nền tảng GIS và viễn thám sẽ giúp các cơ quan quản lý rừng địa phương có thể áp dụng thường xuyên và hiệu quả hơn trong công tác theo dõi diễn biến rừng.

Nhìn chung, các kiến nghị trên không chỉ góp phần hoàn thiện hướng nghiên cứu trong tương lai mà còn tạo tiền đề để chuyển hóa các kết quả khoa học của luận án thành công cụ hỗ trợ thiết thực cho quản lý và sử dụng bền vững tài nguyên rừng trong bối cảnh biến đổi khí hậu và thực hiện các cam kết giảm phát thải của Việt Nam.

DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ

+ 01 (Một) bài báo đăng trên Tạp chí Khoa học quốc tế có chỉ số SCOPUS

1. **Ho Dinh Bao**, Nguyen Thi Thanh Huong (2025), “Estimating forest above ground biomass in Dak Lak province, Vietnam”, *Asian Journal of Forestry*, Volume 9, Number 1, June 2025, DOI: 10.13057/asianjfor/r090112 (SCOPUS, Q3).

+ 02 Bài báo đăng trên Tạp chí Khoa học trong nước thuộc hệ thống tính điểm của Hội đồng Giáo sư Nhà nước

2. **Hồ Đình Bảo**, Nguyễn Thị Thanh Hương (2023). “Sử dụng ảnh viễn thám để ước tính sinh khối rừng trên mặt đất: Thách thức và triển vọng” *Tạp chí Khoa học Lâm nghiệp*, Số 6/2023, tr. 125-147.
3. **Hồ Đình Bảo**, Nguyễn Thế Hiển, Lê Thị Mai Trang, Nguyễn Thị Thanh Hương (2025). “So sánh liệu quả phân loại thảm phủ tỉnh Đắk Lắk từ dữ liệu ảnh viễn thám bằng thuật toán học máy” *Tạp chí Khoa học Lâm nghiệp*, Số 1/2025: 61-75. DOI: <https://doi.org/10.70169/VJFS.1014>.

+ 01 (một) báo cáo được trình bày tại Hội thảo Quốc tế về Đa dạng sinh học, Khoa học và Công nghệ - BioSat, tháng 01 năm 2024, Pattaya, Thái Lan.

4. **Ho Dinh Bao**, Nguyen Thi Thanh Huong, Nguyen Thi Tuyet Nhung (2024). “Using satellite images and statistics for Estimating Aboveground Forest Biomass change in Dak Lak province – Vietnam” *International conference on Biodiversity, Science, and Technology - BioSat*, Jan 2024, Pattaya, Thailand.

TÀI LIỆU THAM KHẢO

Tiếng Việt

- [1]. **Ngô Đức Anh, Đoàn Thị The, Nguyễn Thị Thu Thủy (2021)**, “Thành lập bản đồ lớp phủ rừng từ ảnh vệ tinh Landsat 8 bằng phương pháp phân loại dựa trên đối tượng kết hợp giải đoán bằng mắt thường”, *Tạp chí Khoa học và Công nghệ Lâm nghiệp*, số 3 - 2021.
- [2]. **Phạm Tuấn Anh và Bảo Huy (2016)**, “Mô hình ước tính sinh khối rừng sử dụng biến số đầu vào cộng đồng có khả năng đo đạc”, *Tạp chí Nông nghiệp và PTNT*, số 23(2016): 98-107.
- [3]. **Hồ Đình Bảo, Nguyễn Thế Hiền, Lê Thị Mai Trang, Nguyễn Thị Thanh Hương (2025)**, “So sánh hiệu quả phân loại thảm phủ tỉnh Đắk Lắk từ dữ liệu ảnh viễn thám bằng thuật toán học máy”, *Tạp chí Khoa học Lâm nghiệp*, Số 1/2025: 61-75. DOI: <https://doi.org/10.70169/VJFS.1014>.
- [4]. **Bộ Chính trị (2024)**, *Nghị quyết số 57-NQ/TW về đột phá phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia (NQ-57-TW)*. Công Thông tin điện tử Chính phủ.
- [5]. **Chính phủ nước Cộng hòa Xã hội Chủ nghĩa Việt Nam (2022)**, *Nghị định số 06/2022/NĐ-CP của Chính phủ: Quy định giảm nhẹ phát thải khí nhà kính và bảo vệ tầng Ô-Dôn*, Thủ tướng Chính phủ, ngày 07-01-2022
- [6]. **Chính phủ nước Cộng hòa Xã hội Chủ nghĩa Việt Nam (2022)**, *Quyết định số 888/QĐ-TTg của Thủ tướng Chính phủ: Phê duyệt Đề án về những nhiệm vụ, giải pháp triển khai kết quả Hội nghị lần thứ 26 các bên tham gia Công ước khung của Liên hợp quốc về biến đổi khí hậu*, Thủ tướng Chính phủ, ngày 25-07-2022.
- [7]. **Chính phủ nước Cộng hòa Xã hội Chủ nghĩa Việt Nam (2022)**, *Quyết định số 896/QĐ-TTg của Thủ tướng Chính phủ: Phê duyệt Chiến lược quốc gia về biến đổi khí hậu giai đoạn đến năm 2050*, Thủ tướng Chính phủ, ngày 26-07-2022.
- [8]. **Chính phủ nước Cộng hòa Xã hội Chủ nghĩa Việt Nam (2024)**, *Chỉ thị số 13/CT-TTg về tăng cường công tác quản lý rừng và các-bon nhằm thực hiện đóng góp do*

quốc gia tự quyết định, Thủ tướng chính phủ, ngày 02/04/2024, Công Thông tin điện tử Chính phủ.

- [9]. **Chính phủ nước Cộng hòa Xã hội Chủ nghĩa Việt Nam (2025)**, *Quyết định số 232/QĐ-TTg về phê duyệt Đề án thành lập và phát triển thị trường các-bon Việt Nam*. Thủ tướng Chính phủ, ngày 24/01/2025, Công Thông tin điện tử Chính phủ.
- [10]. **Chính phủ nước Cộng hòa Xã hội Chủ nghĩa Việt Nam (2025)**, *Nghị định số 119/2025/NĐ-CP của Chính phủ: Sửa đổi, bổ sung một số điều của Nghị định số 06/2022/NĐ-CP ngày 07 tháng 01 năm 2022 của Chính phủ quy định giảm nhẹ phát thải khí nhà kính và bảo vệ tầng ô-dôn*, Thủ tướng Chính phủ, ngày 09-06-2025.
- [11]. **Đinh Bá Duy, (2024)**, *Nghiên cứu biến động sinh khối thực vật và năng lực hấp thụ carbon tại các hệ sinh thái rừng Cao nguyên Kon Hà Nừng, tỉnh Gia Lai*, Báo cáo khoa học, Đề tài nghiên cứu khoa học Trung tâm nhiệt đới Việt - Nga.
- [12]. **Võ Đại Hải (2009)**, “Nghiên cứu khả năng hấp thụ các-bon của rừng trồng Bạch đàn *Urophylla* ở Việt Nam”, *Tạp chí NN & PTNT*, số 1(2009): 102-106
- [13]. **Võ Đại Hải và Đặng Thịnh Triều (2012)**, *Nghiên cứu khả năng hấp thụ các-bon của rừng tự nhiên lá rộng thường xanh, bán thường xanh và rụng lá ở Tây Nguyên*, Báo cáo khoa học đề tài nghiên cứu khoa học Bộ Nông nghiệp và Phát triển Nông thôn.
- [14]. **Cao Thị Hoài, Nguyễn Thị Thanh Hương (2025)**, “So sánh phương pháp phân loại thảm phủ sử dụng đất bằng thuật toán Random Forest sử dụng chỉ số NDVI với tổ hợp chỉ số NDVI, DEM, NDWI, NDSI tại huyện Đắk Glong và huyện Krông Nô – tỉnh Đắk Nông”, *Tạp chí Khoa học Lâm Nghiệp*, 10.70169/VJFS.1060.
- [15]. **Bảo Huy (2012)**, *Xác định lượng CO₂ hấp thụ của RTX vùng Tây Nguyên làm cơ sở tham gia chương trình giảm thiểu khí phát thải từ suy thoái và mất rừng*, Đề tài Khoa học và Công nghệ cấp Bộ, Đắk Lắk, 2012.
- [16]. **Bảo Huy (2013)**, *Mô hình sinh trắc và viễn thám – GIS để xác định CO₂ hấp thụ của RTX vùng Tây Nguyên*, NXB. Khoa học và Kỹ thuật, Hà Nội.
- [17]. **Bảo Huy (2016)**, “Phương pháp thiết lập mô hình ước tính sinh khối, carbon rừng theo kiểu rừng và vùng sinh thái: Trường hợp RTX vùng nam trung bộ”, *Tạp chí Nông nghiệp và PTNT*, số 10 (2016): 121-130.

- [18]. **Nguyễn Thị Thanh Hương (2022)**, *Phương pháp ứng dụng viễn thám và GIS trong Lâm nghiệp*. Nhà xuất bản Đại học Huế, 2022.
- [19]. **Nguyễn Thị Thanh Hương, Cao Thị Hoài, Nguyễn Quang Toan, Hồ Đình Bảo, Nguyễn Thế Hiển, Ngô Thế Sơn, Phan Thị Hằng, Nguyễn Đức Định, Đặng Thị Thúy Kiều (2025)**. *Ứng dụng dữ liệu viễn thám và GIS để phân vùng khô hạn, đánh giá một số loại sử dụng đất làm cơ sở định hướng phát triển nông lâm nghiệp thích ứng với biến đổi khí hậu tại tỉnh Đắk Lắk*. Đề tài Khoa học và công nghệ cấp tỉnh Đắk Lắk (2023-2025).
- [20]. **Nguyễn Thị Thanh Hương và Đoàn Minh Trung (2018)**, “Áp dụng thuật toán phân loại Random Forest để xây dựng bản đồ sử dụng đất/thảm phủ tỉnh Đắk Lắk dựa vào ảnh vệ tinh Landsat 8 OLI”, *Tạp chí Nông nghiệp & Phát triển nông thôn*, 13/2018: 122-129.
- [21]. **Mai Thị Huyền, Phan Trọng Thế, Phạm Thị Lộc (2023)**, “Đánh giá độ chính xác của các phương pháp phân loại thảm phủ dựa trên ảnh Sentinel-2 và Landsat 9”, *Tạp chí Khoa học và Công nghệ Lâm nghiệp*, Tập 12, số 4 (2023). <https://doi.org/10.55250/jo.vnuf.12.4.2023.060-069>.
- [22]. **Đỗ Thị Nhung, Đặng Đỗ Lâm Phương, Nguyễn Thị Diễm My, Bùi Quang Thành, Phạm Ngọc Hải, Phạm Văn Mạnh (2024)**, “Đánh giá trữ lượng carbon tích lũy trong hệ sinh thái rừng nhiệt đới có giá trị bảo tồn cao: Nghiên cứu trường hợp tại tỉnh Bắc Kạn, Việt Nam”, *VNU Journal of Science: Earth and Environmental Sciences*, Vol. 40, No. 2 (2024) 92-110. <https://doi.org/10.25073/2588-1094/vnuees.5069>.
- [23]. **Nguyễn Thị Ngọc Quyên, Nguyễn Công Tài Anh, Bùi Tá Long, Nguyễn Kim Lợi (2015)**, “Ứng dụng viễn thám và GIS trong thành lập bản đồ thảm phủ lưu vực Srepok vùng Tây Nguyên”, *Tạp chí Khoa học Nông nghiệp Việt Nam*, 2016; tập 14, số 5: 734-743.
- [24]. **Vũ Thị Phương Thảo và Souksakone Sengchanh (2022)**, “Khả năng xây dựng bản đồ lớp phủ mặt đất từ dữ liệu viễn thám Sentinel-2 theo phương pháp phân loại Random Forest trên nền tảng điện toán đám mây”, *Tạp chí Khoa học Đo đạc và Bản đồ*, Số 52-6/2022.

