Remote Sensing Applications: Society and Environment 42 (2026) 102075

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Deep Learning–eXtreme Gradient Boosting ensemble approach for reliable additive mapping of above- and belowground carbon pools in tropical dry dipterocarp forests from Sentinel-2 imagery

Bao Huy [a,b,*], Krishna P. Poudel [c], Hailemariam Temesgen [b], Andreas Christian Braun [d], Nguyen Quy Truong [a], Pham Cong Tri [a], Nguyen The Hien [a], Nguyen Quy Khiem [a]

[a] Forest Resources and Environment Management Consultancy (FREM), 06 Nguyen Hong, Buon Ma Thuot, Dak Lak, Viet Nam
[b] Department of Forest Engineering, Resources and Management, Oregon State University (OSU), Corvallis, OR, 97333, USA
[c] Department of Forestry, Mississippi State University, P.O. Box 9681, Mississippi State, MS, 39762, USA
[d] Kassel Institute for Sustainability, University of Kassel, Kassel, 34109, Germany

ARTICLE INFO

Keywords:
Additive carbon pools mapping
Deep Learning-XGBoost ensemble
MARS
Sentinel-2
Spatial cross-validation
SUR

ABSTRACT

Accurate prediction of forest carbon pools is essential for assessing carbon sequestration, yet reliable additive mapping of above- and belowground carbon pools and their total (AGCP, BGCP, and TGCP, respectively) in tropical dry dipterocarp forests (DDF) remains limited. This study evaluated a novel Deep Learning–eXtreme Gradient Boosting (DLXG) ensemble against two conventional approaches - Multivariate Adaptive Regression Splines (MARS) and Seemingly Unrelated Regression (SUR) - for additive mapping of AGCP, BGCP, and TGCP. The models were trained and spatially cross-validated using carbon data from 142 ground plots (500 m^2) in the largest preserved DDF in Vietnam, combined with Sentinel-2 imagery. Principal component analysis was applied to identify the most informative predictors from 65 Sentinel-2-derived spectral, vegetation, and texture covariates. The DLXG framework was optimized by jointly tuning the deep neural network architecture and the hyperparameters of the eXtreme Gradient Boosting (XGBoost) meta-learner, and by implementing a residual-stacking strategy in which out-of-fold deep learning predictions captured complex nonlinear relationships and served as inputs to XGBoost to reduce prediction residuals. Spatial cross-validation across the heterogeneous DDF landscape demonstrated that both SUR and MARS produced less reliable estimates than the DLXG approach. The best DLXG model achieved fit indices of 0.92, 0.89, and 0.92 and RMSE values of 21.0, 2.9, and 23.8 Mg ha^{-1} for the additive mapping of AGCP, BGCP, and TGCP, respectively. To support the application, a user-friendly executable package was developed to operationalize the DLXG, enabling reliable additive mapping of DDF carbon pools across diverse spatial and temporal contexts.

---

* Corresponding author. Forest Resources and Environment Management Consultancy (FREM), 06 Nguyen Hong, Buon Ma Thuot, Dak Lak, Viet Nam.

E-mail addresses: baohuy.frem@gmail.com (B. Huy), Krishna.Poudel@msstate.edu (K.P. Poudel), Temesgen.Hailemariam@oregonstate.edu (H. Temesgen), Andreas.Braun@uni-kassel.de (A.C. Braun), QuyTruong.Ng@gmail.com (N.Q. Truong), pcotri@ymail.com (P.C. Tri), nguyenthe_hien@live.com (N.T. Hien), QuyKhiem.frem@gmail.com (N.Q. Khiem).

https://doi.org/10.1016/j.rsase.2026.102075
Received 7 December 2025; Received in revised form 25 April 2026; Accepted 19 May 2026

1. Introduction

Measuring forest carbon pools is key to understanding sequestration rates. Accurate mapping informs tropical forest management and the carbon cycle (Ayushi et al., 2024), supports climate policies (Zhang et al., 2012; Li et al., 2024), and influences carbon credit issuance, where minimizing uncertainty is crucial (Cuesta et al., 2023). Of the five forest carbon pools (IPCC et al., 2006), the aboveground and belowground carbon pools (AGCP and BGCP) are most critical for storage and mitigation. Thus, evaluating the mapping methods is essential for simultaneously and optimally predicting AGCP, BGCP, and their total (TGCP = AGCP + BGCP) in tropical forests.

The tropical dry dipterocarp forests (DDFs) are widespread in Southeast Asia (Huy et al., 2019) and thrive under drought, fire, waterlogging, and low precipitation, often on poor, shallow soils (Huy et al., 2018, 2019). Managing DDFs under such conditions is vital for ecological protection, biodiversity, ecosystem services, and carbon storage (Huy et al., 2019). Existing biomass-carbon models focus on tree-level aboveground biomass (AGB) (Chave et al., 2005; Basuki et al., 2009; Huy et al., 2016, 2019) and on the joint prediction of AGB and tree-level belowground biomass (BGB) (Kralicek et al., 2017). Despite their significant carbon stocks, remote sensing-based biomass estimation in DDF is scarce (Villegas et al., 2023), with few studies using remote sensing to predict AGB, and even fewer targeting BGB or simultaneous AGB-BGB prediction.

Remote sensing plays a key role in forest AGB estimation across scales (Sainuddin et al., 2023; Tian et al., 2023; Anees et al., 2024), with methods evolving under IPCC guidelines (IPCC, 2019) and supported by sensors such as Sentinel-2, microwave, and LiDAR across empirical to hybrid modeling approaches (Sainuddin et al., 2023; Tian et al., 2023). This shift advances forest ecology (Jucker et al., 2017; Feilhauer et al., 2021) and enables the mapping of carbon pools for credit programs. Biomass estimation benefits from free, high-resolution imagery and multi-source data (Nasset et al., 2020; Cuesta et al., 2023; Li et al., 2024). With no consensus on best practices, optimal model choice may outweigh more field data (Fassnacht et al., 2024). Two main approaches to model development are parametric regression (Zaki et al., 2018; Tian et al., 2023) and non-parametric methods like machine learning (ML), kNN, random forest (RF), and deep learning (DL) (Zhang et al., 2012, 2024; Naimi et al., 2021; Tang et al., 2022; Zeraatpisheh et al., 2022; Rana et al., 2023; Li et al., 2024; Matso et al., 2024; Nazeri et al., 2024; Yun et al., 2024; Tan et al., 2025).

Parametric regression, often assuming linear or nonlinear forms, is widely used to link forest attributes with remote sensing data (Jucker et al., 2017; Askar et al., 2018; Sun et al., 2021; Tian et al., 2023; Ma et al., 2024), using predictors like vegetation indices and canopy height (Fisher, 2023). Despite Heiskanen's (2006) comparison of linear and nonlinear regression approaches, many studies have traditionally relied on linear models. However, such models often underperform when complex, nonlinear interactions exist between forest biomass and remotely sensed variables (Lu et al., 2016; Tang et al., 2022; Rana et al., 2023). Consequently, nonlinear or non-parametric approaches are increasingly adopted for biomass estimation (Lu et al., 2016). Adding topographic and climatic variables has improved model fit (Sun et al., 2021; Strunk et al., 2022). In contrast, nonparametric methods - especially artificial neural networks (ANNs) - offer improved accuracy and lower bias (Joshi et al., 2006), and are increasingly applied to biomass, canopy cover, and degradation mapping (Navarro et al., 2019; Nasiri et al., 2022; Tang et al., 2022; Dalagnol et al., 2023; Rana et al., 2023; Villegas et al., 2023). Random forest and eXtreme Gradient Boosting (XGBoost) also perform well in AGB prediction due to their ensemble-learning design (Esteban et al., 2019; Li et al., 2020; Su et al., 2020). More recently, deep learning methods have further improved biomass carbon predictions by leveraging their adaptability and rich parameterization (Yun et al., 2024; Tan et al., 2025).

Mapping biomass carbon in tropical forests is challenging due to their structural complexity and diversity (Ayushi et al., 2024). Most efforts focus on AGB, which closely relates to AGCP (Esteban et al., 2019; Ayushi et al., 2024), with studies across diverse tropical regions - e.g., Indonesian private forests (Askar et al., 2018), Tanzanian Miombo woodlands (Nasset et al., 2020), Senegalese mangroves (Navarro et al., 2019), and Colombian dry forests (Villegas et al., 2023). In contrast, BGB - despite its correlation with AGB and contribution to total biomass - remains underexplored (Dlamini et al., 2025). Some studies mapped AGB and BGB separately (Cao et al., 2014) or estimated total biomass by combining both (Gizachew et al., 2016). The limited BGB focus stems from scarce field models (Kralicek et al., 2017; Huy et al., 2024) and a weak understanding of BGB - remote sensing links (Dlamini et al., 2025). Consequently, BGB is often inferred from AGB using conversion factors (IPCC et al., 2006). Recent work has advanced the use of remote sensing and ML for BGB estimation (Zhao et al., 2025) however, accurate BGB assessment is still hindered by the technical constraints of direct subsurface sensing (Dlamini et al., 2025).

Additive approaches for simultaneously mapping forest carbon pools remain limited (Seely et al., 2023). Since TGCP is the sum of AGCP and BGCP, preserving this additivity in predictions is essential. Simultaneous models maintain this consistency while capturing biological relationships between pools, improving both accuracy and interpretability. Developing such additive models using satellite and field data is therefore crucial. Three methods can simultaneously model AGCP, BGCP, and TGCP from satellite indices and ground data: (1) Parametric regression like seemingly unrelated regression (SUR) accounts for interdependent carbon pools (Parresol, 2001; Huy et al., 2023); (2) Non-parametric regression such as Multivariate Adaptive Regression Splines (MARS) captures nonlinear relationships without predefined forms (Friedman, 1991; Safari et al., 2017); (3) Deep learning simultaneous models (DLSMs) extend neural networks to predict multiple outputs while preserving additivity, showing strong results in biomass estimation (Aggarwal, 2023; Perbet et al., 2024; Yun et al., 2024; Huy et al., 2024, 2025). Huy et al. (2025) developed a multi-input, multi-output DLSM for tropical stand-level AGB and BGB. Building on this, DLSMs can be adapted for additive mapping of AGCP, BGCP, and TGCP. While SUR and MARS have been applied to biomass modeling (Filippi et al., 2014; Kralicek et al., 2017; Safari et al., 2017) and DLSMs to component-level prediction (Huy et al., 2024, 2025), no study has yet combined these methods to comprehensively model and map all three pools additively from satellite data.

Ensemble learning method (Tajik et al., 2020; Zhao et al., 2025), such as Deep Learning-XGBoost (He et al., 2022), effectively reduces bias, uncertainty, and overfitting while improving model fit and robustness. Such ensembles generalize well to unseen data, making them ideal for complex tropical forest ecosystems. Deep learning captures nonlinear, high-dimensional relationships, while XGBoost corrects systematic errors and adds regularization (Zhang et al., 2022). As a result, these hybrid models offer more accurate and stable predictions and are increasingly used in forest structure and biomass estimation (Zhang et al., 2022).

The objective of this research was to develop, spatially cross-validate, and assess the reliability of a novel Deep Learning–eXtreme

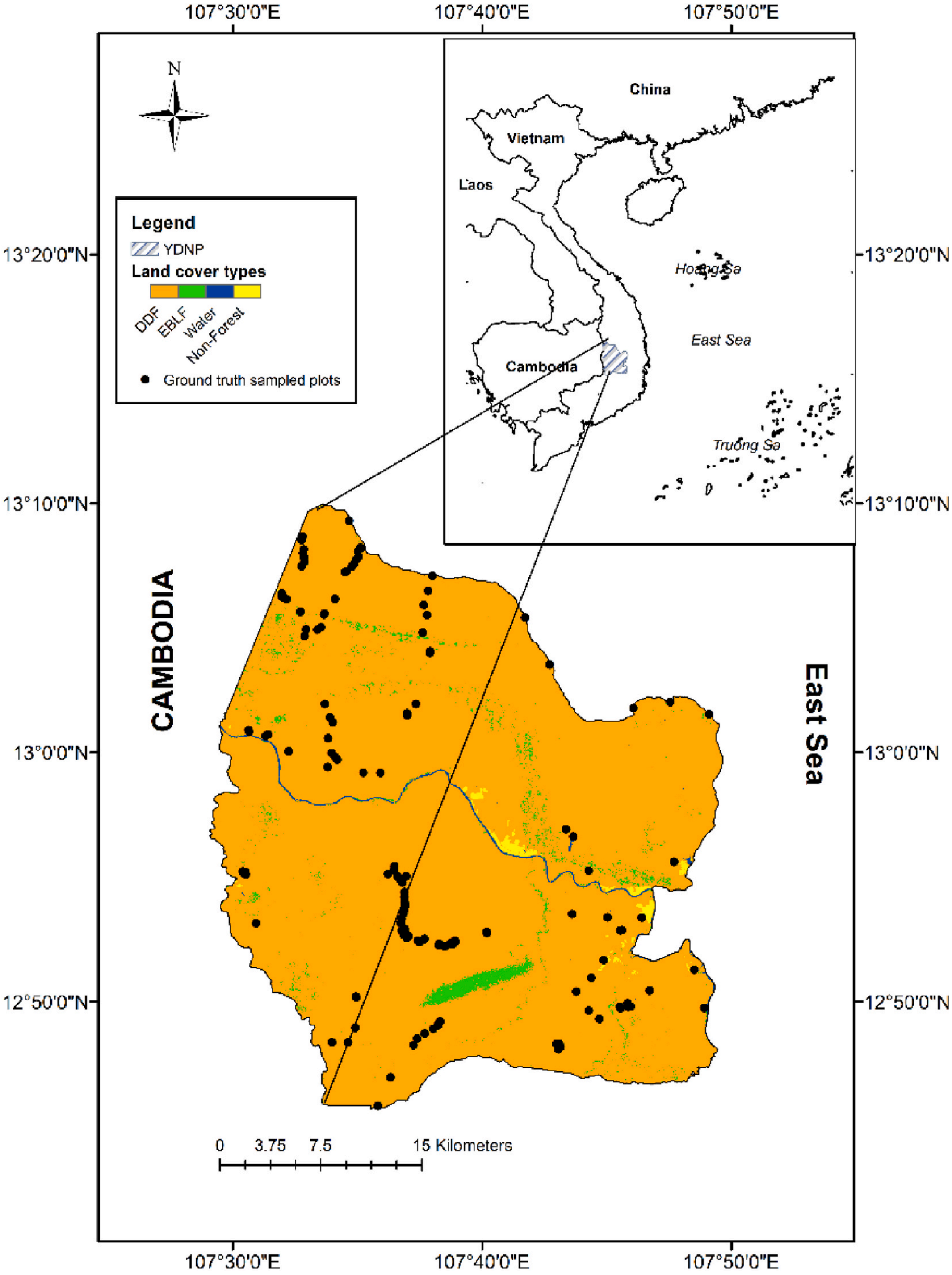


Fig. 1. Location of the tropical Dry Dipterocarp Forest (DDF) study area in Yok Don National Park (YDNP), Vietnam, and the distribution of ground truth sampling plots within the DDF.
Note: EBLF refers to Evergreen Broad-Leaf Forest.

Table 1

Summary statistics of DDF carbon pools and stand, ecological, and environmental factors at the plot level.

ID	Variables	Min.	Mean	Max.	Std.
Forest carbon pool response variables:					
1	AGCP (Mg ha^{-1})	3.9	38.6	121.2	24.8
2	BGCP (Mg ha^{-1})	0.7	7.0	17.6	3.8
3	TGCP (Mg ha^{-1})	4.7	45.6	138.8	28.5
Stand variables:					
4	Dg (Quadratic mean diameter, cm)	6.9	19.1	38.0	6.1
5	Hg (Mean height (m) with the tree of Dg)	5.3	10.0	16.3	2.0
6	G (Stand basal area, m^2 ha^{-1})	0.2	16.8	44.6	9.0
7	V (Stand volume, m^3 ha^{-1})	0.4	98.4	333.2	67.5
8	N (Density with D $\geq$ 5 cm, trees ha^{-1})	40	635	2060	374
Ecological and environmental variables:					
9	P (Mean annual precipitation, mm year^{-1})	1723	1878	1998	74
10	T (Mean annual temperature, ^{0}C year^{-1})	24.9	25.6	26.3	0.3
11	EL (Elevation, m)	144	206	304	29
12	SL (Slope, degree)	0	1	12	2
13	SG (Soil group, categorical variable)	FA, OA			

Note: Statistical metrics were calculated based on 142 ground-truth sampled plots. DDF: Dry dipterocarp forest, AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total ground carbon pools = AGCP + BGCP, D: Diameter at breast height. SG: FA: Ferric Acrisols, OA: Orthic Acrisols. For SG: 83 plots for FA, 59 plots for OA.

Gradient Boosting (DLXG) ensemble, compare it with the conventional MARS and SUR approaches using Sentinel-2 predictors, and identify the most effective method for the additive mapping of AGCP, BGCP, and TGCP in DDF. The sparse, heterogeneous canopy structure of DDF results in high variability in Sentinel-2 spectral reflectance due to exposed ground surfaces, complicating the use of plot-derived spectral data for carbon pool modeling. Accordingly, we hypothesize that conventional regression models may be less reliable for simultaneous prediction tasks, thereby warranting the exploration of a novel approach such as the DLXG ensemble. For example, the deep learning additive modeling system reduced mean absolute percentage error by 2.9%, 14.0%, and 2.4% compared with weighted nonlinear SUR, and by 4.3%, 11.6%, and 4.1% compared with MARS for the simultaneous prediction of AGB, BGB, and total using forest stand, ecological, and environmental covariates (Huy et al., 2025).

This study contributes to the development and documentation of a deep neural network architecture within the DLSM framework and its ensemble integration with XGBoost, demonstrating its application for the simultaneous prediction of AGCP, BGCP, and TGCP using remote sensing imagery across extensive DDF ecosystems in Vietnam and neighboring Southeast Asian countries.

2. Materials and methods

2.1. Study area and characteristics of DDF

The study area, covering over 1000 km^2 within Yok Don National Park (YDNP) in Vietnam's Central Highlands, is the largest preserved area of DDF in Vietnam and a representative example for Southeast Asia (Fig. 1). YDNP lies in a relatively arid region with low annual rainfall (<2000 mm year^{-1}) and high temperatures, exceeding 26 °C (Fick and Hijmans, 2017). The area is primarily composed of Ferric and Orthic Acrisols (FAO-UNESCO, 2005) and has generally flat terrain below 305 m, with scattered low mountain ranges (Table 1). The DDF is characterized (Table 1) by a sparse tree density (N), averaging 600 trees per hectare (with diameter at breast height, D $\geq$ 5 cm). The average stand volume (V) is 100 m^3 ha^{-1}, though some stands exceed 300 m^3 ha^{-1}. Dominated by Dipterocarpaceae species, which account for nearly 70% of forest density, the forest is fully deciduous during the dry season (January - May). The most abundant species include *Dipterocarpus tuberculatus* Roxb. (35% N), *Shorea obtusa* Wall. ex Blume (18% N), and *Pentacme siamensis* (Miq.) Kurz (13% N), while *Dipterocarpus obtusifolius* Teijsm. ex Miq. has the lowest proportion (3% N). Additionally, *Terminalia chebula* Retz. (Combretaceae family) constitutes 5% of the density. Other valuable timber species, such as *Dalbergia oliveri* Gamble ex Prain and *Pterocarpus macrocarpus* Kurz, further contribute to the forest's species diversity.

