Title: Carbon bias from grassy tree misclassification: revealing forest structural heterogeneity by integrating GEDI and Sentinel-2

Author:

Aiyu Zheng[1]✉(az3506@nyu.edu, ORCID: 0000-0002-1162-6944)

Yi Yin[1] (yi.yin@nyu.edu, ORCID: 0000-0003-4750-4997)

Mingzhen Lu[1] (ml9120@nyu.edu, ORCID: 0000-0002-8707-8745)

Author Institute:

[1]Department of Environmental Studies, New York University, 79 Washington Square East Floors 5-6, New York, NY 10003, USA

Abstract

Tropical forests store roughly half of terrestrial carbon, yet carbon estimates in regenerating and disturbed forests remain highly uncertain. A major source of bias is the prevalence of fast-growing, canopy-forming monocots—such as bamboo, palms, and bananas—that are often misclassified as trees. These "grassy trees" achieve canopy dominance but lack secondary growth, violating woody allometries used in most biomass models. Although NASA's GEDI mission has transformed large-scale biomass mapping with spaceborne LiDAR, its products rely on coarse plant functional types (PFTs), causing grassy-tree-dominated canopies to be absorbed into evergreen broadleaf tree (EBT) classes. Using a texture-based Sentinel-2 classifier, we isolated bamboo-dominated forests within GEDI EBT products in Xishuangbanna, China. GEDI observations show that bamboo canopies are structurally distinct from tree-dominated forests and lead to systematic carbon overestimation of 20–44 Mg C ha^{-1} relative to empirical benchmarks. Our framework improves carbon accounting in structurally heterogeneous forests while remaining adaptable for place-based management.

Keywords: Spaceborne LiDAR, Sentinel-2, Sub-PFT stratification, Carbon Uncertainty, Grassy Trees, Bamboo forests, Forest Monitoring

1. Introduction

Tropical forests are a crucial component of the global carbon cycle, storing approximately 50% of terrestrial carbon. While old-growth forests hold large carbon stocks and biodiversity, regenerating and frequently disturbed forests have expanded from 16.2% to 28.9% of total tropical forest area over the past 30 years, sequestrating carbon at rates five times higher than intact forests (Pan et al. 2024). However, their carbon dynamics remain highly uncertain (±20%) despite their increasing importance in land carbon sinks. This uncertainty is amplified in regrowing forests because disturbance and recovery generate fine-scale structural and compositional heterogeneity that challenges mapping and biomass inference.

A prominent contributor to this heterogeneity is the prevalence of canopy-forming monocots—large monocots such as bamboo, palms, and bananas—whose canopy architecture and biomass allocation systematically depart from the tree-dominated canopies that plant functional type (PFT) labels like “evergreen broadleaf trees” implicitly represent. These “grassy trees” have reached tree-like canopy dominance but are constrained by the lack of secondary growth like grasses with densely packed herbaceous tissues, often have hollow stems and lower aboveground biomass density than typical trees (Avalos et al. 2022; Zheng and Lu 2025). They can dominate disturbed tropical canopies (e.g., bamboo covering up to 21.3% of southwestern Amazonian forests; (de Carvalho et al. 2013) and may sequester carbon at rates several times higher than typical regrowth forests (Heinrich et al. 2021; Nath et al. 2015). Differences in above- and belowground biomass of grassy trees (Yuen et al. 2017; Navarro et al. 2008; Kotowska 2015; Song et al. 2017) further complicate tree-based biomass estimations where they are abundant.

Recent advancements in high-resolution remote sensing offer a path forward in addressing these challenges arising from complex canopy structures associated with grassy tree abundance in disturbed and fragmented landscapes. Sentinel-2 from the European Space Agency's Copernicus programme (10-m; global since 2015; (Drusch et al. 2012)), coupled with texture-based analysis techniques, provides new opportunities to identify forest areas dominated by bamboo and palms. Unlike traditional spectral classification, texture-based approaches such as the Gray-Level Co-occurrence Matrix (GLCM) capture canopy structural patterns that distinguish grassy tree canopies from those formed by broadleaf woody trees (Humeau-Heurtier 2019). For example, sub-meter-resolution Maxar imagery shows that grassy trees such as bamboo form elongated or patchy canopy patterns, in contrast to the rounded crowns of broadleaf woody trees (see Fig.1 and Fig.A.1).

These canopy texture cues provide the horizontal context that complements vertical insights from LiDAR missions such as NASA's Global Ecosystem Dynamics Investigation (GEDI), which delivers global products for canopy height (L2A), canopy cover and vertical structure (L2B), and aboveground biomass density (L4A) (Dubayah et al. 2022). These structural insights are crucial to address two primary barriers that have prevented the integration of grassy trees into forest biomass assessment in previous research. First, forest inventories have historically overlooked these groups due to ambiguous growth forms, leading to underrepresentation in ground-truth data (Fadrique et al. 2020). Second, remote sensing has struggled to accurately identify these functionally unique species within forest canopies. Grassy trees share spectral properties similar to those of broadleaf trees (Li et al. 2019; Yusof et al. 2021), while lower-resolution sensors such as Landsat (30 m) and MODIS (250 m) lack the spatial detail needed to detect them (Qi et al. 2022; Venkatappa et al. 2020; Jia et al. 2020). This

spectral and spatial ambiguity has compounded existing ground-based biases, reinforcing uncertainties in forest carbon estimates.