- [25]. **Nguyễn Văn Thị, Nguyễn Đình Dương, Trần Quang Bảo (2018)**, “Quan hệ giữa tán xạ ngược của ảnh Radar Sentinel-1 với chỉ số NDVI của ảnh quang học Sentinel-2: Trường hợp nghiên cứu cho đối tượng rừng khộp tại tỉnh Đắk Lắk”, *Tạp chí Khoa học và Công nghệ Lâm nghiệp*, Số 3-2018.
- [26]. **Huỳnh Nhân Trí (2014)**, *Xây dựng các cơ sở khoa học và thực tiễn để giám sát lượng CO₂ hấp thụ của RTX ở Tây Nguyên*, Luận án Tiến Sĩ chuyên ngành Lâm học, Viện Khoa học Lâm nghiệp Việt Nam.
- [27]. **Nguyễn Thanh Tuấn, Nguyễn Văn Phú, Nguyễn Văn Quý, Hoàng Thị Phương Nhung (2022)**, “Ứng dụng một số thuật toán máy học và ảnh Landsat 8 ước tính trữ lượng carbon trên mặt đất rừng tự nhiên lá rộng thường xanh tại tỉnh Bình Phước”, *VNU Journal of Science: Earth and Environmental Sciences*, Vol. 38, No. 4 (2022) 89-104. <https://doi.org/10.25073/2588-1094/vnuees.4890>.
- [28]. **UBND tỉnh Đắk Lắk (2017)**, *Báo cáo số liệu rừng và đất lâm nghiệp năm 2016 trên địa bàn tỉnh Đắk Lắk*, Số 75/BC-UBND, ngày 05/04/2017.
- [29]. **UBND tỉnh Đắk Lắk (2025)**, *Quyết định công bố số liệu hiện trạng diện tích rừng và đất chưa thành rừng trên địa bàn tỉnh Đắk Lắk có đến ngày 31/12/2024*, Số 383/QĐ-UBND, ngày 21/02/2025.

Tiếng anh

- [30]. **Abbas S., Wong M.S., Wu J., Shahzad, Irteza S.M. (2020)**, “Approaches of Satellite Remote Sensing for the Assessment of Above-Ground Biomass across Tropical Forests: Pan-tropical to National Scales”, *Remote Sens.*, 2020, 12, 3351.
- [31]. **Allison P. D. (2012)**, *Multiple regression: A primer*, Thousand Oaks, CA: Pine Forge Press.
- [32]. **Anadita K., Sinha A.K., Jeganathan C. (2024)**, “Machine learning approach to biomass estimation - Intergrating satellite and ground data in Sal forests of Jharkhand”, *Journal of the Indian society of Remote sensing*, DOI: <https://doi.org/10.1007/s12524-024-02061-3>.

- [33]. **Antunes R., Costa G., Feitosa R., Bias, E. and Júnior L. (2024).** Leveraging SAR and optical remote sensing for enhanced biomass estimation in the Amazon with Random Forest and XGBoost models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 5, X-3, 21–28. <https://doi.org/10.5194/isprs-annals-X-3-2024-21-2024>.
- [34]. **Astrup R., Ducey M.J., Granhus A., Ritter T., Von Lupke N. (2014),** “Approaches for estimating stand-level volume using terrestrial laser scanning in a single-scan mode” *Can. J. For. Res.*, 2014, 44, 666–676.
- [35]. **Avitabile V., Herold M., Heuvelink G. B. M., Lewis S. L., Phillips O. L., Asner G. P., et al. (2016),** “An integrated pan-tropical biomass map using multiple reference datasets”, *Global Change Biology*, 22(4), 1406–1420. <https://doi.org/10.1111/gcb.13139>
- [36]. **Badreldin N., Sanchez-Azofeifa A. (2015),** “Estimating Forest Biomass Dynamics by Integrating Multi-Temporal Landsat Satellite Images with Ground and Airborne LiDAR Data in the Coal Valley Mine, Alberta, Canada”, *Remote Sens.*, 2015, 7, 2832–2849.
- [37]. **Bao H., Nguyen Q.T., Nguyen Q.K., Krishna P.P., Temesgen H. (2022),** “Deep learning models for improved reliability of tree aboveground biomass prediction in the tropical evergreen broadleaf forest:”, *Forest Ecology and Management*, 2022, 508, 120031.
- [38]. **Basuki T.M., van Laake P.E., Skidmore A.K., Hussin Y.A. (2009),** “Allometric equations for estimating the above-ground biomass in the tropical lowland Dipterocarp forests”, *Forest Ecology and Management*, 257, 1684-1694. DOI 10.1016/j.foreco.2009.01.027
- [39]. **Beaudoin A., Bernier P.Y., Guindon L., Villemaire P., Guo X.J., Stinson G., Bergeron T., Magnussen S., Hall R.J. (2014),** “Mapping attributes of Canada’s forests at moderate resolution through kNN and MODIS imagery”, *Can. J. For. Res.*,

- 2014, 44, 521–532.
- [40]. **Belgiu M., & Drăgut L. (2016)**, “Random forest in remote sensing: A review of applications and future directions”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- [41]. **Bouvet A., Mermoz S., Toan T.L., Villard L., Mathieu R., Naidoo L., Asner G.P. (2018)**, “An above-ground biomass map of African savannahs and woodlands at 25 m resolution derived from ALOS PALSAR”, *Remote Sens. Environ.*, 2018, 206, 156–173.
- [42]. **Breiman L. (2001)**, *Random forests*, *Mach. Learn.* 2001, 45, 5–32.
- [43]. **Brown M.L., Canham C.D., Murphy L., Donovan T.M. (2018)**, “Timber harvest as the predominant disturbance regime in northeastern US forests: Effects of harvest intensification”, *Ecosphere*, 2018, 9, e02062.
- [44]. **Carreiras J. M. B., Pereira J. M. C., and Pereira J. S. (2012)**, “Estimating the above-ground biomass in miombo savanna woodlands (Mozambique, East Africa) using L-band synthetic aperture radar data”, *Remote Sensing*, 4(12), 3787–3813. <https://doi.org/10.3390/rs4123787>
- [45]. **Chave J., Mechain M.R., Burquez A., Chidumayo E., Colgan M.S., et al. (2014)**, “Improved allometric models to estimate the aboveground biomass of tropical trees”, *Global change biology*, 20(2014): 3177-3190. Doi: 10.1111/gcb.12629.
- [46]. **Chen Q., Laurin G.V., Valentini R. (2015)**, “Uncertainty of remotely sensed aboveground biomass over an African tropical forest: Propagating errors from trees to plots to pixels”, *Remote Sens. Environ.*, 2015, 160, 134–143.
- [47]. **Chen X., Luo M., Larjavaara M. (2023)**, “Effects of climate and plant functional types on forest above-ground biomass accumulatio”, *Carbon Balance and Management*, 18, 5. <https://doi.org/10.1186/s13021-023-00225-1>.
- [48]. **Chopping M., Schaaf C.B., Zhao F., Wang Z.S., Nolin A.W., Moisen G.G.,**

- Martonchik J.V., Bull M. (2011)**, “Forest structure and aboveground biomass in the southwestern United States from MODIS and MISR”, *Remote Sens. Environ.*, 2011, 115, 2943–2953.
- [49]. **Chopping M., Wang Z.S., Schaaf C., Bull M.A., Duchesne R.R. (2022)**, “Forest aboveground biomass in the southwestern United States from a MISR multi-angle index, 2000–2015”, *Remote Sens. Environ.*, 2022, 275, 112964.
- [50]. **Claverie M., Ju J., Masek J.G., Dungan J.L., Vermote E.F., Roger J.C., Skakun S.V., Justice C. (2018)**, “The Harmonized Landsat and Sentinel-2 surface reflectance data set”, *Remote Sens. Environ.*, 2018, 219, 145–161.
- [51]. **Congalton R. G., and Green K. (2009)**, *Assessing the accuracy of remotely sensed data: Principles and practices*, 2nd Edition. Lewis Publishers, Boca Raton.
- [52]. **Cohen J. (1988)**, *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum Associates.
- [53]. **Cutler D. R., Edwards Jr. T. C., Beard K. H., Cutler A., Hess K. T., Gibson J., and Lawler J. J. (2007)**, “Random forests for classification in ecology”, *Ecology*, 88(11), 2783–2792. <https://doi.org/10.1890/07-0539.1>
- [54]. **Dang T.N.A., Nandy S., Srinet R., Nguyen V.L., Ghosh S., Kumar A.S. (2019)**, “Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam”, *Ecological Informatics*, Volume 50, 2019, Pages 24-32, <https://doi.org/10.1016/j.ecoinf.2018.12.010>.
- [55]. **David R.M., Rosser N.J., Donoghue D.N.M. (2022)**, “Improving above ground biomass estimates of Southern Africa dryland forests by combining Sentinel-1 SAR and Sentinel-2 multispectral imagery”, *Remote Sensing of Environment*, Volume 282, 2022, 113232, <https://doi.org/10.1016/j.rse.2022.113232>.
- [56]. **Dietz J., Kuyah S. (2011)**, *Guidelines for establishing regional allometric equations for biomass estimation through destructive sampling*. World Agroforestry Center (ICRAF). Protocol CBP 1.3. Available at

- <http://reddcommunity.org/sites/default/files/field/publications/allometry2013.pdf>.
- [57]. **Dillabaugh K.A., King D.J. (2008)**, “Riparian marshland composition and biomass mapping using Ikonos imagery”, *Can. J. Remote Sens.*, 2008, 34, 143–158.
- [58]. **Dong Y., Zhang H., Chen J., Zhao J., and Li J. (2024)**, “Integrating optical and radar remote sensing for improved forest aboveground biomass estimation using machine learning”, *Forests*, 15(2), 214. <https://doi.org/10.3390/f15020214>
- [59]. **Dormann C. F., Elith J., Bacher S., Buchmann C., Carl G., Carré G., Lautenbach S. (2013)**, “Collinearity: a review of methods to deal with it and a simulation study”, *Ecography*, 36(1), 27–46.
- [60]. **Dube T., Mutanga O. (2015)**, “Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas”, *ISPRS J. Photogramm.*, 2015, 108, 12–32.
- [61]. **Durante P., Martín-Alcón S., Gil-Tena A., Algeet N., Tomé J.L., Recuero L., Palacios-Orueta A., Oyonarte C. (2019)**, “Improving Aboveground Forest Biomass Maps: From High-Resolution to National Scale”, *Remote Sens.*, 2019, 11, 795.
- [62]. **Dutcă I., Mather R., Ioraş F. (2020)**, “Sampling trees to develop allometric biomass models: How does tree selection affect model prediction accuracy and precision?”, *Ecological Indicators*, 117 (2020): 106553, 1-12 pp.
- [63]. **Ebi K.L., Vanos J., Baldwin J.W., Bell J.E., Hondula D.M., Errett N.A., Hayes K., Reid C.E., Saha S., Spector J. (2021)**, “Extreme Weather and Climate Change: Population Health and Health System Implications”, *Fielding, J.E., Ed., Annual Review of Public Health: San Mateo, CA, USA, 2021, Volume 42*, pp. 293–315.
- [64]. **Fang J.Y., Zhu J.X., Li P., Ji C.J., Zhu J.L., Jiang L., Chen G.P., Cai Q., Su H.J., Feng Y.H. (2021)**, “Carbon Budgets of Forest Ecosystems in China”, *Science Press: Beijing, China, 2021*.