2.2. Field measurements

We measured 142 purposively selected sample plots, each with a radius of 12.62 m and covering an area of 500 m^2 (Fig. 1). The sample plot data were collected in December 2024 and January 2025, closely aligning with the acquisition of Sentinel-2 imagery in January 2025 for this study. The sample plots were selected to represent a range of forest conditions, including variations in density, maturity, disturbance, and recovery levels (Table 1). The distributions of sample plots based on stand factors generally follow a normal, bell-shaped curve (Fig. 2). However, the non-random sampling strategy may introduce bias, as it may not fully represent the range of forest conditions across the study area. Nevertheless, the relatively large sample size distributed across the entire study area and the near-normal distribution of stand attributes help mitigate this potential bias. For each sample plot, we recorded the coordinates using a global navigation satellite system (GNSS). To minimize GNSS spatial errors, which were approximately 3 m per plot, we ensured that the selected plots had consistent forest conditions within a 10 m radius beyond the plot borders (Safari et al., 2017).

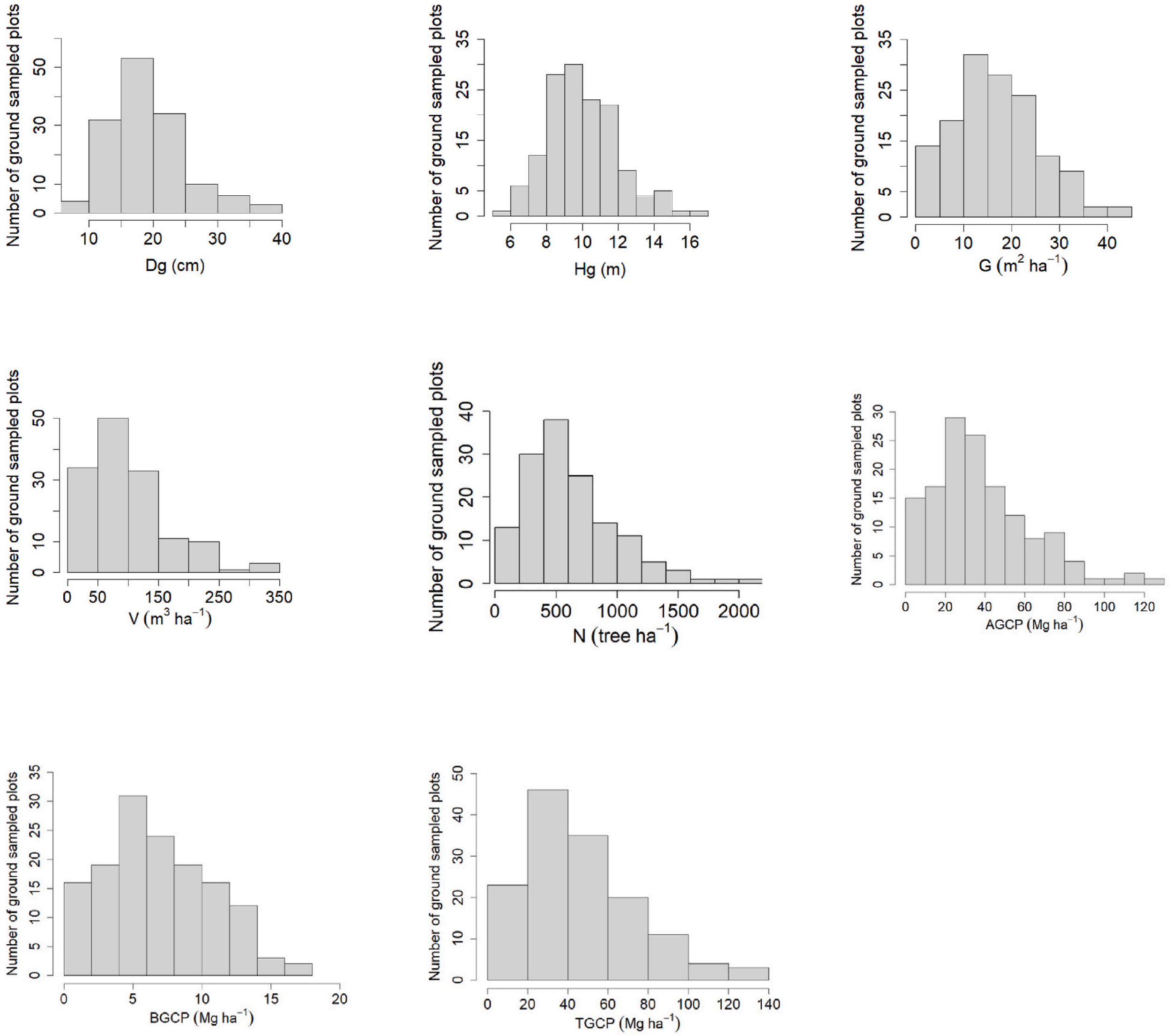


Fig. 2. Distributions of stand variables and forest carbon pools in ground-truth sampled plots
Note: n = 142 ground-truth sampled plots. Dg: Quadratic mean diameter, Hg: Mean height with the tree of Dg, G: Basal area, V: Stand volume, N: Tree density with diameter at breast height (D) ≥ 5 cm, AGCP: Aboveground carbon pool, BGCP: belowground carbon pool, TGCP: Total ground carbon pool = AGCP + BGCP.

The plot-level values for AGCP, BGCP, and TGCP were determined using a collection of allometric models. First, all trees with D ≥ 5 cm were measured, and their species were identified. Then, a tree height-diameter (H/D) database for tropical DDF in the Central Highlands ecoregion of Vietnam (Huy et al. 2016, 2019, 2024) was used to develop an H/D modeling system. A weighted, nonlinear, mixed-effects power model was developed to estimate tree heights (H) for each dominant species across all plots, incorporating random effects for the five dominant species and the remaining mixed species (Cosenza et al., 2024). Additionally, plot-level quadratic mean diameter (Dg) values were calculated, and the corresponding mean tree heights (Hg) were estimated using H/D regression models (Huy et al., 2025). The values of N and stand basal area (G) were derived from field measurements, while stand V was obtained by summing individual tree volumes, which were calculated from D and height (H) using a form factor of 0.45 (Huy et al., 2025). A set of ecological and environmental factors was recorded or extracted at each plot's coordinates (Table 1). Mean annual temperature (T, ^{o}C year^{-1}) and mean annual precipitation (P, mm year^{-1}), averaged over 30 years (1970–2000), were extracted from 30-s (~1 km) resolution climate raster data (Fick and Hijmans, 2017). Soil groups (SG) were derived from the FAO-UNESCO (2005) soil map, while elevation (EL) and slope (SL) were recorded for each plot. Next, the preceding variables were used as inputs to a previously developed deep learning additive model (DLAM, fit indices (FI) = 0.986, 0.989, and 0.987; RMSE = 22.8, 1.6, and 16.2 Mg ha^{-1} for AGB, BGB, and total biomass, respectively) (Huy et al., 2025), which simultaneously predicts these three components at the plot level. Finally,

Table 2
Potential Sentinel-2 imagery predictors, encompassing multispectral bands, vegetation indices, and textural features in this study.

ID	Potential predictors	Bands/Formulas	References
Spectral Bands (SBs):			
1 - 10	BLUE, GREEN, RED, RED.EDGE1, RED.EDGE2, RED.EDGE3, NIR1 (Near Infrared 1), NIR2, SWIR1 (Short Wave InfraRed 1), and SWIR2	B2, B3, B4, B5, B6, B7, B8, B8A, B11 and B12	Navarro et al. (2019); Ahmadi et al. (2020); Sun et al. (2021); Fang et al. (2023); Cosenza et al. (2024); Zhang et al. (2024)
Vegetation indices (VIs):			
11	ARVI (Atmospherically Resistant Vegetation Index)	(B8 - 2×B4 + B2)/(B8 + 2×B4 + B2)	Tang et al. (2022); Zhang et al., (2024)
12	AVI (Advanced Vegetation Index)	$(B8 \times (1 - B4) \times (B8 - B4))^{1/3}$	Banerjee et al. (2014); Geo University (GU) (2025)
13	CIgreen (Green Chlorophyll Index)	B8/B3 - 1	Fang et al. (2023)
14	CIre (Red-edge Chlorophyll Index)	B7/B5 - 1	Fang et al. (2023)
15	DVI (Difference Vegetation Index)	B8 - B4	Heiskanen (2006); Ahmadi et al. (2020); Nasiri et al. (2022); Fang et al. (2023); Ma et al. (2023)
16	EVI (Enhanced Vegetation Index)	2.5 × (B8 - B4)/(B8 + 6×B4 - 7.5×B2 +1)	Gascon and Eva (2014); Sun et al. (2021); Tang et al. (2022); Ma et al. (2023); Zhang et al., (2024)
17	GNDVI (Green Normalized Difference Vegetation Index)	(B8 - B3)/(B8 + B3)	Navarro et al. (2019); Nasiri et al. (2022)
18	MCARI (Modified Chlorophyll Absorption Ratio Index)	((B5 – B4) – 0.2×(B5 – B3))× (B5/B4)	Fang et al. (2023)
19	MTCI (Meris Terrestrial Chlorophyll Index)	(B6 – B5)/(B5 - B4)	Fang et al. (2023)
20	NBRI (Normalized Burn Reflectance Index)	(B8 – B12)/(B8 + B12)	Cosenza et al. (2024)
21	NDI45 (Normalized Difference Index 4 and 5 Red Edge)	(B5 – B4)/(B5 + B4)	Nasiri et al. (2022)
22	NDII (Normalized Difference Infrared Index)	(B8 – B11)/(B8 + B11)	Fang et al. (2023)
23	NDRE (Normalized Difference Red Edge index)	(B8 – B5)/(B8 + B5)	Fang et al. (2023); Nasiri et al. (2022)
24	NDVI (Normalized Difference Vegetation Index)	(B8 - B4)/(B8 + B4)	Heiskanen (2006); Gascon and Eva (2014); Navarro et al. (2019); Ahmadi et al. (2020); Huang et al. (2021); Sun et al. (2021); Nasiri et al. (2022); Tang et al. (2022); Ma et al. (2023); Zhang et al. (2024)
25	NDVI2 (Normalized Difference Vegetation Index 2)	(B8A – B4)/(B8A + B4)	Nasiri et al. (2022)
26	PVI (Perpendicular Vegetation Index)	0.939×B8 - 0.344×B4 + 0.9	Ma et al. (2023)
27	RECI (Red-Edge Chlorophyll Index)	(B6/B5) - 1	Cosenza et al. (2024)
28	RENDI (Red-Edge Normalized Difference Index)	(B7 – B5)/(B7 + B5)	Cosenza et al. (2024)
29	RVI (Ratio Vegetation Index)	B8/B4	Gascon and Eva (2014); Ahmadi et al. (2020); Ma et al. (2023); Cosenza et al. (2024)
30	SAPI (Scatter-Adjusted Pigment Index)	B8A/(B3 × B5)	Cosenza et al. (2024)
31	SAVI (Soil Adjusted Vegetation Index)	((B8 - B4)/(B8 + B4 + 0.428)) × 1.428	Heiskanen (2006); Gascon and Eva (2014); Navarro et al. (2019); Ahmadi et al. (2020); Sun et al. (2021)
32	SIPI (Structure Insensitive Pigment Index)	(B8 – B2)/(B8 + B4)	Tang et al. (2022)
33	SR45 (Simple Ratio 45)	B4/B5	Cosenza et al. (2024)
34	SR63 (Simple Ratio 63)	B6/B3	Cosenza et al. (2024)
35	TVI (Transformed Vegetation Index)	$\sqrt{(NDVI + 0.5)}$	Ma et al. (2023)
Textural features (TFs): Using 5 texture metrics: Mean (M), Variance (V), Entropy (E), Contrast (C), and Homogeneity (H) derived from 3 key-selected spectral bands GREEN, RED.EDGE3, and SWIR2, and 3 key-selected vegetation indices NDVI, SAPI, and SR45			
36 - 41	M (Mean): M_GREEN, M_RED.EGDE3, M_SWIR2, M_NDVI, M_SAPI, M_SR45	$M = \frac{1}{N}\sum_{i=1}^{N} P(i)$	Gascon and Eva (2014); Najafi et al. (2019); Fang et al. (2023); Ayushi et al. (2024); Zhang et al. (2024)
42 - 47	V (Variance): V_GREEN, V_RED.EGDE3, V_SWIR2, V_NDVI, V_SAPI, V_SR45	$V = \sum_{i=1}^{N} (P(i) - M)^2 .P(i)$	Gascon and Eva (2014); Fang et al. (2023); Ayushi et al. (2024); Zhang et al. (2024)
48 - 53	E (Entropy): E_GREEN, E_RED.EGDE3, E_SWIR2, E_NDVI, E_SAPI, E_SR45	$E = -\sum_i \sum_j P(i,j).\log_2 P(i,j)$	Gascon and Eva (2014); Najafi et al. (2019); Ayushi et al. (2024); Zhang et al. (2024)
54 - 59	C (Contrast): C_GREEN, C_RED.EGDE3, C_SWIR2, C_NDVI, C_SAPI, C_SR45	$C = \sum_i \sum_j (i - j)^2 .P(I,j)$	Gascon and Eva (2014); Najafi et al. (2019); Ayushi et al. (2024)
60 - 65	H (Homogeneity): H_GREEN, H_RED.EGDE3, H_SWIR2, H_NDVI, H_SAPI, H_SR45	$H = \sum_i \sum_j \frac{P(I,j)}{1 + \lvert i - j \rvert}$	Gascon and Eva (2014); Najafi et al. (2019); Ayushi et al. (2024)

where P(i): The pixel value at position *i*, N: The total number of pixels in the moving window = 3 × 3 pixels, P(i, j): The normalized probability gray-level co-occurrence matrix (GLCM) value, i: Gray-level of the reference pixel, j: Gray-level of the neighboring pixel, $\log_2$: Logarithm base 2, P(i, j) = Number of occurrences of (i, j)/Total number of pixel pairs in the kernel.

AGCP, BGCP, and TGCP for each plot were obtained by multiplying the predicted plot-level AGB, BGB, and their total by a carbon fraction of 0.47 (IPCC et al., 2006). This value is appropriate for the study area, where reported carbon fractions range from 0.46 to 0.48 (Huy et al., 2023). Statistical summaries of these response variables are presented in Table 1.

2.3. Remote sensing data

In this study, the Sentinel-2 imagery acquired on January 21, 2025, was used, as it was completely cloud-free and effectively captured the canopy condition of the DDF. To generate a mask for the DDF area under investigation, supervised classification using the Maximum Likelihood method was applied to Sentinel-2 bands, with regions of interest (ROIs) defined based on 142 ground-sampled plots.

Three categories of potential Sentinel-2 imagery predictors were used: spectral bands (SBs), vegetation indices (VIs), and texture features (TFs), totaling 65 potential predictors (Table 2). The 10 potential SBs were used (Navarro et al., 2019; Cosenza et al., 2024) (Table 2). We derived 25 potential VIs from the spectral bands, as detailed in Table 2 (Huang et al., 2021; Sun et al., 2021; Villegas et al., 2023; Ayushi et al., 2024). Remote sensing-derived texture features have been increasingly used for forest biomass modeling (Ahmad et al., 2021), which we use here as an additional source of covariates. Texture features can enhance the characterization of vegetation structure and are crucial for improving the accuracy of forest biomass estimation in large, heterogeneous areas such as the studied DDF, overcoming the limitations of relying solely on SBs and VIs (Fang et al., 2023; Zhang et al., 2024).

Texture features can be generated from SBs and VIs using gray-level co-occurrence matrices (GLCMs) (Ayushi et al., 2024), which facilitate the computation of texture features. GLCMs require a predefined window size. We evaluated three options - 3 × 3, 5 × 5, and 7 × 7 pixels (Zhang et al., 2024) - and selected the 3 × 3 window. This choice was based on its superior sensitivity in capturing texture metrics strongly associated with forest carbon pools in DDF areas compared with larger window sizes. Our selection is consistent with previous findings reported by Tian et al. (2023) and Zhang et al. (2024). This study utilized five key texture functions: mean (M), variance (V), entropy (E), contrast (C), and homogeneity (H) (Gascon and Eva, 2014; Fang et al., 2023; Ayushi et al., 2024) applied to key spectral bands and key vegetation indices, as summarized in Table 2. This resulted in 30 texture features (5 texture functions x 6 spectral bands and vegetation indices). TFs were generated using GLCMs (Haralick, 1979; Najafi et al., 2019) with the 'glcm' package (Zvoleff, 2020; Ayushi et al., 2024) in R (R Core Team, 2023).

The modeling dataset, consisting of three response variables – ground truth AGCP, BGCP, and TGCP - and 65 potential Sentinel-2 predictors, was generated by overlaying a vector layer containing the coordinates of 142 ground-sample plots with carbon pool data in conjunction with 65 potential Sentinel-2 raster predictors. The plots were buffered to a 500 m^2 circular area with a radius of 12.62 m and overlaid on raster predictor layers. Using the 'Zonal Statistics' tool in QGIS (2023), raster values were extracted for each plot buffer as the weighted mean of all intersecting pixels (Cosenza et al., 2024). To further minimize the influence of GNSS positional error (∼3 m), each 12.62 m radius plot was established within a homogeneous forest condition extending at least 10 m beyond the plot boundary (i.e., a minimum homogeneous radius of 22.62 m).

To explore the multivariate structure of the dataset and support the selection of Sentinel-2 predictors for modeling AGCP, BGCP, and TGCP, Principal Component Analysis (PCA) was performed. The analysis included 68 continuous variables (three carbon pool variables and 65 Sentinel-2 covariates) and generated linear combinations that explained most of the variance in the dataset. PCA was implemented using the R package 'prcomp' (R Core Team, 2023). PCA was used solely to assess inter-variable relationships and guide predictor selection. It was conducted prior to model development and was not embedded within the spatial cross-validation workflow. Because PCA was not used for feature transformation during model training, it did not affect the training–validation separation and therefore did not introduce data leakage. Although PCA-based dimensionality reduction may reduce direct interpretability, predictor selection was guided by both statistical structure and ecological relevance to the carbon pools. Variables were retained based on their principal component loadings and geometric relationships in the PCA biplot.

2.4. Regression models

We implemented four SUR models (Parresol, 2001; Kralicek et al., 2017; Huy et al., 2023) to simultaneously model the response variables AGCP, BGCP, and TGCP using selected Sentinel-2-derived metrics. The first pair of systems (SUR 1 and SUR 2) uses linear trend functions for each carbon pool variable: SUR 1 does not apply any transformations to the included predictors in the system, while SUR 2 applies log transformations to each predictor. The second pair of systems (SUR 3 and SUR 4) utilizes nonlinear trend functions: SUR 3 employs a power function, while SUR 4 uses a third-order polynomial. Heteroscedasticity in the errors was accounted for by fitting these models with optimal weights (Huy et al., 2016). The SUR modeling systems were fit using the SAS 'Proc Model' procedure (SAS Institute Inc, 2014).

In addition to the parametric SUR models, we developed a nonparametric MARS-based modeling system to simultaneously predict AGCP, BGCP, and TGCP. The MARS modeling system is crafted as an extension employing tensor product splines. It systematically assesses the number of hinge functions, their associated coefficients, optimal predictive covariates, and knot placement for each spline degree based on the given data. This approach is consistent with the methodology outlined by Friedman and Silverman (1989) and Friedman (1991). The MARS modeling system for AGCP, BGCP, and TGCP is structured as follows:

$$\mathrm{AGCP} = a_1 + \sum_{f=1}^{m} b_{1f}.\max\left(0, X_{jf} - c_{jf};\ 0, c_{jf} - X_{jf}\right) + \varepsilon_1 \tag{1}$$

$$\mathrm{BGCP} = a_2 + \sum_{f=1}^{m} b_{2f}.\max\left(0, X_{jf} - c_{jf};\ 0, c_{jf} - X_{jf}\right) + \varepsilon_2 \tag{2}$$

$$\mathrm{TGCP} = \mathrm{a}_3 + \sum_{\mathrm{f}=1}^{\mathrm{m}} \mathrm{b}_{3\mathrm{f}}.\max\left(0, \mathrm{X}_{\mathrm{jf}} - \mathrm{c}_{\mathrm{jf}};\ 0, \mathrm{c}_{\mathrm{jf}} - \mathrm{X}_{\mathrm{jf}}\right) + \varepsilon_3 \tag{3}$$

where AGCP, BGCP, and TGCP represent the plot-level above- and belowground carbon pools and their total, respectively; a_1, a_2, and a_3 are the intercepts; b_{1f}, b_{2f}, and b_{3f} are coefficients for the f^{th} hinge function; X_{jf} represents the j^{th} predictor from Sentinel-2 imagery selected for the f^{th} hinge function; the constant c_{jf} is a breakpoint value linked to the j^{th} predictor in sequential order for the f^{th} hinge function; and ε_1, ε_2, and ε_3 account for the errors of the component models.

