



# Ứng dụng công nghệ phản xạ GNSS-R và học máy trong ước tính mật độ sinh khối

Nguyễn Phương Bắc<sup>1</sup>, Vũ Phương Lan<sup>1\*</sup>, Hà Minh Cường<sup>2</sup>, Nguyễn Vũ Hà<sup>1</sup>

<sup>1</sup>Trường ĐH Khoa học Tự Nhiên, ĐHQG HN, 334 Nguyễn Trãi, Thanh Xuân, Hà Nội, Việt Nam

<sup>2</sup>Trường ĐH Công nghệ, ĐHQG HN, 144 Xuân Thủy, Cầu Giấy, Hà Nội, Việt Nam

Email tác giả liên hệ: [vuphuonglan@hus.edu.vn](mailto:vuphuonglan@hus.edu.vn)

<https://doi.org/10.5281/zenodo.18186255>

## Tóm tắt:

Ước tính sinh khối là một nhiệm vụ thiết yếu trong quản lý rừng bền vững và giám sát môi trường toàn cầu, đặc biệt trong bối cảnh biến đổi khí hậu. Các phương pháp truyền thống dựa trên điều tra thực địa thường tốn nhiều thời gian, chi phí và khó triển khai trên diện rộng, trong khi các kỹ thuật viễn thám quang học và radar khẩu độ tổng hợp (SAR) dù đã chứng minh hiệu quả nhưng vẫn còn những hạn chế nhất định. Gần đây, công nghệ viễn thám cơ hội sử dụng tín hiệu phản xạ GNSS (GNSS-R) nổi lên như một phương pháp tiềm năng nhờ ưu điểm hoạt động trong mọi điều kiện thời tiết, cung cấp dữ liệu liên tục và chi phí thấp. Dữ liệu GNSS-R phản xạ từ thảm thực vật mang thông tin về cấu trúc và mật độ của chúng, cho phép ước tính các thông số sinh vật lý, bao gồm sinh khối. Nghiên cứu này tập trung vào việc đánh giá tiềm năng của phương pháp học máy trong việc ước tính sinh khối thực vật tại khu vực phía Bắc Việt Nam, sử dụng dữ liệu phản xạ GNSS-R thu được từ vệ tinh CyGNSS. Kết quả cho thấy mối tương quan rất cao (lên tới 0,99) giữa các đặc trưng tín hiệu GNSS-R và mật độ sinh khối. Nghiên cứu đã khẳng định GNSS-R kết hợp học máy có thể trở thành công cụ hiệu quả cho giám sát sinh khối nhanh chóng và liên tục, đồng thời mở ra hướng ứng dụng trong quản lý tài nguyên rừng và nghiên cứu biến đổi khí hậu ở quy mô quốc gia và toàn cầu.

**Từ khóa:** GNSS-R, Sinh khối, Rừng, học máy

Ngày nhận bài: 10/09/2025

Ngày sửa lại: 28/09/2025

Ngày chấp nhận đăng: 30/09/2025

Ngày xuất bản: 31/12/2025

## Application of GNSS Reflectometry and Machine Learning in Estimation Vegetation Biomass

Phuong Bac Nguyen<sup>1</sup>, Phuong Lan Vu<sup>1\*</sup>, Minh Cuong Ha<sup>2</sup>, Vu Ha Nguyen<sup>1</sup>

<sup>1</sup>VNU University of Science, 334 Nguyen Trai, Thanh Xuan, Ha Noi, Viet Nam

<sup>2</sup>VNU University of Engineering and Technology, 144 Xuan Thuy, Cau Giay, Ha Noi, Viet Nam

Corresponding Author Email: [vuphuonglan@hus.edu.vn](mailto:vuphuonglan@hus.edu.vn)

## Abstract:

Biomass estimation plays a vital role in sustainable forest management and global environmental monitoring, especially under the challenges of climate change. Traditional field-based surveys are costly, labor-intensive, and difficult to implement at large scales, while optical remote sensing and synthetic aperture radar (SAR) methods, though effective, remain constrained by specific limitations. Recently, Global Navigation Satellite System Reflectometry (GNSS-R) has emerged as a promising alternative, offering continuous, low-cost, and all-weather observations. Reflected GNSS signals from vegetation provide structural and density-related information that can be used to infer biophysical parameters such as biomass. This study evaluates the potential of machine learning algorithms for estimating vegetation biomass in northern Vietnam using GNSS-R data collected by the CyGNSS satellite constellation. Results indicate a strong correlation (up to 0.99) between GNSS-R signal features and biomass density. The findings highlight the effectiveness of GNSS-R combined with machine learning for rapid, scalable, and continuous biomass monitoring, with implications for forest resource management and climate change research.

**Keywords:** GNSS-R, biomass, forest, machine learning

Submission received: 10/09/2025

Revised: 28/09/2025

Accepted: 30/09/2025

Published: 31/12/2025

## 1. Mở đầu

Tài nguyên rừng đóng vai trò nền tảng trong sự phát triển bền vững của các quốc gia. Vượt ra ngoài chức năng cung cấp lâm sản, rừng là hệ sinh thái then chốt duy trì sự cân bằng sinh thái toàn cầu, đảm bảo tính đa dạng sinh học và điều hòa khí hậu [1]. Trong bối cảnh biến đổi khí hậu gia tăng, việc quản lý và theo dõi tài nguyên rừng một cách hiệu quả trở nên cấp bách. Một trong những tham số quan trọng nhất được sử dụng trong lĩnh vực này là sinh khối trên mặt đất (Aboveground Biomass - AGB). AGB không chỉ là chỉ số trực tiếp phản ánh trữ lượng carbon của một hệ sinh thái rừng, mà còn là đại diện cho năng suất sinh học và sức khỏe



của hệ sinh thái đó [2]. Các nghiên cứu đã chứng minh rằng AGB có mối liên hệ chặt chẽ với chu trình carbon toàn cầu và là một biến số thiết yếu trong việc định lượng và giám sát các mục tiêu giảm phát thải từ mất rừng và suy thoái rừng (REDD+). Do đó, việc ước tính và giám sát AGB một cách chính xác là nền tảng cho công tác quản lý lâm nghiệp bền vững và bảo tồn tài nguyên môi trường.

Các phương pháp truyền thống như khảo sát thực địa tỏ ra không hiệu quả do tốn kém chi phí, thời gian và khó tiếp cận các khu vực hẻo lánh, đặc biệt là ở những khu rừng nhiệt đới rậm rạp [1]. Để khắc phục hạn chế này, các công nghệ viễn thám như radar và LiDAR đã được ứng dụng rộng rãi [3–5]. Tuy nhiên, dữ liệu radar thường bị bão hòa khi sinh khối vượt quá một ngưỡng nhất định (thường khoảng 100 tấn/ha), làm giảm độ chính xác của ước tính ở các khu vực rừng có sinh khối cao [6,7]. Trong khi đó, tín hiệu LiDAR lại dễ dàng bị ảnh hưởng bởi điều kiện thời tiết như mây, sương mù, làm giảm hiệu quả thu thập dữ liệu tại các khu vực khí hậu nhiệt đới ẩm ướt như Việt Nam [8]. Do đó, việc nghiên cứu và phát triển các phương pháp tích hợp dữ liệu từ nhiều nguồn khác nhau là cần thiết để nâng cao độ chính xác và khả năng ứng dụng trong giám sát AGB trên diện rộng.