- [65]. **FAO (2020)**, *Global Forest Resources Assessment 2020 – Key findings*, Rome, <https://doi.org/10.4060/ca8753en>.
- [66]. **Farr T. G., Rosen P. A., Caro E., Crippen R., Duren R., Hensley S., Kobrick M., et al. (2007)**, “The Shuttle Radar Topography Mission”, *Reviews of Geophysics*, 45(2), RG2004. <https://doi.org/10.1029/2005RG000183>.
- [67]. **Fassnacht F.E., Hartig F., Latifi H., Berger C., Hernandez J., Corvalan P., Koch B. (2014)**, “Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass”, *Remote Sens. Environ.*, 2014, 154, 102–114.
- [68]. **Ferreira I.J.M., Campanharo W.A., Fonseca M.G., Escada M.I.S., Nascimento M.T., Villela D.M., Brancalion P., Magnago L.F.S., Anderson L.O., Nagy L. (2023)**, “Potential aboveground biomass increase in Brazilian Atlantic Forest fragments with climate change”, *Glob. Change Biol.*, 2023, 29, 3098–3113.
- [69]. **Foody G.M. (2002)**, “Status of land cover classification accuracy assessment”, *Remote Sensing of Environment*, 80: 185-201. DOI: [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4).
- [70]. **Foody G.M., Boyd D.S., Cutler M.E.J. (2003)**, “Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions”, *Remote Sensing of Environment*, Volume 85, Issue 4, 2003, Pages 463-474, ISSN 0034-4257, [https://doi.org/10.1016/S0034-4257\(03\)00039-7](https://doi.org/10.1016/S0034-4257(03)00039-7).
- [71]. **Fremout T., Vinatea J.C.D., Thomas E., Huaman-Zambrano W., Salazar-Villegas M., de la Fuente D.L., Bernardino P.N., Atkinson R., Csaplovics E., Muys B. (2022)**, “Site-specific scaling of remote sensing-based estimates of woody cover and aboveground biomass for mapping long-term tropical dry forest degradation status”, *Remote Sens. Environ.*, 2022, 276, 113040.
- [72]. **Frumkin H., Haines A. (2019)**, “Global Environmental Change and Noncommunicable Disease Risks”, *In Annual Review of Public Health, Fielding, J.E.*,

- Ed.*, 2019, Volume 40, pp. 261–282.
- [73]. **Gao X., Huete A.R., Ni W., Miura T. (2000)**, “Optical–Biophysical Relationships of Vegetation Spectra without Background Contamination”, *Remote Sens. Environ.*, 2000, 74, 609–620.
- [74]. **Gao Y.K., Lu D.S., Li G.Y., Wang G.X., Chen Q., Liu L.J., Li D.Q. (2018)**, “Comparative Analysis of Modeling Algorithms for Forest Aboveground Biomass Estimation in a Subtropical Region”, *Remote Sens.*, 2018, 10, 627.
- [75]. **Géron A. (2019)**, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, Published by O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
- [76]. **Ghasemi N., Sahebi M., Mohammadzadeh A. (2011)**, “A review on biomass estimation methods using synthetic aperture radar data”, *Int. J. Geomat. Geosci.*, 2011, 1, 776–778.
- [77]. **Gholizadeh H., Maghsoudi Y., and Alavipanah S.K. (2020)**, “Evaluating the capability of Sentinel-2 data to estimate aboveground biomass in Persian oak forests of Iran using machine learning algorithms”, *International Journal of Applied Earth Observation and Geoinformation*, 89, 102095.
- [78]. **Ghosh S.M., Behera M.D. (2021)**, “Aboveground biomass estimates of tropical mangrove forest using Sentinel-1 SAR coherence data—The superiority of deep learning over a semi-empirical model”, *Comput. Geosci.*, 2021, 150, 104737.
- [79]. **Han H., Wan R., Li B. (2022)**, “Estimating Forest Aboveground Biomass Using Gaofen-1 Images, Sentinel-1 Images, and Machine Learning Algorithms: A Case Study of the Dabie Mountain Region, China”, *Remote Sens.*, 2022, 14, 176. <https://doi.org/10.3390/rs14010176>
- [80]. **Han Y., and Joe I. (2024)**, “Enhancing Machine Learning Models Through PCA, SMOTE-ENN, and Stochastic Weighted Averaging”, *Applied Sciences*, 14(21), 9772. <https://doi.org/10.3390/app14219772>.

- [81]. **Hansen M.C, Potapov P.V., Moore R., Hancher M., Turubanova S.A., Tyukavina A., Thau D., Stehman S.V., et al. (2013)**, High-Resolution Global Maps of 21st-Century Forest Cover Change, *Science*, 342, 850-853 (2013). DOI:10.1126/science.1244693
- [82]. **Harris N. L., Gibbs D. A., Baccini A., Birdsey R.A., De Bruin S., Farina M., Fatoyinbo L., Hansen M.C., Herold M., Houghton R.A., Potapov P.V., Turubanova S. (2021)**, Global maps of twenty-first century forest carbon fluxes, *Nature Climate Change*, 11, 234–240. <https://doi.org/10.1038/s41558-020-00976-6>.
- [83]. **Henry M., Jara M.C., Réjou-Méchain M., Piotto D., Fuentes J.M.M., Wayson C., Guier F.A., et al. (2015)**, “Recommendations for the use of tree models to estimate national forest biomass and assess their uncertainty”, *Annals of Forest Science*, Issue 6, 72(2015): 769 – 777.
- [84]. **Hirsch R.M., Slack J.R., Smith R.A. (1982)**, “Techniques of trend analysis for monthly water quality data”, *Water Resources Research*, 18(1), 107–121.
- [85]. **Ho T.D.T., Le Toan T., Rocca F., Tebaldini S., Villard L., & others. (2016)**, SAR tomography for the retrieval of forest biomass and height: Cross-validation at two tropical forest sites in French Guiana, *Remote Sensing of Environment*, 175, 138–147. <https://doi.org/10.1016/j.rse.2015.12.037>
- [86]. **Hoegh-Guldberg O., Northrop E., Lubchenco J. (2019)**, *The ocean as a solution to climate change: Five opportunities for action*, Washington, DC: World Resources Institute, Retrieved from https://oursharedseas.com/wp-content/uploads/2019/11/Hoegh_Guldberg_et_al_2019_The_ocean_is_key_to_achieving_climate_and_societal_goals-1.pdf
- [87]. **Houghton R.A., Hall F., Goetz S.J. (2009)**, “Importance of biomass in the global carbon cycle”, *J. Geophys. Res.-Biogeosci.*, 2009, 114, G00E03.
- [88]. **Huang T. (2023)**, “Comparing algorithms for estimation of aboveground biomass”, *Forests*, 14(9), 1742.

- [89]. **Huang X.D., Ziniti B., Torbick N., Ducey M.J. (2018)**, “Assessment of Forest above Ground Biomass Estimation Using Multi-Temporal C-band Sentinel-1 and Polarimetric L-band PALSAR-2 Data”, *Remote Sens.*, 2018, 10, 1424.
- [90]. **Huang Y., Liu D., Song X., He H. (2019)**, “Integration of multi-resource remotely sensed data and allometric models to estimate mangrove aboveground biomass: A case study in Hainan Island, China”, *Forests*, 10(11), 1011. <https://doi.org/10.3390/f10111011>.
- [91]. **Huang Y.P., Chen J.S. (2013)**, “Advances in the estimation of forest biomass based on SAR data”, *Remote Sens. Nat. Resour.*, 2013, 25, 7–13.
- [92]. **Huy B., Poudel K.P., Temesgen H. (2016a)**, “Aboveground biomass equations for evergreen broadleaf forests in South Central Coastal ecoregion of Viet Nam: Selection of eco-regional or pantropical models”, *For. Ecol. and Mgmt.*, 376: 276-282.
- [93]. **IPCC (2019)**, *Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Volume 4 Agriculture, Forestry and Other Land Use*, https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch04_Forest%20Land.pdf
- [94]. **IPCC (2023)**, *Summary for Policymakers*, In: *Climate Change 2023: Synthesis Report*. Intergovernmental Panel on Climate Change.
- [95]. **Istrefi K., Breidenbach J., Kändler G. (2019)**, “The coefficients of log-transformed allometric biomass equations differ between species”, *New Zealand Journal of Forestry Science*, 49:8. <https://doi.org/10.33494/nzjfs492019x51x>.
- [96]. **Jenkins J.C., Chojnacky D.C., Heath L.S. Birdsey R.A. (2003)**, “National-Scale Biomass Estimators for United States Tree Species”, *Journal of Forest Science*, 49(1): 12-35.
- [97]. **Jensen J. R. (1995)**, *Introductory Digital Image Processing: A Remote Sensing Perspective (3rd ed.)*, Upper Saddle River, NJ: Prentice-Hall. ISBN 0-13-145361-0.

- [98]. **Jin X.L., Li Y., Yu X.B. (2024)**, “UAV-RGB-image-based aboveground biomass equation for planted forest in semi-arid Inner Mongolia, China”, *Ecological Informatics*, 81 (2024) 02574.
- [99]. **Jolliffe I.T., Cadima J. (2016)**, “Principal component analysis: a review and recent developments”, *Philos Trans A Math Phys Eng Sci.*, 2016 Apr 13;374(2065):20150202. doi: 10.1098/rsta.2015.0202. PMID: 26953178; PMCID: PMC4792409.
- [100]. **Joshi N., Mitchard E.T.A., Woo N., Torres J., Moll-Rodriguez J.B., Ehammer A., De Sy V. (2017)**, *Understanding ‘saturation’ of radar signals over forests*, Scientific Reports, 7(1), 3505. <https://doi.org/10.1038/s41598-017-03469-3>
- [101]. **Kandel P., Karki R., Sharma R. (2021)**, “Estimation of aboveground biomass in subtropical community forests of Nepal using Landsat 8 data and Random Forest regression”, *Journal of Forestry Research*, 32, 1713–1725.
- [102]. **Kendall M.G. (1975)**, *Rank Correlation Methods (4th ed.)*, Griffin, London.
- [103]. **Kim J.H. (2019)**, “Multicollinearity and misleading statistical results”, *Korean Journal of Anesthesiology*, 72(6), 558–569.
- [104]. **Knott J.A., Liknes G.C., Giebink C.L., Oh S., Domke G.M., McRoberts R.E., Quirino V.F., Walters B.F. (2023)**, “Effects of outliers on remote sensing-assisted forest biomass estimation: A case study from the United States national forest inventory”, *Methods Ecol. Evol.*, 2023, 00, 1–16.
- [105]. **Koetz B., Sun G.Q., Morsdorf F., Ranson K.J., Kneubuhler M., Itten K., Allgower, B. (2007)**, “Fusion of imaging spectrometer and LIDAR data over combined radiative transfer models for forest canopy characterization”, *Remote Sens., Environ.* 2007, 106, 449–459.
- [106]. **Kralicek K., Huy B., Poudel K.P., Temesgen H., Salas C. (2017)**, “Simultaneous estimation of above- and below-ground biomass in tropical forests of Viet Nam”, *Forest Ecology and Management*, 390 (2017): 147–156.

- [107]. **Kuhn M., Johnson K. (2013)**, *Applied Predictive Modeling*, Springer, New York.
- [108]. **Kutner M. H., Nachtsheim C. J., Neter J. (2004)**, *Applied Linear Statistical Models (5th ed.)*, McGraw-Hill/Irwin.
- [109]. **Laurin G.V., Pirotti F., Callegari M., Chen Q., Cuozzo G., Lingua E., Notarnicola C., Papale D. (2017)**, “Potential of ALOS2 and NDVI to Estimate Forest Above-Ground Biomass, and Comparison with Lidar-Derived Estimates”, *Remote Sens.*, 2017, 9, 18.
- [110]. **Lefsky M.A., Keller M., Pang Y., de Camargo P.B., Hunter M.O. (2007)**, “Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms”, *J. Appl. Remote Sens.*, 2007, 1, 013537.
- [111]. **Levine J., Valpine P.D., Battles J. (2020)**, “Generalized additive models reveal among-stand variation in live tree biomass equations”, *Canadian Journal of Forest Research*.
- [112]. **Leys C., Ley C., Klein O., Bernard P., Licata L. (2013)**, “Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median”, *Journal of Experimental Social Psychology*, 49(4), 764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- [113]. **Li C., Li Y.C., Li M.Y. (2019)**, “Improving Forest Aboveground Biomass (AGB) Estimation by Incorporating Crown Density and Using Landsat 8 OLI Images of a Subtropical Forest in Western Hunan in Central China”, *Forests*, 2019, 10, 104.
- [114]. **Li D.R., Wang C.W., Hu Y.M., Liu S.G. (2012)**, “General Review on Remote Sensing-Based Biomass Estimation”, *Geomat. Inform. Sci.*, Wuhan Univ. 2012, 37, 631–635.
- [115]. **Li J.R., and Mao X.G. (2020)**. Comparison of Canopy Closure Estimation of Plantations Using Parametric, Semi-Parametric, and Non-Parametric Models Based on GF-1 Remote Sensing Images. *Forests* 2020, 11, 597.