The hinge function is a fundamental component of the MARS modeling system and is what allows MARS to fit nonlinear patterns in the data. The depth of this exploration (e.g., second- third- or greater interactive effects) is determined by the 'degree' hyperparameter in the R implementation of MARS in the package 'earth' (R Core Team, 2023; Milborrow et al., 2024). To determine the most appropriate 'degree' value, we employed spatial cross-validation across values ranging from 1 to 6. When 'degree' is specified, the 'earth' function automatically manages the selection of optimal predictors, choosing the ideal number of terms, and estimating the values of the intercept, coefficients, and knots for different hinge functions (Friedman, 1993; R Core Team, 2023). Additionally, to address the heteroscedasticity of errors, MARS models incorporated a weight variable, specifically $1/W^n$, with n ranging between −5 and +5, where W is the most influential Sentinel-2 predictor identified via PCA.

2.5. Ensemble Deep Learning-XGBoost (DLXG)

This study develops a novel DLSM-XGBoost ensemble (DLXG) that: (1) employs a DLSM to capture complex nonlinear relationships between carbon pools and PCA-selected Sentinel-2 predictors in DDFs; and (2) integrates XGBoost to reduce bias, enhance robustness, and improve the reliability of additive mapping for AGCP, BGCP, and TGCP. Its regularization capabilities help reduce overfitting, making it effective for biomass mapping (Zhang et al., 2022).

DLSM is a deep neural network–based modeling system introduced by Huy et al. (2022, 2024, 2025). Unlike conventional univariate approaches, DLSM simultaneously predicts multiple interrelated outputs from multiple inputs while preserving additivity. The network transforms inputs through hidden layers and iteratively updates weights to minimize prediction error, thereby capturing interdependencies among variables. Model performance depends on architecture, hyperparameters, algorithms, and data quality (LeCun et al., 2015; Huy et al., 2022, 2024, 2025).

The DLSM training process is described as follows. Let the training dataset consist of multi-input features $X \in \mathbb{R}^{j.p}$, where j is the number of ground-sampled plots and p is the number of input selected optimal Sentinel-2 predictors (features); the corresponding multi-output targets are denoted as $Y \in \mathbb{R}^{j.k}$, where k is the number of forest carbon pool components - AGCP, BGCP and TGCP - that satisfy an additivity constraint (i.e., TGCP = AGCP + BGCP).

The DLSM learns a mapping:

$$f : \mathbb{R}^{j.p} \rightarrow \mathbb{R}^{j.k} \tag{4}$$

parameterized by trainable network weights and biases. These parameters are estimated by minimizing a multivariate loss function:

$$L = L_{\mathrm{pred}}(\mathrm{Y}, \widehat{\mathrm{Y}}) + \lambda L_{add} \tag{5}$$

Table 3
Optimal architectures, hyperparameters, and functions for the deep neural network and XGBoost model were determined through spatial cross-validation.

ID	Items	Min.	Max.	Selected optimal
For the deep neural network:				
1	Optimization algorithm	Adam, RMSprop, Stochastic Gradient Descent (SGD)		Adam
2	Learning rate	0.001	0.01	0.003
3	Hidden layers	2	5	2
4	Neurons per hidden layer	32	256	64, 128 for 2 hidden layers
5	Activation function	Rectified Linear Unit (ReLU), Sigmoid, and Tanh		ReLU
6	Dropout	0.1	0.5	0.1
7	L2_Lambda	0.00001	0.01	0.00001
8	Epochs	500	1000	500
9	Batch (sample size)	16	64	32
10	Patience	50	150	100
For XGBoost model:				
1	n estimators	100	1000	300
2	Max depth	1	5	3
3	Learning rate	0.01	0.3	0.05
4	Subsample	0.5	1.0	0.8
5	Colsample by tree	0.5	1.0	0.7
6	Reg alpha	0	3	2
7	Reg lambda	0	7	5
8	Objective			reg:squarederror

where $L_{\text{pred}}\ (\text{Y}, \widehat{\text{Y}})$ represents the prediction error across all output variables, L_{add} enforces additive consistency among the predicted carbon pools, and λ is a regularization coefficient that controls the strength of the additivity constraint.

DLSM requires specifying a loss function for forward and backward propagation to optimize the network (Chollet, 2018; Huy et al., 2022) and to guide the search for optimal hyperparameters. To ensure that the developed modeling system supports additive prediction of AGCP, BGCP, and TGCP, we adopted the additive symmetric mean absolute percentage error (Tofallis, 2014) as the loss function ($\text{SMAPE}_{\text{loss}}$).

$$\text{SMAPE}_{\text{loss (\%)}} = \frac{100}{3\text{V}} \sum_{\text{j}=1}^{\text{V}} \left\{ \frac{\left|\text{Y}_{\text{Aj}} - \widehat{\text{Y}}_{\text{Aj}}\right|}{\frac{\left|\text{Y}_{\text{Aj}}\right| + \left|\widehat{\text{Y}}_{\text{Aj}}\right|}{2}} + \frac{\left|\text{Y}_{\text{Bj}} - \widehat{\text{Y}}_{\text{Bj}}\right|}{\frac{\left|\text{Y}_{\text{Bj}}\right| + \left|\widehat{\text{Y}}_{\text{Bj}}\right|}{2}} + \frac{\left|\text{Y}_{\text{Tj}} - \widehat{\text{Y}}_{\text{Tj}}\right|}{\frac{\left|\text{Y}_{\text{Tj}}\right| + \left|\widehat{\text{Y}}_{\text{Tj}}\right|}{2}} \right\} \tag{6}$$

where Y_{Aj}, $\text{Y}_{\text{Bj},}$ Y_{Tj} and $\widehat{\text{Y}}_{\text{Aj}}$, $\widehat{\text{Y}}_{\text{Bji}}$, $\widehat{\text{Y}}_{\text{Tj}}$ are observed, and simultaneously predicted AGCP, BGCP, and TGCP for the j^{th} sample plot, respectively. V is the number of sample plots in the block(s) used for validation or training.

XGBoost (Chen and Guestrin, 2016; Diamantopoulou et al., 2025) is an advanced implementation of gradient boosted decision trees. It improves upon traditional gradient boosting by incorporating several innovations, including regularization to reduce overfitting, parallelized tree construction, handling of missing values, and optimized resource use. Because of these features, XGBoost has become one of the most widely used machine learning algorithms for both regression and classification tasks, especially for structured/tabular data.

XGBoost is summarized as follows. The model is constructed as an additive ensemble of regression trees:

$$\widehat{\text{Y}}_{\text{j,k}} = \sum_{\text{t}=1}^{\text{T}} \text{f}_{\text{t}}\left(\text{X}_{\text{j,p}}\right) \tag{7}$$

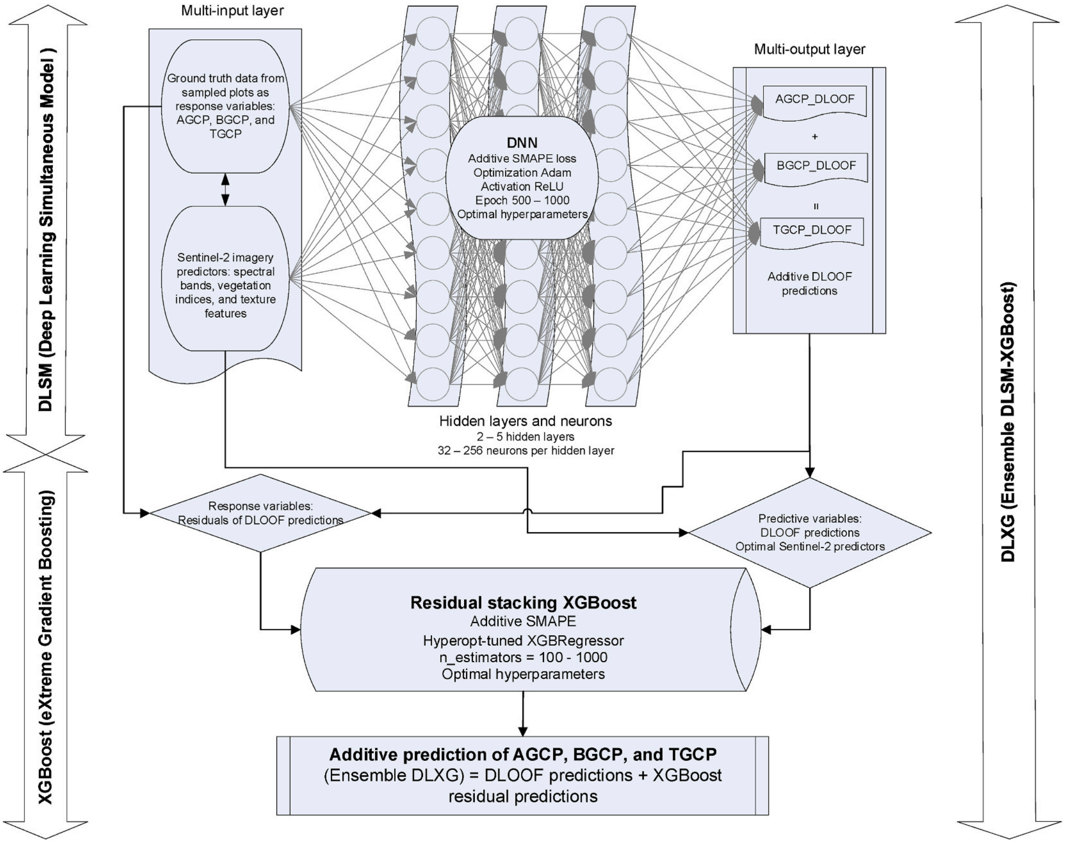


Fig. 3. Ensemble Deep Learning-XGBoost (DLXG) simulation. The main architecture and hyperparameters of the Deep Neural Network (DNN) used in the multi-input–multi-output training of the Deep Learning Simultaneous Model (DLSM), combined with a residual stacking XGBoost model for additive prediction of aboveground, belowground, and total carbon pools (AGCP, BGCP, and TGCP, respectively). Sentinel-2 imagery predictors served as the foundation for simultaneous carbon mapping.
Note: DLOOF predictions: Deep Learning out-of-fold predictions (i.e., predictions on validation data during spatial cross-validation); ReLU: Rectified Linear Unit; SMAPE: Symmetric Mean Absolute Percentage Error.

where $\widehat{Y}_{j,k}$ is the final prediction for sample plot j and output carbon pool component k; f_t is the t$^{\text{th}}$ regression tree, T is the total number of boosting rounds (i.e., number of trees in the ensemble); and $X_{j,p}$ is the *p*-dimensional vector of selected optimal Sentinel-2 predictors for sample plot *j*.

At each boosting step, XGBoost adds a new tree f_t by minimizing the regularized objective:

$$\mathscr{L} = \sum_{j=1}^{n} l\left(Y_{j,k}, \widehat{Y}_{j,k}\right) + \Omega(f_t) \tag{8}$$

where n is the total number of sample plots; $Y_{j,k}$: the obsevered value of carbon pool component k of sample plot j; and l(.) is the prediction loss; and $\Omega(f_t)$ is the penalty imposed on tree complexity.

XGBoost models were trained using the mean squared error (MSE) objective, while hyperparameter selection and model ranking were based on additive SMAPE to ensure balanced predictive accuracy across AGCP, BGCP, and TGCP.

After constructing tree f_t, the model prediction is updated according to

$$\widehat{Y}_{j,k}^{t} \leftarrow \widehat{Y}_{j,k}^{(t-1)} + \eta f_t\left(X_{j,p}\right) \tag{9}$$

where η is the learning rate. Each new tree f_t is trained to correct the residual errors of the previous ensemnble, and after T boosting rounds, the final model prediction becomes the additive sum of all trees.

In this study, we optimized a multi-input–multi-output deep neural network to develop the DLSM using Sentinel-2 predictors selected via PCA. Optimization included architecture design, training algorithms, activation functions, and hyperparameter tuning (Seely et al., 2023; Perbet et al., 2024), ensuring alignment with dataset characteristics (Huy et al., 2024, 2025). The selected settings are summarized in Table 3. L2 regularization was applied in each hidden layer to penalize large weights, and Dropout rates (0.1- 0.5) were defined based on established deep learning practice, where probabilities within this range are commonly adopted to balance regularization strength and model stability. In our framework, dropout was treated as a tunable regularization hyperparameter and optimized jointly with other architectural components (e.g., network depth, number of hidden layers, number of neurons, and L2 regularization). The optimal configuration was selected based on performance under spatial cross-validation to ensure robust generalization, as presented in Table 3. The XGBoost model (Chen and Guestrin, 2016; Diamantopoulou et al., 2025) was optimized using Bayesian optimization with the Hyperopt library to identify optimal hyperparameters (Table 3). Final architectures and hyperparameters for both models were selected through spatial cross-validation based on the lowest additive SMAPE (Table 3).

In this work, the DLXG modeling system was implemented as a stacked ensemble model (Yao et al., 2022; Kablan et al., 2023). DLXG integrated DLSM with XGBoost to improve additive predictive accuracy for above- and belowground carbon pools. The DLSM, serving as base models, captured complex nonlinear relationships between forest carbon pools and Sentinel-2 predictors, while XGBoost acted as a meta-learner to model residuals from the DLSM predictions. This residual-stacking approach leveraged the complementary strengths of both methods. Although DLSM produced high-quality predictions, it can leave systematic errors, particularly at higher target values. By using XGBoost to learn and correct these residuals, the DLXG system mitigated overfitting, reduced bias, and improved overall predictive accuracy, establishing a robust ensemble framework.

The ensemble DLXG workflow: 1) Generation of DLSM out-of-fold (DLOOF) predictions (Yao et al., 2022; Kablan et al., 2023) (i.e., predictions on validation data during spatial cross-validation): Using the optimized architecture, functions, and hyperparameters, the DLSMs generated DLOOF predictions for the three additive components of forest carbon pools - AGCP, BGCP, and TGCP (denoted AGCP_DLOOF, BGCP_DLOOF, and TGCP_DLOOF, respectively) - based on the optimal Sentinel-2 predictors. Multiple DLSMs were trained and selected using 5-fold spatial cross-validation with two repeated runs per fold, resulting in a total of 10 realizations. In each iteration, models were trained on approximately 80% of the dataset (four folds) for 500 epochs with a batch size of 32, computing $SMAPE_{loss}$ at each epoch. Model performance was assessed on the held-out approximately 20% of the dataset (validation fold), and the best model in each iteration was selected based on the lowest validation additive $SMAPE_{loss}$ while ensuring comparable training loss. Additive AGCP_DLOOF, BGCP_DLOOF, and TGCP_DLOOF predictions were then generated for the validation data, ensuring that each sample was predicted only by a model that had not seen it during training, thereby preventing data leakage and providing an unbiased performance estimate. Finally, after completing the 5-fold spatial cross-validation, DLOOF predictions for all samples were obtained from the five unseen validation sets and subsequently averaged across the two runs, yielding the final set of simultaneous AGCP_DLOOF, BGCP_DLOOF, and TGCP_DLOOF predictions, which were then used as inputs to the XGBoost meta-learner. 2) Residual Stacking with XGBoost: Residuals of the DLOOF predictions (true values minus DLOOF predictions) were calculated. Using the selected hyperparameters, XGBoost, acting as a meta-learner, was trained to predict these residuals using both the optimal Sentinel-2 predictors and the DLOOF predictions as input features. By learning these residuals, XGBoost corrected systematic errors from the base DLSMs, thereby improving overall predictive accuracy. The best XGBoost meta-learner was selected using the identical 5-fold spatial cross-validation with two repeated runs (10 realizations) as that applied in the base DLSMs. In each realization, four folds (approximately 80% of the dataset) were used for training, and one fold (approximately 20%) was reserved for validation. The optimal XGBoost model was identified based on the lowest SMAPE across the validation folds, ensuring minimal relative prediction error. Finally, the additive predictions of AGCP, BGCP, and TGCP from the ensemble DLXG model were obtained by combining the DLOOF-based predictions with the corresponding XGBoost residual predictions. Performance metrics of the best-selected ensemble DLXG modeling system were then computed based on the corresponding validation dataset. This workflow can be summarized as follows: generate DLOOF predictions using DLSM, employ XGBoost to correct the residual errors, and then combine them to produce

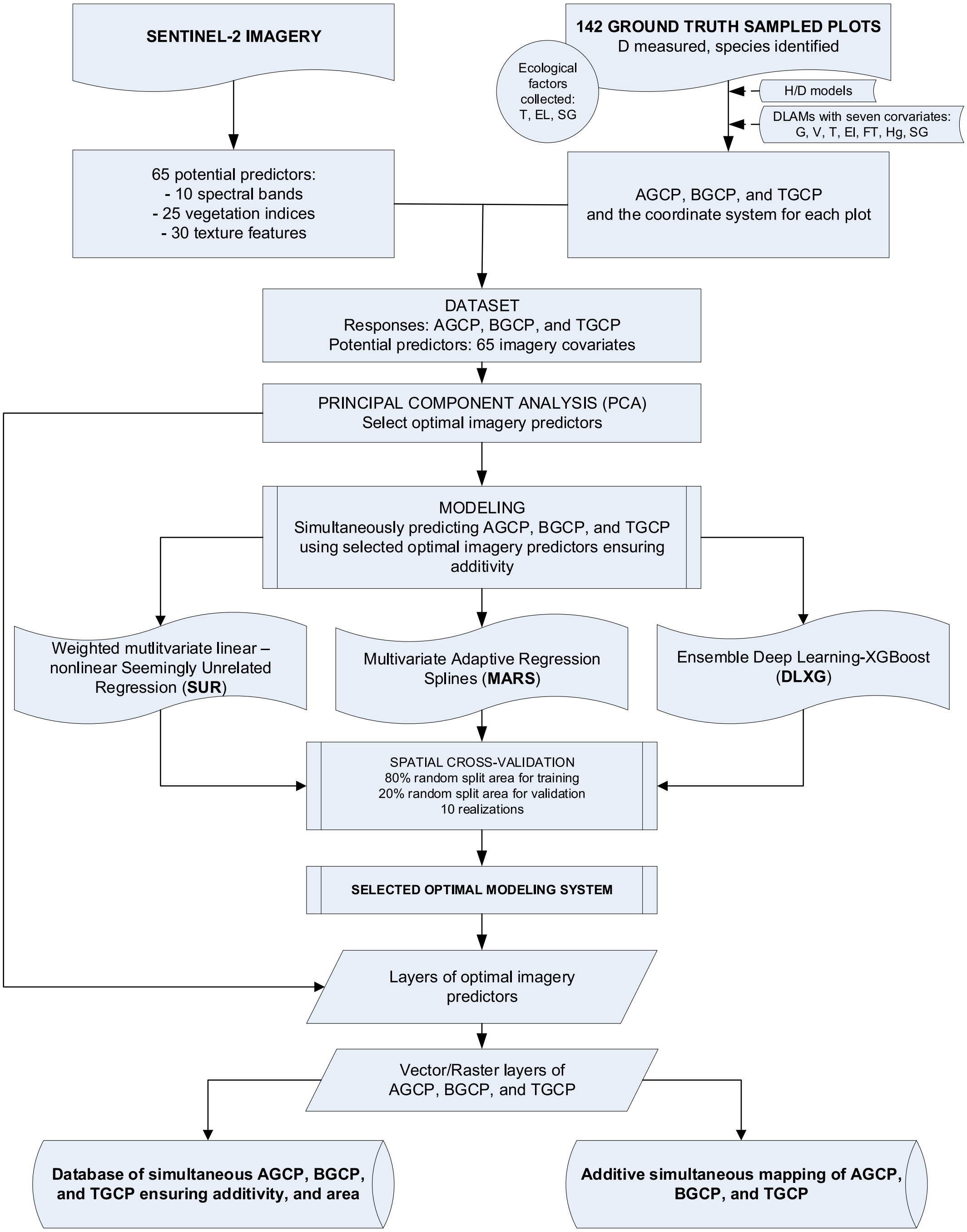


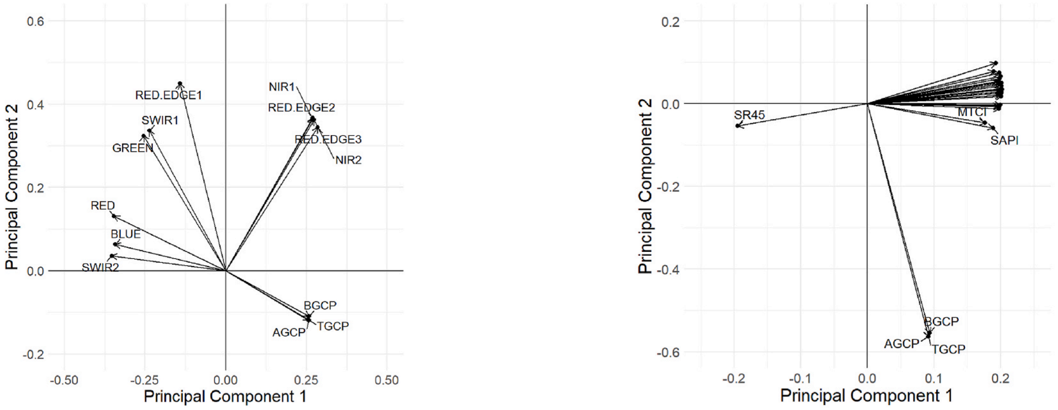
Fig. 4. Workflow of the methodology

Note: D: Diameter at breast height, H: Tree height (H), DLAMs: Deep learning additive models for simultaneous prediction of stand level above-belowground biomass using 7 covariates: G (Stand basal area), V (Stand volume), T (Mean annual temperature), EL (Elevation), FT (Forest type), Hg (Mean height with the tree of Dg), and SG (Soil group), AGCP: Aboveground carbon pool, BGCP: belowground carbon pool, TGCP: Total ground carbon pools (TGCP = AGCP + BGCP).

the final simultaneous predictions (Fig. 3).