Phương pháp viễn thám mới tận dụng tín hiệu phản xạ từ hệ thống vệ tinh định vị toàn cầu (GNSS-R) đang nổi lên như một công nghệ viễn thám thụ động đầy tiềm năng, thu hút sự quan tâm đáng kể trong lĩnh vực quan trắc môi trường và quản lý tài nguyên thiên nhiên [9,10]. Về bản chất, GNSS-R hoạt động như một hệ thống radar đa tĩnh (multistatic radar), sử dụng các tín hiệu vệ tinh định vị toàn cầu có sẵn ở băng tần L làm nguồn phát. Kỹ thuật này khai thác năng lượng phản xạ từ bề mặt Trái Đất để thực hiện các phép đo liên tục, không bị cản trở bởi điều kiện thời tiết [11]. Tín hiệu ở băng tần L ( $\approx 1-2$  GHz) có đặc tính vật lý quan trọng là độ nhạy cao đối với độ ẩm bề mặt và khả năng xuyên qua thảm thực vật [12]. Những ưu điểm về tính liên tục, chi phí thấp và khả năng hoạt động trong mọi điều kiện thời tiết đã khiến GNSS-R trở thành một công cụ đầy hứa hẹn trong việc ước tính các tham số địa vật lý quan trọng. Nhiều nhiệm vụ không gian sử dụng công nghệ này đã được triển khai, tiêu biểu như UK-DMC-1 và TDS-1, tập trung chủ yếu vào quan trắc đại dương, đặc biệt là ước tính tốc độ gió trên biển và độ cao sóng hiệu dụng [13,14]. Các sứ mệnh gần đây và chuyên biệt hơn, điển hình là CyGNSS (Cyclone Global Navigation Satellite System), tuy ban đầu được thiết kế để theo dõi bão nhiệt đới và tốc độ gió, nhưng đã nhanh chóng mở rộng phạm vi ứng dụng sang lĩnh vực quan trắc trên đất liền, mở ra nhiều cơ hội nghiên cứu mới [15]. Trong những năm gần đây, nhiều nghiên cứu đã tập trung vào việc khai thác dữ liệu GNSS-R trên đất liền để ước tính độ ẩm đất [16,17], sinh khối thực vật trên mặt đất [18], theo dõi băng tuyết và độ cao nước trong nội địa [19–22]. Sự phát triển này cũng được hỗ trợ bởi các nhiệm vụ mới như FengYun-3 [23] và Bufeng-1 [24], tiếp tục mở rộng phạm vi và chất lượng dữ liệu GNSS-R toàn cầu.

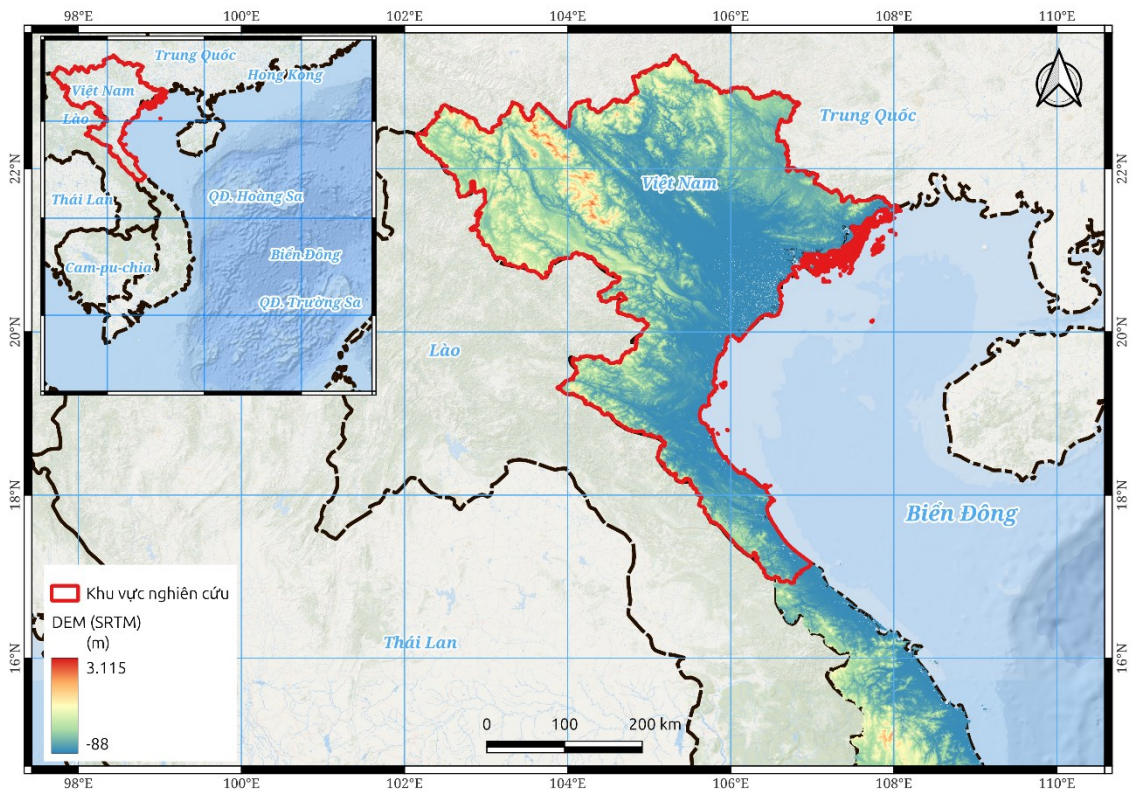
Kể từ khi được lắp đặt trên Trạm Vũ trụ Quốc tế (ISS) vào năm 2018, GEDI (Global Ecosystem Dynamics Investigation) đã trở thành nguồn dữ liệu chủ yếu trong nghiên cứu sinh khối rừng toàn cầu. Với khả năng đo chiều cao tán rừng và cấu trúc thẳng đứng ở độ phân giải điểm  $\sim 25$  m, GEDI đã được ứng dụng rộng rãi để xây dựng và kiểm chứng bản đồ AGB ở nhiều quy mô khác nhau. Các nghiên cứu đã chứng minh rằng sản phẩm GEDI Level 4A Mật độ sinh khối trên mặt đất (AGBD) có độ chính xác cao khi được hiệu chỉnh bằng dữ liệu thực địa, đồng thời đóng vai trò tham chiếu quan trọng trong hiệu chuẩn và đánh giá các bản đồ sinh khối tích hợp từ viễn thám quang học (Sentinel-2, Landsat), radar (Sentinel-1, ALOS PALSAR), và các mô hình kết hợp [8,25]. Tuy nhiên, dữ liệu GEDI có tính rời rạc theo vệt quét và phạm vi ứng dụng ở các khu vực nhiệt đới gió mùa, đặc biệt là Đông Nam Á, vẫn còn hạn chế. Do đó, việc tích hợp GEDI với các nguồn dữ liệu viễn thám khác và áp dụng phương pháp học máy tiên tiến là cần thiết để nâng cao độ chính xác và khả năng khái quát hóa trong ước tính sinh khối.

Nghiên cứu này ứng dụng dữ liệu CyGNSS và GEDI dựa trên các thuật toán học máy để xây dựng mô hình ước tính sinh khối cho khu vực miền bắc Việt Nam. Sự kết hợp này tận dụng đồng thời độ nhạy của CyGNSS đối với độ ẩm và đặc tính bề mặt, cùng thông tin cấu trúc tán rừng theo phương thẳng đứng do GEDI cung cấp [8]. Cách tiếp cận này mở ra tiềm năng cải thiện giám sát môi trường và hỗ trợ quản lý tài nguyên bền vững ở quy mô lớn.