However, GEDI LiDAR products have limitations because they stratify global forests into five PFTs—evergreen broadleaf trees (EBT), deciduous broadleaf trees, evergreen needleleaf trees, deciduous needleleaf trees, and a combined grass–shrub–woodland class—informed by MODIS land cover inputs (Kellner et al. 2023). While suitable for global-scale modeling, this coarse classification can obscure critical structural differences among vegetation subtypes, especially in regions with high functional heterogeneity. Applying a single PFT label (e.g., EBT) across diverse canopy structures—such as tree-dominated, palm-rich, or bamboo-dominated forests—can introduce systematic biases in GEDI-based biomass estimates (Bruening et al. 2023; Lin et al. 2025). These issues are particularly acute in tropical Asia and Africa, where calibration data are sparse, and canopy forms often depart from the tree-based assumptions embedded in GEDI biomass models (Dubayah et al. 2022).

To address these limitations, we combine GEDI with complementary datasets to improve grassy tree stratification and quantify carbon-estimation bias from mis-labeling grassy trees as EBT. While GEDI provides high-fidelity vertical structure metrics, it lacks the horizontal context needed to disaggregate compositionally diverse forest types. Sentinel-2, despite its limited sensitivity to vertical structure, adds valuable context through its consistent spatial resolution and spectral–textural richness. When paired with texture-based approaches, it can reveal patterns that help distinguish vegetation types otherwise lumped under the same PFT (Humeau-Heurtier 2019; Descals 2023).

We conceptualize this integration in Fig.1, where Sentinel-derived compositional context helps subdivide GEDI's broad PFT class, EBT, into grassy trees (bamboo) and woody trees,

identifying areas where structural and biomass deviations are likely and thus require calibration refinement. We then build a specific sub-PFT (e.g., bamboo) classifier for mapping that is designed to generalize. Together, Fig.1 presents the core workflow of this study—combining structure and composition via texture-enhanced satellite imagery and GEDI LiDAR—to stratify heterogeneous forests more accurately and quantify functional misclassification biases in carbon accounting.