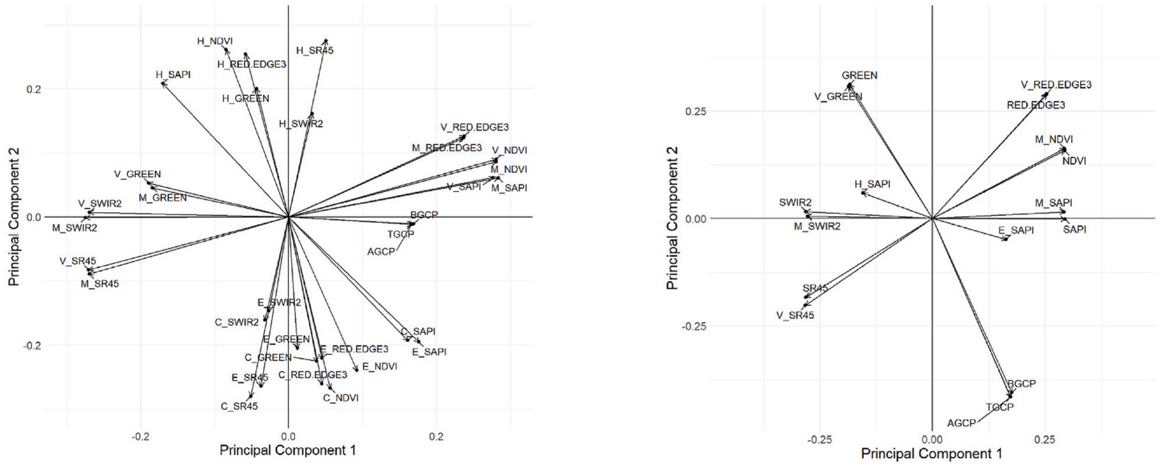
The ensemble DLXG modeling system for additively and simultaneously predicting AGCP, BGCP, and TGCP is formulated in the following equations.

The DLSM $f(\cdot)$ function generates out-of-fold DLOOF predictions for the three additive components - $\widehat{\mathrm{AGCP}}_{\mathrm{DLOOF}}(j)$, $\widehat{\mathrm{BGCP}}_{\mathrm{DLOOF}}(j)$ and $\widehat{\mathrm{TGCP}}_{\mathrm{DLOOF}}(j)$ - on the unseen spatial-validation datasets using the predictors set X_j, where X_j denotes the optimal Sentinel-2



PCA Step 1 selected three key Sentinel-2 spectral bands SWIR2, GREEN, and RED.EDGE3 from 10 bands that significantly influenced the response variables AGCP, BGCP, and TGCP.

PCA Step 2 selected three key Sentinel-2 vegetation indices SR45, SAPI, and NDVI from 25 indices that significantly influenced the response variables AGCP, BGCP, and TGCP.



PCA Step 3 selected eight key Sentinel-2 texture features V_SR45, M_SWIR2, V_GREEN, H_SAPI, V_RED.EDGE3, M_NDVI, M_SAPI, and E_SAPI from 30 features that significantly influenced the response variables AGCP, BGCP, and TGCP.

PCA Step 4 selected six optimal Sentinel-2 predictors V_SR45, SWIR2, V_GREEN, V_RED.EDGE3, NDVI, and SAPI from 14 key-selected variables: 3 key spectral bands, 3 key vegetation indices, and 8 key texture features that significantly influenced the response variables AGCP, BGCP, and TGCP.

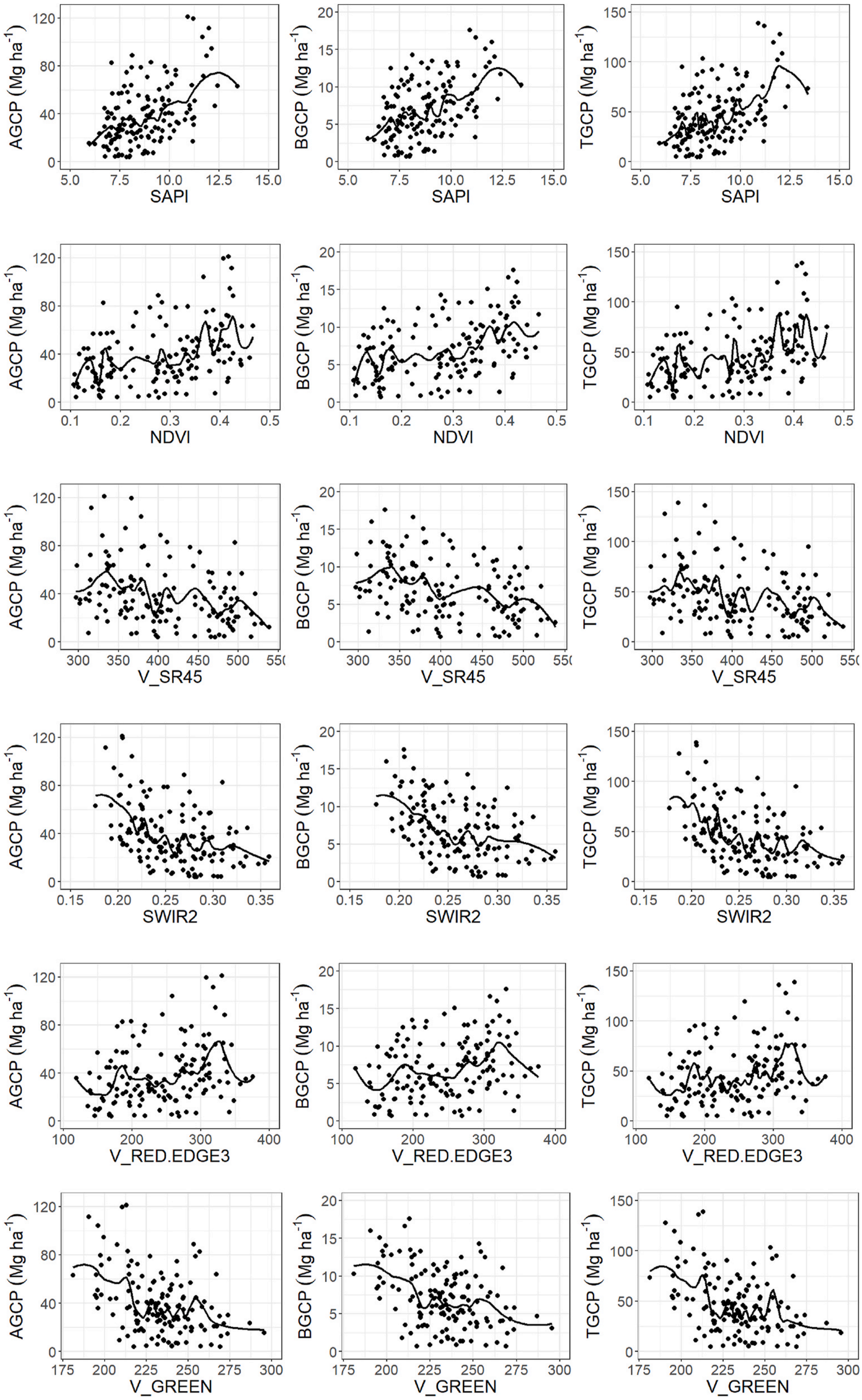
Fig. 5. Four-step Principal Component Analysis (PCA) process for identifying optimal Sentinel-2 predictors. Plot of variable weights in Principal Component 1 (PC1) vs. Principal Component 2 (PC2).
Note: AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total carbon pool. The 65 acronyms of spectral bands (10), vegetation indices (25), and texture features (30) derived from Sentinel-2 imagery are listed in Table 2.

predictors for the j$^{\text{th}}$ sample plot:

$$\widehat{\text{AGCP}}_{\text{DLOOF}}(j) = f_{AGCP}(\text{X}_j) \tag{10}$$

$$\widehat{\text{BGCP}}_{\text{DLOOF}}(j) = f_{BGCP}(\text{X}_j) \tag{11}$$

$$\widehat{\text{TGCP}}_{\text{DLOOF}}(j) = f_{TGCP}(\text{X}_j) \tag{12}$$



(caption on next page)

Fig. 6. Scatter plots with LOESS (Locally Estimated Scatterplot Smoothing) curves showing the complex non-linear relationships between above- and belowground, and total carbon pools (AGCP, BGCP, and TGCP, respectively), and six optimal Sentinel-2 imagery predictors (SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3).

Table 4
Spatial cross-validation outcomes for parametric modeling systems of weighted multivariate linear - nonlinear SUR for simultaneously predicting AGCB, BGCP, and TGCP. The functional form and Sentinel-2 selected predictors (SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN) for each system are also given.

SUR modeling systems	Weight	Bias (%)	RMSE (Mg ha^{-1})	SMAPE (%)	FI
SUR-1:					
AGCP = a_1 + b_{11}×SAPI + b_{12}×NDVI + b_{13}×V_SR45 + b_{14}×SWIR2 + b_{15}× V_RED.EDGE3 + b_{16}× V_GREEN	1/SAPI$^{-2.8}$	−59.85	23.5	48.41	0.252
BGCP = a_2 + b_{21}×SAPI + b_{22}×NDVI + b_{23}×V_SR45 + b_{24}×SWIR2 + b_{25}× V_RED.EDGE3 + b_{26}× V_GREEN	1/SAPI$^{-1.1}$	−54.70	3.5	44.67	0.238
TGCP = AGCP + BGCP	1/SAPI$^{3.0}$	−58.64	27.0	47.76	0.251
SUR-2:					
AGCP = a_1 + b_{11}×log(SAPI) + b_{12}×log(NDVI) + b_{13}×log(V_SR45) + b_{14}×log(SWIR2) + b_{15}×log(V_RED.EDGE3) + b_{16}×log(V_GREEN)	1/SAPI$^{-2.8}$	−78.51	22.9	49.32	0.238
BGCP = a_2 + b_{21}×log(SAPI) + b_{22}×log(NDVI) + b_{23}×log(V_SR45) + b_{24}×log(SWIR2) + b_{25}×log(V_RED.EDGE3) + b_{26}×log(V_GREEN)	1/SAPI$^{-1.1}$	−72.71	3.4	44.69	0.228
TGCP = AGCP + BGCP	1/SAPI$^{-3.0}$	−77.02	26.2	48.57	0.237
SUR-3:					
AGCP = a_1 × SAPI^{b11} × NDVI^{b12} × V_SR45^{b13} × SWIR2^{b14} × V_RED.EDGE3^{b15}× V_GREEN^{b16}	1/SAPI$^{-0.5}$	−10.87	24.9	52.62	0.177
BGCP = a_2 × SAPI^{b21} × NDVI^{b22} × V_SR45^{b23} × SWIR2^{b24} × V_RED.EDGE3^{b25}× V_GREEN^{b26}	1/SAPI$^{0.5}$	60.02	7.2	159.73	−2.442
TGCP = AGCP + BGCP	1/SAPI$^{-0.5}$	0.57	30.4	57.39	0.116
SUR-4:					
AGCP = a_1 + b_{11}×SAPI + b_{12}×SAPI^3 + b_{13}×NDVI + b_{14}×NDVI^3 + b_{15}×V_SR45 + (b_{16}/10000)×V_SR45^3 + b_{17}×SWIR2 + b_{18}×SWIR2^3 + b_{19}×V_RED.EDGE3 + (b_{110}/10000)× V_RED.EDGE3^3 + b_{111}×V_GREEN + (b_{112}/10000)× V_GREEN^3	1/SAPI$^{5.3}$	−61.85	25.0	51.00	0.192
BGCP = a_2×1000 + b_{21}×SAPI + b_{22}×SAPI^3 + b_{23}×NDVI + b_{24}×NDVI^3 + b_{25}×V_SR45 + (b_{26}/10000)×V_SR45^3 + b_{27}×SWIR2 + b_{28}×SWIR2^3 + b_{29}×V_RED.EDGE3 + (b_{210}/10000)× V_RED.EDGE3^3 + b_{211}×V_GREEN + (b_{212}/10000)× V_GREEN^3	1/SAPI$^{1.8}$	−55.63	3.7	45.89	0.215
TGCP = AGCP + BGCP	1/SAPI$^{-3.2}$	−60.41	28.6	50.12	0.195

Note: SUR: Seemingly Unrelated Regression. Spatial cross-validation: The study area was divided into 25 equal geographic blocks. For each of 10 realizations, 5 blocks (≈20%) were randomly selected and assigned as the validation set, while the remaining 20 blocks (≈80%) were used for calibration. Error metrics were averaged across realizations, weighted by the number of sampled plots in the validation blocks, while FIs were averaged over the training blocks. AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total carbon pool. SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3). log: logarithm Neper. **Bold**: The best modeling system.

DLOOF prediction residuals $\left(r_j^{AGCP}, r_j^{BGCP}, r_j^{TGCP}\right)$ are used as the response variables for the XGBoost model:

$$r_j^{AGCP} = \mathrm{AGCP}_{true}(\mathrm{j}) - \widehat{\mathrm{AGCP}}_{\mathrm{DLOOF}}(j) \tag{13}$$

$$r_j^{BGCP} = \mathrm{BGCP}_{true}(\mathrm{j}) - \widehat{\mathrm{BGCP}}_{\mathrm{DLOOF}}(j) \tag{14}$$

$$r_j^{TGCP} = \mathrm{TGCP}_{true}(\mathrm{j}) - \widehat{\mathrm{TGCP}}_{\mathrm{DLOOF}}(j) \tag{15}$$

XGBoost meta-model (residual learning): The meta-features $\mathrm{Z_j}$ supplied to XGBoost consist of both the optimal Sentinel-2 predictor set $\mathrm{X_j}$ and the corresponding $\mathrm{DLOOF_j}$ predictions for the j$^{\mathrm{th}}$ sample plot:

$$Z_j = \left[X_j, \widehat{\mathrm{AGCP}}_{\mathrm{DLOOF}}(j), \widehat{\mathrm{BGCP}}_{\mathrm{DLOOF}}(j), \widehat{\mathrm{TGCP}}_{\mathrm{DLOOF}}(j)\right] \tag{16}$$

The XGBoost $g(\cdot)$ function generates the residual corrections $\widehat{r_j}$ for the three additive components for the j$^{\mathrm{th}}$ sample plot:

$$\widehat{r}_j^{AGCP} = g_{AGCP}\left(\mathrm{Z}_j\right) \tag{17}$$

Table 5
Estimated parameters of the best seemingly unrelated regression modeling system (SUR-1) in weighted linear form for simultaneously predicting AGCP, BGCP, and TGCP based on six selected predictors of Sentinel-2 using the entire dataset.

SUR-1 modeling system	Parameters		Std. Error
AGCP = a_1 + b_{11}×SAPI + b_{12}×NDVI + b_{13}×V_SR45 + b_{14}×SWIR2 + b_{15}× V_RED.EDGE3 + b_{16}× V_GREEN	a_1*	−99.51	54.66
	b_{11}	10.12	3.95
	b_{12}*	129.9	92.4
	b_{13}*	0.04256	0.1035
	b_{14}*	8.545	80.890
	b_{15}*	−0.2156	0.1041
	b_{16}*	0.2022	0.1522
BGCP = a_2 + b_{21}×SAPI + b_{22}×NDVI + b_{23}×V_SR45 + b_{24}×SWIR2 + b_{25}× V_RED.EDGE3 + b_{26}× V_GREEN	a_2*	−5.348	8.804
	b_{21}	1.325	0.611
	b_{22}*	16.23	15.05
	b_{23}*	−0.00369	0.01680
	b_{24}*	3.576	13.514
	b_{25}*	−0.03132	0.01630
	b_{26}*	0.01956	0.02460
TGCP = AGCP + BGCP.			

Note: AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total carbon pool. SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3). *: Parameters with p_value > 0.05.

Table 6
Spatial cross-validation outcomes for MARS models were obtained across various ‘degrees’. These models were employed to simultaneously predict AGCP, BGCP, and TGCP, utilizing six selected predictors – SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3 and V_GREEN - extracted from Sentinel-2 bands, vegetation indices, and textural features.