## 2. Khu vực nghiên cứu và dữ liệu

### 2.1 Khu vực nghiên cứu

Khu vực nghiên cứu được lựa chọn ở miền Bắc Việt Nam (Hình 1), giới hạn trong tọa độ địa lý 16,9 đến 23,4 độ vĩ Bắc và 102,1 đến 108 độ kinh Đông, nơi có điều kiện địa lý và khí hậu điển hình cho vùng nhiệt đới gió mùa. Đặc trưng của khu vực là mùa hè nóng ẩm, mưa nhiều, mùa đông khô lạnh, kèm theo độ ẩm không khí cao và tần suất mây che phủ lớn. Những yếu tố này thường gây trở ngại cho viễn thám quang học và LiDAR, nhưng lại tạo môi trường phù hợp để đánh giá tiềm năng của công nghệ GNSS-R, vốn ít chịu tác động của điều kiện khí quyển. Miền Bắc Việt Nam có sự đa dạng về địa hình, trải dài từ đồng bằng châu thổ, trung du đến vùng núi cao. Quy mô nghiên cứu rộng lớn cho phép khai thác lợi thế của dữ liệu GNSS-R vốn có độ phân giải không gian thấp, nhưng phù hợp để phân tích trên diện rộng và đảm bảo độ tin cậy thống kê.



Hình 1: Khu vực nghiên cứu

Bên cạnh đó, miền Bắc Việt Nam sở hữu sự đa dạng cao về thảm thực vật, bao gồm rừng tự nhiên thường xanh, rừng rụng lá, rừng ngập mặn ven biển và rừng trồng. Điều này cho phép đánh giá khả năng của phương pháp trong việc phân biệt và ước tính AGB của các loại cấu trúc rừng khác nhau, từ đó chứng minh tính khả thi của mô hình thực nghiệm của nghiên cứu.

### 2.2. Dữ liệu sử dụng

#### 2.2.1. Dữ liệu vệ tinh CyGNSS

CyGNSS là một sứ mệnh không gian tiên phong của NASA, được triển khai nhằm cung cấp các phép đo vận tốc gió bề mặt đại dương, đặc biệt tập trung vào các xoáy thuận nhiệt đới.



Chòm sao này bao gồm tám vệ tinh nhỏ trên quỹ đạo Trái đất thấp (LEO), khai thác tín hiệu L-band từ vệ tinh GPS phản xạ lại từ bề mặt biển, sau đó phân tích đặc trưng tín hiệu thông qua bản đồ trễ - Doppler (Delay-Doppler Map, DDM) để suy ra độ nhám mặt nước và ước tính tốc độ gió bề mặt [15,26].

Với chu kỳ lặp quan sát chỉ vài giờ tại khu vực nhiệt đới và cận nhiệt đới, CyGNSS đã góp phần cải thiện đáng kể năng lực dự báo và mô phỏng cường độ bão [27]. Tất cả các sản phẩm dữ liệu CyGNSS, bao gồm dữ liệu thô mức 1 và các sản phẩm như sức gió bề mặt mức 2 đã được xử lý, đều được truy cập miễn phí, tạo điều kiện thuận lợi cho các nghiên cứu về động lực khí quyển - đại dương. Trong nghiên cứu này, 1.977 tập dữ liệu từ 8 vệ tinh CyGNSS thu thập trong năm 2024 đã được sử dụng để xây dựng mô hình dự đoán sinh khối cho toàn bộ khu vực nghiên cứu.

### 2.2.2. Dữ liệu vệ tinh GEDI

GEDI là hệ thống LiDAR toàn sóng (full-waveform LiDAR) tiên tiến, được phát triển với mục tiêu giải quyết những bất định lớn trong việc định lượng và hiểu rõ chu trình carbon trên đất liền. Thiết bị này cung cấp những quan sát có độ phân giải cao đầu tiên về cấu trúc dọc của rừng trên phạm vi toàn cầu, bao phủ các hệ sinh thái rừng nhiệt đới và ôn đới nằm trong khoảng vĩ độ 51.6°B và 51.6°N [8]. Phạm vi quan sát này vẫn bao trùm phần lớn các khu vực rừng quan trọng nhất thế giới – nơi lưu trữ phần lớn sinh khối carbon toàn cầu.

Trong số các sản phẩm dữ liệu của GEDI, Level 4A có ý nghĩa khoa học đặc biệt quan trọng, cung cấp ước tính AGBD cho từng điểm đo. Dữ liệu L4A được tạo ra bằng cách xử lý tín hiệu LiDAR thông qua các mô hình hồi quy tiên tiến, thường kết hợp phương pháp thống kê và thuật toán học máy, được huấn luyện trên dữ liệu sinh khối thực địa và tham số cấu trúc rừng từ GEDI. Nhờ độ chính xác cao ở cấp điểm, GEDI L4A đã trở thành nguồn dữ liệu chuẩn để kiểm chứng các bản đồ AGB toàn cầu và khu vực [25], đặc biệt là các sản phẩm tích hợp từ viễn thám quang học, radar (như Sentinel-1/2, ALOS PALSAR) và dữ liệu địa hình [28,29].

Trong nghiên cứu này, dữ liệu GEDI được sử dụng làm dữ liệu tham chiếu cho tập huấn luyện mô hình dự đoán sinh khối. Theo đó, một tập hợp gồm 340 tập dữ liệu GEDI mức 4A, được thu thập trong năm 2024, đã được xử lý và cắt theo khu vực nghiên cứu. Chúng tôi ưu tiên sử dụng.

## 3. Phương pháp nghiên cứu

### 3.1. Ước tính độ nhám bề mặt từ dữ liệu CyGNSS

Cơ chế cho phép GNSS-R trở nên nhạy cảm với sinh khối mặt đất xuất phát từ sự suy giảm cường độ tín hiệu phản xạ từ bề mặt đất khi có sự che phủ của thảm thực vật [12]. Về mặt nguyên lý, khi sóng GNSS phát ra từ vệ tinh đi qua tầng khí quyển và chạm tới bề mặt Trái Đất, tín hiệu phản xạ sẽ bị hấp thụ, tán xạ hoặc làm suy yếu bởi các thành phần của thảm thực vật, đặc biệt là lá, cành và thân cây. Mức độ suy giảm này phụ thuộc mạnh vào đặc điểm cấu trúc, độ dày và hàm lượng nước của thảm thực vật, qua đó tạo ra mối liên hệ gián tiếp giữa tín hiệu GNSS-R và sinh khối mặt đất.

Một trong những đại lượng quan trọng dùng để định lượng sự tương tác này là độ nhám bề mặt ( $\Gamma$ ), được xác định theo phương trình [30]:

$$\Gamma_{r_l} = \frac{(4\pi)^2 P_{r_l}^r (R_{ts} + R_{sr})^2}{P_r^t G^t G^r \lambda^2} \quad (1)$$

Trong đó  $P_{r_l}^r$  là công suất tán xạ,  $R^t$  là khoảng cách từ điểm phản chiếu tới máy phát,  $R^r$  là khoảng cách từ máy thu đến điểm phản chiếu,  $\lambda$  là bước sóng của tín hiệu GPS L1 ( $\lambda = 0.19m$ ),  $G^r$  là độ lợi anten máy thu theo hướng điểm phản chiếu, và  $G^t P^t$  là công suất bức xạ đẳng hướng tương đương máy phát (EIRP).

Đại lượng  $\Gamma_{r_l}$  phản ánh tỷ lệ năng lượng tín hiệu GNSS bị phản xạ so với năng lượng tín hiệu tới, do đó cung cấp chỉ báo tổng hợp về tính chất bề mặt bao gồm cả thảm thực vật. Hình

2 minh họa giá trị trung bình dạng lưới của  $\Gamma$  trên lưới với độ phân giải không gian 0.03 độ (tương đương khoảng 3km) để thể hiện sự biến thiên theo khu vực.

Các khu vực có giá trị AGB cao thường đi kèm với giá trị  $\Gamma$  thấp, nghĩa là nơi sinh khối dày đặc làm giảm khả năng phản xạ tín hiệu GNSS-R; ngược lại, các khu vực có thảm thực vật thưa thớt hoặc bề mặt đất trống thường cho giá trị  $\Gamma$  cao hơn [18]. Mối quan hệ nghịch đảo này cung cấp cơ sở khoa học quan trọng cho việc ứng dụng GNSS-R trong ước tính sinh khối, đồng thời gợi ý rằng dữ liệu GNSS-R có thể bổ sung hữu ích cho các sản phẩm viễn thám quang học hoặc radar truyền thống.