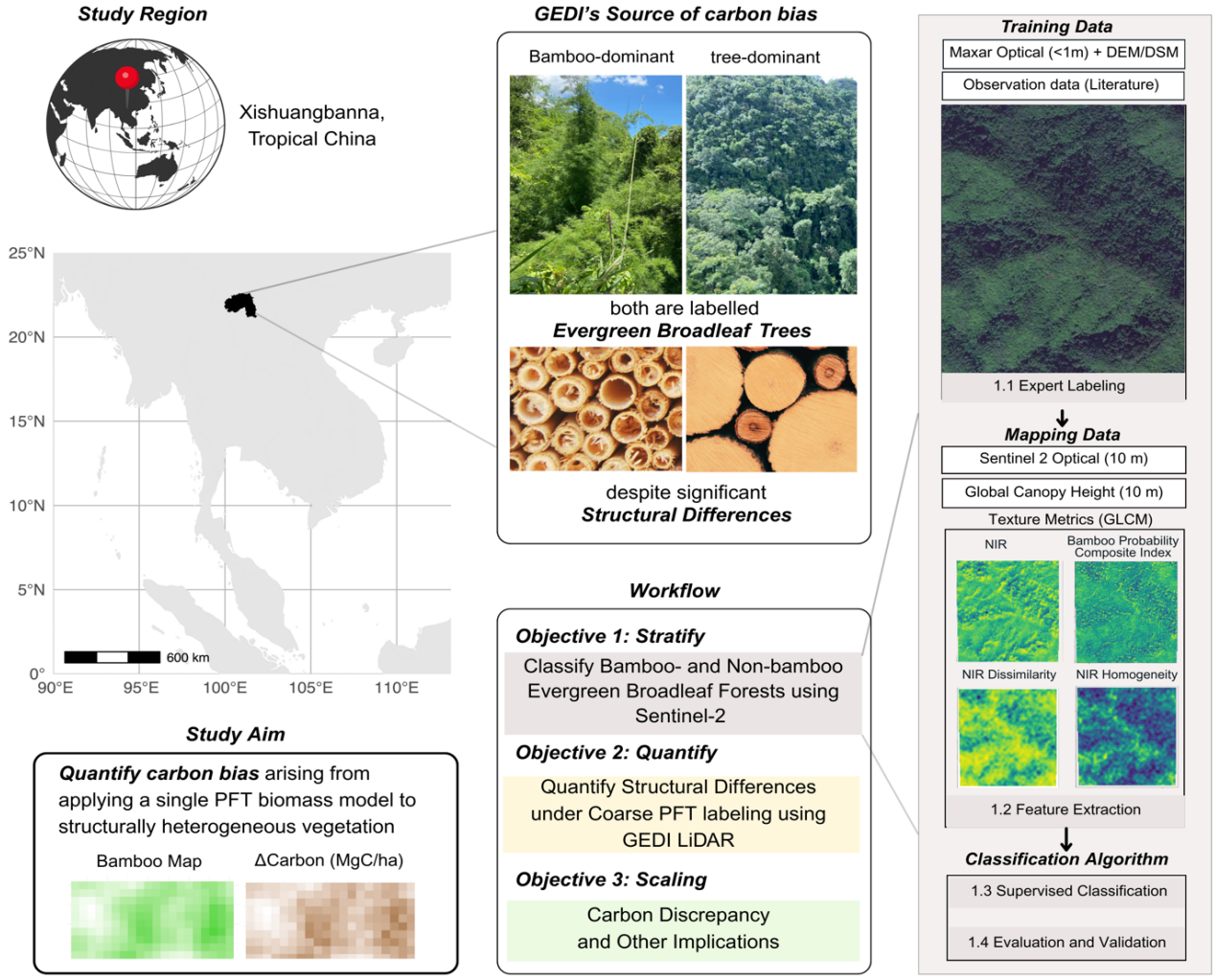


Fig 1. Conceptual workflow of this study, combining texture cues from Sentinel-2 with vertical structure from a global canopy height model (Lang et al. 2022) and GEDI LiDAR metrics to quantify carbon bias in functionally heterogeneous evergreen broadleaf tree (EBT) forests with high bamboo abundance in Xishuangbanna.

In this study, we focus on bamboo-dominated forests in Xishuangbanna, southwestern China—a region with high bamboo richness, natural abundance, and cultivation, yet currently subsumed under GEDI's evergreen broadleaf tree (EBT) class. We use this region to demonstrate how combining texture-enhanced Sentinel-2 classification with GEDI structural metrics refines carbon assessments and maps structurally distinct sub-PFT canopies. While this study centers on bamboo, the workflow is designed to be extended to other grassy-tree-dominated systems such as palm forests, as well as other structurally heterogeneous forests, like those with high liana abundance. Specifically, our objectives are outlined in Fig.1:

1. **Stratify** bamboo-dominated forests from tree-dominated EBT forests and other land-use types using a texture-enhanced Sentinel-2 classification.
2. **Quantify** structural and biomass differences in GEDI metrics between classified bamboo and non-bamboo EBT forests to evaluate bias from GEDI's tree-centric EBT biomass model.
3. **Scale** aboveground biomass density to aboveground carbon density and map the spatial pattern of carbon bias in forests containing sub-PFT bamboo canopies.

2. Methods

2.1 Study Region

Xishuangbanna, on the northern edge of tropical Asia, hosts true tropical rainforests despite its monsoonal climate (Zhu et al. 2006). A biodiversity hotspot with over 5,000 higher plant species in just 19,690 km², the region has experienced extensive land-use change over recent decades (Li et al. 2008). Forest cover has declined from ~60% in the 1950s due to the expansion of rubber plantations, cash crops, and shifting cultivation (Zhang and Cao 1995; Senf et al. 2013)(Zhang and Cao 1995); (Zhang and Cao 1995; Senf et al. 2013), giving rise to widespread secondary

vegetation—especially bamboo forests (*Dendrocalamus membranaceus*, *Cephalostachyum pergracile*, etc., Yang et al. 2008)—now common along forest edges and within fragmented landscapes (bamboo in Fig.1).

Xishuangbanna presents an ideal setting to develop and evaluate our method, given its highly heterogeneous forest matrix, strong bamboo presence, and the cultural significance of bamboo among the local Dai population (Wang et al. 2002). We focus on this region to assess how the broad classification of bamboo as evergreen broadleaf trees in GEDI products can affect biomass estimation and carbon accounting.

2.2 Remote Sensing and Analytical Workflow

To accomplish the three objectives of our study, we designed an integrated remote sensing and analytical workflow that includes training data compilation, texture analysis, classification and accuracy assessment, evaluation of GEDI's limitations, and the quantification of carbon discrepancies (Fig.1). Below, we describe each component of this workflow in detail.

2.2.1 Datasets and Preprocessing

We aligned GEDI footprints to the 10-m Sentinel-2 grid by assigning each footprint to the 10-m pixel containing its geolocation (footprint center). Land-cover labels were extracted at that pixel and linked to footprint-level GEDI metrics for subsequent analyses.

Sentinel-2 (Optical, 10 m resolution): We obtained Sentinel-2 imagery from Google Earth Engine (COPERNICUS/S2_harmonized) and generated four seasonal composites (January–March, April–June, July–September, and October–December) for the year 2021, selecting scenes with less than 5% cloud cover. Each composite included nine spectral bands (B2–B8, B11, B12) and was processed using QA60 and s2cloudless cloud masks to reduce atmospheric noise.

149 ***The Global Canopy Height Model*** (10 m, Lang et al. 2022): a probabilistic deep learning model
150 that fuses sparse height data from GEDI LiDAR with dense optical satellite images from
151 Sentinel-2.
152 ***GEDI (LiDAR, 25 m resolution):*** GEDI datasets provided canopy height and relative height
153 metrics (Level 2A), canopy cover and vertical structure (Level 2B), and aboveground biomass
154 density (Level 4A). All GEDI footprints (25-m resolution) were aligned to 10-meter Sentinel-2
155 composites for value extraction and analysis.