Degree	Responses	Bias (%)	RMSE (Mg ha^{-1})	SMAPE (%)	FI	Number of selected terms	Number of selected predictors
1	AGCP	−51.02	24.6	55.93	0.352	10/17	5/6 (V_GREEN, SWIR2, NDVI, V_RED.EDGE3, SAPI)
	BGCP	−45.23	3.7	49.43	0.321		
	TGCP	−49.86	28.2	54.80	0.349		
2	AGCP	−46.93	24.0	55.42	0.438	**12/20**	**5/6 (SWIR2, V_GREEN, V_RED.EDGE3, NDVI, V_SR45)**
	BGCP	−42.87	3.7	50.12	0.388		
	TGCP	−46.09	27.6	54.61	0.433		
3	AGCP	−49.60	26.0	59.52	0.442	11/20	5/6 (SWIR2, V_GREEN, V_RED.EDGE3, NDVI, V_SR45)
	BGCP	−46.07	3.9	51.50	0.386		
	TGCP	−48.83	29.8	58.46	0.436		
4	AGCP	−48.99	26.0	59.38	0.443	11/19	6/6 (SWIR2, V_RED.EDGE3, V_GREEN, NDVI, SAPI, V_SR45)
	BGCP	−45.16	3.9	52.78	0.387		
	TGCP	−48.16	29.9	58.44	0.437		
5	AGCP	−48.99	26.0	59.38	0.443	11/19	6/6 (SWIR2, V_RED.EDGE3, V_GREEN, NDVI, SAPI, V_SR45)
	BGCP	−45.16	3.9	52.78	0.387		
	TGCP	−48.16	29.9	58.44	0.437		
6	AGCP	−48.99	26.0	59.38	0.443	11/19	6/6 (SWIR2, V_RED.EDGE3, V_GREEN, NDVI, SAPI, V_SR45)
	BGCP	−45.16	3.9	52.78	0.387		
	TGCP	−48.16	29.9	58.44	0.437		

Note: MARS: Multivariate Adaptive Regression Splines. Spatial cross-validation: The study area was divided into 25 equal geographic blocks. For each of 10 realizations, 5 blocks (≈20%) were randomly selected and assigned as the validation set, while the remaining 20 blocks (≈80%) were used for calibration. Error metrics were averaged across realizations, weighted by the number of sampled plots in the validation blocks, while FIs were averaged over the training blocks. AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total carbon pool. SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3). Weight = 1/SAPI^{-4}. **Bold:** The best degree for the MARS model.

$$\hat{r}_j^{BGCP} = g_{BGCP}(\mathrm{Z}_j) \tag{18}$$

$$\hat{r}_j^{TGCP} = g_{TGCP}(\mathrm{Z}_j) \tag{19}$$

The final DLXG predictions for the additive carbon pools - $\widehat{\mathrm{AGCP}}$, $\widehat{\mathrm{BGCP}}$, and $\widehat{\mathrm{TGCP}}$:

$$\widehat{AGCP}_{DLXG}(j) = \widehat{\mathrm{AGCP}}_{\mathrm{DLOOF}}(j) + \hat{r}_{j,X}^{AGCP} \tag{20}$$

$$\widehat{BGCP}_{DLXG}(j) = \widehat{\mathrm{BGCP}}_{\mathrm{DLOOF}}(j) + \hat{r}_{j,X}^{BGCP} \tag{21}$$

Table 7
Term/hinge functions and their coefficients for the best-selected MARS modeling system with degree = 2, terms = 12, and five selected Sentinel-2 predictors: SWIR2, V_GREEN, V_RED.EDGE3, NDVI, and V_SR45 using the entire dataset.

Terms/hinge functions with knot values	AGCP	BGCP	TGCP
	Coefficients		
(Intercept)	211.37	31.766	243.10
h(SWIR2 - 0.2054)	−4907.38	−700.374	−5606.58
h(0.2411 - SWIR2)	−3284.29	−492.230	−3776.24
h(SWIR2 - 0.2411)	4786.32	677.884	5462.87
h(NDVI - 0.40257)	−3656.29	−485.558	−4139.18
h(0.2231 - SWIR2) × h(330.946 - V_RED.EDGE3)	−66.45	−8.119	−74.54
h(0.2411 - SWIR2) × h(V_GREEN - 218.257)	−92.20	−11.187	−103.41
h(0.2411 - SWIR2) × h(218.257 - V_GREEN)	88.52	10.886	99.39
h(0.2411 - SWIR2) × h(NDVI - 0.386508)	64641.28	8508.223	73105.30
h(230.5 - V_GREEN) × h(NDVI - 0.357261)	−29.57	−3.193	−32.75
h(230.5 - V_GREEN) × h(0.357261 - NDVI)	−31.70	−3.671	−35.35
h(230.5 - V_GREEN) × h(V_SR45 - 365.948)	0.05	0.006	0.06

Note: MARS: Multivariate Adaptive Regression Spline. h(.): hinge function and its knot values. AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total carbon pool. SWIR2 (Band 12), V_GREEN (Texture variance derived from GREEN, Band 3), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), and Weight = $1/\text{SAPI}^{-4}$.

$$\widehat{TGCP}_{DLXG}(j) = \widehat{\text{TGCP}}_{\text{DLOOF}}(j) + \widehat{r}^{TGCP}_{j,X} \tag{22}$$

The DLXG system effectively leverages DLSMs for capturing complex nonlinear patterns and XGBoost for correcting residuals, ensuring that the additive predictions of forest carbon pools are unbiased and robustly validated through spatial cross-validation. This workflow enables accurate and reliable additive predictions of above- and belowground carbon pools (Fig. 3).

The algorithm for DLSM, XGBoost, and DLXG was implemented in the Python programming language (Python, 2022). Various libraries were employed in this process, including Keras (2022), Pandas (McKinney and Pandas Development Team, 2022), and the TensorFlow backend (Abadi et al., 2016; TensorFlow, 2023), XGBoost (Chen and Guestrin, 2016), all of which played a crucial role in facilitating multi-input - multi-output DLXG operations.

2.6. Spatial cross-validation and model selection

Forest ecological and environmental variables are often spatially autocorrelated (Le Rest et al., 2014), necessitating spatial cross-validation to ensure model generalizability. However, many forest biomass carbon mapping studies overlook this spatial autocorrelation, resulting in overoptimistic assessments of model predictive performance (Ploton et al., 2020).

To account for spatial autocorrelation in both the Sentinel-2 predictor fields and the ground-sampled carbon plots, and to assess model transferability to spatially independent areas, we used a spatial cross-validation strategy rather than conventional random cross-validation (Roberts et al., 2017; Valavi et al., 2019; Ploton et al., 2020). The extensive study area (>1000 km^2) was partitioned into equally sized geographic blocks defined from plot coordinates with a 5 × 5 grid, yielding distinct 25 blocks. For each of 10 realizations, 5 blocks (≈20% of the sampled plots) were randomly selected and assigned as the validation set, while the remaining 20 blocks (≈80% of the sampled plots) were used for calibration; plots within the same block were therefore never split between calibration and validation. This blocked random partitioning was repeated 10 times with different random seeds to generate spatially independent and robust estimates of model performance. Model fitting and prediction were performed independently within each realization. This spatial cross-validation strategy reduces optimistic bias caused by spatial autocorrelation and yields a more reliable assessment of the models' capacity to generalize to new, spatially independent regions (Roberts et al., 2017).

For implementing spatial cross-validation in the SUR modeling systems, blocked sampling was conducted in SAS using the procedures 'Proc Syslin', 'Proc Surveyselect', and 'Proc Rank' (SAS Institute Inc, 2024). For the MARS models and multivariate predictions, the 'earth' package (Milborrow et al., 2024) in R was used, with spatial blocked sampling implemented using the 'blockCV' package (Valavi et al., 2019).

Specifically, for the ensemble DLXG modeling, spatial cross-validation (Roberts et al., 2017; Meyer et al., 2019) was implemented using 'Kmeans' clustering on plot coordinates to divide the study area into five spatially distinct blocks. A 5-fold spatial cross-validation with two repeated runs was then applied, where in each iteration one block (≈20% of the sampled plots) served as the validation set and the remaining four blocks (≈80%) as the training set, resulting in 10 unique spatial splits (5 folds × 2 runs) for each modeling stage. This spatial cross-validation was first applied to the base DLSM to generate deep learning out-of-fold (DLOOF) predictions on the held-out validation sets, using different random seeds to enhance robustness and assess model stability. To ensure consistency, the residual-stacked XGBoost meta-learner was subsequently trained using the DLOOF outputs combined with the optimal Sentinel-2 predictors as inputs, under the identical 5-fold spatial cross-validation scheme with two repeated runs as applied to the base DLSMs for identifying the best-performing DLXG model. This principled approach was consistently applied to the SUR and MARS spatial cross-validation procedures as well. The entire procedure was implemented in Python, primarily using the 'scikit-learn' library for spatial cross-validation and clustering functions (Pedregosa et al., 2011). Spatial grouping was achieved using 'Kmeans' clustering

Table 8

The best-performing ensemble DLXG modeling systems - comprising the Deep Learning Simultaneous Model (DLSM) and XGBoost - and their statistical metrics for the additive mapping of AGCP, BGCP, and TGCP using different combinations of optimal Sentinel-2 predictors, including spectral bands, vegetation indices, and textural features (SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN). Spatial cross-validation results are provided.

DLXG modeling systems	Bias (%)	RMSE (Mg ha^{-1})	SMAPE (%)	FI
DLXG-1: 6 predictors: SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, V_GREEN				
AGCP	−20.46	21.0	32.76	0.919
BGCP	−10.13	2.9	29.60	0.887
TGCP	−18.73	23.8	32.05	0.916
DLXG-2: 5 predictors: SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3				
AGCP	−26.07	22.8	36.64	0.890
BGCP	−13.53	3.0	31.69	0.867
TGCP	−23.97	25.7	35.80	0.887
DLXG-3: 4 predictors: SAPI, NDVI, V_SR45, SWIR2				
AGCP	−25.40	23.1	36.36	0.884
BGCP	−15.22	3.2	33.31	0.852
TGCP	−23.65	26.2	35.89	0.880
DLXG-4: 3 predictors: SAPI, NDVI, V_SR45				
AGCP	−40.95	12.7	40.08	0.833
BGCP	−37.41	2.3	36.18	0.794
TGCP	−40.60	15.0	39.57	0.828
DLXG-5: 3 predictors: V_SR45, V_RED.EDGE3, V_GREEN				
AGCP	−20.54	22.5	35.69	0.903
BGCP	−6.98	2.9	29.62	0.860
TGCP	−18.34	25.1	34.36	0.899
DLXG-6: 2 predictors: SAPI, NDVI				
AGCP	−16.74	21.0	35.86	0.801
BGCP	−11.49	2.7	31.58	0.750
TGCP	−15.81	23.5	35.05	0.795
DLXG-7: 1 predictor: SAPI				
AGCP	−9.29	24.8	34.07	0.768
BGCP	−2.59	3.3	32.81	0.714
TGCP	−8.12	28.0	33.84	0.761
DLXG-8: 1 predictor: NDVI				
AGCP	−33.11	24.2	37.35	0.737
BGCP	−23.71	2.9	30.57	0.682
TGCP	−31.46	27.0	36.30	0.730
DLXG-9: 1 predictor: V_SR45				
AGCP	−53.05	25.7	45.21	0.744
BGCP	−55.94	4.1	43.05	0.709
TGCP	−53.38	29.5	44.82	0.741
DLXG-10: 1 predictor: SWIR2				
AGCP	−37.11	28.9	43.04	0.776
BGCP	−28.58	3.6	36.53	0.722
TGCP	−35.66	32.4	42.00	0.769
DLXG-11: 1 predictor: V_RED.EDGE3				
AGCP	−14.35	32.8	53.38	0.772
BGCP	−9.01	4.5	47.39	0.723
TGCP	−13.34	37.2	52.39	0.765
DLXG-12: 1 predictor: V_GREEN				
AGCP	−11.22	24.7	34.53	0.765
BGCP	−5.37	3.2	30.73	0.706
TGCP	−10.19	27.8	33.86	0.758

Note: Spatial cross-validation: A 5-fold spatial cross-validation was applied using KMeans clustering on plot coordinates to divide the study area into five spatially distinct blocks with two repeated runs (5 × 2 = 10 realizations). In each iteration, one spatial block (≈20% of the sampled plots) was used for validation and four blocks (≈80%) for training. The best DLSMs were selected in each realization based on the lowest validation $\text{SMAPE}_{\text{loss}}$ achieved within 500 training epochs, and were used to generate deep learning out-of-fold (DLOOF) predictions on the held-out validation sets. After completing the 5-fold cross-validation, DLOOF predictions for all samples were obtained from the five unseen validation sets and averaged across the two runs. These DLOOF predictions were then combined with the optimal Sentinel-2 predictors as inputs for residual-stacked XGBoost modeling. The XGBoost meta-learner adopted the identical 5-fold spatial cross-validation strategy with two runs to select the best-performing ensemble DLXG model, based on the lowest validation SMAPE across all validation sets. Performance metrics of the best-selected DLXG modeling systems were then computed based on the corresponding validation dataset, while FIs were derived from the corresponding training subsets. AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total carbon pools. SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3). **Bold:** The best-selected DLXG model system was determined based on the component-metrics combinations.

Table 9

Comparison of statistical metrics for the best-performing single Deep Learning Simultaneous Model (DLSM), single XGBoost model, and Deep Learning-XGBoost ensemble model (DLXG) in simultaneously predicting AGCP, BGCP, and TGCP using six optimal Sentinel-2 predictors, including spectral bands, vegetation indices, and textural features (SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN). Spatial cross-validation results are reported.

Modeling systems	Bias (%)	RMSE (Mg ha^{-1})	SMAPE (%)	FI
DLSM:				
AGCP	−15.79	21.4	30.58	0.220
BGCP	−13.39	2.9	29.85	0.282
TGCP	−15.26	24.1	30.42	0.233
XGBoost:				
AGCP	−25.41	20.7	34.82	0.486
BGCP	−21.30	2.7	30.71	0.432
TGCP	−24.64	23.3	34.14	0.479
DLXG ensemble:				
AGCP	−20.46	21.0	32.76	0.919
BGCP	−10.13	2.9	29.60	0.887
TGCP	−18.73	23.8	32.05	0.916

Note: Spatial cross-validation: A 5-cluster spatial cross-validation was applied, in which, for each of 10 realizations, one spatial block (≈20% of the sampled plots) was randomly selected for validation, while the remaining four blocks (≈80%) were used for training. This procedure yielded 10 independent, spatially distinct validation runs. The best DLSM and XGBoost models were selected across all realizations based on the lowest validation SMAPE. The spatial cross-validation procedure and selection of the best-performing ensemble DLXG model are summarized in Table 8. Performance metrics of the best-selected modeling systems were then computed based on the corresponding validation dataset, while FIs were derived from the corresponding training subsets. AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total carbon pools. SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3). **Bold:** The best-selected model system was determined based on the component-metrics combinations.

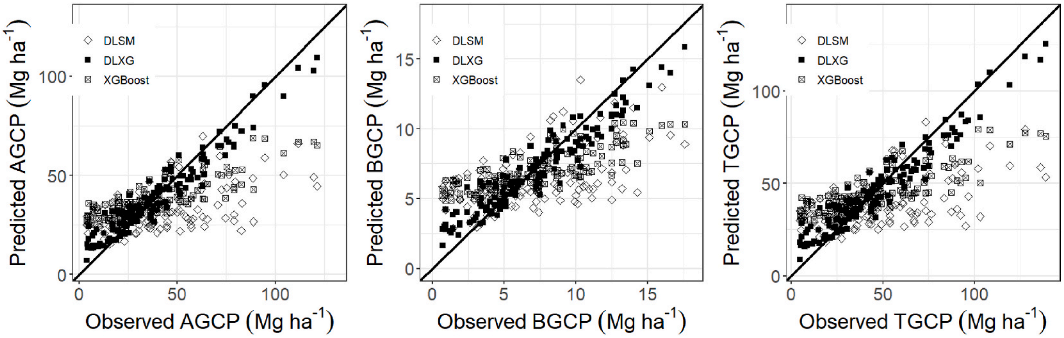


Fig. 7. Predicted versus observed values with the 1:1 reference line for the best-performing single Deep Learning Simultaneous Model (DLSM), single XGBoost model, and ensemble Deep Learning-XGBoost model (DLXG) for simultaneously predicting above-belowground, and total carbon pools (AGCP, BGCP, and TGCP, respectively). Predictions were based on six optimal Sentinel-2 predictors, including spectral bands, vegetation indices, and textural features: SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3), using the entire dataset.

on plot coordinates, followed by 'GroupKFold' from 'sklearn.model_selection' to ensure spatially grouped samples within each fold.

The best modeling systems were selected based on four evaluation metrics: bias percentage (Bias), root mean square error (RMSE), symmetric mean absolute percentage error (SMAPE) (Tofallis, 2014), and fit index (FI, R^2), computed as follows:

$$\text{Bias }(\%) = \frac{1}{R}\sum\nolimits_1^R \frac{100}{V}\sum_{j=1}^{V}\frac{Y_j - \widehat{Y}_j}{Y_i} \tag{23}$$

$$\text{RMSE }\left(\text{Mg ha}^{-1}\right) = \frac{1}{R}\sum\nolimits_1^R \sqrt{\frac{1}{V}\sum_{j=1}^{V}\left(Y_j - \widehat{Y}_j\right)^2} \tag{24}$$

$$\text{SMAPE }(\%) = \frac{1}{R}\sum\nolimits_1^R \frac{100}{V}\sum_{j=1}^{V}\frac{\left|Y_j - \widehat{Y}_j\right|}{\frac{\left|Y_j\right| + \left|\widehat{Y}_j\right|}{2}} \tag{25}$$

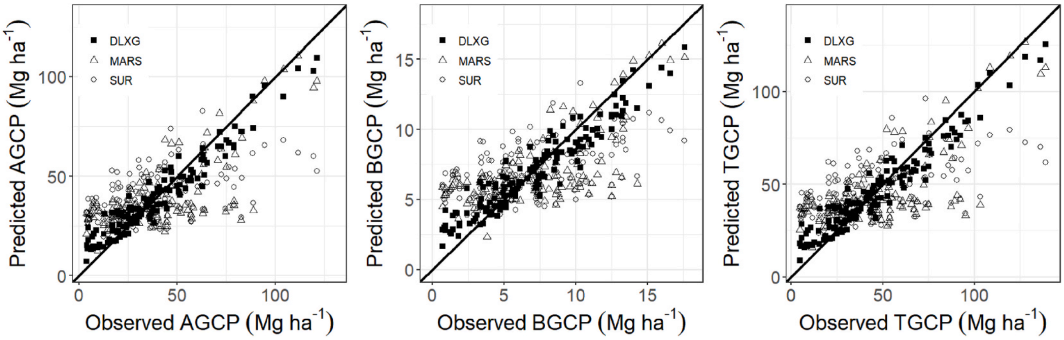


Fig. 8. Comparison of predicted versus observed values with the 1:1 reference line for the best-selected ensemble Deep Learning-XGBoost model (DLXG), multivariate adaptive regression splines (MARS), and seemingly unrelated regression (SUR) modeling systems, using six optimal Sentinel-2 predictors, including spectral bands, vegetation indices, and textural features: SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED. EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3). The plots show predicted vs. observed values for the simultaneous prediction of above-belowground, and total carbon pools (AGCP, BGCP, and TGCP, respectively). Based on the entire dataset.

Table 10

Comparison of the best-selected modeling systems - DLXG (Deep Learning-XGBoost ensemble), MARS (Multivariate Adaptive Regression Splines), and SUR (Seemingly Unrelated Regression) - for the additive mapping of AGCP, BGCP, and TGCP are based on Sentinel-2 spectral bands, vegetation indices, and textural features, incorporating six optimal predictors: SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN. Spatial cross-validation results are provided.