Tuy nhiên, ước tính định lượng sinh khối từ dữ liệu GNSS-R vẫn là một thách thức lớn. Cường độ tín hiệu GNSS-R thu được không chỉ phụ thuộc vào sinh khối mà còn chịu ảnh hưởng của nhiều yếu tố khác như góc tới và góc phản xạ của tín hiệu, độ nhám bề mặt, độ ẩm đất, điều kiện khí quyển, tần số và cấu hình vệ tinh. Các yếu tố này có thể gây ra biến động đáng kể trong tín hiệu phản xạ, gây khó khăn trong việc tách đóng góp của sinh khối từ dữ liệu phản xạ.

Độ nhám bề mặt được ước tính từ dữ liệu CyGNSS theo công thức (1), với tất cả các giá trị đã được quy đổi về đơn vị dB nhằm đảm bảo tính đồng nhất và khả năng so sánh giữa các khu vực [30]:

$$\Gamma_{rL} = 10 \cdot \log_{10}(P_{rL}^r) + 20 \cdot \log_{10}(4\pi) + 20 \cdot \log_{10}(R_{ts} + R_{sr}) - 10 \cdot \log_{10}(P_r^t G^t) - 10 \cdot \log_{10}(G^r) - 20 \cdot \log_{10}(\lambda) \quad (2)$$

Tín hiệu GNSS băng tần L chịu ảnh hưởng đáng kể của điều kiện bề mặt, đặc biệt là khi tồn tại các vùng mặt nước hoặc bề mặt ẩm ướt. Do đó, độ nhám bề mặt đóng vai trò như một chỉ số quan trọng phản ánh trực tiếp mức suy hao tín hiệu GNSS do sự tán xạ khuếch tán của sóng vô tuyến tại bề mặt phản xạ.

### 3.2. Thuật toán Random Forest trong ước tính mật độ sinh khối

Để giải quyết mối quan hệ phi tuyến phức tạp giữa các biến giải thích và AGBD từ dữ liệu CyGNSS, nghiên cứu này áp dụng mô hình rừng ngẫu nhiên dưới dạng hồi quy, một phương pháp học máy dựa trên kỹ thuật tập hợp (ensemble learning) có khả năng mô phỏng các quan hệ đa chiều và phi tuyến một cách linh hoạt [31].

Trong khuôn khổ này, giá trị sinh khối AGBD được biểu diễn như một hàm của tập hợp các biến giải thích  $X = x_1, x_2, \dots, x_p$  được trích xuất từ dữ liệu CyGNSS, bao gồm các đặc trưng phản xạ tín hiệu GNSS-R, các tham số hình học (góc phản xạ, góc tới), và các yếu tố môi trường phụ trợ:

$$AGBD = f(X) + \varepsilon \quad (3)$$

Trong đó  $f(X)$  được ước lượng bằng thuật toán RF, là trung bình của B cây hồi quy độc lập:

$$\hat{f}(x) = \frac{1}{B} \sum T_b(x) \quad (4)$$

với  $T_b$  là cây hồi quy thứ  $b$ , được xây dựng trên một mẫu bootstrap, ngẫu nhiên rút ra từ tập huấn luyện gốc. Mỗi cây được xây dựng bằng cách chia tách đệ quy các nút để giảm sai số dự báo, và tại mỗi nút chia, chỉ một tập con ngẫu nhiên gồm  $m$  biến (với  $m < p$ ) được xem xét để lựa chọn tiêu chí tách tốt nhất. Cơ chế lựa chọn ngẫu nhiên này giúp giảm tương quan giữa các cây, từ đó cải thiện độ chính xác tổng thể của mô hình.

### 3.3 Đánh giá độ chính xác của mô hình

Hiệu suất của mô hình Random Forest (RF) được đánh giá thông qua chỉ số thống kê và kiểm tra độc lập trên tập dữ liệu kiểm chứng nhằm đảm bảo tính tin cậy và khả năng tổng quát hóa của mô hình. Bộ dữ liệu được chia thành hai phần: huấn luyện và kiểm tra, được lựa chọn ngẫu nhiên nhưng bảo đảm duy trì phân bố của biến mục tiêu (AGB).

Để định lượng mức độ phù hợp giữa giá trị sinh khối ước tính và giá trị quan sát thực tế, nghiên cứu sử dụng sai số trung phương:

$$RMSE = \sqrt{\frac{1}{n} \sum (\hat{Y}_i - Y_i)^2} \quad (5)$$

và hệ số tương quan Pearson:

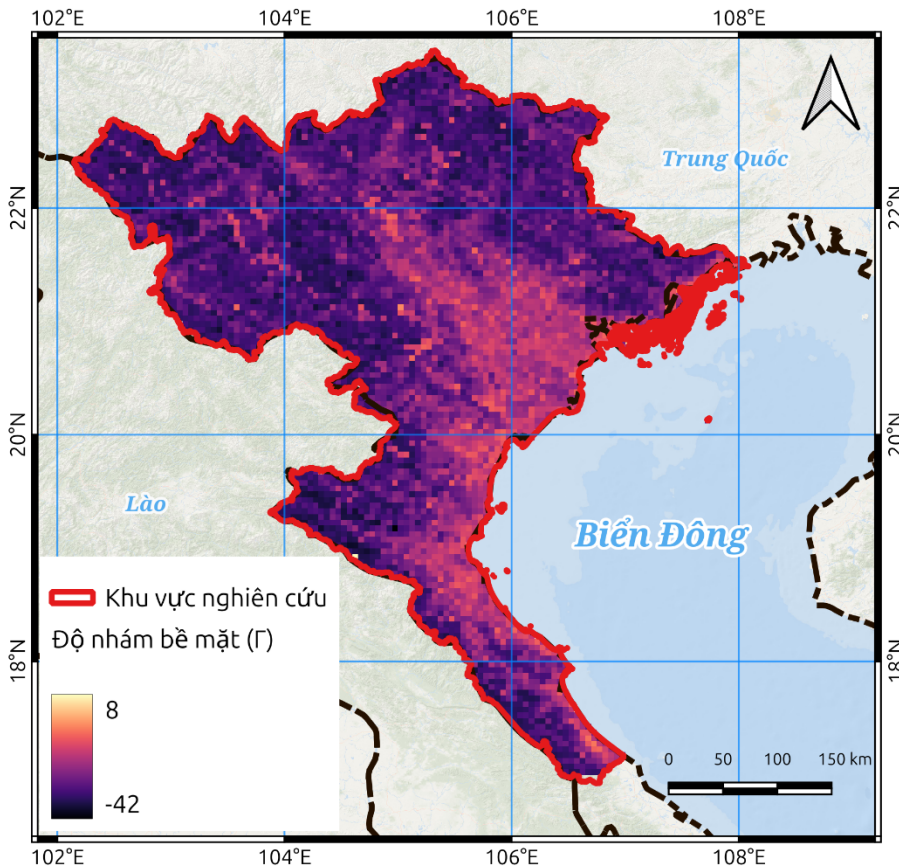
$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

Bên cạnh các chỉ số sai số, nghiên cứu cũng tiến hành cũng tính toán độ quan trọng của biển.