- [116]. **Li L. and Zhang D. (2024)**, “Forest carbon offset protocols in compliance carbon markets: comparative review”, *ScienceDirect*, (2024). Volume 165, August 2024, 103253. <https://doi.org/10.1016/j.forpol.2024.103253>
- [117]. **Li T., Li M.Y., Ren F., Tian L. (2022)**, “Estimation and Spatio-Temporal Change Analysis of NPP in Subtropical Forests: A Case Study of Shaoguan, Guangdong, China”, *Remote Sens.*, 2022, 14, 2541.
- [118]. **Liang X.L., Kankare V., Hyyppa J., Wang Y.S., Kukko A., Haggren H., Yu X.W., Kaartinen H., Jaakkola A., Guan F.Y. (2016)**, “Terrestrial laser scanning in forest inventories”, *ISPRS J. Photogramm.*, 2016, 115, 63–77.
- [119]. **Liu, H.Q.; Huete, A.R. (1995)**, A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and Remote Sensing* 1995, 33, 457-465
- [120]. **Lu D., Chen Q., Wang G., Liu L., Li G., Moran E. (2014)**, “A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems”, *International Journal of Digital Earth*, 9(1), 63–105. <https://doi.org/10.1080/17538947.2014.990526>
- [121]. **Lu D.S., Chen Q., Wang G.X., Moran E., Batistella M., Zhang M.Z., Laurin G.V., David S. (2012)**, “Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates”, *Int. J. For. Res.*, 2012, 436537. <https://doi.org/10.1155/2012/436537>.
- [122]. **Lucas R.M., Cronin N., Lee A., Moghaddam M., Witte C., Tickle P. (2006)**, “Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia”, *Remote Sensing of Environment*, 100(3), 407–425. <https://doi.org/10.1016/j.rse.2005.10.019>
- [123]. **Ludwig M., Moreno A., Hölzel N., Pebesma E., Meyer H. (2023)**, “Assessing and improving the transferability of current global spatial prediction models”, *Global Ecology and Biogeography*, volume - 32, DOI: 10.1111/geb.13635.

- [124]. **Luo M., Wang Y.F., Xie Y.H., Zhou L., Qiao J.J., Qiu S.Y., Sun Y.J. (2021)**, “Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass”, *Forests*, 2021, 12, 216.
- [125]. **Luo S.Z., Wang C., Xi X.H., Pan F.F., Qian M.J., Peng D.L., Nie S., Qin H.M., Lin Y. (2017)**, “Retrieving aboveground biomass of wetland *Phragmites australis* (common reed) using a combination of airborne discrete-return LiDAR and hyperspectral data”, *Int. J. Appl. Earth Obs.*, 2017, 58, 107–117.
- [126]. **Ma T., Zhang C., Ji L., Zuo Z., Beckline M., Hu Y., Li X., Xiao X. (2024)**, “Development of forest aboveground biomass estimation, its problems and future solutions: A review”, *Ecological Indicators*, 159, 111653. <https://doi.org/10.1016/j.ecolind.2024.111653>
- [127]. **Mallinis G., and Koutsias N. (2008)**, “Spectral and spatial-based classification for broad-scale land cover mapping based on logistic regression”, *Sensors*, 8: 8067 – 8085.
- [128]. **Mancino G., Ferrara A., Padula A., Nolè A. (2020)**, “Cross-Comparison between Landsat 8 (OLI) and Landsat 7 (ETM+) Derived Vegetation Indices in a Mediterranean Environment”, *Remote Sensing*, 12(2), 291. <https://doi.org/10.3390/rs12020291>.
- [129]. **Markus T., Neumann T., Martino A., Abdalati W., Brunt K., Csatho B., Farrell S., Fricker H., Gardner A., Harding D. (2017)**, “The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation”, *Remote Sens. Environ.*, 2017, 190, 260–273.
- [130]. **Mendez-González J., Montero-Moya C., Martínez-Barrón E. (2012)**, “Allometric models for estimating above- and below-ground biomass in restored mixed-species Atlantic Forest stands”, *Biota Neotropica*, 12(3), 1-8. <https://doi.org/10.1590/1676-06032013008413>.
- [131]. **Messinger M., Asner G.P., Silman M. (2016)**, “Rapid Assessments of Amazon

- Forest Structure and Biomass Using Small Unmanned Aerial Systems”, *Remote Sens.*, 2016, 8, 615.
- [132]. **Mitchard E.T.A., Saatchi S.S., Lewis S.L., Feldpausch T.R., Woodhouse I.H., Sonké B., Rowland C., Meir P. (2011)**, “Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest–savanna boundary region of central Africa using multi-temporal L-band radar backscatter”, *Remote Sensing of Environment*, Volume 115, Issue 11, 2011, Pages 2861-2873, <https://doi.org/10.1016/j.rse.2010.02.022>.
- [133]. **Mo L., Zohner C.M., Reich P.B., et al. (2023)**, “Integrated global assessment of the natural forest carbon potential”, *Nature*, 624, 92–101 (2023). <https://doi.org/10.1038/s41586-023-06723-z>.
- [134]. **Molotoks A., Stehfest E., Doelman J., Albanito F., Fitton N., Dawson T.P., Smith P. (2018)**, “Global projections of future cropland expansion to 2050 and direct impacts on biodiversity and carbon storage”, *Glob. Change Biol.*, 2018, 24, 5895–5908.
- [135]. **Moore A.W. (2017)**, *Cross-validation for detecting and preventing overfitting*, School of Computer Science. Carnegie Mellon University. Available on-line: <https://www.autonlab.org/media/tutorials/overfit10.pdf>.
- [136]. **Morel A.C., Fisher J.B., Malhi Y. (2012)**, “Evaluating the Potential to Monitor Aboveground Biomass in Forest and Oil Palm in Sabah, Malaysia, for 2000–2008 with Landsat ETM+ and ALOS-PALSAR”, *International Journal of Remote Sensing*, 33 (11): 3614–3639. doi:10.1080/01431161.2011.631949.
- [137]. **Mutanga O., Shoko C., Adelabu S, Bangira T. (2016)**, “Remote sensing of aboveground forest biomass: A review”, *Tropical Ecology.*, 57. 125-132.
- [138]. **Muukkonen P., Heiskanen J. (2005)**, “Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data”, *Remote Sens. Environ.*, 2005, 99, 434–447.

- [139]. **Navulur K. (2007)**, *Multispectral image analysis using the object-oriented paradigm*, CRC Press.
- [140]. **Nerem R.S., Beckley B.D., Fasullo J.T., Hamlington B.D., Masters D., Mitchum G.T. (2018)**, Climate-change-driven accelerated sea-level rise detected in the altimeter era. *Proc. Natl. Acad. Sci. USA* 2018, 115, 2022–2025.
- [141]. **Næsset E., Gobakken T., Bollandsas O.M., Gregoire T.G., Nelson R., Stahl G. (2013)**, “Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway”, *Remote Sens. Environ.*, **2013**, *130*, 108–120.
- [142]. **Nguyen A. & Saha S. (2024)**. Machine Learning and Multi-source Remote Sensing in Forest Carbon Stock Estimation: A Review. *39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Tackling Climate Change with Machine Learning*. <https://doi.org/10.48550/arXiv.2411.17624>
- [143]. **Nguyen T.T.H. (2011)**, *Forest Remote Sensing: Using multi-data sources for inventory of Natural broad-leaved ever-green forests in the Central Highlands of Vietnam*, Lambert Academic Publishing, Germany.
- [144]. **Nguyen H.T.T., Doan T.M., Tomppo E., McRoberts R.E. (2020)**, “Land Use/Land Cover Mapping Using Multitemporal Sentinel-2 Imagery and Four Classification Methods—A Case Study from Dak Nong, Vietnam”, *Remote Sensing*, *12*(9), 1367. <https://doi.org/10.3390/rs12091367>.
- [145]. **O’Brien R.M. (2007)**, “A caution regarding rules of thumb for variance inflation factors”, *Quality & Quantity*, *41*(5), 673–690.
- [146]. **Omar H., Misman M.A., Kassim A.R. (2017)**, “Synergetic of PALSAR-2 and Sentinel-1A SAR Polarimetry for Retrieving Aboveground Biomass in Dipterocarp Forest of Malaysia”, *Applied Sciences.*, *7*. 1-20. [10.3390/app7070675](https://doi.org/10.3390/app7070675).
- [147]. **Ou G.L., Li C., Lv Y.Y., Wei A.C., Xiong H.X., Xu H., Wang G.X. (2019)**, “Improving Aboveground Biomass Estimation of *Pinus densata* Forests in Yunnan

- Using Landsat 8 Imagery by Incorporating Age Dummy Variable and Method Comparison”, *Remote Sens.*, 2019, 11, 738.
- [148]. **Pan Y.D., Birdsey R.A., Fang J.Y., Houghton R., Kauppi P.E., Kurz W.A., Phillips O.L., Shvidenko A., Lewis S.L., Canadell J.G. (2011)**, “A Large and Persistent Carbon Sink in the World’s Forests”, *Science*, 2011, 333, 988–993.
- [149]. **Patel P., Srivastava H., Panigrahy S., Parihar J., (2006)**, “Comparative evaluation of the sensitivity of multi-polarized multi-frequency SAR backscatter to plant density”, *International Journal of Remote Sensing - INT J REMOTE SENS.* 27. 293-305. 10.1080/01431160500214050.
- [150]. **Payne N.J., Cameron D.A., Leblanc J.D., Morrison I.K. (2019)**, “Carbon storage and net primary productivity in Canadian boreal mixedwood stands”, *J. For. Res.*, 2019, 30, 1667–1678.
- [151]. **Penner M., Power K., Muhairwe C., Tellier R., Wang Y. (1997)**, *Canada’s forest biomass resources: deriving estimates from Canada’s forest inventory*. Pacific Forestry Centre Canadian Forest Services. Information Report BC-X-370.
- [152]. **Picard N, Rutishauser E., Ploton P., Ngomanda A., Henry M. (2015)**, “Should tree biomass allometry be restricted to power models?”, *For. Ecol. and Manag.*, 353: 156-163. DOI 10.1016/j.foreco.2015.05.035.
- [153]. **Picard N., Saint-André L., Henry M. (2012)**, *Manual for building tree volume and biomass allometric equations: from field measurement to prediction*, Food and Agricultural Organization of the United Nations, Rome, and Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, 215 pp.
- [154]. **Popescu S.C. (2007)**, “Estimating biomass of individual pine trees using airborne lidar”, *Biomass Bioenerg.*, 2007, 31, 646–655.
- [155]. **Price B., Gomez A., Mathys L., Gardi O., Schellenberger A., Ginzler C., Thurig E. (2017)**, “Tree biomass in the Swiss landscape: Nationwide modelling for