Approaches	Modeling systems	Components	Bias (%)	RMSE (Mg ha^{-1})	SMAPE (%)	FI
Deep Learning-XGBoost ensemble	**The best-selected DLXG-1**	AGCP	−20.46	21.0	32.76	0.919
		BGCP	−10.13	2.9	29.60	0.887
		TGCP	−18.73	23.8	32.05	0.916
Non-parametric nonlinear regression MARS	The best-selected modeling system with a degree of 2 and 12 terms	AGCP	−46.93	24.0	55.42	0.438
		BGCP	−42.87	3.7	50.12	0.388
		TGCP	−46.09	27.6	54.61	0.433
Parametric linear regression SUR	The best-selected modeling system in the weighted linear SUR-1	AGCP	−59.85	23.5	48.41	0.252
		BGCP	−54.70	3.5	44.67	0.238
		TGCP	−58.64	27.0	47.76	0.251

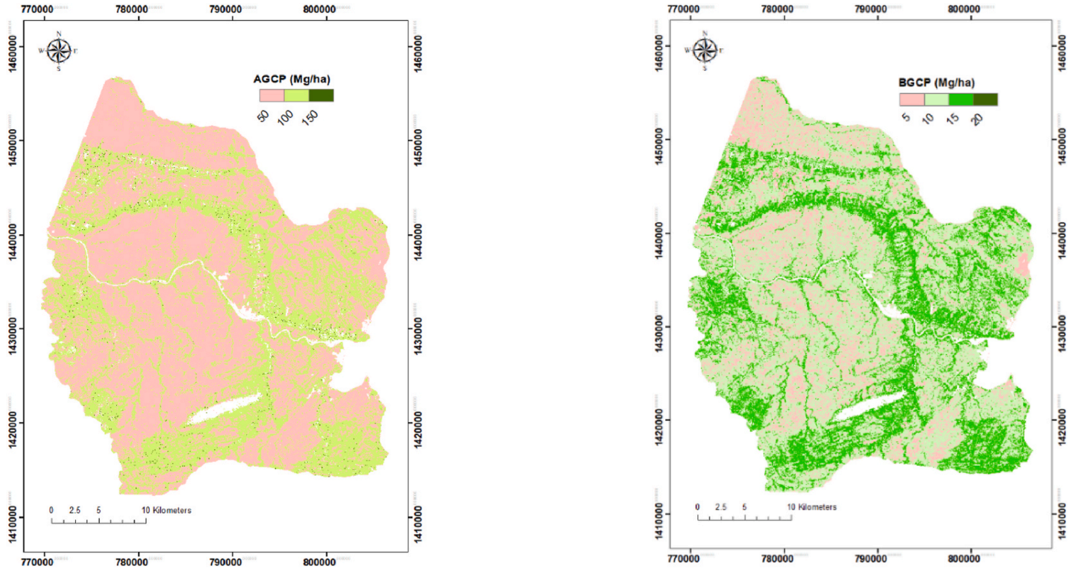
Note: AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total carbon pools. SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3). **Bold:** The optimal modeling system among the three approaches.

Table 11

Total AGCP, BGCP, and TGCP by class and summary statistics of carbon pools in the studied dry dipterocarp forest in YDNP. Values derived from additive mapping generated by the best-selected ensemble Deep Learning-XGBoost model (DLXG-1) based on six optimal Sentinel-2 predictors: SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN.

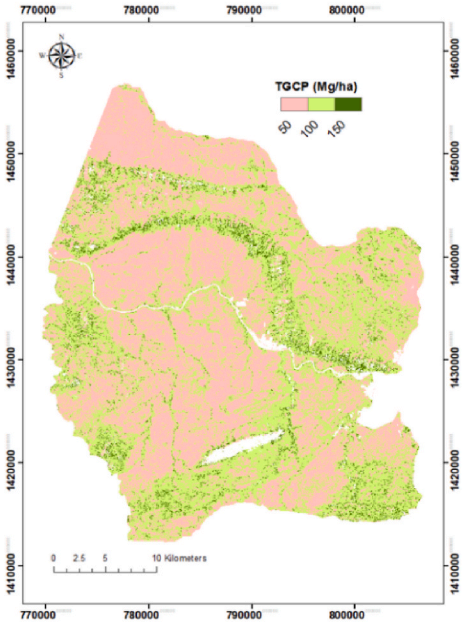
TGCP			AGCP		BGCP		
class (Mg ha^{-1})	Area (ha)	Total (Mg)	Area (ha)	Total (Mg)	class (Mg ha^{-1})	Area (ha)	Total (Mg)
0 – 50	55,633	1,962,309	66,546	2,162,030	0 - 5	19,852	81,548
50 - 100	49,510	3,592,956	42,137	2,925,170	5 - 10	59,963	426,971
100 - 150	3955	422,503	415	42,863	10 - 15	29,225	338,299
					15 - 20	58	888
Total	109,098	5,977,768	109,098	5,130,063		109,098	847,705
Mean (Mg ha^{-1})		54.8		47.0			7.8
Std. (Mg ha^{-1})		23.8		21.0			2.8
Min. (Mg ha^{-1})		0.4		0.3			0.1
Max. (Mg ha^{-1})		141.8		124.1			17.7

Note: The calculated statistics were based on 10,909,761 pixels, each representing 100 m^2. YDNP: Yok Don National Park, Vietnam. AGCP: Aboveground carbon pool, BGCP: Belowground carbon pool, TGCP: Total carbon pool (TGCP = AGCP + BGCP). SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3).



AGCP map

BGCP map



TGCP map

Fig. 9. Simultaneous predictive mapping of above-belowground carbon pools and total (AGCP, BGCP, and TGCP, respectively) in the dry dipterocarp forest (DDF) of Yok Don National Park (YDNP), Vietnam, based on the best-selected ensemble Deep learning-XGBoost model system (DLXG-1). This modeling system utilized six optimal Sentinel-2 predictors, including spectral bands, vegetation indices, and texture features: SAPI (Scatter-Adjusted Pigment Index), NDVI (Normalized Difference Vegetation Index), V_SR45 (Texture variance derived from Simple Ratio 45), SWIR2 (Band 12), V_RED.EDGE3 (Texture variance derived from RED.EDGE3, Band 7), and V_GREEN (Texture variance derived from GREEN, Band 3).

$$\mathrm{FI} = \frac{1}{\mathrm{R}}\sum\nolimits_{1}^{\mathrm{R}}\left\{1 - \frac{\sum_{j=1}^{\mathrm{T}}\left(\mathrm{Y}_j - \widehat{\mathrm{Y}}_i\right)^2}{\sum_{j=1}^{\mathrm{T}}\left(\mathrm{Y}_j - \overline{\mathrm{Y}}\right)^2}\right\} \tag{26}$$

where R represents the number of realizations (set to 10). V represents the number of sampled plots in the validation area block, and T denotes the number of plots in the training, or the total area blocks, as applicable. Y_j, $\widehat{\mathrm{Y}}_j$ and $\overline{\mathrm{Y}}$ correspond to the j$^{\mathrm{th}}$ plot's observed, predicted, and mean values of AGCP, BGCP, and TGCP, respectively.

To compare different modeling systems, four performance measures - FI, Bias, RMSE, and SMAPE - were evaluated across three model components: AGCP, BGCP, and TGCP, yielding 12 component-metric combinations from spatial cross-validation. The superior modeling system was identified by the largest number of component-metric combinations that exhibited better performance. In cases where multiple models achieved the same number of superior component-metric combinations, graphical analyses were performed to evaluate their effectiveness further. These visual evaluations included plots comparing predicted versus observed values for AGCP, BGCP, and TGCP, offering deeper insights into model performance and potential systematic deviations.

Using the best-performing model and raster layers of the optimal Sentinel-2 predictors, we conducted additive mapping of AGCP, BGCP, and TGCP through a Python-based implementation. The overall methodology is illustrated in Fig. 4.

3. Results

3.1. *The selection of optimal Sentinel-2 imagery predictors*

This study implemented a four-step PCA to select the optimal Sentinel-2 predictors for the AGCP, BGCP, and TGCP modeling systems in the DDF. The relationships and magnitudes of associations between the three response variables and potential Sentinel-2 predictors derived from PCA are presented in Fig. 5.

Using variable weights in Principal Component 1 (PC1) and trends from the PC1 vs. PC2 plot (Fig. 5), we identified Sentinel-2 predictors influencing AGCP, BGCP, and TGCP. Selection criteria were: (1) variables positively correlated with responses and with high absolute weights, grouped near responses; (2) variables showing strong inverse relationships (180° opposite trends); (3) for correlated clusters, selecting the variable with the highest absolute weight to reduce redundancy; (4) independent predictors with substantial contributions to PC1 were included to represent the dominant variance structure of the dataset.

PCA Step 1 analyzed relationships between 10 Sentinel-2 spectral bands (SBs) (Table 2) and the response variables (AGCP, BGCP, TGCP). From the PC1 vs. PC2 plot (Fig. 5) and PC1 weights, three key SBs were selected: SWIR2 (weight = −0.353), GREEN (−0.254), and RED.EDGE3 (+0.284).

PCA Step 2 examined 25 Sentinel-2 vegetation indices (VIs) (Table 2) against the responses. Three key VIs were selected: SR45 (−0.194), SAPI (+0.189), and NDVI (+0.202).

PCA Step 3 assessed 30 texture features (TFs) (Table 2), derived from five metrics (mean (M), variance (V), entropy (E), contrast (C), homogeneity (H)), calculated on the six selected SBs and VIs. Eight key TFs were chosen: V_SR45 (−0.270), M_SWIR2 (−0.275), V_GREEN (−0.189), H_SAPI (−0.170), V_RED.EDGE3 (+0.238), M_NDVI (+0.281), M_SAPI (+0.284), and E_SAPI (+0.176).

PCA Step 4 combined the 14 key variables from steps 1–3 and identified six optimal Sentinel-2 predictors, ranked by absolute PC1 weight: SAPI (+0.295), NDVI (+0.293), V_SR45 (−0.283), SWIR2 (−0.282), V_RED.EDGE3 (+0.253), and V_GREEN (−0.186).

The four PCA steps identified six optimal Sentinel-2 predictors, ranked by influence on AGCP, BGCP, and TGCP: SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN. These were used in the DLXG, MARS, and SUR models. SAPI and NDVI, with the highest component weights, also served as weight variables.

Fig. 6 contains scatter plots showing substantial variation in AGCP, BGCP, and TGCP values in relation to the six optimal Sentinel-2 predictors. This variation reflects spectral differences resulting from the uneven distribution of the DDF canopy. Given the complexity of these relationships, as illustrated by the LOESS (Locally Estimated Scatterplot Smoothing) curves (Fig. 6), conventional regression methods may struggle to capture them effectively.

3.2. *Conventional regression systems*

Using the six Sentinel-2 predictors selected through a four-step PCA (SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN), we developed weighted multivariate linear and non-linear SUR modeling systems to simultaneously predict AGCP, BGCP, and TGCP while ensuring additivity. The spatial cross-validation results for the examined SUR modeling systems are presented in Table 4. Of the four spatial cross-validated SUR modeling systems, the SUR-1 weighted linear model outperformed the others in 12 component–metric combinations, leading to its selection as the best model. However, despite being the best SUR modeling system among those compared, it exhibited poor fit, with fit indices below 0.3 (Table 4). Table 5 presents the estimated parameters of the SUR-1 model for the simultaneous prediction of AGCP, BGCP, and TGCP using the six optimal Sentinel-2 predictors and the full dataset. The results further reveal the poor performance of SUR-1, with many parameters exhibiting p-values greater than 0.05.

Notably, this result confirms the hypothesis that the substantial variability in the spectral reflectance of DDF foliage observed in Sentinel-2 imagery (Fig. 6) hinders the ability of conventional regression models, such as SUR, to achieve an adequate level of fit when estimating forest carbon pools.

We utilized MARS as a non-parametric method for developing modeling systems to simultaneously predict AGCP, BGCP, and TGCP. In contrast to SUR (e.g., Table 4), MARS does not require the manual specification of a trend function, and the algorithm automatically selects predictor variables. These models incorporated SBs, VIs, and TFs via the six predictors selected through the four-step PCA described previously. The spatial cross-validation results for the MARS modeling systems with degrees ranging from 1 to 6 are presented in Table 6. Based on 12 component-metric combinations from the spatial cross-validation of the three component models, the MARS modeling system with a degree of 2 achieved the best overall performance. Beyond degree 4, up to degree 6, the model metrics and MARS model components showed no further changes. The optimal model MARS with a degree of 2 utilized 12 terms and 5 key predictors: SWIR2, V_GREEN, V_RED.EDGE3, NDVI, and V_SR45 (Table 6). The degree value of 2 in this optimal MARS system indicates nonlinear interactive effects between the Sentinel-2 predictors and forest carbon pool components in this DDF, which are difficult to accurately capture without modeling approaches that explicitly account for such interactions (Fig. 6). Spatial cross-validation showed that the optimal MARS modeling system yielded SMAPE values exceeding 50% across all component equations (Table 6). In contrast, the best-selected MARS models exhibited moderate FI values of approximately 0.4 for the simultaneous prediction of AGCP, BGCP, and TGCP in DDF (Table 6). After spatially validating the performance of the MARS modeling systems, the entire dataset was used to fit the best-selected system (Table 7).

3.3. DLXG systems

The ensemble DLXG systems, designed for simultaneous mapping and additive predictions of AGCP, BGCP, and TGCP in DDF based on different combinations of the six optimal Sentinel-2 predictors, underwent spatial cross-validation, with the corresponding results reported in Table 8.

Among these, DLXG-1 included all six optimal Sentinel-2 predictors - SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN - ranked from highest to lowest component weight based on PCA analysis. For each successive DLXG system, the number of predictors was reduced by sequentially eliminating those with the lowest to highest PCA component weight values. DLXG-7 retained only a single Sentinel-2 predictor, SAPI, which had the highest PCA component weight. Additionally, six single-predictor DLXG systems were developed and spatially cross-validated, resulting in a total of 12 model systems (Table 8).

Table 8 summarizes the spatial cross-validation statistics for each DLXG model. Among them, DLXG-1 consistently ranked as the top performer, outperforming the others across 12 component-metric combinations. DLXG-1 simultaneously predicts AGCP, BGCP, and TGCP based on two vegetation indices (SAPI and NDVI), three texture features (V_SR45, V_RED.EDGE3, and V_GREEN), and a single spectral band (SWIR2). Incorporating SBs, VIs, and TFs improved model performance for the additive mapping of AGCP, BGCP, and TGCP. DLXG-1 demonstrated strong goodness-of-fit, achieving the highest FI values of 0.919, 0.887, and 0.916, with corresponding SMAPE results of 32.76%, 29.60%, and 32.05%, and RMSE values of 21.0, 2.9, and 23.8 Mg ha^{-1} for the simultaneous prediction of AGCP, BGCP, and TGCP, respectively (Table 8).

Notably, as the number of optimal Sentinel-2 predictors decreased from six to one, the reliability of the DLXG models declined accordingly. DLXG-5, which incorporated only three texture features - V_SR45, V_RED.EDGE3, and V_GREEN - still demonstrated strong reliability, ranking second overall, with FI values of 0.903, 0.860, and 0.899 for AGCP, BGCP, and TGCP predictions, respectively. This finding highlights the essential role of texture features in predicting forest carbon in DDF. In contrast, DLXG-6, which included only two vegetation indices - SAPI and NDVI - ranked third, with FI values of 0.801, 0.750, and 0.795, indicating the significant contribution of vegetation indices as Sentinel-2 predictors.

When the DLXG systems retained only a single predictor from SBs, VIs, or TFs, their reliability declined substantially. This underscores the importance of integrating SBs, VIs, and TFs to effectively capture spectral variability in Sentinel-2 imagery and optimize DLXG model performance in DDF.

Uncertainty estimates, expressed as mean ± standard deviation (SD) across spatial cross-validation realizations, showed that the best-performing DLXG-1 model, incorporating all six optimal Sentinel-2 predictors (SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN), achieved FI of 0.934 ± 0.020, 0.902 ± 0.020, and 0.930 ± 0.020, and RMSE values of 23.1 ± 4.4, 3.6 ± 0.7, and 26.5 ± 5.1 Mg ha^{-1} for the addtive prediction and simultaneous mapping of AGCP, BGCP, and TGCP.

3.4. Selecting and applying the optimal modeling system

Given the limitations of the individual modeling systems - DLSM and XGBoost - in reliably predicting AGCP, BGCP, and TGCP in DDF using six optimal Sentinel-2 predictors, the ensemble DLXG system was developed to enhance predictive performance by leveraging their complementary strengths. As shown in Table 9, both DLSM and XGBoost exhibited relatively low FI scores, indicating suboptimal model fit and limited robustness across components. Specifically, DLSM yielded notably low FI values (up to ≈0.28), while XGBoost achieved moderately higher values (up to ≈0.49). In contrast, the ensemble DLXG model, designed to combine the two individual models (DLSM and XGBoost), consistently outperformed both, surpassing the single models across all 12 component–metric combinations and achieving the highest FI scores (up to ≈0.92) across all three components, thereby demonstrating superior goodness of fit and generalization capability.

These spatial cross-validation results confirmed the DLXG ensemble as a more robust and accurate solution for predicting the simultaneous biomass carbon components in DDF regions characterized by high variability in the relationship between Sentinel-2-derived spectral features and forest carbon pools (Fig. 6). This result was consistent with the predicted vs. observed plots of AGCP, BGCP, and TGCP (Fig. 7) obtained from the two single models (DLSM and XGBoost) and the ensemble DLXG system using six optimal Sentinel-2 predictors.

The development and spatial cross-validation results of the three modeling systems - the parametric method SUR, the non-parametric method MARS, and the DLXG approach - indicate that DLXG effectively captured the complex relationships between Sentinel-2 data and forest biomass carbon pools (Fig. 6) while delivering consistent and reliable predictive performance (Fig. 8). When comparing the spatial cross-validation results of the best modeling systems from the three approaches - DLXG-1, SUR-1 (a parametric linear regression) using six selected Sentinel-2 predictors (SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN), and MARS (a non-parametric nonlinear regression) using five selected predictors (SWIR2, V_GREEN, V_RED.EDGE3, NDVI, and V_SR45) with degree = 2 and terms = 12 - it became evident that the DLXG approach outperformed both the SUR and MARS modeling systems, surpassing them across all 12 component–metric combinations (Table 10).

The best DLXG-1 modeling system achieved the highest goodness-of-fit, with FI values of up to 0.92, whereas the MARS and SUR systems showed considerably lower FIs, below 0.44 and 0.26, respectively. DLXG-1 reduced SMAPE by 22.66%, 20.52%, and 22.56% compared to the best MARS model, and by 15.65%, 15.07%, and 15.71% compared to the best SUR model for the simultaneous prediction of AGCP, BGCP, and TGCP, respectively (Table 10).

Reinforcing this finding, Fig. 8 presents predicted versus observed values for AGCP, BGCP, and TGCP, showing that the predictions of the best DLXG model follow the 1:1 line fairly well across low-high ranges, clearly demonstrating its superior performance compared to the best MARS and SUR modeling systems. Based on our findings, we recommend using DLXG-1 for the simultaneous predictive mapping of above- and belowground carbon pools, as well as their total, in tropical DDF, employing six optimal Sentinel-2 imagery predictors - SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN.

Applying the best-selected DLXG-1 to the six optimal Sentinel-2 predictors across the studied DDF area in YDNP (109,098 ha) resulted in an estimated total carbon sequestration of approximately 6.0 million metric tons (Table 11), equivalent to more than 22.0 million metric tons of CO_2. This approach also allowed for the simultaneous prediction of AGCP, BGCP, and TGCP while ensuring additivity, as presented in Table 11 at the time of the study. These results revealed that the tropical DDF in the study area has averaged values of 47.0 Mg ha^{-1} for AGCP, 7.8 Mg ha^{-1} for BGCP, and 54.8 Mg ha^{-1} for TGCP (Table 11). Additionally, we simultaneously generated additive maps for AGCP, BGCP, and TGCP, which are displayed in Fig. 9.

4. Discussion

4.1. Performance of the DLXG ensemble vs. conventional regression models

The key contribution of this study is the development of a novel DLXG ensemble - a hybrid of DLSM and XGBoost - that enhances the reliability of simultaneous mapping of forest carbon pools using optimal Sentinel-2 predictors in DDF. Although deep learning has demonstrated strong performance in multi-output prediction (Huy et al., 2024, 2025), its application to carbon mapping in DDF using remote sensing data such as Sentinel-2 remains limited, yielding very low FI values (below 0.29) (Table 9). In contrast, machine learning algorithms such as XGBoost achieve moderate FI values (around 0.49) (Table 9). This is consistent with the findings of Yadegari et al. (2025), who reported the highest machine learning performance (including XGBoost) with FI = 0.47 but still a high RMSE of ≈79 Mg ha^{-1} for AGB (equivalent to 37 Mg ha^{-1} for AGCP) using features derived from satellite data, including Sentinel-2. The DLXG ensemble addresses the spectral variability of tropical DDFs (Fig. 6) by: (1) simultaneously optimizing the deep neural network architecture of the base DLSM and the hyperparameters of the XGBoost meta-learner, and (2) implementing a residual-stacking ensemble in which out-of-fold predictions from the DLSMs capture complex nonlinear relationships and serve as inputs to the XGBoost model, thereby systematically correcting residual errors and enhancing the overall additive predictive accuracy of forest carbon pools in DDF. Compared with conventional machine learning reported by Yadegari et al. (2025) and Tikuye and Ray (2025), which achieved FIs of 0.43 – 0.47, and those tested in this study, where the best MARS and SUR models attained FIs below 0.44 and 0.26, respectively (Table 10), the best DLXG modeling system achieved FI value exceeding 0.91 and an RMSE of 21.0 Mg ha^{-1} for AGCP (Table 8, Tables 9, and Table 10). These findings underscore its potential as a reliable tool for carbon monitoring and credit development in tropical forests using Sentinel-2 data.