## 4. Kết quả

### 4.1 Bản đồ phân bố độ nhám bề mặt

Bản đồ phân bố không gian của độ nhám bề mặt trên khu vực nghiên cứu được thể hiện trong Hình 2. Có thể nhận thấy các vùng đồng bằng ven biển và khu vực ven sông suối lớn, đặc biệt là vùng châu thổ sông Hồng và vùng ven biển vịnh Bắc Bộ, có xu hướng xuất hiện các giá trị  $\Gamma$  cao hơn, phản ánh bề mặt có mức tán xạ mạnh do sự hiện diện của nước hoặc đất bùn ẩm. Ngược lại, các khu vực miền núi phía Tây Bắc và vùng rừng núi Hoàng Liên Sơn chủ yếu thể hiện giá trị  $\Gamma$  thấp hơn, tương ứng với bề mặt rừng rậm hoặc địa hình gồ ghề nhưng ít mặt nước, dẫn đến tín hiệu phản xạ yếu và suy hao nhỏ hơn điều này khá tương đồng với nghiên cứu của Pilikos và nnc, 2024 [18]



Hình 2: Phân bố độ nhám bề mặt ( $\Gamma$ ) thể hiện trên lưới 0.03 độ

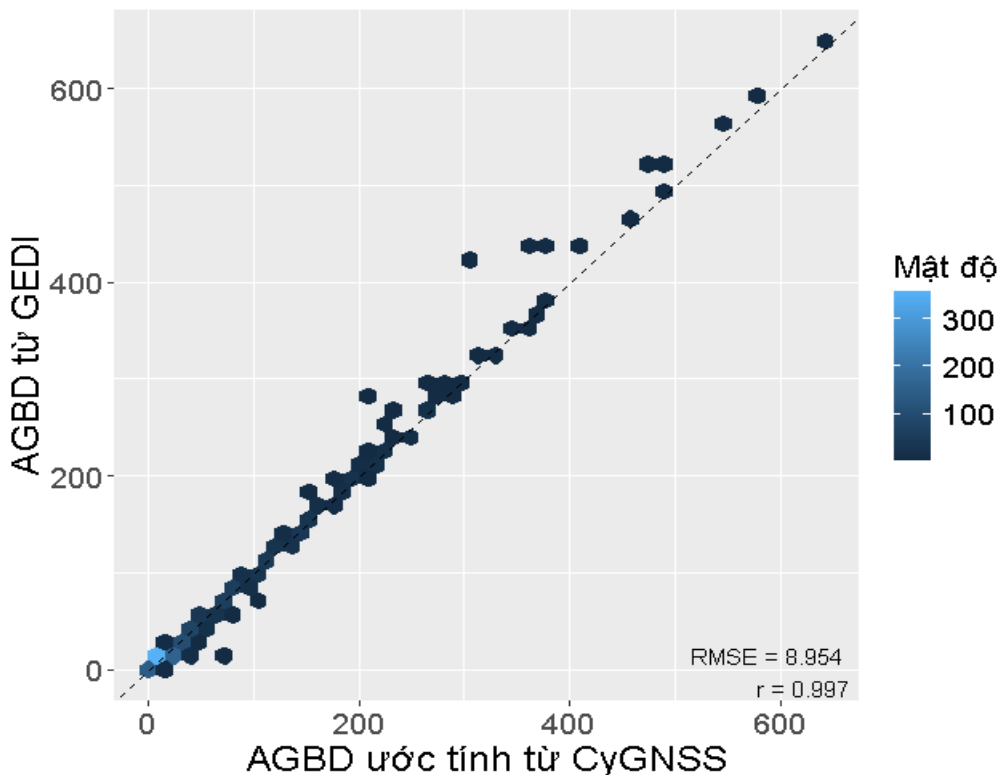
Sự khác biệt rõ rệt về phân bố độ nhám này minh chứng cho vai trò của yếu tố bề mặt trong điều biến cường độ tín hiệu GNSS. Đặc biệt, các giá trị  $\Gamma$  cao tập trung ở khu vực đồng bằng và ven biển gợi ý mức đóng góp lớn của độ nhám bề mặt vào quá trình ước tính sinh khối, phù hợp với kết quả phân tích tầm quan trọng của biến (variable importance) trong mô hình RF (Hình 4). Điều này cho thấy thông tin từ CyGNSS không chỉ phản ánh điều kiện thủy văn mà còn cung cấp tín hiệu gián tiếp để ước tính các đặc tính thảm thực vật và sinh khối trên đất liền.

#### 4.2 So sánh kết quả của mô hình dự báo với dữ liệu GEDI

Tập dữ liệu huấn luyện được xây dựng bằng cách ghép nối không gian giữa các điểm đo AGBD từ GEDI và các quan sát CyGNSS trong phạm vi dung sai không gian  $0.0001^\circ$  ( $\sim 10$  m). Để tránh hiện tượng đa cộng tuyến và các tín hiệu vị trí giả, toàn bộ các biến liên quan đến kinh độ và vĩ độ được loại bỏ khỏi tập biến giải thích. Tập biến đầu vào còn lại gồm 72 biến CyGNSS, bao gồm các chỉ số backscatter, đặc trưng Delay–Doppler, tham số phản xạ bề mặt, thông tin lớp phủ đất MODIS và dữ liệu địa hình SRTM.

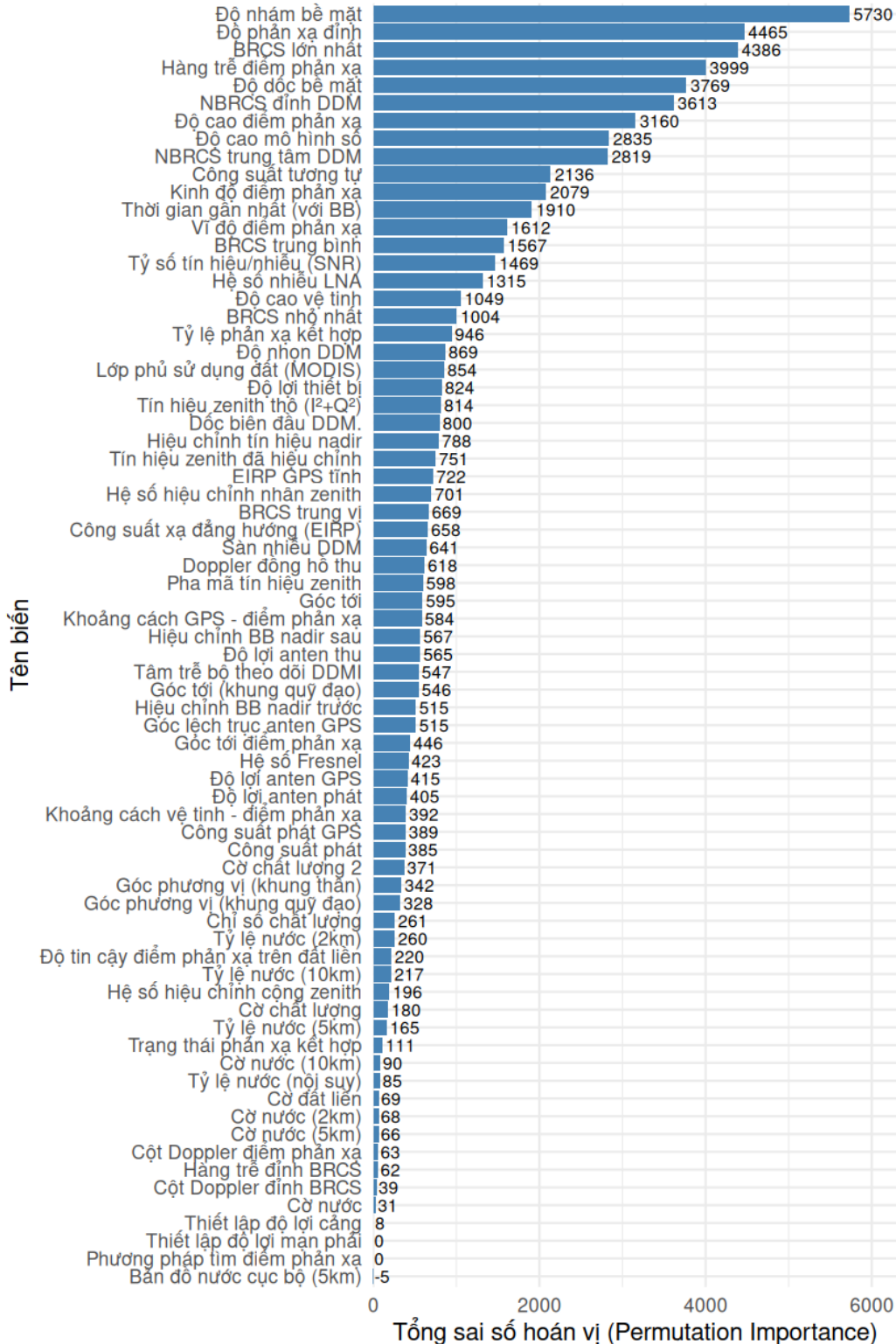
Mô hình được huấn luyện bằng gói ranger [32] trong môi trường caret [33] của R. Tập dữ liệu được chia 70% huấn luyện và 30% kiểm định. Tham số mtry (số biến được xem xét tại mỗi nút) được dò từ 0 đến 22 và min.node.size (kích thước nút tối thiểu) được dò trong tập giá trị {1,5,10,15,20} thông qua 10-fold cross-validation, chọn cấu hình cho RMSE thấp nhất. Số lượng cây được cố định ở  $B=1000$ . Độ quan trọng của biến được đánh giá theo phương pháp hoán vị (permutation importance) của [31]