- improved accounting for forest and non-forest trees”, *Environ. Monit. Assess.*, 2017, 189, 1–14.
- [156]. **Qasim M., Csaplovics E., Villegas M.H.S. (2023)**, “Forest biomass assessment combining field inventorying and remote sensing data”, *Open Geosciences*, 2023, 15: 20220553. <https://doi.org/10.1515/geo-2022-0553>.
- [157]. **Richards J.A. (2013)**, *Remote Sensing Digital Image Analysis*, Springer, Berlin.
- [158]. **Refaeilzadeh P., Tang L., Liu H. (2009)**, *Cross-validation*. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 532–538). Springer. https://doi.org/10.1007/978-0-387-39940-9_565
- [159]. **Saatchi S.S., Harris N.L., Brown S., Lefsky M., Mitchard E.T.A., Salas W. et al. (2011)**, “Benchmark map of forest carbon stocks in tropical regions across three continents”, *Proc. Natl. Acad. Sci. U.S.A.*, 108 (24) 9899–9904, <https://doi.org/10.1073/pnas.1019576108>.
- [160]. **Sadeghi Y., St-Onge B., Leblon B., Prieur J.F., Simard M. (2018)**, “Mapping boreal forest biomass from a SRTM and TanDEM-X based on canopy height model and Landsat spectral indices”, *Int. J. Appl. Earth Obs.*, 2018, 68, 202–213.
- [161]. **Sandberg G., Ulander L.M.H., Fransson J.E.S., Holmgren J., Le Toan T. (2011)**, “L- and P-band backscatter intensity for biomass retrieval in hemiboreal forest”, *Remote Sens. Environ.*, 2011, 115, 2874–2886.
- [162]. **Santoro M., Cartus O., Carvalhais N., Rozendaal D. M., Avitabile V., Araza A., Willcock S. (2021)**. The global forest above-ground biomass pool for 2010 estimated from highresolution satellite observations. *Earth System Science Data*, 13(8), 3927–3950.
- [163]. **Sarker L.R., Nichol J.E. (2011)**, “Improved forest biomass estimates using ALOS AVNIR-2 texture indices”, *Remote Sens. Environ.*, 2011, 115, 968–977.
- [164]. **Schimel D., Pavlick R., Fisher J.B., Asner G.P., Saatchi S., Townsend P.**

- Miller C., Frankenberg C., Hibbard K., Cox P. (2015)**, “Observing terrestrial ecosystems and the carbon cycle from space”, *Glob. Change Biol.*, 2015, 21, 1762–1776.
- [165]. **Sen P.K. (1968)**, “Estimates of the regression coefficient based on Kendall’s tau”, *Journal of the American Statistical Association*, 63(324), 1379–1389.
- [166]. **Sibanda M., Mutanga O., Rouget M. (2015)**, “Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments” *ISPRS J. Photogramm.*, 2015, 110, 55–65.
- [167]. **Simard M., Fatoyinbo L., Smetanka C., Rivera-Monroy V.H., Castaneda-Moya E., Thomas N., Van der Stocken T. (2019)**, “Mangrove canopy height globally related to precipitation, temperature and cyclone frequency”, *Nat. Geosci.*, 2019, 12, 40–45.
- [168]. **Singh M., Malhi Y., Bhagwat S. (2014)**, “Biomass estimation of mixed forest landscape using a Fourier transform texture-based approach on very-high-resolution optical satellite imagery”, *Int. J. Remote Sens.*, 2014, 35, 3331–3349.
- [169]. **Smith B., Knorr W., Widlowski J.L., Pinty B., Gobron N. (2008)**, “Combining remote sensing data with process modelling to monitor boreal conifer forest carbon balances”, *For. Ecol. Manag.*, 2008, 255, 3985–3994.
- [170]. **Song J., Liu X., Adingo S., Guo Y., Li Q. (2024)**, “A Comparative Analysis of Remote Sensing Estimation of Aboveground Biomass in Boreal Forests Using Machine Learning Modeling and Environmental Data”, *Sustainability*, 2024, 16, 7232. <https://doi.org/10.3390/su16167232>.
- [171]. **Sohrabi A., Safari A., Powell S. (2018)**, “Comparison of satellite-based estimates of aboveground biomass in coppice oak forests using parametric, semiparametric and nonparametric modeling methods”, *Journal of Applied Remote Sensing*, 12(4), 046026.
- [172]. **Stas S. (2011)**, *Aboveground Biomass and Carbon Stocks in A Secondary Forest in Comparison with Adjacent Primary Forest on Limestone in Seram, the Moluccas*,

- Indonesia. Bogor, Indonesia: Center for International Forestry Research (CIFOR), Retrieved from https://www1.cifor.org/fileadmin/subsites/colupsia/documents/Aboveground_biomass_Suzanne.pdf
- [173]. **Statistical Horizons (2020)**, *Dealing with multicollinearity*, Retrieved from <https://statisticalhorizons.com>
- [174]. **Stovall A.E.L., Vorster A.G., Anderson R.S., Evangelista P.H., Shugart H.H. (2017)**, “Non-destructive aboveground biomass estimation of coniferous trees using terrestrial LiDAR”, *Remote Sens. Environ.*, 2017, 200, 31–42.
- [175]. **Talebiesfandarani S., and Shamsoddini A. (2022)**, “Global aboveground biomass estimation using feature selection and deep learning models”, *Remote Sensing Applications: Society and Environment*, 28, 100868.
- [176]. **Thuy P.T. (CIFOR-ICRAF). (2023)**. *Forest carbon market in Vietnam: opportunities and challenges* (Occasional Paper). Center for International Forestry Research.
- [177]. **Tian L., Tao Y., Fu W.X., Li T., Ren F., Li M.Y. (2022)**, “Dynamic Simulation of Land Use/Cover Change and Assessment of Forest Ecosystem Carbon Storage under Climate Change Scenarios in Guangdong Province, China”, *Remote Sens.* 2022, 14, 2330.
- [178]. **Tian L., Wu X., Tao Y., Li M., Qian C., Liao L., Fu W. (2023)**, “Review of Remote Sensing-Based Methods for Forest Aboveground Biomass Estimation: Progress, Challenges, and Prospects”, *Forests*, 14(6), 1086. <https://doi.org/10.3390/f14061086>
- [179]. **Wai P., Su H., Li M. (2022)**, “Estimating Aboveground Biomass of Two Different Forest Types in Myanmar from Sentinel-2 Data with Machine Learning and Geostatistical Algorithms”, *Remote Sens.* 2022, 14, 2146. <https://doi.org/10.3390/rs14092146>.

- [180]. **Wang Y., Fang H.L., Zhang Y.H., Li S.J. (2020)**, “Retrieval of Forest LAI Using Airborne LVIS and Spaceborne GLAS Waveform LiDAR Data”, *Remote Sens. Technol. Appl.* 2020, 35, 1004–1014.
- [181]. **Wang Z., Lu Y., Zhao G., Sun C., Zhang F., He S. (2022)**, “Sugarcane Biomass Prediction with Multi-Mode Remote Sensing Data Using Deep Archetypal Analysis and Integrated Learning”, *Remote Sensing*, 14(19), 4944. <https://doi.org/10.3390/rs14194944>.
- [182]. **Waring R.H., Coops N.C., Landsberg J.J. (2010)**, “Improving predictions of forest growth using the 3-PGS model with observations made by remote sensing”, *For. Ecol. Manag.* 2010, 259, 1722–1729.
- [183]. **Wood S.N. (2017)**, *Generalized Additive Models: An Introduction with R, Second Edition (2nd ed.)*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>
- [184]. **Xiao J.F., Chevallier F., Gomez C., Guanter L., Hicke J.A., Huete A.R., Ichii K., Ni W.J., Pang Y., Rahman A.F. (2019)**, “Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years”, *Remote Sens. Environ.* 2019, 233, 111383.
- [185]. **Xing Y.Q., You H.T., Huo D., Sun X.T., Wang R. (2014)**, “Research Progress in Estimating Forest Tree Height Using Small Footprint Lidar Data”, *World Forestry Res.* 2014, 27, 29–34.
- [186]. **Xu X.L., Cao M.K. (2006)**, “An Analysis of the Applications of Remote Sensing Method to the Forest Biomass Estimation”, *J. Geo-Inf. Sci.* 2006, 8, 122–128.
- [187]. **Yan X.D., Shugart H.H. (2005)**, “FAREAST: A forest gap model to simulate dynamics and patterns of eastern Eurasian forests”, *J. Biogeogr.* 2005, 32, 1641–1658.
- [188]. **Ye Z., Long J., Zheng H., Liu Z., Zhang T., Wang Q. (2023)**, “Mapping Growing Stem Volume Using Dual-Polarization GaoFen-3 SAR Images in Evergreen Coniferous Forests”, *Remote Sens.* 2023, 15, 2253. <https://doi.org/10.3390/rs15092253>.

- [189]. **Yu Y.F., Saatchi S., Heath L.S., LaPoint E., Myneni R., Knyazikhin Y. (2010)**, “Regional distribution of forest height and biomass from multisensor data fusion”, *J. Geophys. Res.-Biogeosci.* 2010, 115, G00E12.
- [190]. **Yue S., & Pilon P. (2004)**, “A comparison of the power of the t test, Mann–Kendall and bootstrap tests for trend detection”, *Hydrological Sciences Journal*, 49(1), 21–37.
- [191]. **Zaki N.A.M., Abd Latif Z., Suratman M.N., Zainal M.Z. (2016)**, “Aboveground biomass and carbon stocks modelling using non-linear regression model”, *In Proceedings of the 8th IGRSM International Conference and Exhibition on Geospatial and Remote Sensing (IGRSM)*, Kuala Lumpur, Malaysia, 13–14 April 2016.
- [192]. **Zeng Y.L., Hao D.L., Huete A., Dechant B., Berry J., Chen J.M., Joiner J., Frankenberg C., Bond-Lamberty B., Ryu Y. (2022)**, “Optical vegetation indices for monitoring terrestrial ecosystems globally”, *Nat. Rev. Earth Environ.* 2022, 3, 447–493.
- [193]. **Zhang G., Ganguly S., Nemani R.R., White M.A., Milesi C., Hashimoto H., Wang W.L., et al. (2014)**, “Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data”, *Remote Sens. Environ.* 2014, 151, 44–56.
- [194]. **Zhang M.Z., Lin H., Zeng S.Q., Li J.P., Shi J.N., Wang G.X. (2013)**, “Impacts of Plot Location Errors on Accuracy of Mapping and Scaling Up Aboveground Forest Carbon Using Sample Plot and Landsat TM Data”, *IEEE Geosci. Remote Sens.* 2013, 10, 1483–1487.
- [195]. **Zhang R., Zhou X., Ouyang Z., Avitabile V., Qi J., Chen J., Giannico V. (2019)**, “Estimating aboveground biomass in subtropical forests of China by integrating multisource remote sensing and ground data”, *Remote Sensing of Environment* 232 (2019) 111341. <https://doi.org/10.1016/j.rse.2019.111341>.

- [196]. **Zhang X., Pang Y., Xie Z., Li C., Liu Q. (2021)**, “Estimation of forest aboveground biomass using optical and SAR data: A case study in the Greater Khingan Mountains, China”, *Remote Sensing*, 13(4), 722. <https://doi.org/10.3390/rs13040722>
- [197]. **Zhao K.G., Popescu S., Nelson R. (2009)**, “Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers”, *Remote Sens. Environ.* 2009, 113, 182–196.
- [198]. **Zheng D., Rademacher J., Chen J., Crow T., Bresee M., Le Moine J., Ryu S. (2004)**, “Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA”, *Remote Sensing of Environment* 2004, 93:402–411.