Various studies have combined remote sensing with conventional regressions to predict AGB. Tang et al. (2022) used Stepwise Linear Regression with vegetation indices to predict AGB in *Pinus* forests, yielding a low R^2 of 0.32. Similarly, Rana et al. (2023) applied weighted least-squares regression to RapidEye imagery in Nepal and reported an R^2 of 0.34. In contrast, higher R^2 values were achieved in some cases: Askar et al. (2018) obtained $R^2 = 0.79$ using Sentinel-2, and Sun et al. (2021) reported $R^2 = 0.74$ using combined Sentinel-1 and -2 data. Fisher (2023) used LiDAR-derived canopy height and linear regression in South Africa, with R^2 ranging from 0.16 to 0.75. Gascon and Eva (2014) highlighted that non-linear regression can improve model performance. Overall, both linear and non-linear models often struggle to capture the complex relationships between tropical biomass and remotely sensed data, suggesting a need for flexible approaches such as MARS or deep learning.

SUR (Parresol, 2001; Huy et al., 2023) enables simultaneous mapping of above- and belowground biomass carbon but is rarely applied in this context. Like other conventional regressions, SUR struggles with multi-input, multi-output modeling due to challenges in defining appropriate equation forms and selecting optimal covariates (Kralicek et al., 2017; Huy et al., 2023). In this study, SUR showed poor spatial cross-validation performance, with FI values below 0.3 for all components, whereas the DLXG model achieved FI values of up to 0.9 (Table 10).

The non-parametric MARS method (Friedman, 1991) can model AGCP, BGCP, and TGCP simultaneously, supporting additivity without requiring variable normalization or predefined functional forms. Like DLXG, MARS handles complex multi-input/output relationships, but it also performs automatic variable selection (Friedman, 1993; R Core Team, 2023; Huy et al., 2025), which DLXG lacks. MARS has shown mixed performance in forest carbon mapping: Filippi et al. (2014) reported high accuracy for AGB estimation, while Safari et al. (2017) found lower R^2 values (0.34 - 0.58) using Landsat 8. In this study, MARS achieved moderate predictive accuracy (FIs = 0.39 – 0.44 for all components; Table 10) but was outperformed by DLXG (FIs = 0.89 – 0.92; Table 10). Given its flexibility, further exploration of MARS for simultaneous forest carbon prediction under complex tropical conditions is warranted.

A key contribution of this study is the explicit inclusion of the belowground carbon pool, which is often overlooked in remote sensing studies that typically focus on AGB. Despite its strong correlation with AGB, BGCP remains understudied due to its limited spectral visibility. By incorporating BGCP into the best-performing simultaneous model (DLXG-1), this study demonstrates that BGCP can be reliably predicted (FI = 0.89, RMSE = 2.9 Mg ha^{-1}) using Sentinel-2 imagery. These results highlight the strength of simultaneous modeling in capturing belowground carbon dynamics from remotely sensed data.

While the hybrid DLXG model outperformed conventional parametric and nonparametric regression methods in simultaneously mapping and predicting AGCP, BGCP, and TGCP, a key limitation of this study is the time-intensive process required to design the architecture of the multi-input–multi-output deep neural network. Specifically, identifying the optimal hyperparameters of DLSM and XGBoost for the best-performing ensemble DLXG model involved extensive evaluation of numerous combinations of activation functions, learning algorithms, hyperparameter settings, and network structures.

4.2. Sentinel-2 imagery predictors for mapping carbon pools in DDF

Due to the high spectral variability of Sentinel-2 imagery over heterogeneous DDF canopies, developing DLXG models for simultaneous carbon pool prediction required integrating predictors across three categories: SBs, VIs, and TFs (Table 8). PCA helped identify six optimal predictors - SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN. Models using only VIs, TFs, or single-category predictors showed markedly reduced reliability (Table 8), consistent with the findings of Fang et al. (2023) and Zhang et al. (2024). Notably, variance-based texture features (Villegas et al., 2023) significantly enhanced prediction performance under complex DDF conditions.

Remotely sensed VIs have been widely used to estimate forest AGB (Sun et al., 2021), and model performance improves with the inclusion of texture measures (Gascon and Eva, 2014). Fang et al. (2023) showed that VIs-derived textures outperformed spectral bands and their textures for estimating growing stock volume. In contrast, this study highlights the critical role of texture features from both SBs and VIs, particularly V_SR45, V_RED.EDGE3, and V_GREEN - in predicting forest carbon pools in DDF (Table 8).

NDVI is the most widely used vegetation index for modeling forest biomass from satellite data (Huang et al., 2021), followed by GNDVI and EVI (Navarro et al., 2019; Askar et al., 2018; Tang et al., 2022). However, its significance varies; Ghimire et al. (2024) reported a non-significant NDVI-AGB relationship in some cases. In this study, NDVI was a key predictor of DDF carbon, ranking second after SAPI. Among spectral bands, SWIR2 also played a critical role.

4.3. Variation of forest carbon pools in DDF

According to IPCC et al. (2006), the AGCP of tropical DDF ranges from 47 to 75 Mg ha^{-1}, with an average of 61 Mg ha^{-1}. In contrast, this study found that AGCP varied from 0.3 to 124.1 Mg ha^{-1}, with an average of 47.0 Mg ha^{-1} across 1000 km^2 of the largest DDF area in Vietnam, located within YDNP (Table 11). This indicates that DDFs in Vietnam exhibit greater carbon variability than those in tropical regions worldwide. Some forest stands have low biomass yet are fully mature due to extremely harsh ecological conditions, resulting in stunted yet mature forests with limited total biomass accumulation. Conversely, certain areas exhibit relatively high carbon biomass due to more favorable conditions for carbon accumulation.

In summary, the DDF ecosystem is harsh, with highly variable environmental conditions that can lead to significant fluctuations in carbon levels. Additionally, young and mid-aged forest stands emerge from regeneration in gaps and clearings left by fallen parent trees, eventually developing into diverse forest formations that accumulate varying levels of aboveground and belowground carbon biomass.

4.4. Application of the best-selected DLXG

The best-selected DLXG-1 is recommended for the additive prediction and simultaneous mapping of DDF forest carbon pools, including AGCP, BGCP, and TGCP. The model system incorporates six optimal Sentinel-2 predictors: SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN. By updating Sentinel-2 imagery time series and recalculating predictor layers, forest carbon stock can be continuously monitored and updated over time.

For DDF, January represents the optimal period - the transition from the rainy to the dry season - for acquiring Sentinel-2 imagery. Rainy-season images are often cloud-covered and unusable, while dry-season imagery, although cloud-free, captures leafless canopies due to dipterocarp leaf fall, which distorts spectral signals. January offers cloud-free conditions with intact canopies, providing the most accurate spectral representation of DDF biomass carbon.

To facilitate the application of the best-performing DLXG system across different spatial and temporal contexts, a Python script was developed. To further enhance accessibility, a user-friendly executable package was then created to run the DLXG-1 model using the six optimal Sentinel-2 predictor layers - SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN. This package easily and rapidly enables simultaneous prediction and mapping of AGCP, BGCP, and total TGCP by simply double-clicking, ensuring additivity across large-scale tropical DDFs over time using Sentinel-2 time series data - without requiring professional Python programming skills. The guidelines and executable package can be downloaded from: https://baohuy-frem.org/deep-learning-for-forest-biomass-prediction/

5. Conclusions

This study confirmed the hypothesis that substantial variation in Sentinel-2 spectral reflectance across the DDF canopy limits the reliability of conventional regression models - whether linear, nonlinear, parametric, or non-parametric. In contrast, the innovatively developed ensemble DLXG model effectively overcame this limitation, demonstrating enhanced performance in the additive prediction and simultaneous mapping of AGCP, BGCP, and TGCP by integrating Sentinel-2 imagery with ground-based observations across diverse DDF ecosystems.

The results indicated that the parametric linear SUR and non-parametric nonlinear MARS models performed less reliably in the additive mapping of forest carbon pools. In contrast, the ensemble DLXG model outperformed both in spatial cross-validation by incorporating three Sentinel-2 data categories (SBs, VIs, and TFs), comprising six optimal predictors (SAPI, NDVI, V_SR45, SWIR2, V_RED.EDGE3, and V_GREEN). The best-performing ensemble DLXG-1 model achieved FI values of up to 0.92, 0.89, and 0.92, with corresponding RMSE values of 21.0, 2.9, and 23.8 Mg ha^{-1} for the additive mapping of AGCP, BGCP, and TGCP, respectively. We recommend DLXG-1 for accurate simultaneous prediction and additive mapping of above- and belowground carbon pools in tropical DDF. This optimal approach leverages Sentinel-2 imagery time series and the DLXG-1 model, enabling periodic updates of DDF biomass carbon data and maps by refreshing the satellite-based predictors.

Simultaneously mapping AGCP and BGCP in tropical forests remains challenging but essential for carbon management. This study developed optimized DLXG models for DDF, leveraging satellite imagery and biological relationships to enhance accuracy, interpretability, and additive consistency across carbon pool estimates.

Key innovations of this study include: (1) the development of a novel hybrid ensemble DLXG framework through the joint optimization of the base DLSM and XGBoost meta-learner architectures, in which out-of-fold DLSM predictions were used as inputs to a residual-stacking ensemble that systematically corrected residual errors and enhanced the overall additive predictive accuracy of forest carbon pools in DDF; and (2) the improvement of the DLXG model's predictive ability by integrating three categories of Sentinel-2 predictors - SBs, VIs, and TFs - selected via PCA within an ensemble modeling framework.

CRediT authorship contribution statement

Bao Huy: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Krishna P. Poudel:** Writing – review & editing, Validation, Methodology. **Hailemariam Temesgen:** Writing – review & editing, Validation, Supervision, Methodology. **Andreas Christian Braun:** Writing – review & editing, Validation, Methodology. **Nguyen Quy Truong:** Writing – original draft, Software, Methodology. **Pham Cong Tri:** Resources, Investigation, Formal analysis, Data curation. **Nguyen The Hien:** Resources, Investigation, Formal analysis, Data curation. **Nguyen Quy Khiem:** Writing – original draft, Project administration, Investigation, Formal analysis, Data curation.

Ethical statement

The authors declare that this manuscript is original, has not been published previously, and is not under review elsewhere. All authors have approved the final version. If accepted, the article will not be published in the same form in any language, including electronically, without written permission from the copyright holder.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors sincerely thank Yok Don National Park for supporting our research team's collection of ground truth data. We are especially grateful to its Director, Pham Tuan Linh, for his guidance, as well as to technical staff members Luu Thanh and Mai Van Hoa for their active participation in fieldwork within the tropical dry dipterocarp forest. We also appreciate the logistical support provided by the park's ranger stations.

Abbreviations

AGB	Aboveground biomass. AGB (Mg ha^{-1}). Tree AGB (kg).
AGCP	Aboveground carbon pool (Mg ha^{-1})
ANN	Artificial neural network
BGB	Tree belowground biomass. Tree BGB (kg $tree^{-1}$). BGB (Mg ha^{-1})
BGCP	Belowground carbon pools (Mg ha^{-1})
C	Contrast
D	Diameter at breast height (cm)
DDF	Dry Dipterocarp Forest
Dg	Quadratic mean diameter at breast height (cm)
DL	Deep Learning
DLAMs	Deep Learning Additive Models
DLOOF	Deep learning simultaneous model out-of-fold
DLSM	Deep Learning Simultaneous Model
DLXG	Ensemble Deep Learning Simultaneous Model (DLSM) and eXtreme Gradient Boosting (XGBoost)
E	Entropy
EL	Elevation (m)
FAO	Food and Agriculture Organization of the United Nations
FI	Fit Index or R^2
FT	Forest Type
G	Stand basal area (m^2 ha^{-1})
GLCMs	Gray-Level Co-occurrence Matrices
GNSS	Global Navigation Satellite System
H	Homogeneity
Hg	Mean height of trees corresponding to Dg (m)
IPCC	The Intergovernmental Panel on Climate Change
kNN	k-nearest neighbors
LiDAR	Light Detection and Ranging
LOESS	Locally Estimated Scatterplot Smoothing
M	Mean
MARS	Multivariate Adaptive Regression Splines
ML	Machine Learning
N	Tree density (tree ha^{-1})
P	Mean annual precipitation (mm $year^{-1}$)
PC	Principal Component
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
RF	Random Forest
RMSE	Root mean squared error (Mg ha^{-1})
ROIs	Regions of Interest
SBs	Spectral Bands
SG	Soil Group
SL	Slope (degree)
SMAPE	Symmetric Mean Absolute Percentage Error (%)
SUR	Seemingly Unrelated Regression
T	Mean annual temperature (^oC $year^{-1}$)
TFs	Texture Features
TGCP	Total carbon pool (Mg ha^{-1}) (TGCP = AGCP + BGCP)
UNESCO	United Nations Educational, Scientific and Cultural Organization
V	Variance
VIs	Vegetation Indices
XGBoost	eXtreme Gradient Boosting
YDNP	Yokdon National Park

Data availability

Data will be made available on request.

References

Abadi, M., Barham, P., Chen, J., et al., 2016. TensorFlow: a system for large-scale machine learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi.

Aggarwal, C.C., 2023. Neural networks and deep learning. A Textbook, second ed. Springer, Nature Switzerland AG, Gewerbestrasse. https://doi.org/10.1007/978-3-031-29642-0. 11, 6330 Cham, Switzerland, 541 pp. ISBN 978-3-031-29642-0 (eBook).

Ahmad, A., Gilani, H., Ahmad, S.R., 2021. Forest aboveground biomass estimation and mapping through high-resolution optical satellite imagery – a literature review. Forests 12, 914. https://doi.org/10.3390/f12070914, 2021.

Ahmadi, K., Kalantar, B., Saeidi, V., Harandi, E.K.G., Janizadeh, S., Ueda, N., 2020. Comparison of machine learning methods for mapping the stand characteristics of temperate forests using multi-spectral Sentinel-2 data. Remote Sens. 12, 3019. https://doi.org/10.3390/rs12183019, 2020.

Anees, S.A., Mehmood, K., Khan, W.R., Sajjad, M., Alahmadi, T.A., Alharbi, S.A., Luo, M., 2024. Integration of machine learning and remote sensing for above ground biomass estimation through Landsat-9 and field data in temperate forests of the himalayan region. Ecol. Inform. 82, 102732. https://doi.org/10.1016/j.ecoinf.2024.102732, 2024.

Askar, A., Nuthammachot, N., Phairuang, W., Wicaksono, P., Sayektiningsih, T., 2018. Estimating aboveground biomass on private forest using Sentinel-2 imagery. J. Sens. 2018, 1–11. https://doi.org/10.1155/2018/6745629 article ID 6745629.

Ayushi, K., Babu, K.N., Ayyappan, N., Nair, J.R., Kakkara, A., Reddy, C.F., 2024. A comparative analysis of machine learning techniques for aboveground biomass estimation: a case study of the Western Ghats, India. Ecol. Inform. 80, 102479. https://doi.org/10.1016/j.ecoinf.2024.102479, 2024.

Banerjee, K., Panda, S., Bandyopadhyay, J., Jain, M.K., 2014. Forest canopy density mapping using advanced geospatial technique. IJISET - Int. J. Innov. Sci., Eng. Technol. 1 (7), 358–363. www.ijiset.com.

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. For. Ecol. Manag. 257, 1684–1694.

Cao, L., Coops, N.C., Innes, J., Dai, J., She, G., 2014. Mapping above- and below-ground biomass components in subtropical forests using small-footprint lidar. Forests 5, 1356–1373. https://doi.org/10.3390/f5061356, 2014.

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Folster, H., Fromard, F., Higuchi, N., Kira, T., 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145, 87–99, 2005.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chollet, F., 2018. Deep Learning with Python, p. 386. Manning, Shelter Island, NY, USA.

Cosenza, D.N., Saarela, S., Strunk, J., Korhonen, L., Maltamo, M., Packalen, P., 2024. Effects of model-overfit on model-assisted forest inventory in boreal forests with remote sensing data. Forestry: Int. J. For. Res. 2024, 1–15. https://doi.org/10.1093/forestry/cpae055.

Cuesta, F., Calderón-Loor, M., Rosero, P., Miron, N., Sharf, A., Proaño-Castro, C., Andrade, F., 2023. Mapping above-ground carbon stocks at the landscape scale to support a carbon compensation mechanism: the Chocó andino case study. Forests 14, 1903. https://doi.org/10.3390/f14091903, 2023.

Dalagnol, R., Wagner, F.H., Galvao, L.S., Braga, D., Osborn, F., Sagang, L.B., Bispo, P.D.C., Payne, M., Junior, C.S., Favrichon, S., Silgueiro, V., Anderson, L.O., de Aragao, L.E.O.E.C., Fensholt, R., Brandt, M., Ciais, P., Saatchi, S., 2023. Mapping tropical forest degradation with deep learning and planet NICFI data. Rem. Sens. Environ. 298, 113798. https://doi.org/10.1016/j.rse.2023.113798, 2023.

Diamantopoulou, M.J., Şahi̇n, A., Zçeli̇k, R., 2025. Innovative approaches to estimating sessile oak stem volume: comparing additive models and ensemble machine learning. Turk. J. Agric. For. 49, 740–754. https://doi.org/10.55730/1300-011X.3301, 2025.

Dlamini, C.M., Odindi, J., Matongera, T.N., Mutanga, O., 2025. Exploring the utility of remote sensing technology in vegetation below ground biomass (BGB) estimation: a critical review of methods and challenges. Front. Rem. Sens. 6, 1668676. https://doi.org/10.3389/frsen.2025.1668676.

Esteban, J., McRoberts, R.E., Fernández-Landa, A., Tome, J.L., Næsset, E., 2019. Estimating forest volume and biomass and their changes using random forests and remotely sensed data. Remote Sens. 11, 1944. https://doi.org/10.3390/rs11161944, 2019.

Fang, G., He, X., Weng, Y., Fang, L., 2023. Texture features derived from Sentinel-2 vegetation indices for estimating and mapping forest growing stock volume. Remote Sens. 15 (11), 2821. https://doi.org/10.3390/rs15112821, 2023.

FAO-UNESCO, 2005. Soil Map of the World, Digitized by ESRI. Soil Climate Map, USDA-NRCS, Soil Science Division, World Soil Resources. Washington D.C.

Fassnacht, F.E., Hartig, F., Latifi, H., Bergerd, C., Hernándeze, J., Corvalán, P., Koch, B., 2024. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. Rem. Sens. Environ. 154, 102–114. https://doi.org/10.1016/j.rse.2014.07.028, 2014.

Feilhauer, H., Zlinszky, A., Kania, A., Foody, G.M., Doktor, D., Lausch, A., Schmidtlein, S., 2021. Let your maps be fuzzy! - class probabilities and floristic gradients as alternatives to crisp mapping for remote sensing of vegetation. Remote Sens. Ecol. Conserv. 7 (2), 292–305. https://doi.org/10.1002/rse2.188.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315.

Filippi, A.M., Guneralp, I., Randall, J., 2014. Hyperspectral remote sensing of aboveground biomass on a river meander bend using multivariate adaptive regression splines and stochastic gradient boosting. Rem. Sens. Lett. 5 (5), 432–441. https://doi.org/10.1080/2150704X.2014.915070.