Kết quả xây dựng mô hình cho thấy các điểm dữ liệu phân bố tập trung sát đường đồng nhất, phản ánh mức độ tương đồng cao giữa AGBD dự báo và GEDI. Giá trị  $RMSE = 8.954$  Mg/ha và hệ số tương quan Pearson  $r=0.997$  chứng minh rằng mô hình Random Forest đã đạt được độ chính xác rất cao, với sai số trung bình thấp và mối quan hệ tuyến tính gần như hoàn hảo.



Hình 3: Tương quan giữa kết quả mô hình dự báo với dữ liệu GEDI

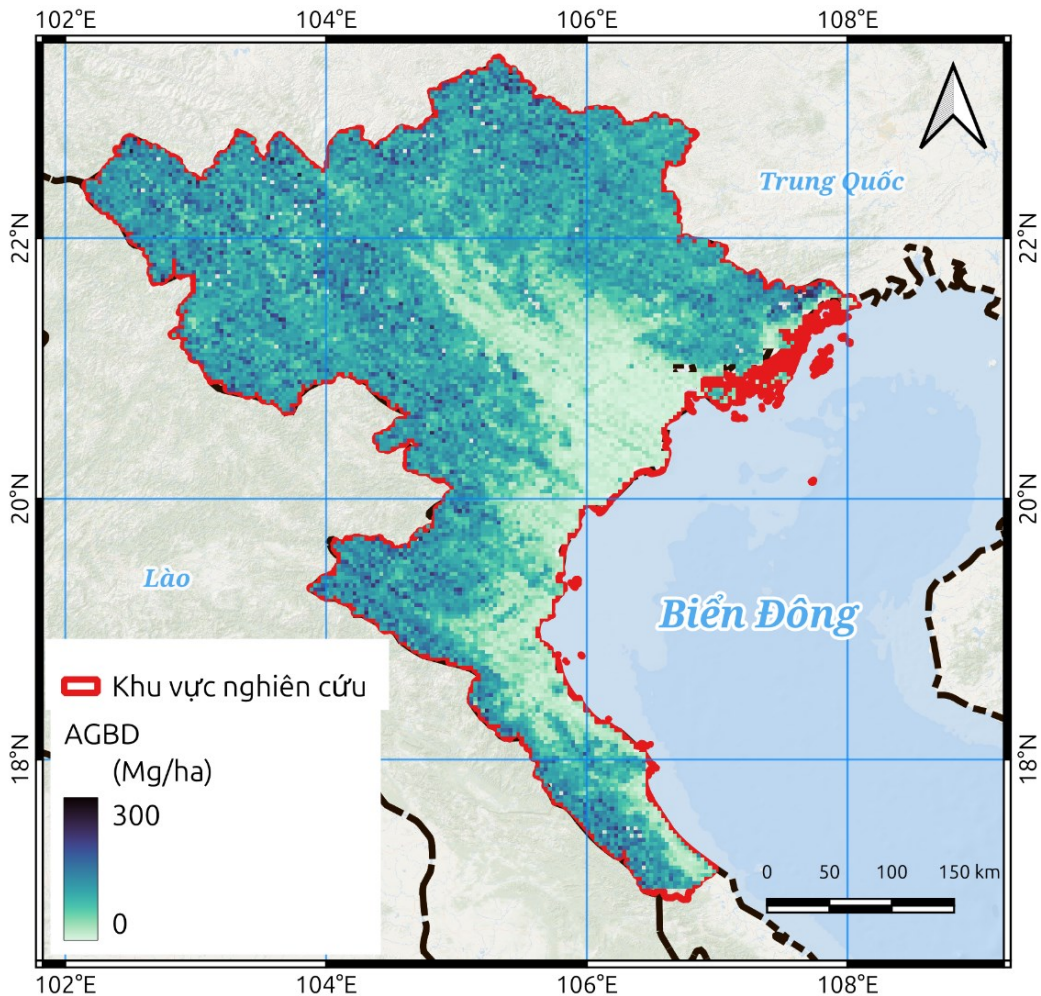
Đáng chú ý, không xuất hiện sai lệch hệ thống đáng kể ở các khoảng giá trị khác nhau: cả vùng sinh khối thấp (<100 Mg/ha) lẫn cao (>400 Mg/ha) đều thể hiện sự tương đồng tốt. Điều này cho thấy mô hình không chỉ dự báo tốt trên các khu vực có mật độ sinh khối trung bình mà còn duy trì độ chính xác ở những khu vực rừng giàu sinh khối – vốn thường khó ước tính do tín hiệu GNSS-R bị suy giảm mạnh.



Hình 4: Tầm quan trọng của các biến trong mô hình

### 4.3. Kết quả ước tính mật độ sinh khối

Hình 5 trình bày phân bố không gian của AGBD (Mg/ha) ước tính từ dữ liệu CyGNSS kết hợp mô hình Random Forest cho khu vực miền Bắc Việt Nam. Kết quả cho thấy sự biến thiên không gian rõ rệt của AGBD. Các vùng có giá trị cao (>250 Mg/ha) tập trung chủ yếu ở các dãy núi và khu vực rừng tự nhiên thuộc vùng Tây Bắc và Đông Bắc, nơi độ che phủ rừng còn tốt và thảm thực vật phát triển dày đặc. Ngược lại, những khu vực có giá trị AGBD thấp (<100 Mg/ha) xuất hiện phổ biến ở đồng bằng ven biển, các vùng đất canh tác và khu dân cư, đặc biệt dọc Đồng bằng sông Hồng và ven biển Bắc Trung Bộ, nơi rừng tự nhiên đã bị thay thế bởi đất nông nghiệp hoặc đô thị.



Hình 5: Kết quả ước tính AGBD từ dữ liệu CyGNSS

Một số khu vực trung du và miền núi có mức AGBD trung bình (100–200 Mg/ha) phản ánh sự chuyển tiếp từ rừng tự nhiên sang rừng trồng hoặc đất nông nghiệp, thể hiện ảnh hưởng của hoạt động quản lý rừng và sử dụng đất. Ngoài ra, những mảng giá trị cao rải rác dọc biên giới Việt–Lào và Việt–Trung cũng phù hợp với các vùng rừng phòng hộ và rừng đặc dụng đã được bảo vệ tốt. Kết quả này không chỉ minh họa khả năng của dữ liệu GNSS-R trong việc mô phỏng các mẫu phân bố sinh khối trên diện rộng, mà còn cung cấp thông tin hữu ích phục vụ quy hoạch quản lý rừng, ước tính trữ lượng carbon và giám sát biến đổi khí hậu ở cấp khu vực.