156 **2.2.2 Training data**

157 We compiled a training dataset encompassing diverse vegetation and land-use categories,
158 including bamboo, tree-dominated evergreen forests, rubber plantations, fruit and tea plantations,
159 farms, construction areas, and water bodies. Labels were manually interpreted from 2021 Google
160 Earth imagery (licensed Maxar submeter imagery), chosen for its minimal cloud cover and
161 alignment with GEDI and Sentinel-2 temporal windows.
162 We reliably identified bamboo based on its distinct texture and canopy structure (Fig.1),
163 supported by expert knowledge of its landscape distribution—such as along riparian corridors,
164 within rubber plantations, and near forest edges—and by regional vegetation accounts (Zhu
165 2006; Zhu et al. 2015) describing *Dendrocalamus membranaceus* forests. We identified other
166 vegetation types using a combination of phenological signals (e.g., seasonal reflectance in rice
167 paddies), spatial context (e.g., farms near villages), auxiliary map labels (e.g., "Xishuangbanna
168 Tropical Botanical Garden"), and literature-based spatial descriptions (Min et al. 2019; Li et al.
169 2008), drawing from multiple lines of evidence.
170 In Appendix A, we summarized the training data's sample size and geographic coverage
171 in Table A.1 and Fig A.1 and detailed the visual interpretation criteria used for each land-use

type in Table A.2. Fig A.2 shows the strong spectral separability of the October-September composite for Xishuangbanna in 2021.

2.2.3 Feature Extraction

For each labeled polygon, we extracted a comprehensive suite of features that capture spectral, textural, and contextual information (Fig. A.3). *Spectral features* include surface reflectance from selected Sentinel-2 bands (e.g., red, red-edge, SWIR), along with vegetation indices that characterize greenness, moisture, and phenology: Normalized Difference Vegetation Index (NDVI = (NIR – Red)/(NIR+ Red)), and the Bamboo Phenological Characteristic Index (BPCI = Red Edge 1/NIR-NIR/SWIR 1) (Huang et al. 2024). These features were computed from the seasonal composite imagery described earlier.

2.2.4 Texture Metrics

We applied Gray-Level Co-occurrence Matrix (GLCM) techniques to characterize the spatial distribution of pixel intensities, capturing distinct texture patterns associated with bamboo's unique leaf morphology and clumping architecture. Metrics such as pixel dissimilarity, homogeneity, and entropy have proven useful in distinguishing different species within mixed canopies (Mohammadpour et al. 2022). To identify the most informative spectral regions, we examined how texture metrics varied across different vegetation types—especially between bamboo-, tree-dominated forests, and rubber plantations. This analysis reduces dimensional redundancy and minimizes overfitting risks by selecting texture features (Band 8 dissimilarity and B8 homogeneity computed using a 5×5-pixel sliding window) that are less correlated yet discriminative.

2.2.5 Classification algorithm and accuracy assessment

We trained and evaluated two machine learning classifiers: Random Forest (RF) and Convolutional Neural Networks (CNN) (detailed in Appendix B). RF is robust to noise and well-suited for integrating heterogeneous feature sets, while CNN enables pixel-based classification with spatial context sensitivity. To compare their performance, we withheld a subset of ground-truth data and applied both classifiers independently.

To assess classification accuracy, we conducted visual cross-validation by randomly sampling 100 points within each land-use class and verifying their true identity in Google Earth Pro (v7.3.6, Google Earth 2025) using high-resolution imagery from 2020 to 2022. Confusion matrices summarizing the results of this photo-interpretation-based validation are presented in Table C.1. We report standard accuracy metrics for the better-performing RF classifier: overall accuracy (the proportion of correctly classified cases), class-wise sensitivity (recall; true positive rate) and specificity (true negative rate), and the F1 score (which penalizes both missed detections and false positives).

2.2.6 GEDI overlay

To evaluate the limitations of GEDI-derived products in distinguishing bamboo-dominated forests, we overlaid GEDI Level 2A (canopy height), Level 2B (canopy cover and vertical structure), and Level 4A (aboveground biomass) with our classified land cover map (Figure C.1). For each GEDI footprint intersecting a bamboo-classified pixel, we extracted its associated canopy height, structural, and biomass metrics. For comparison, the same set of variables was extracted from footprints located within tree-dominated forests and other land-use classes.

To account for classification uncertainty—particularly confusion between bamboo and tree categories—we treated the validation-derived misclassification rates as an observation model (confusion matrix; Table C.1) and applied a Bayesian measurement-error correction to

infer the posterior true area of each land-cover class (detailed in Appendix C). This yields classification-error-adjusted estimates of the total extent of bamboo-dominated forests and other vegetation types across Xishuangbanna.

At the footprint level, we propagated this classification uncertainty by assigning each GEDI footprint a posterior probability of belonging to bamboo versus tree conditional on its mapped label, and then reweighted footprint-level summaries accordingly. This probabilistic reweighting enables more robust estimation of biomass distributions for bamboo- and tree-dominated forests and quantifies the extent to which biomass estimates are biased when bamboo-dominated canopies are treated as evergreen broadleaf forests under the default GEDI L4A EBT biomass model.