Fisher, J.T., 2023. Optimizing aboveground carbon mapping in Afrotemperate forests to fulfill IPCC carbon reporting standards. For. Ecol. Manag. 552, 121583. https://doi.org/10.1016/j.foreco.2023.121583, 2024.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19 (1), 1–141. https://doi.org/10.1214/aos/1176347963. http://projecteuclid.org/euclid.aos/1176347963.

Friedman, J.H., 1993. Fast MARS. Technical Report, No. 110. Department of Statistics. Stanford University, USA, p. 19. https://purl.stanford.edu/vr602hr6778.

Friedman, J.H., Silverman, B.W., 1989. Flexible parsimonious smoothing and additive modeling. Technometrics 31 (1), 3–21. http://links.jstor.org/sici?sici=0040-1706%28198902%2931%3A1%3C3%3AFPSAAM%3E2.0.CO%3B2-Z.

Gascon, L.H., Eva, H., 2014. Field guide for forest mapping with high resolution satellite data. Monitoring deforestation and forest degradation in the context of the UN-REDD programme. The Tanzania REDD+ initiative. Technical Reports. Joint Research Center (JRC). Publications Office of the European Union, Luxembourg, p. 79 sdoi:10.2788/657954.

Geo University (GU), 2025. Remote sensing satellite data and spectral indices. Available at: https://www.geo.university/pages/spectral-indices-with-multispectral-satellite-data. (Accessed 12 March 2025).

Ghimire, S., Joshi, R., Gautam, J., Bhatta, B., 2024. Modeling forest above-ground biomass of teak (*Tectona grandis* L. f.) using field measurement and Sentinel-2 imagery. J. Sens. 2024. https://doi.org/10.1155/2024/9910094. Article ID 9910094, 12 pages.

Gizachew, B., Solberg, S., Næsset, E., Gobakken, T., Bollandsås, O.M., Breidenbach, J., Zahabu, E., Mauya, E.W., 2016. Mapping and estimating the total living biomass and carbon in low-biomass woodlands using Landsat 8 CDR data. Carbon Bal. Manag. 11, 13. https://doi.org/10.1186/s13021-016-0055-8, 2016.

Haralick, R.M., 1979. Statistical and structural approaches to texture. Proc. IEEE 67, 786–804, 1979.

He, X., Li, S., He, X.T., Wang, W., Zhang, X., Wang, B., 2022. A novel ensemble learning model combined XGBoost with deep neural network for credit scoring. J. Inf. Technol. Res. 15 (1), 18. https://doi.org/10.4018/JITR.299924.

Heiskanen, J., 2006. Estimating aboveground tree biomass and leaf area index in a mountain birch forest using Aster satellite data. Int. J. Rem. Sens. 27 (6), 1135–1158.

Huang, S., Tang, L., Hupy, J.P., Wang, Y., Shao, G., 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J. For. Res. 32 (1), 1–6. https://doi.org/10.1007/s11676-020-01155-1, 2021.

Huy, B., Khiem, N.Q., Truong, N.Q., Poudel, K.P., Temesgen, H., 2023. Additive modeling systems to simultaneously predict aboveground biomass and carbon for *Litsea glutinosa* of agroforestry model in tropical highlands. For. Sys. 32 (1), e006. https://doi.org/10.5424/fs/2023321-19780.

Huy, B., Poudel, K.P., Kralicek, K., Hung, N.D., Khoa, P.V., Phương, V.T., Temesgen, H., 2016. Allometric equations for estimating tree aboveground biomass in tropical dipterocarp forests of Viet Nam. Forests 7 (180), 1–19. https://doi.org/10.3390/f7080180, 2016. http://www.mdpi.com/1999-4907/7/8/180. ISSN 1999-4907.

Huy, B., Poudel, K.P., Temesgen, H., Salas-Eljatib, C., Truong, N.Q., Khiem, N.Q., 2025. Comparing statistical and deep learning approaches for simultaneous prediction of stand-level above- and belowground biomass in tropical forests. Sci. Total Environ. 958, 177869. https://doi.org/10.1016/j.scitotenv.2024.177869, 2025.

Huy, B., Tinh, N.T., Poudel, K.P., Frank, B.M., Temesgen, H., 2019. Taxon-specific modeling systems for improving reliability of tree aboveground biomass and its components estimates in tropical dry dipterocarp forests. For. Ecol. Manag. 437 (2019), 156–174. ISSN: 0378-1127.

Huy, B., Tri, P.C., Triet, T., 2018. Assessment of enrichment planting of teak (*Tectona grandis*) in degraded dry deciduous dipterocarp forest in the central highlands, Vietnam, Southern Forests: a. J. For. Sci. 80 (1), 75–84. https://doi.org/10.2989/20702620.2017.1286560.

Huy, B., Truong, N.Q., Khiem, N.Q., Poudel, K.P., Temesgen, H., 2022. Deep learning models for improved reliability of tree aboveground biomass prediction in the tropical evergreen broadleaf forests. For. Ecol. Manag. 508, 1–13. https://doi.org/10.1016/j.foreco.2022.120031 (2022) 120031.

Huy, B., Truong, N.Q., Poudel, K.P., Temesgen, H., Khiem, N.Q., 2024. Multi-output deep learning models for enhanced reliability of simultaneous tree above- and below-ground biomass predictions in tropical forests of Vietnam. Comput. Electron. Agric. 222, 109080. https://doi.org/10.1016/j.compag.2024.109080, 2024.

IPCC, 2006. In: Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K. (Eds.), IPCC Guidelines for National Greenhouse Gas Inventories. Prepared by the National Greenhouse Gas Inventories Programme. IGES, Japan. Published.

IPCC, 2019. Refinement to the 2006 IPCC guidelines for national greenhouse gas inventories. Volume 4, Chapter 2: Generic Methodologies Applicable to Multiple Land-use Categories. Last accessed August 2019. https://www.ipcc-nggip.iges.or.jp/public/2019rf/index.html.

Joshi, C., Leeuw, J.D., Skidmore, A.K., van Duren, I.C., van Oosten, H., 2006. Remotely sensed estimation of forest canopy density: a comparison of the performance of four methods. Int. J. Appl. Earth Obs. Geoinf. 8 (2006), 84–95.

Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., Dalponte, M., Van Ewijk, K.Y., Forrester, D.I., Haeni, M., Higgins, S.I., Holdaway, R.J., Iida, Y., Lorimer, C., Marshall, P.L., Momo, S., Moncrieff, G.R., Ploton, P., Poorter, L., Rahman, K.A., Schlund, M., Sonke, B., Sterck, F.J., Trugman, A.T., Usoltsev, V.A., Vanderwel, M.C., Waldner, P., Wedeux, B.M.M., Wirth, C., Woll, H., Woods, M., Xiang, W., Zimmermann, N.E., Coomes, D., 2017. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. Glob. Change Biol. (23), 177–190. https://doi.org/10.1111/gcb.13388, 2017.

Kablan, R., Miller, H.A., Suliman, S., Frieboes, H.B., 2023. Evaluation of stacked ensemble model performance to predict clinical outcomes: a COVID-19 study. Int. J. Med. Inf. 175, 105090. https://doi.org/10.1016/j.ijmedinf.2023.105090, 2023.

Keras, 2022. K Keras – simple. Flexible. Powerful. https://keras.io/.

Kralicek, K., Huy, B., Poudel, K.P., Temesgen, H., Salas, C., 2017. Simultaneous estimation of above- and belowground biomass in tropical forests of Viet Nam. For. Ecol. Manag. 390, 147–156, 2017. http://www.sciencedirect.com/science/article/pii/S0378112716307411.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., Bretagnolle, V., 2014. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. Global Ecol. Biogeogr. (23), 811–820. https://doi.org/10.1111/geb.12161, 2014.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (2015), 436–444. https://doi.org/10.1038/nature14539.

Li, H., Hiroshima, T., Li, X., Hayashi, M., Kato, T., 2024. High-resolution mapping of forest structure and carbon stock using multi-source remote sensing data in Japan. Rem. Sens. Environ. 312, 114322. https://doi.org/10.1016/j.rse.2024.114322.

Li, Y., Li, M., Li, C., Liu, Z., 2020. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. Sci. Rep. 10, 9952. https://doi.org/10.1038/s41598-020-67024-3, 2020.

Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., Moran, E., 2016. 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.

Ma, T., Hu, Y., Wang, J., Beckline, M., Pang, D., Chen, L., Ni, X., LiA, X., 2023. Novel vegetation index approach using Sentinel-2 data and random forest algorithm for estimating forest stock volume in the Helan Mountains, Ningxia, China. Remote Sens. 15, 1853. https://doi.org/10.3390/rs15071853, 2023.

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. Ecol. Indic. 159 (2024), 111653.

Matso, N.M., Ong, H.B., Barcellano, E.V., 2024. Mapping and estimating forest stand volume using machine learning methods and multi-spectral Sentinel 2 data. Euro. J. Theor. Appl. Sci. 2 (2), 635–647. https://doi.org/10.59324/ejtas.2024.2(2).55.

McKinney, W., 2022. Pandas: Powerful Python Data Analysis Toolkit. Release 1.4.4. Pandas, p. 3743. Pandas Development Team.

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications - moving from data reproduction to spatial prediction. Ecol. Model. 411, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815.

Milborrow, S., Hastie, T., Tibshirani, R., 2024. Earth: Multivariate Adaptive Regression Splines (R Package). CRAN. cran.r-project.org+1.

Naimi, S., Ayoubi, S., Zeraatpisheh, M., Dematte, J.A.M., 2021. Ground observations and environmental covariates integration for mapping of soil salinity: a machine learning-based approach. Remote Sens. 13, 4825. https://doi.org/10.3390/rs13234825.

Najafi, P., Navid, H., Feizizadeh, B., Eskandari, I., Blaschke, T., 2019. Fuzzy object-based image analysis methods using Sentinel-2A and Landsat-8 data to map and characterize soil surface residue. Remote Sens. 11, 2583. https://doi.org/10.3390/rs11212583, 2019.

Nasiri, V., Darvishsefat, A.A., Arefi, H., Griess, V.C., Sadeghi, S.M.M., Borz, S.A., 2022. Modeling forest canopy cover: a synergistic use of Sentinel-2, aerial photogrammetry data, and machine learning. Remote Sens. 14, 1453. https://doi.org/10.3390/rs14061453, 2022.

Nasset, E., McRoberts, R.E., Pekkarinen, A., Saatchi, S., Santoro, M., Trier, O.D., Zahabu, E., Gobakken, T., 2020. Use of local and global maps of forest canopy height and aboveground biomass to enhance local estimates of biomass in miombo woodlands in Tanzania. Int. J. Appl. Earth Obs. Geoinform. 89, 102109. https://doi.org/10.1016/j.jag.2020.102109, 2020.

Navarro, J.A., Algeet, N., Fernández-Landa, A., Esteban, J., Rodríguez-Noriega, P., Guillén-Climent, M.L., 2019. Integration of UAV, Sentinel-1, and Sentinel-2 data for mangrove plantation aboveground biomass monitoring in Senegal. Remote Sens. 11, 77. https://doi.org/10.3390/rs11010077, 2019.

Nazeri, P., Ayoubi, S., Khademi, H., Afshar, F.A., Mousavi, S.R., 2024. Machine learning-based soil aggregation assessment under four scenarios in northwestern Iran. Int. Agrophys. 38, 293–310. https://doi.org/10.31545/intagr/188506.

Parresol, B.R., 2001. Additivity of nonlinear biomass equations. Can. J. For. Res. 31 (5), 865–878.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12 (85), 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html.

Perbet, P., Guindon, L., Cote, J.F., Beland, M., 2024. Evaluating deep learning methods applied to Landsat time series subsequences to detect and classify boreal forest disturbances events: the challenge of partial and progressive disturbances. Rem. Sens. Environ. 306 (2024), 114107.

Ploton, P., Mortier, F., Réjou-Méchain, M., et al., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 4540. https://doi.org/10.1038/s41467-020-18321-y.

Python, 2022. Python Packaging User Guide. https://packaging.python.org/.

QGIS, 2023. A free and open source geographic information system. https://qgis.org/en/site/. Access on January 1st, 2023.

R Core Team, 2023. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.r-project.org/index.html.

Rana, P., Popescu, S., Tolvanen, A., Gautam, B., Srinivasan, S., Tokola, T., 2023. Estimation of tropical forest aboveground biomass in Nepal using multiple remotely sensed data and deep learning. Int. J. Rem. Sens. 44 (17), 5147–5171. https://doi.org/10.1080/01431161.2023.2240508.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881.

Safari, A., Sohrabi, H., Powell, S., Shataee, S., 2017. A comparative assessment of multi-temporal Landsat 8 and machine learning algorithms for estimating aboveground carbon stock in coppice oak forests. Int. J. Rem. Sens. 38 (22), 6407–6432. https://doi.org/10.1080/01431161.2017.1356488.

Sainuddin, F.V., Mathew, S.M., Saranya, K.R.L., Suthari, S., Asok, S.V., Reddy, C.S., 2023. Chapter 16. From field to map: a review of state-of-the-art approaches to estimate above-ground biomass integrating remote sensing techniques. In: Suthari, S. (Ed.), Exploring Emerging Techniques in Plant Sciences, Vaagdevi Colleges. Telangana, India, pp. 168–181, 978-93-5406-563-7.

SAS Institute Inc, 2014. SAS/ETS® 13.2 User's guide. Chapter 19: the MODEL Procedure. SAS Institute Inc, Cary, NC, pp. 1067–1373.

SAS Institute Inc, 2024. SAS/STAT® User's Guide — Chapter 29 for PROC SYSLIN and Chapter 95 for PROC SURVEYSELECT. https://support.sas.com.

Seely, H., Coops, N.C., White, J.C., Montwe, S.D., Winiwarter, L., Ragab, A., 2023. Modelling tree biomass using direct and additive methods with point cloud deep learning in a temperate mixed forest. Sci. Rem. Sens. 8, 100110. https://doi.org/10.1016/j.srs.2023.100110, 2023.

Strunk, J.L., Bell, D.M., Gregory, M.J., 2022. Pushbroom photogrammetric heights enhance state-level forest attribute mapping with Landsat and environmental gradients. Remote Sens. 14 (14), 3433. https://doi.org/10.3390/rs14143433, 2022.

Su, H., Shen, W., Wang, J., Ali, A., Li, M., 2020. Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. Forest Ecosyst. 7, 64. https://doi.org/10.1186/s40663-020-00276-7, 2020.

Sun, S., Wang, Y., Song, Z., Chen, C., Zhang, Y., Chen, X., Chen, W., Yuan, W., Wu, X., Ran, X., Wang, Y., Li, Q., Wu, L., 2021. Modelling aboveground biomass carbon stock of the Bohai Rim Coastal wetlands by integrating remote sensing, terrain, and climate data. Remote Sens. 13, 4321. https://doi.org/10.3390/rs13214321, 2021.

Tajik, S., Ayoubi, S., Zeraatpisheh, M., 2020. Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, northern Iran. Geoderma Reg. 20, e00256. https://doi.org/10.1016/j.geodrs.2020.e00256.

Tan, H., Kou, W., Xu, W., Wang, L., Wang, H., Lu, N., 2025. Improved estimation of aboveground biomass in rubber plantations using deep learning on UAV multispectral imagery. Drones 9, 32. https://doi.org/10.3390/drones9010032, 2025.

Tang, J., Liu, Y., Li, L., Liu, Y., Wu, Y., Xu, H., Ou, G., 2022. Enhancing aboveground biomass estimation for three *Pinus* forests in Yunnan, SW China, using Landsat 8. Remote Sens. 14, 4589. https://doi.org/10.3390/rs14184589, 2022.

TensorFlow, 2023. Create production-grade machine learning models with TensorFlow. https://www.TensorFlow.org/. (Accessed 10 January 2023).

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, 1086. https://doi.org/10.3390/f14061086, 2023.

Tikuye, B.G., Ray, R.L., 2025. Estimating aboveground biomass using environmental covariates and a machine-learning approach in the Lower Brazos River basin, Texas, USA. Appl. Comput. Geosci. 27, 100289. https://doi.org/10.1016/j.acags.2025.100289, 2025.

Tofallis, C., 2014. A better measure of relative prediction accuracy for model selection and model estimation. J. Oper. Res. Soc. 66, 1352–1362. Available at: SSRN: https://ssrn.com/abstract=2635088.

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. BlockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol. Evol. 10, 225–232. https://doi.org/10.1111/2041-210X.13107.

Villegas, M.H.S., Qasim, M., Csaplovics, E., González-Martinez, R., Rodriguez-Buritica, S., Abril, L.N.R., Villegas, B.S., 2023. Examining the potential of sentinel imagery and ensemble algorithms for estimating aboveground biomass in a tropical dry forest. Remote Sens. 15, 5086. https://doi.org/10.3390/rs15215086, 2023.

Yadegari, L.Z., Sohrabi, H., Quirós, E., Immitzer, M., 2025. Mapping above-ground biomass in old-growth deciduous forests using open-access satellite data, field plots, and machine learning algorithms. Adv. Space Res. https://doi.org/10.1016/j.asr.2025.09.013. ISSN 0273-1177.

Yao, J., Zhang, X., Luo, W., Liu, C., Ren, L., 2022. Applications of stacking/blending ensemble learning approaches for evaluating flash flood susceptibility. Int. J. Appl. Earth Obs. Geoinf. 112, 102932. https://doi.org/10.1016/j.jag.2022.102932, 2022.

Yun, T., Li, J., Ma, L., Zhou, J., Wang, R., Eichhorn, M.P., Zhang, H., 2024. Status, advancements and prospects of deep learning methods applied in forest studies. Int. J. Appl. Earth Obs. Geoinf. 131, 103938. https://doi.org/10.1016/j.jag.2024.103938, 2024.

Zaki, N.A.M., Latif, Z.A., Suratman, M.N., 2018. Modelling above-ground live trees biomass and carbon stock estimation of tropical lowland dipterocarp forest: integration of field-based and remotely sensed estimates. Int. J. Rem. Sens. 39 (8), 2312–2340. https://doi.org/10.1080/01431161.2017.1421793, 2018.

Zeraatpisheh, M., Garosi, Y., Owliaie, H.R., Ayoubi, S., Taghizadeh-Mehrjardi, R., Scholten, T., Xu, M., 2022. Improving the spatial prediction of soil organic carbon using environmental covariates selection: a comparison of a group of environmental covariates. Catena 208, 105723. https://doi.org/10.1016/j.catena.2021.105723.

Zhang, X., Shen, H., Huang, T., Wu, Y., Guo, B., Liu, Z., Luo, H., Tang, J., Zhou, H., Wang, L., Xu, W., Ou, G., 2024. Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery. Ecol. Indic. 159, 111752. https://doi.org/10.1016/j.ecolind.2024.111752, 2024.

Zhang, X., Zhao, Y., Ashton, M.S., Lee, X., 2012. Measuring carbon in forests. In: Ashton, M.S., Tyrrel, M.L., Spalding, D., Gentry, B. (Eds.), Managing Forest Carbon in a Changing Climate. Springer, Dordrecht, Heidelberg, London, New York, pp. 139–164. https://doi.org/10.1007/978-94-007-2232-3, 411.

Zhang, Y., Liu, J., Shen, W., 2022. A review of ensemble learning algorithms used in remote sensing applications. Appl. Sci. 12, 8654. https://doi.org/10.3390/app12178654, 2022.

Zhao, Y., Zhang, L., Lei, S., Liao, L., Zhang, C., 2025. Machine learning-based prediction of belowground biomass from aboveground biomass and soil properties. Environ. Model. Software 185, 106313. https://doi.org/10.1016/j.envsoft.2024.106313.

Zvoleff, A., 2020. Package 'glcm'. In: Calculate Textures from Grey-Level Co-ccurence Matrices (Glcms). Available online: https://cran.r-project.org/web/packages/glcm/index.html. (Accessed 30 November 2024).