## 5. Thảo luận

Kết quả nghiên cứu cho thấy dữ liệu CyGNSS có tiềm năng ứng dụng đáng kể trong ước tính các đặc trưng bề mặt đất liền, đặc biệt là AGBD. Việc khai thác tín hiệu tán xạ GNSS băng tần L cho phép ghi nhận những biến đổi vi mô của bề mặt, từ đó cung cấp các biến giải thích mới mẻ và độc lập so với các nguồn dữ liệu viễn thám truyền thống như quang học hay SAR.

Một trong những phát hiện quan trọng của nghiên cứu là độ nhám bề mặt ( $\Gamma$ ) đóng vai trò chi phối trong mô hình RF, thể hiện qua giá trị tầm quan trọng cao nhất trong toàn bộ tập hợp biến đầu vào. Điều này phù hợp với cơ sở vật lý của tín hiệu GNSS: độ nhám bề mặt quyết định trực tiếp đến cơ chế tán xạ khuếch tán, dẫn đến sự suy hao năng lượng phản xạ và thay đổi cường độ tín hiệu được vệ tinh thu nhận [30]. Bản đồ không gian của  $\Gamma$  cũng cho thấy sự phân hóa rõ rệt giữa các vùng đồng bằng ven biển – nơi giá trị  $\Gamma$  cao – và các vùng núi phía tây bắc – nơi  $\Gamma$  thấp. Sự phân hóa này không chỉ phản ánh đặc điểm địa hình và thủy văn mà còn liên quan chặt chẽ tới cấu trúc thảm thực vật và hàm lượng ẩm của đất, hai yếu tố vốn được biết đến là quyết định mật độ sinh khối.

Khi so sánh với các nghiên cứu trước đây sử dụng dữ liệu quang học hoặc radar, cách tiếp cận dựa trên CyGNSS thể hiện một số ưu thế nổi bật. Tín hiệu L-band có khả năng xuyên tán tốt qua thảm thực vật và ít bị bão hòa ở vùng sinh khối cao, giúp giảm sai số ở các khu rừng rậm nhiệt đới. Hơn nữa, CyGNSS có tần suất lặp cao, cho phép theo dõi sự thay đổi của AGBD trong các khoảng thời gian ngắn, đặc biệt hữu ích trong giám sát động thái mùa hoặc sau các sự kiện cực đoan. Tuy nhiên, cũng cần ghi nhận những hạn chế như độ phân giải không gian của CyGNSS ( $\sim 0.03^\circ$ ) còn tương đối thô, khiến việc áp dụng ở quy mô nhỏ gặp khó khăn; ngoài ra, ảnh hưởng của điều kiện khí quyển và sự phức tạp của bề mặt nước ven sông/biển có thể gây nhiễu cho tín hiệu phản xạ.

Kết quả nghiên cứu cho thấy mô hình rừng ngẫu nhiên phù hợp với tập dữ liệu có tính phi tuyến cao. Mặc dù vậy, cơ sở vật lý của các mối quan hệ này vẫn cần được nghiên cứu rõ hơn, đặc biệt khi áp dụng cho các hệ sinh thái hoặc điều kiện khí hậu khác. Ngoài ra, việc kết hợp CyGNSS với các nguồn dữ liệu khác như GEDI, ICESat-2 hoặc dữ liệu khí tượng có thể giúp khắc phục độ phân giải không gian và nâng cao độ chính xác của mô hình. Bên cạnh những kết quả hứa hẹn, vẫn còn một số thách thức, đặc biệt là độ phân giải không gian của dữ liệu. Các nghiên cứu trong tương lai cần tập trung vào việc tích hợp các bộ dữ liệu đa nguồn và phát triển các hệ thống giám sát gần thời gian thực để hỗ trợ quản lý tức thời.

## 6. Kết luận

Nghiên cứu này khẳng định tiềm năng của dữ liệu phản xạ GNSS từ sứ mệnh CyGNSS kết hợp mô hình học máy trong ước tính mật độ sinh khối trên mặt đất (AGBD) tại khu vực miền Bắc Việt Nam. Dữ liệu AGBD từ GEDI được sử dụng làm tham chiếu để huấn luyện và kiểm chứng mô hình, trong đó kết quả hồi quy đạt độ chính xác rất cao ( $R = 0,99$ ). Các bản đồ AGBD được xây dựng phản ánh rõ sự phân hóa không gian của thảm thực vật và thể hiện khả năng nắm bắt hiệu quả mối quan hệ phi tuyến tính giữa tín hiệu phản xạ và mật độ sinh khối. Phân tích tầm quan trọng của biến cho thấy độ nhám bề mặt ( $\Gamma$ ) đóng vai trò chi phối trong dự báo AGBD, qua đó củng cố cơ sở vật lý về sự suy giảm và khuếch tán tín hiệu GNSS-R trên bề mặt phản xạ. Kết quả này không chỉ cung cấp một phương pháp tiếp cận mới trong phép đo sinh khối từ vệ tinh mà còn mang lại ý nghĩa thực tiễn cho quản lý bền vững tài nguyên rừng và xây dựng chiến lược thích ứng với biến đổi khí hậu. Kết quả không chỉ đóng góp thêm bằng chứng cho khả năng ứng dụng đa dạng của CyGNSS mà còn mở ra hướng tiếp cận mới trong giám sát rừng và đánh giá các chỉ số sinh thái phục vụ quản lý tài nguyên và ứng phó biến đổi khí hậu.

## Lời cảm ơn

Nghiên cứu này được tài trợ bởi đề tài mã số TN.24.08 của trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội.

Nhóm tác giả xin cảm ơn Khoa Địa lý, trường Đại học Khoa học Tự nhiên đã cung cấp hệ thống máy trạm và thiết bị đo GNSS GR30 để đo đạc và xử lý dữ liệu cho nghiên cứu này.  
**Cam kết của các tác giả**

Tất cả các tác giả có tên trong bài báo cam kết sự đồng thuận và không có xung đột lợi ích trong công bố khoa học tại bài báo này.