2.2.7 Aboveground biomass bias from GEDI EBT model

GEDI Level 4A aboveground biomass density (AGBD, Mg ha^{-1}) estimates in Xishuangbanna are derived from a global EBT biomass model parameterized on GEDI relative height (RH) metrics (Kellner et al. 2023):

$$AGBD = 1.113 \times \left(-104.965 + 6.802\sqrt{RH50 + 100} + 3.995 \times \sqrt{RH98 + 100}\right)^2 (Eq.\ 1)$$

RH metrics summarize the vertical structure of a LiDAR footprint as percentiles of the cumulative GEDI waveform energy returned from height z. RH98 approximates the upper canopy envelope (near-maximum vegetation height), whereas RH50 represents the median height of returned vegetation material, capturing the vertical distribution of canopy structure.

However, this EBT model is applied uniformly across evergreen broadleaf forests, including bamboo-dominated stands that differ fundamentally from trees in growth form and vertical structure. To test whether this unstratified application induces systematic overestimation and to quantify its effect size, we (1) compared GEDI Level 2A/2B structural metrics between

bamboo- and tree-dominated areas classified by our random-forest land-cover map, and (2) used empirical bamboo biomass benchmarks from the same ecoregion (Yuen et al. 2017) as external references for bias quantification (see Table D.1). Given strong management and land-use contrasts, we further separated bamboo into forest versus fallow contexts to obtain context-specific benchmarks and bias estimates.

2.3 GEDI structural comparison

We extracted GEDI Level 2A and 2B metrics from footprints whose surrounding area was classified as bamboo or non-bamboo EBT forests (trees). We compared their canopy height (RH98), relative height profile, plant area index (PAI), and plant area volume index (PAVD) throughout the canopy, to reveal structural differences that are masked when all these forests are labelled as a single EBT class in GEDI-based products. These structural contrasts motivate subsequent bias analysis.

Empirical benchmarks for bamboo aboveground biomass and carbon density

Established methods for estimating individual bamboo biomass often are typically based on diameter and height of clumpy tropical bamboos (Xayalath et al. 2019; Camargo García et al. 2023; Huy et al. 2019). Unlike trees, bamboo culms do not exhibit continuous secondary growth, and stand-level biomass depends strongly on age structure and culm turnover rather than height alone (Yen 2016). These demographic properties cannot be retrieved from GEDI waveform metrics. Our objective is therefore not to recalibrate bamboo allometry, but to evaluate whether a tree-based RH model produces systematic overestimation on bamboo and to quantify its effect size relative to empirically observed ranges.

Accordingly, we used empirical aboveground carbon density estimates synthesized by (Yuen et al. 2017) for tropical bamboos in Indochina (Myanmar, Laos, Thailand, Vietnam, and

India), a region biogeographically adjacent to Xishuangbanna and characterized by comparable species composition. Because Xishuangbanna and Indochina share a contiguous Indo-Burma floristic context with overlapping bamboo taxa and broadly similar monsoonal tropical–subtropical climates (Zhu et al. 2006), we expect these synthesized carbon-density benchmarks to be broadly transferable to bamboo-dominated stands in our study region.

Reported summary statistics (mean, standard deviation, and max) were used to derive confidence intervals for comparison with GEDI L4A footprint-level estimates (detailed in Appendix D). To reflect agroforestry mosaics abundant with secondary forests in Xishuangbanna, bamboo pixels were classified as forest or fallow based on a 3×3 neighborhood rule: bamboo cells surrounded by ≥ 5 bamboo or tree neighbors were classified as forest bamboo; otherwise as fallow bamboo.

For bamboo footprints, we compared GEDI L4A AGBD estimates with empirical bamboo benchmarks stratified by bamboo context $i \in \{forest, fallow\}$, and quantified aboveground carbon bias as:

$$\Delta C_i = AGC_{EBT,i} - AGC_{empirical,i} \ (Eq.\ 2)$$

where aboveground carbon density (AGC) was derived by applying a constant carbon fraction (CF = 0.5) to AGBD for both bamboo and trees. Although CF varies among species and tissues, it has been inconsistently treated across the bamboo literature, whereas CF = 0.5 is widely used as a default in forest carbon accounting (Pan et al. 2011). We therefore adopt a single CF to ensure comparability between GEDI-derived AGBD and empirical benchmarks, noting that alternative defaults (e.g., CF = 0.47; IPCC 2006) would linearly rescale ΔC without affecting its sign. Positive ΔC values indicate overestimation of bamboo carbon by GEDI's EBT model relative to empirical benchmarks.

Spatial extrapolation of carbon bias across Xishuangbanna

To characterize the spatial distribution of aboveground carbon bias (ΔC) across bamboo-rich landscapes in Xishuangbanna, footprint-level estimates must be extrapolated to continuous pixel-based surfaces. To ensure internal consistency with GEDI's operational biomass products, this extrapolation was conducted entirely within the framework of the GEDI EBT biomass model (Eq. 1). Because the EBT model requires both RH50 and RH98 as inputs, we first quantified the statistical relationship between RH50 and RH98 separately for bamboo and tree footprints using GEDI Level 2A data. We applied the class-specific RH50–RH98 relationships to the 10-m global canopy height model (Lang et al. 2022) as an operational approximation to enable internally consistent regional mapping within our framework. Because the Lang et al. (2022) product is generated by fusing sparse GEDI Level 2 RH98 observations with dense Sentinel-2 optical data, it provides a GEDI-anchored proxy for canopy stature; nevertheless, it is not a one-to-one substitute for footprint-level RH98. Any residual mismatch between this 10-m canopy height layer and GEDI RH98 over overlapping footprints in Xishuangbanna is therefore treated as a limitation of our extrapolation. The estimated RH50 and RH98 were then used to compute pixel-level AGBD and aboveground carbon density using the EBT biomass model (Eq. 1).