PHỤ LỤC

1. Phụ lục 1: Dữ liệu mô hình hóa

TT	AGB	logAGB	B7	MSI	NDWI	VH	elevation	slope	Year
1	68,86	1,84	0,06	0,39	-0,76	0,44	0,78	0,08	2020
2	96,07	1,98	0,05	0,43	-0,73	0,47	0,94	0,08	2020
3	117,58	2,07	0,05	0,38	-0,78	0,60	0,84	0,28	2020
4	131,95	2,12	0,05	0,41	-0,77	0,55	0,80	0,70	2020
5	133,98	2,13	0,04	0,30	-0,79	0,28	0,82	0,47	2020
6	136,79	2,14	0,05	0,41	-0,79	0,52	0,94	0,15	2020
7	150,89	2,18	0,08	0,55	-0,75	0,39	0,91	0,00	2020
8	159,07	2,20	0,07	0,52	-0,72	0,33	0,79	0,45	2020
9	161,73	2,21	0,05	0,41	-0,78	0,36	0,89	0,19	2020
10	172,76	2,24	0,05	0,40	-0,79	0,63	0,96	0,00	2020
11	189,71	2,28	0,05	0,42	-0,77	0,48	0,72	0,28	2020
12	190,20	2,28	0,09	0,58	-0,63	0,45	0,98	0,29	2020
13	318,30	2,50	0,05	0,42	-0,78	0,40	0,93	0,36	2020
14	6,73	0,83	0,07	0,56	-0,68	0,26	0,43	0,55	2021
15	16,62	1,22	0,09	0,53	-0,76	0,79	0,45	0,48	2021
16	31,75	1,50	0,06	0,54	-0,68	0,79	0,53	0,71	2021
17	37,19	1,57	0,05	0,46	-0,72	0,22	0,52	0,54	2021
18	42,78	1,63	0,08	0,50	-0,76	0,82	0,44	0,59	2021
19	43,44	1,64	0,08	0,55	-0,69	0,68	0,52	0,76	2021
20	45,70	1,66	0,07	0,52	-0,69	0,87	0,41	0,80	2021
21	46,09	1,66	0,08	0,57	-0,70	0,51	0,92	0,74	2021
22	47,10	1,67	0,06	0,53	-0,71	0,64	0,47	0,10	2021
23	48,72	1,69	0,06	0,45	-0,76	0,82	0,44	0,12	2021
24	48,89	1,69	0,09	0,54	-0,74	0,62	0,42	0,35	2021
25	53,79	1,73	0,07	0,49	-0,77	0,74	0,42	0,51	2021
26	57,87	1,76	0,08	0,51	-0,77	0,88	0,38	0,57	2021
27	62,92	1,80	0,08	0,63	-0,67	0,67	0,41	0,16	2021
28	107,08	2,03	0,05	0,53	-0,72	0,43	0,47	0,76	2021
29	109,39	2,04	0,08	0,47	-0,76	0,27	0,72	1,00	2021
30	127,11	2,10	0,10	0,56	-0,77	1,00	0,76	0,62	2021
31	141,47	2,15	0,05	0,42	-0,79	0,00	0,85	0,73	2021
32	165,87	2,22	0,07	0,62	-0,66	0,48	1,00	0,57	2021
33	17,14	1,23	0,05	0,45	-0,75	0,44	0,49	0,00	2022
34	138,35	2,14	0,05	0,41	-0,76	0,37	0,66	0,15	2022
35	140,89	2,15	0,05	0,40	-0,75	0,42	0,49	0,19	2022
36	164,73	2,22	0,05	0,42	-0,76	0,24	0,47	0,38	2022
37	1,07	0,03	0,12	0,64	-0,63	0,47	0,00	0,17	2023
38	1,32	0,12	0,08	0,57	-0,73	0,56	0,43	0,35	2023
39	1,75	0,24	0,09	0,62	-0,65	0,53	0,07	0,18	2023
40	7,64	0,88	0,12	0,80	-0,65	0,47	0,03	0,13	2023

TT	AGB	logAGB	B7	MSI	NDWI	VH	elevation	slope	Year
41	9,17	0,96	0,09	0,73	-0,62	0,38	0,06	0,23	2023
42	27,30	1,44	0,08	0,61	-0,64	0,38	0,16	0,53	2023
43	52,17	1,72	0,16	0,83	-0,56	0,33	0,42	0,43	2023
44	57,21	1,76	0,06	0,49	-0,69	0,76	0,89	0,43	2023
45	67,84	1,83	0,08	0,58	-0,70	0,62	0,42	0,38	2023
46	127,37	2,11	0,04	0,46	-0,73	0,37	0,55	0,32	2023
47	137,61	2,14	0,05	0,45	-0,75	0,62	0,69	0,38	2023
48	179,14	2,25	0,06	0,51	-0,76	0,67	0,66	0,05	2023
49	191,00	2,28	0,04	0,43	-0,74	0,19	0,88	0,40	2023
50	191,00	2,28	0,04	0,43	-0,74	0,19	0,88	0,40	2023
51	196,37	2,29	0,07	0,51	-0,76	0,39	0,38	0,12	2023
52	214,58	2,33	0,06	0,48	-0,72	0,69	0,62	0,51	2023
53	219,58	2,34	0,05	0,51	-0,69	0,42	0,69	0,55	2023
54	359,73	2,56	0,06	0,46	-0,72	0,85	0,87	0,74	2023
55	384,53	2,58	0,04	0,47	-0,71	0,23	0,91	0,47	2023
56	393,68	2,60	0,06	0,53	-0,70	0,46	0,85	0,23	2023
57	644,79	2,81	0,06	0,54	-0,70	0,67	0,56	0,55	2023
58	66,59	1,82	0,08	0,70	-0,63	0,50	0,55	0,06	2024
59	74,72	1,87	0,06	0,46	-0,74	0,41	0,48	0,18	2024
60	113,34	2,05	0,05	0,42	-0,75	0,50	0,54	0,32	2024
61	120,45	2,08	0,07	0,58	-0,66	0,57	0,55	0,64	2024
62	123,24	2,09	0,06	0,48	-0,73	0,50	0,47	0,52	2024
63	131,49	2,12	0,07	0,59	-0,67	0,43	0,66	0,43	2024
64	132,71	2,12	0,05	0,41	-0,76	0,11	0,44	0,58	2024
65	173,35	2,24	0,05	0,42	-0,75	0,50	0,54	0,32	2024
66	188,80	2,28	0,06	0,52	-0,71	0,58	0,56	0,11	2024
67	236,27	2,37	0,06	0,50	-0,68	0,48	0,64	0,36	2024
68	250,31	2,40	0,04	0,40	-0,77	0,27	0,49	0,31	2024
69	280,03	2,45	0,05	0,52	-0,68	0,37	0,61	0,35	2024
70	395,89	2,60	0,07	0,59	-0,65	0,61	0,65	0,53	2024

2. Phụ lục 2: Dữ liệu đánh giá độc lập

TT	AGB	logAGB	B7	MSI	VH	elevation	slope	Year
1	13,73	1,14	0,05	0,45	0,36	0,50	0,25	2013
2	72,67	1,86	0,06	0,48	0,39	0,08	0,07	2013
3	76,72	1,88	0,06	0,42	0,31	0,45	0,25	2013
4	77,50	1,89	0,06	0,49	0,40	0,20	0,82	2013
5	80,20	1,90	0,06	0,41	0,71	0,10	0,39	2013
6	80,63	1,91	0,06	0,48	0,31	0,32	0,84	2013
7	82,16	1,91	0,07	0,46	0,48	0,02	0,00	2013
8	87,83	1,94	0,07	0,44	0,32	0,37	0,56	2013
9	96,99	1,99	0,06	0,43	0,48	0,35	0,70	2013
10	103,50	2,01	0,07	0,47	0,17	0,25	0,72	2013
11	114,03	2,06	0,06	0,43	0,73	0,10	0,41	2013
12	114,35	2,06	0,07	0,46	0,48	0,02	0,00	2013
13	118,58	2,07	0,07	0,47	0,45	0,45	0,42	2013
14	132,15	2,12	0,07	0,52	0,36	0,29	0,72	2013
15	136,38	2,13	0,07	0,44	0,49	0,05	0,46	2013
16	141,02	2,15	0,11	0,82	0,42	0,02	0,55	2013
17	142,50	2,15	0,07	0,42	0,63	0,21	0,65	2013
18	147,42	2,17	0,07	0,48	0,39	0,20	0,20	2013
19	148,38	2,17	0,05	0,45	0,44	0,46	0,10	2013
20	150,09	2,18	0,05	0,42	0,61	0,20	0,15	2013
21	152,10	2,18	0,07	0,47	0,44	0,20	0,03	2013
22	155,72	2,19	0,05	0,44	0,45	0,87	0,55	2013
23	157,75	2,20	0,06	0,49	0,33	0,00	0,09	2013
24	168,30	2,23	0,05	0,41	0,63	0,44	0,18	2013
25	168,48	2,23	0,06	0,44	0,60	0,26	0,29	2013
26	172,01	2,24	0,08	0,49	0,20	0,27	1,00	2013
27	193,90	2,29	0,07	0,51	0,59	0,04	0,59	2013
28	196,40	2,29	0,05	0,42	0,19	0,55	0,49	2013
29	198,03	2,30	0,07	0,55	0,38	0,04	0,38	2013
30	199,21	2,30	0,06	0,40	0,00	0,40	0,71	2013
31	204,72	2,31	0,07	0,48	0,49	0,10	0,24	2013
32	209,83	2,32	0,06	0,38	0,46	0,28	0,25	2013
33	225,41	2,35	0,06	0,45	0,10	0,76	0,75	2013
34	225,44	2,35	0,06	0,41	0,49	0,22	0,27	2013
35	226,80	2,36	0,05	0,44	0,43	0,95	0,60	2013
36	227,55	2,36	0,06	0,52	0,64	0,11	0,55	2013
37	236,35	2,37	0,05	0,43	0,35	0,23	0,06	2013
38	250,66	2,40	0,05	0,41	0,65	0,58	0,58	2013
39	264,15	2,42	0,04	0,42	0,30	0,58	0,62	2013
40	273,99	2,44	0,05	0,44	0,45	0,21	0,31	2013
41	282,23	2,45	0,06	0,39	0,16	0,39	0,84	2013
42	330,69	2,52	0,04	0,42	0,57	0,85	0,63	2013

TT	AGB	logAGB	B7	MSI	VH	elevation	slope	Year
43	336,40	2,53	0,04	0,41	1,00	0,89	0,75	2013
44	373,80	2,57	0,05	0,47	0,58	0,51	0,31	2013
45	471,93	2,67	0,07	0,40	0,21	0,32	0,56	2013
46	473,39	2,68	0,04	0,38	0,54	0,67	0,83	2013
47	551,24	2,74	0,06	0,45	0,34	1,00	0,73	2013

3. Phụ lục 3: Code xử lý tải ảnh vệ tinh

```
// Tải ranh giới tỉnh Đắk Lắk
// =====
var roi = ee.FeatureCollection("FAO/GAUL/2015/level1")
  .filter(ee.Filter.eq('ADM1_NAME', 'Dak Lak'));
Map.centerObject(roi, 8);
// =====
// === 1. DỮ LIỆU ĐIỂM HUẤN LUYỆN (AGB) =====
// =====
var points = points.filterBounds(roi).filter(ee.Filter.notNull(['AGB_D_H', 'Year']));
// Tạo trường logAGB = log(AGB_D_H + 1)
points = points.map(function(f){
  var agb = ee.Number(f.get('AGB_D_H'));
  return f.set('logAGB', agb.add(1).log());
});
var label = 'logAGB';
// =====
// === 2. HÀM TẠO STACK ẢNH CHO 1 NĂM =====
// =====
function makeStack(year) {
  // ----- ẢNH QUANG HỌC (Landsat 8) -----
  var landsat = ee.ImageCollection("LANDSAT/LC08/C02/T1_L2")
    .filterBounds(roi)
    .filter(ee.Filter.calendarRange(year, year, 'year'))
    .map(function(img){
      var qa = img.select('QA_PIXEL');
      var cloudMask = qa.bitwiseAnd(1 << 3).eq(0).and(qa.bitwiseAnd(1 << 4).eq(0));
      var img_sr = img.updateMask(cloudMask)
        .multiply(0.0000275).add(-0.2) // scale Landsat SR
        .select(['SR_B2','SR_B3','SR_B4','SR_B5','SR_B6','SR_B7'],
          ['B2','B3','B4','B5','B6','B7']);
      return img_sr;
    })
    .median();
  // ----- CHỈ SỐ THỰC VẬT -----
  var ndvi = landsat.normalizedDifference(['B5','B4']).rename('NDVI');
  var evi = landsat.expression(
    '2.5 * ((B5 - B4) / (B5 + 6*B4 - 7.5*B2 + 1))',
```

```

    {'B2': landsat.select('B2'),
     'B4': landsat.select('B4'),
     'B5': landsat.select('B5')}).rename('EVI');
var savi = landsat.expression(
  '((B5 - B4) / (B5 + B4 + 0.5)) * 1.5',
  {'B4': landsat.select('B4'),
   'B5': landsat.select('B5')}).rename('SAVI');
var arvi = landsat.expression(
  '(B5 - (2*B4 - B2)) / (B5 + (2*B4 - B2))',
  {'B2': landsat.select('B2'),
   'B4': landsat.select('B4'),
   'B5': landsat.select('B5')}).rename('ARVI');
var sipi = landsat.expression(
  '(B5 - B3) / (B5 - B2)',
  {'B2': landsat.select('B2'),
   'B3': landsat.select('B3'),
   'B5': landsat.select('B5')}).rename('SIPI');
var ndwi = landsat.normalizedDifference(['B3','B5']).rename('NDWI');
var msi = landsat.expression('B6 / B5', {
  'B6': landsat.select('B6'),
  'B5': landsat.select('B5')
}).rename('MSI');
var gndvi = landsat.normalizedDifference(['B5','B3']).rename('GNDVI');
var sr = landsat.expression('B5 / B4', {
  'B5': landsat.select('B5'),
  'B4': landsat.select('B4')
}).rename('SR');
// ----- ĐÌA HÌNH -----
var srtm = ee.Image("USGS/SRTMGL1_003");
var elev = srtm.select('elevation').rename('elevation');
var slope = ee.Terrain.slope(srtm).rename('slope');
// ----- RADAR (Sentinel-1 VV, VH) -----
var s1 = ee.ImageCollection('COPERNICUS/S1_GRD')
  .filterBounds(roi)
  .filter(ee.Filter.calendarRange(year, year, 'year'))
  .filter(ee.Filter.eq('instrumentMode', 'IW'))
  .filter(ee.Filter.eq('resolution_meters', 10))
  .filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VV'))
  .filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VH'))
  .filter(ee.Filter.eq('orbitProperties_pass', 'DESCENDING'))
  .select(['VV','VH'])
  .median();
// ----- GỘP TẤT CẢ THÀNH STACK -----
var stack = landsat
  .addBands([ndvi, evi, savi, arvi, sipi, ndwi, msi, gndvi, sr])
  .addBands([elev, slope])
  .addBands(s1);