## Tài liệu tham khảo

- [1] Pan Y, RA Birdsey, J Fang, R Houghton, PE Kauppi, WA Kurz, OL Phillips, A Shvidenko, SL Lewis, JG Canadell, P Ciais, RB Jackson, SW Pacala, AD McGuire, S Piao, A Rautiainen, S Sitch, D Hayes, "A Large and Persistent Carbon Sink in the World's Forests", *Science*, vol. 333 no. 6045, pp. 988, 2011.
- [2] Gibbs HK, S Brown, JO Niles, JA Foley, "Monitoring and estimating tropical forest carbon stocks: making REDD a reality", *Environmental Research Letters*, vol. 2 no. 4, pp. 045023, 2007.
- [3] Xu X, J Yang, S Qi, Y Ma, W Liu, L Li, X Lu, Y Liu, "Estimation of Forest Aboveground Biomass Using Sentinel-1/2 Synergized with Extrapolated Parameters from LiDAR Data and Analysis of Its Ecological Driving Factors", *Remote Sensing*, vol. 17 no. 14, 2025.
- [4] Du L, Y Pang, Q Wang, C Huang, Y Bai, D Chen, W Lu, D Kong, "A LiDAR biomass index-based approach for tree- and plot-level biomass mapping over forest farms using 3D point clouds", *Remote Sensing of Environment*, vol. 290, pp. 113543, 2023.
- [5] Pan L, L Liu, AG Condon, GM Estavillo, RA Coe, G Bull, EA Stone, L Petersson, V Rolland, "Biomass Prediction with 3D Point Clouds from LiDAR", In: *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022. pp. 1716.
- [6] Imhoff ML, "Radar backscatter and biomass saturation: ramifications for global biomass inventory", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33 no. 2, pp. 511, 1995.
- [7] Kasischke ES, JM Melack, MC Dobson, "The use of Imaging radars for ecological applications: A review", *Remote Sensing of Environment*, vol. 59, pp. 141, 1997.
- [8] Dubayah R, JB Blair, S Goetz, L Fatoyinbo, M Hansen, S Healey, M Hofton, G Hurtt, J Kellner, S Luthcke, J Armston, H Tang, L Duncanson, S Hancock, P Jantz, S Marselis, PL Patterson, W Qi, C Silva, "The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography", *Science of Remote Sensing*, vol. 1, pp. 100002, 2020.
- [9] Motte E, A Egido, N Roussel, K Boniface, F Frappart, "9 - Applications of GNSS-R in Continental Hydrology", In: (Baghdadi, N, M Zribi ed.), *Land Surface Remote Sensing in Continental Hydrology*, Elsevier, 2016. pp. 281.
- [10] Zuffada C, C Chew, SV Nghiem, "Global Navigation Satellite System Reflectometry (GNSS-R) algorithms for wetland observations", In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017. pp. 1126.
- [11] Jin S, A Camps, Y Jia, F Wang, M Martin-Neira, F Huang, Q Yan, S Zhang, Z Li, K Edokossi, D Yang, Z Xiao, Z Ma, W Bai, "Remote sensing and its applications using GNSS reflected signals: advances and prospects", *Satellite Navigation*, vol. 5 no. 1, pp. 19, , 2024.
- [12] Chew C, E Small, "Description of the UCAR/CU Soil Moisture Product", *Remote Sensing*, vol. 12 no. 10, 2020.
- [13] Hammond ML, G Foti, C Gommenginger, M Srokosz, "Temporal variability of GNSS-Reflectometry ocean wind speed retrieval performance during the UK TechDemoSat-1 mission", *Remote Sensing of Environment*, vol. 242, pp. 111744, 2020.
- [14] Xu L, W Wan, X Chen, S Zhu, B Liu, Y Hong, "Spaceborne GNSS-R Observation of Global Lake Level: First Results from the TechDemoSat-1 Mission", *Remote Sensing*, vol. 11 no. 12, 2019.
- [15] Ruf CS, R Atlas, PS Chang, MP Clarizia, JL Garrison, S Gleason, SJ Katzberg, Z Jelenak, JT Johnson, SJ Majumdar, A O'brien, DJ Posselt, AJ Ridley, RJ Rose, VU Zavorotny, "New Ocean Winds Satellite Mission to Probe Hurricanes and Tropical Convection", *Bulletin of the American Meteorological Society*, vol. 97 no. 3, pp. 385, 2016.
- [16] Ha M-C, J Darrozes, M Llubes, M Grippa, G Ramillien, F Frappart, F Baup, HT Tagesson, E Mougin, I Guiro, L Kergoat, HD Nguyen, L Seoane, G Dufrechou, P-L Vu, "GNSS-R monitoring of soil moisture



- dynamics in areas of severe drought: example of Dahra in the Sahelian climatic zone (Senegal)", *European Journal of Remote Sensing*, vol. 56 no. 1, pp. 2156931, 2023.
- [17] Hoang TP, MC Ha, PL Vu, J Darrozes, PB Nguyen, "Integrating SMAP and CYGNSS data for daily soil moisture and agricultural drought monitoring in Nghe An province, Vietnam", *Remote Sensing Applications: Society and Environment*, vol. 39, pp. 101664, 2025.
- [18] Pilikos G, MP Clarizia, N Floury, "Biomass Estimation with GNSS Reflectometry Using a Deep Learning Retrieval Model", *Remote Sensing*, vol. 16 no. 7, 2024.
- [19] Vu PL, M Cuong, PB Nguyen, HD Nguyen, T Bao, H Dinh, TH Nguyen, G Şerban, M Zelenakova, J Darrozes, "Demonstrating the Potential of Low-Cost GNSS Receiver for tidal monitoring, storms, and flood detecting: example of 2022 Noru Storm in Thua Thien Hue province, Vietnam", *Acta Montanistica Slovaca*, vol. 28 no. 4, pp. 1034, 2023.
- [20] Chen Y, Q Yan, "Unlocking the potential of CYGNSS for pan-tropical inland water mapping through multi-source data and transformer", *International Journal of Applied Earth Observation and Geoinformation*, vol. 133, pp. 104122, 2024.
- [21] Ma W, L Huang, X Wu, S Jin, W Bai, X Li, "Evaluation of CYGNSS Observations for Snow Properties, a Case Study in Tibetan Plateau, China", *Remote Sensing*, vol. 14 no. 15, 2022.
- [22] Mayers D, C Ruf, "Measuring Ice Thickness with Cygnss Altimetry", In: *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018., pp. 8535.
- [23] Yang W, F Guo, X Zhang, Y Zhu, Z Li, Z Zhang, "First quasi-global soil moisture retrieval using Fengyun-3 GNSS-R constellation observations", *Remote Sensing of Environment*, vol. 321, pp. 114653, 2025.
- [24] Jing C, W Li, W Wan, F Lu, X Niu, X Chen, A Rius, E Cardellach, S Ribó, B Liu, Z Guo, Y Nan, "A review of the BuFeng-1 GNSS-R mission: calibration and validation results of sea surface and land surface", *Geo-spatial Information Science*, vol. 27 no. 3, pp. 638, 2024.
- [25] Potapov P, X Li, A Hernandez-Serna, A Tyukavina, MC Hansen, A Kommareddy, A Pickens, S Turubanova, H Tang, CE Silva, J Armston, R Dubayah, JB Blair, M Hofton, "Mapping global forest canopy height through integration of GEDI and Landsat data", *Remote Sensing of Environment*, vol. 253, pp. 112165, 2021.
- [26] Liu Y, I Collett, YJ Morton, "Application of Neural Network to GNSS-R Wind Speed Retrieval", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57 no. 12, pp. 9756, 2019.
- [27] McNoldy B, B Annane, S Majumdar, J Delgado, L Bucci, R Atlas, "Impact of Assimilating CYGNSS Data on Tropical Cyclone Analyses and Forecasts in a Regional OSSE Framework", *Marine Technology Society Journal*, vol. 51 no. 1, pp. 7, 2017.
- [28] Li H, X Li, T Kato, S Inukai, T Hiroshima, "National-scale calibrated GEDI AGBD models for effective assessment of growth conditions across forest strata", *Forest Ecology and Management*, vol. 585, pp. 122657, 2025.
- [29] Nazir A, NP Hanan, Q Yu, H Gilani, "Enhancing GEDI above ground biomass density estimates in contrasting forests of Pakistan", *Forest Ecology and Management*, vol. 587, pp. 122747, 2025.
- [30] Ruf C, D McKague, D Posselt, S Gleason, MP Clarizia, V Zavorotny, T Butler, J Redfern, W Wells, M Morris, J Crespo, C Chew, E Small, D Pascual, T Wang, A Warnock, D Mayer, M Al-Khaldi, A O'Brien, s Munchak, *CYGNSS Handbook (2nd-ed.)*, 2022.
- [31] Breiman L, "Random Forests", *Machine Learning*, vol. 45 no. 1, pp. 5, 2001.
- [32] Wright MN, A Ziegler, "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R", *Journal of Statistical Software*, vol. 77 no. 1, pp. 1, 2017.
- [33] Kuhn M, "Building Predictive Models in R Using the caret Package", *Journal of Statistical Software*, vol. 28 no. 5, pp. 1, 2008.