The analytical form of the EBT biomass model (Eq. 1) does not capture all calibration steps and ancillary corrections used to generate official GEDI Level 4A biomass estimates. To quantify the systematic error introduced by applying this simplified RH-based formulation alone, we compared AGBD estimates derived from Eq. 1—using RH metrics from GEDI Level 2 data—with published Level 4A AGBD values for footprints where Level 2 and Level 4 observations overlapped.

Finally, we applied Eq. 1 across the entire Xishuangbanna region and converted AGBD to aboveground carbon density. To illustrate the spatial magnitude of overestimation, carbon bias was expressed as a percentage of EBT-estimated carbon density and mapped separately for forest and fallow bamboo areas:

$$\%Bias = \frac{\Delta C_i}{AGC_{EBT,i}} \times 100 \ (Eq. 3)$$

This multi-step extrapolation introduces three main sources of uncertainty: (1) misclassification errors from the random-forest land-cover map, (2) uncertainty in the class-specific RH50–RH98 regression models, and (3) systematic error arising from applying the simplified EBT biomass formulation (Eq. 1). These uncertainties were explicitly evaluated and propagated in subsequent analyses (see Results).

3. Results

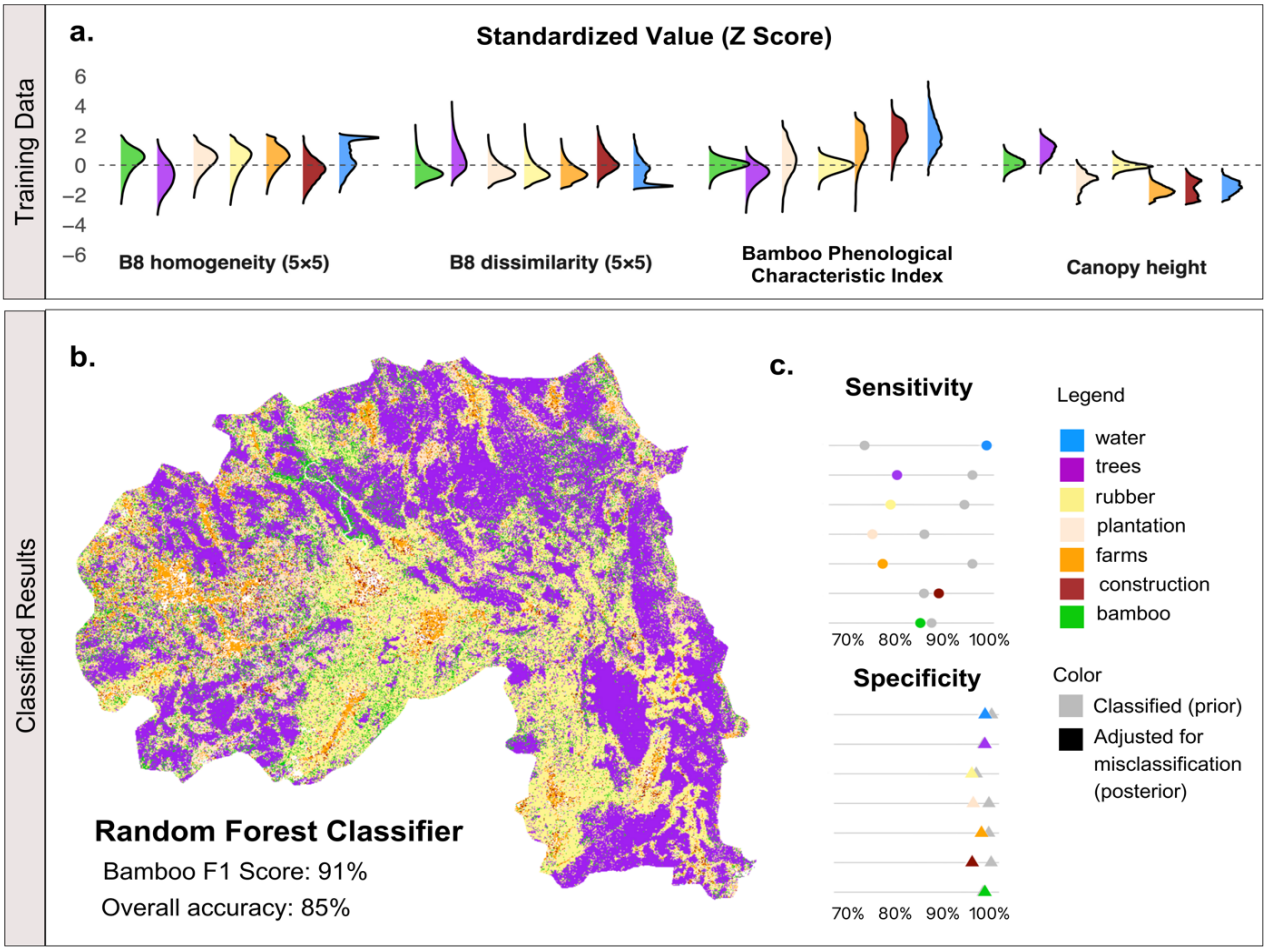
3.1 Reliable sub-PFT land-cover classification enables robust bamboo detection