```

```

return stack.clip(roi);
}
// =====
// === 3. LẤY MẪU CHO MỖI NĂM =====
// =====
function getTrainingForYear(year) {
  var stack = makeStack(year);
  var ptsYear = points.filter(ee.Filter.eq('Year', year));
  var samples = stack.sampleRegions({
    collection: ptsYear,
    properties: ['AGB_D_H', 'logAGB', 'Year'],
    scale: 30,
    geometries: false
  }).filter(ee.Filter.notNull(stack.bandNames()));
  return samples;
}
// =====
// === 4. GỘP TẤT CẢ CÁC NĂM =====
// =====
var yearsTrain = [2020, 2021, 2022, 2023, 2024];
var allSamples = ee.FeatureCollection([]);

yearsTrain.forEach(function(y) {
  var s = getTrainingForYear(y);
  allSamples = allSamples.merge(s);
});
// =====
// === 5. KIỂM TRA & XUẤT RA DRIVE =====
// =====
print('Tổng số điểm mẫu (2020–2024):', allSamples.size());
print(allSamples.limit(10));

Export.table.toDrive({
  collection: allSamples,
  description: 'Samples_Train_20_24_full',
  folder: 'PhD_BAO',
  fileNamePrefix: 'Samples_20_24_full',
  fileFormat: 'CSV'
});
// =====
// 6. TRAIN MÔ HÌNH RANDOM FOREST (ntree = 1500, mtry = 6)
// =====

// Các biến sẽ dùng để dự báo
var predictors = [
  'B7', 'NDWI', 'MSI', 'VH', 'slope', 'elevation'

```

```

];
// Train RF
var predictors = ['B7', 'NDWI', 'MSI', 'VH', 'slope', 'elevation'];
var rf = ee.Classifier.smileRandomForest({
  numberOfTrees: 1500,
  variablesPerSplit: 6,
  bagFraction: 0.7
})
.setOutputMode('REGRESSION')
.train({
  features: allSamples,
  classProperty: 'logAGB',
  inputProperties: predictors
});
print("RF Regression Model:", rf);
// LƯU MÔ HÌNH RF VÀO ASSET//
var trees = ee.FeatureCollection(rf.explain().get('trees'));
Export.table.toAsset({
  collection: trees,
  description: 'Export_RF_Model',
  assetId: 'users/YOUR_USERNAME/rf_model_agb_1500_6'
});

//DỰ BÁO AGB THEO NĂM//
function predictAGB(year) {
  var stack = makeStack(year);
  var img = stack.select(predictors);
  // Dự báo logAGB
  var log_pred = img.classify(rf).rename('logAGB_pred');
  // Chuyển về AGB
  var agb = log_pred.exp().subtract(1).rename('AGB_pred');
  return agb.set('year', year).clip(roi);
}

//TẠO ẢNH COLLECTION//
var yearsPred = ee.List.sequence(2015, 2025);
var agbCollection = ee.ImageCollection(
  yearsPred.map(function(y) {
    return predictAGB(ee.Number(y));
  })
);
print("AGB Predictions (2015–2025):", agbCollection);
//HIỂN THỊ NĂM 2025//
var agb2025 = agbCollection.filter(ee.Filter.eq('year', 2025)).first();
Map.addLayer(agb2025,
  {min: 0, max: 300, palette: ['yellow','green','darkgreen']},
  'AGB 2025');

```

```
//TÀI BẢN ĐỒ AGB//
yearsPred.getInfo().forEach(function(y) {
  var img = agbCollection.filter(ee.Filter.eq('year', y)).first();
  Export.image.toDrive({
    image: img,
    description: 'AGB_' + y,
    folder: 'PhD_BAO',
    fileNamePrefix: 'AGB_RF_' + y,
    region: roi.geometry(),
    scale: 30,
    maxPixels: 1e13
  });
});
```

Phụ lục 3: Code xử lý số liệu trên RStudio

#1. CÀI & NẠP GÓI

```
# =====
install.packages("tidyverse", repos = "https://cloud.r-project.org/")
install.packages("car", repos = "https://cloud.r-project.org/")
install.packages("mgcv", repos = "https://cloud.r-project.org/")
install.packages("randomForest", repos = "https://cloud.r-project.org/")
install.packages("GGally", repos = "https://cloud.r-project.org/")
```

```
library(tidyverse)
library(car)      # VIF
library(mgcv)     # GAM
library(randomForest) # RF
library(GGally)   # correlation plot
```

```
# =====
```

2. ĐỌC DỮ LIỆU

```
# =====
df <- read.csv("D:/NCS2024/Luan_an/Data/Omau_20-24.csv")
df1 <- read.csv("D:/NCS2024/Luan_an/Data/Omau_2013_Nov1.csv")
#view(df)
#view(df1)
```

#3. Kiểm tra dữ liệu

```
str(df)
summary(df)
```

4. HỒI QUY TUYẾN TÍNH

```
# =====
```

```

#####
#Hồi quy tuyến tính từng biến
lr1 <- lm(AGB ~ B7, data=df)
summary(lr1)
# Hồi quy đơn MSI
lr2 <- lm(AGB ~ MSI, data=df)
summary(lr2)
# Hồi quy đơn: AGB ~ NDWI
lr3 <- lm(AGB ~ NDWI, data=df)
summary(lr3)
# Hồi quy đơn VH
lr4 <- lm(AGB ~ VH, data=df)
summary(lr4)
# Hồi quy đơn: AGB ~ ele
lr5 <- lm(AGB ~ elevation, data=df)
summary(lr5)
# Hồi quy đơn: AGB ~ slope
lr7 <- lm(AGB ~ slope, data=df)
summary(lr7)
#####
# Hồi quy đa biến: AGB ~ nhiều biến
data<-df
lm_multi <- lm(AGB ~ B7+ MSI+NDWI+VH+elevation+slope, data=data)
summary(lm_multi)
# Hồi quy đa biến: logAGB ~ nhiều biến
lm_multi2 <- lm(logAGB ~ B7+ MSI+NDWI+VH+elevation+slope, data=data)
summary(lm_multi2)
lm_multi3 <- lm(logAGB ~ GNDVI+B6+B7+ MSI+VH+elevation+slope, data=data)
# Kiểm tra đa cộng tuyến
vif(lm_multi)
# Kiểm tra đa cộng tuyến logAGB
vif(lm_multi2)
vif(lm_multi3)
lm_multi4 <- lm(logAGB ~ B7+ elevation, data=data)
summary(lm_multi4)
#####
# =====
# 5. MÔ HÌNH PHI TUYẾN (GAM)
# =====
gam1 <- gam(logAGB ~ s(B2)+ s(B3)+ s(B4)+ s(B5)+ s(B6)+ s(B7)+ s(NDVI)+ s(GNDVI)
+ s(ARVI) + s(EVI)+ s(SAVI) + s(SIPI)+ s(SR)+ s(NDWI) + s(MSI) + s(VH)+ s(HH)
+s(elevation)+s(slope), data=data)
summary(gam2)

gam2 <- gam(logAGB ~ s(B7)+ s(NDWI) + s(MSI) + s(VH)+s(elevation)+s(slope),
data=data)
summary(gam2)

```

```

# 2. Nạp thư viện vẽ GAM
install.packages("gratia")
library(gratia)
library(ggplot2)

# 3. Vẽ bằng draw() (ggplot style)
draw(gam2) +
  labs(title = "GAM Smooth Terms", x = "Predictors", y = "Effect on log(AGB)") +
  theme_minimal(base_size = 14)
# =====

# 6. RANDOM FOREST
# =====
set.seed(42)
library(randomForest)
library(caret)
library(Metrics)
library(ggplot2)

# 6.1 CHUẨN BỊ DỮ LIỆU
# =====
#vars_selected <- c("ARVI","B2","B3","B4","B5","B6","B7","EVI", "GNDVI", "MSI",
"NDVI", "NDWI","SAVI", "SIPI", "SR","VV","VH", "elevation", "slope")
#vars_selected <- c("ARVI","B2","B3","B4","B5","B6","B7","EVI", "GNDVI", "MSI",
"NDVI", "NDWI","SAVI", "SIPI", "SR", "elevation", "slope")
vars_selected <- c("B7","MSI","NDWI","VH","elevation","slope")
#vars_selected <- c("B7","MSI","NDWI","elevation","slope")
# Chuyển sang số và loại bỏ NA
data_rf <- df[, c("logAGB", vars_selected)]
#data_rf[] <- lapply(data_rf, function(x) as.numeric(as.character(x)))
#data_rf <- na.omit(data_rf)
# =====

# 6.2 PHÂN TẦNG DỮ LIỆU (STRATIFIED SAMPLING)
# =====
set.seed(42)
# Tạo nhóm phân tầng theo giá trị logAGB (5 nhóm)
data_rf$bin <- cut(data_rf$logAGB, breaks = 5, labels = FALSE)
# Tạo index chia train/test có phân tầng
trainIndex <- createDataPartition(data_rf$bin, p = 0.8, list = FALSE)
train_data <- data_rf[trainIndex, ]
test_data <- data_rf[-trainIndex, ]
# Xóa cột phân tầng
train_data$bin <- NULL
test_data$bin <- NULL
# =====

# 6.3 THIẾT LẬP & HUẤN LUYỆN MÔ HÌNH RF
# =====