The random forest classifier achieved strong overall performance in distinguishing bamboo from other land-cover classes across Xishuangbanna (Fig. 2). Feature distributions derived from spectral texture (Band 8—Near Infrared's homogeneity and dissimilarity), canopy height, and the Bamboo Phenological Characteristic Index (BPCI) show clear separation among vegetation classes in the training data (Fig. 2a), indicating that these features capture complementary structural and phenological information relevant for bamboo detection.

The resulting land-cover map reveals spatially coherent patterns of bamboo, trees, rubber plantations, and agricultural land consistent with known landscape organization in the region (Fig. 2b). For instance, bamboo-dominated riparian belts are apparent along the Lancang River corridor, consistent with reported patterns of river-associated bamboo stands in the region. Beyond riparian zones, bamboo also occurs as patchy elements embedded within rubber

plantations and farmland mosaics, consistent with local smallholder management practices. These spatial patterns are further supported by high-resolution aerial maps (Fig A.1).

Classification performance was strong despite the highly fragmented tropical landscape, with an overall accuracy of 85% and a bamboo F1 score of 91% (Fig. 2b), indicating reliable bamboo detection with low omission. Nonetheless, confusion among spectrally and structurally similar vegetation classes—most notably bamboo, trees, and rubber plantations—remained non-negligible, consistent with frequent co-occurrence and fine-scale mixing of these vegetation types in agroforestry mosaics. Class-specific sensitivity and specificity (Fig. 2c) further indicate that residual misclassification persists even under a well-performing classifier. We therefore explicitly propagated this classification uncertainty in subsequent analyses using a Bayesian posterior resampling framework and Monte Carlo reassignment of class labels, rather than treating the classified map as error-free. The posterior updating methods and results are detailed in Appendix C.



344 **Fig 2. Sub-PFT mapping reveals hidden vegetation heterogeneity in Xishuangbanna. a**)
345 Standardized distributions (Z-scores) of four key predictors in the training data—Band 8 (near-
346 infrared) homogeneity and dissimilarity (5×5-pixel window), the Bamboo Phenological
347 Characteristic Index, and canopy height—show that the seven mapped classes occupy distinct
348 regions of feature space, indicating that the training data and selected predictors provide strong
349 discriminatory power for subsequent classification. **b**) Using Sentinel-2 spectral indices and
350 GLCM texture features, we classified sub-PFT vegetation and land-cover types across
351 Xishuangbanna with a Random Forest model. The resulting 10-m map reveals fine-scale mosaics
352 of bamboo patches embedded within forest, rubber, and agricultural matrices—heterogeneity
353 typically masked by the EBT forest label in GEDI-based products. **c**) Class-wise sensitivity