```

```

ntree_val <- 1500
mtry_val <- 6
rf_model <- randomForest(
  logAGB ~ .,
  data = train_data,
  ntree = ntree_val,
  mtry = mtry_val,
  importance = TRUE
)

# =====
# 6.4 ĐÁNH GIÁ TRAIN / TEST
# =====
pred_train <- predict(rf_model, newdata = train_data)
pred_test <- predict(rf_model, newdata = test_data)
R2_train <- cor(train_data$logAGB, pred_train, use="complete.obs")^2
RMSE_train <- sqrt(mean((train_data$logAGB - pred_train)^2, na.rm=TRUE))
MAE_train <- mean(abs(train_data$logAGB - pred_train), na.rm=TRUE)
R2_test <- cor(test_data$logAGB, pred_test, use="complete.obs")^2
RMSE_test <- sqrt(mean((test_data$logAGB - pred_test)^2, na.rm=TRUE))
MAE_test <- mean(abs(test_data$logAGB - pred_test), na.rm=TRUE)
cat("==== HIỆU NĂNG MÔ HÌNH RANDOM FOREST (PHÂN TẦNG) ==== \n")
cat("Số cây (ntree):", ntree_val, "\n")
cat("Số biến (mtry):", mtry_val, "\n\n")
cat(">> TRAIN:\n")
cat("R2 =", round(R2_train, 4), " RMSE =", round(RMSE_train, 4), " MAE =",
round(MAE_train, 4), "\n\n")
cat(">> TEST:\n")
cat("R2 =", round(R2_test, 4), " RMSE =", round(RMSE_test, 4), " MAE =",
round(MAE_test, 4), "\n\n")

# =====
# 6.5 VẼ ĐỒ THỊ MINH HỌA MỐI TƯƠNG QUAN
# =====
# --- Hàm vẽ đồ thị tương quan ---
plot_corr <- function(obs, pred, title_text) {
  df_plot <- data.frame(Observed = obs, Predicted = pred)
  # Tính hệ số đánh giá
  R2 <- round(cor(obs, pred, use = "complete.obs")^2, 3)
  RMSE <- round(sqrt(mean((obs - pred)^2, na.rm = TRUE)), 3)
  MAE <- round(mean(abs(obs - pred), na.rm = TRUE), 3)
  ggplot(df_plot, aes(x = Observed, y = Predicted)) +
    geom_point(color = "#1f77b4", alpha = 0.6, size = 2.8) +
    geom_smooth(method = "lm", color = "red", se = FALSE, linetype = "solid") +
    geom_abline(slope = 1, intercept = 0, color = "darkgreen", linetype = "dashed", size = 1)
+
  labs(

```

```

title = title_text,
subtitle = paste("R2 =", R2, "| RMSE =", RMSE, "| MAE =", MAE),
x = "Observed log(AGB)",
y = "Predicted log(AGB)"
) +
theme_bw(base_size = 13) +
theme(
plot.title = element_text(face = "bold", hjust = 0.5),
plot.subtitle = element_text(hjust = 0.5, color = "gray30"),
panel.grid.minor = element_blank()
)
}

# --- Biểu đồ cho tập TRAIN ---
plot_train <- plot_corr(train_data$logAGB, pred_train,
"Random Forest_Sel_vars")
# --- Biểu đồ cho tập TEST ---
plot_test <- plot_corr(test_data$logAGB, pred_test,
"Testing Data - Random Forest")

# Hiện thị
print(plot_train)
print(plot_test)
# =====
# 6.6 BIỂU ĐỒ TẦM QUAN TRỌNG CỦA BIẾN (VARIABLE IMPORTANCE)
# =====
cat("\n==== TẦM QUAN TRỌNG CỦA BIẾN TRONG MÔ HÌNH ==== \n")
print(importance(rf_model))
varImpPlot(rf_model, main = "Tầm quan trọng của các biến - Random Forest")
# =====
# 7. ĐÁNH GIÁ MÔ HÌNH OLS, GAM, RF
# =====
# Hàm tính RMSE và MAE
rmse <- function(obs, pred) sqrt(mean((obs - pred)^2, na.rm=TRUE))
mae <- function(obs, pred) mean(abs(obs - pred), na.rm=TRUE)
# Dự báo từ các mô hình
pred_lm <- predict(lm_multi2, newdata=data)
pred_gam <- predict(gam2, newdata=data)
pred_rf <- predict(rf_model, newdata=data)
# So sánh
cat("OLS RMSE:", rmse(data$logAGB, pred_lm), "MAE:", mae(data$logAGB, pred_lm),
"\n")
cat("GAM RMSE:", rmse(data$logAGB, pred_gam), "MAE:", mae(data$logAGB,
pred_gam), "\n")
cat("RF RMSE:", rmse(data$logAGB, pred_rf), "MAE:", mae(data$logAGB, pred_rf),
"\n")
# =====
# 7.1 ĐÁNH GIÁ MÔ HÌNH

```

```

# =====
# Hàm tính RMSE, MAE, R2
rmse <- function(obs, pred) sqrt(mean((obs - pred)^2, na.rm=TRUE))
mae <- function(obs, pred) mean(abs(obs - pred), na.rm=TRUE)
rsq <- function(obs, pred) cor(obs, pred, use="complete.obs")^2

# Dự báo từ các mô hình
pred_lm <- predict(lm_multi2, newdata=data)
pred_gam <- predict(gam2, newdata=data)
pred_rf <- predict(rf_model, newdata=data)
# Tính toán chỉ số
results <- data.frame(
  Model = c("OLS", "GAM", "RF"),
  RMSE = c(rmse(data$logAGB, pred_lm),
           rmse(data$logAGB, pred_gam),
           rmse(data$logAGB, pred_rf)),
  MAE = c(mae(data$logAGB, pred_lm),
           mae(data$logAGB, pred_gam),
           mae(data$logAGB, pred_rf)),
  R2 = c(rsq(data$logAGB, pred_lm),
          rsq(data$logAGB, pred_gam),
          rsq(data$logAGB, pred_rf))
)
print(results)
# =====
# 7.2 VỀ BIỂU ĐỒ SO SÁNH
# =====
p1 <- ggplot(data, aes(x = logAGB, y = pred_lm)) +
  geom_point(color = "darkgreen", alpha = 0.6) +
  geom_abline(intercept = 0, slope = 1, color = "red", linetype="dashed") +
  labs(title="OLS", x="logAGB thực tế", y="logAGB ước tính") +
  theme_minimal()
p2 <- ggplot(data, aes(x = logAGB, y = pred_gam)) +
  geom_point(color = "orange", alpha = 0.6) +
  geom_abline(intercept = 0, slope = 1, color = "red", linetype="dashed") +
  labs(title="GAM", x="logAGB thực tế", y="logAGB ước tính") +
  theme_minimal()
p3 <- ggplot(data, aes(x = logAGB, y = pred_rf)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_abline(intercept = 0, slope = 1, color = "red", linetype="dashed") +
  labs(title="Random Forest", x="logAGB thực tế", y="logAGB ước tính") +
  theme_minimal()
# nạp thư viện
install.packages("gridExtra")
library(gridExtra)
# Hiển thị 3 biểu đồ cạnh nhau
grid.arrange(p1, p2, p3, ncol=3)

```

```
#####
# 8. ĐÁNH GIÁ MÔ HÌNH BẰNG K-FOLD CROSS-VALIDATION
# (K = 5 và K = 10)
#####
# Cài đặt và nạp gói
install.packages(c("caret", "mgcv", "randomForest", "Metrics"))
library(caret)
library(mgcv)
library(randomForest)
library(Metrics)

set.seed(123)
# -----
# 8.1 Chuẩn bị dữ liệu
# -----
vars_selected <- c("B7", "NDWI", "VH", "elevation", "slope", "MSI")
data_sel <- df[, c("logAGB", vars_selected)]
data_sel <- na.omit(data_sel)
# -----
# 8.2 Thiết lập hàm đánh giá CV
# -----
run_kfold_cv <- function(model_method, data, k) {
  ctrl <- trainControl(method = "cv", number = k)
  model_cv <- train(
    logAGB ~ .,
    data = data,
    method = model_method, # tên mô hình
    trControl = ctrl,
    tuneLength = 3
  )
  results <- model_cv$results[which.max(model_cv$results$Rsquared), ]
  results$k <- k
  return(results[, c("k", "RMSE", "Rsquared", "MAE")])
}
# -----
# 8.3 Chạy CV cho 3 mô hình: OLS, GAM, RF
# -----
# OLS (Linear Regression)
res_lm_5 <- run_kfold_cv("lm", data_sel, 5)
res_lm_10 <- run_kfold_cv("lm", data_sel, 10)
# GAM (phi tuyến)
res_gam_5 <- run_kfold_cv("gam", data_sel, 5)
res_gam_10 <- run_kfold_cv("gam", data_sel, 10)
# Random Forest (RF)
res_rf_5 <- run_kfold_cv("rf", data_sel, 5)
res_rf_10 <- run_kfold_cv("rf", data_sel, 10)
# -----
```

```

# 8.4 Tổng hợp kết quả so sánh
# -----
results_cv <- rbind(
  cbind(Model = "OLS", res_lm_5),
  cbind(Model = "OLS", res_lm_10),
  cbind(Model = "GAM", res_gam_5),
  cbind(Model = "GAM", res_gam_10),
  cbind(Model = "RF", res_rf_5),
  cbind(Model = "RF", res_rf_10)
)
print(results_cv)
# -----
# 8.5 Biểu đồ so sánh R2 giữa 5-fold và 10-fold
# -----
library(ggplot2)
ggplot(results_cv, aes(x = factor(k), y = Rsquared, fill = Model)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  labs(title = "So sánh hiệu suất mô hình theo K-Fold CV",
       x = "Số K trong K-Fold Cross Validation",
       y = expression(R2)) +
  theme_minimal(base_size = 14)
#####
#####
# 10. ĐÁNH GIÁ MÔ HÌNH RANDOM FOREST BẰNG DỮ LIỆU ĐỘC LẬP
#####
# Cài đặt & nạp thư viện
#install.packages(c("randomForest", "Metrics", "ggplot2"))
library(randomForest)
library(Metrics)
library(ggplot2)
# -----
# 10.1 Đọc dữ liệu
# -----
df1 <- read.csv("D:/NCS2024/Luan_an/Data/Omau_2013_Nov1.csv")
Omau_2013 <- df1
#view(Omau_2013)
# -----
# 10.2 Chọn các biến sử dụng trong mô hình
# -----
vars_selected <- c("B7", "elevation", "MSI", "NDWI", "slope")
train_data <- df[, c("logAGB", vars_selected)]
test_data <- Omau_2013[, c("logAGB", vars_selected)]
view(test_data)
# -----
# 10.3 Huấn luyện mô hình RF trên tập train
# -----
set.seed(42)

```

```

vars_selected <- c("B7", "elevation", "MSI", "NDWI", "slope")
data_rf <- df[, c("logAGB", vars_selected)]
data_rf <- na.omit(data_rf)
# -----
# 10.4 Xây dựng mô hình RF
# -----
set.seed(42)
rf_model <- randomForest(
  logAGB ~ .,
  data = data_rf,
  ntree = 1500,
  mtry=6,
  importance = TRUE
)
print(rf_model)
# -----
# 10.5 Dự báo trên dữ liệu độc lập
# -----
pred_rf <- predict(rf_model, newdata = test_data)
# -----
# 10.6 Tính các chỉ số đánh giá
# -----
RMSE_val <- rmse(test_data$logAGB, pred_rf)
MAE_val <- mae(test_data$logAGB, pred_rf)
R2_val <- cor(test_data$logAGB, pred_rf)^2
Bias_val <- mean(pred_rf - test_data$logAGB)
# Tính slope của đường hồi quy
lm_fit <- lm(pred_rf ~ logAGB, data = test_data)
Slope_val <- coef(lm_fit)[2]
# -----
# 10.7 Biểu đồ thực tế - dự báo
# -----
ggplot(test_data, aes(x = logAGB, y = pred_rf)) +
  geom_point(color = "#1b9e77", alpha = 0.7, size = 3) +
  # Đường x = y (đường 1:1)
  geom_abline(intercept = 0, slope = 1,
              color = "red", linetype = "dashed", linewidth = 1) +
  # Đường hồi quy ước lượng
  geom_smooth(method = "lm", se = FALSE,
              color = "blue", linewidth = 1) +
  labs(
    title = "Kiểm định mô hình Random Forest bằng dữ liệu độc lập",
    x = "log(AGB) thực tế (Observed)",
    y = "log(AGB) ước tính (Predicted)"
  ) +
  annotate("text",
         x = min(test_data$logAGB) + 0.2,

```

```
y = max(pred_rf) - 0.3,  
hjust = 0, vjust = 1, size = 5, color = "black",  
label = paste0("R2 = ", round(R2_val, 3),  
              "\nRMSE = ", round(RMSE_val, 3),  
              "\nMAE = ", round(MAE_val, 3),  
              "\nBias = ", round(Bias_val, 3),  
              "\nSlope = ", round(Slope_val, 3))) +  
theme_minimal(base_size = 14)
```