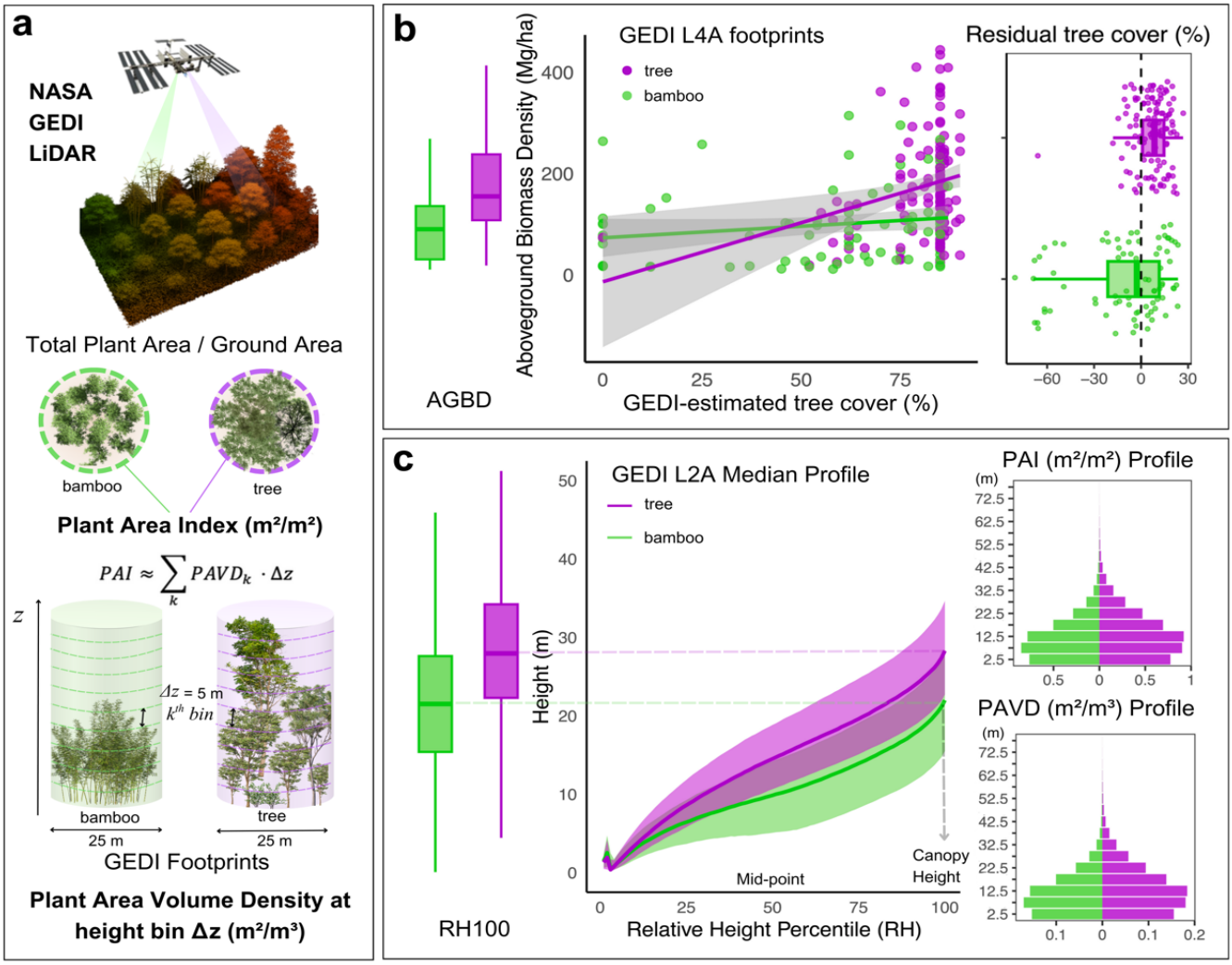
(circles) and specificity (triangles) estimated from 70/30 hold-out validation during training (grey symbols) and from visually interpreted sample points drawn from the classified map (colored symbols), which we used to adjust for misclassification. Bamboo achieved an F1 score of 0.91 (overall accuracy 0.85), enabling reliable downstream quantification of bamboo cover and configuration.

3.2 GEDI reveals systematic structural differences between bamboo- and tree-dominated forests

Building on the validated land-cover classification and explicitly quantified uncertainty, we overlaid the classification map with GEDI LiDAR footprints to test whether bamboo- and tree-dominated forests exhibit systematic differences in canopy structure and biomass-related metrics as directly observed by GEDI.

To aid interpretation of subsequent results, Fig. 3a illustrates how GEDI LiDAR footprint-level observations characterize canopy structure. GEDI footprints sample forest structure within circular columns approximately 25 m in diameter, capturing the vertical distribution of vegetation from the ground to the top of the canopy. Within each footprint, the returned waveform resolves a vertical profile of vegetation material, allowing both canopy height and the allocation of plant material along the vertical axis to be directly observed.

As shown in Fig. 3a, the vertical distribution of plant material is summarized by plant area volume density (PAVD) across successive height bins (Δz), while plant area index (PAI) reflects the cumulative plant area per unit ground area within the footprint. Conceptually, PAVD describes how plant material is distributed with height, whereas PAI indicates how plant biomass accumulates vertically, analogous to leaf area index but incorporating both foliage and woody components detected by the LiDAR waveform. Together, these GEDI-derived quantities provide an intuitive framework for interpreting the structural contrasts between bamboo- and tree-dominated forests shown in Fig. 3b–c and Table 1.



379 **Fig 3. Structural differences between bamboo- and tree-dominated forests from GEDI. (a)**
380 Conceptual illustration of how GEDI waveforms capture vertical canopy structure within each
381 ~25 m footprint, from which canopy height, plant area index (PAI), and plant area volume
382 density (PAVD) profiles are derived (see Wang et al. 2025). **(b)** Bamboo-dominated footprints
383 have lower AGBD than tree-dominated footprints. Residual tree cover is defined as GEDI-
384 reported tree cover minus the tree-cover value expected for a tree-dominated footprint with the
385 same AGBD; values are consistently negative for bamboo, indicating its cover–biomass
386 decoupling relative to tree-dominated forests. **(c)** Vertical structure differs strongly by vegetation
387 type: bamboo has lower canopy heights (RH98) and a lower upper-canopy profile than trees,

388 with broadly overlapping lower–mid canopy layers. PAI and PAVD profiles are consistent with
389 these canopy-structure contrasts.

Panel A. Raw GEDI observations (classified bamboo vs trees)

Metric	n (Bamboo)	Median (Bamboo)	n (Trees)	Median (Trees)	Δ Median (Trees − Bamboo)
Aboveground Biomass Density (Mg ha^{-1})	81	90.29	121	154.41	64.12
Canopy Height (m)	31,067	21.87	123,142	28.39	6.52
Plant Area Index (m^2 m^{-2})	43,538	3.28	171,085	4.08	0.81
Total Vegetation Volume (m^3 m^{-2})	43,508	0.74	171,056	0.90	0.15

Panel B. Misclassification-aware contrasts (posterior Monte Carlo; n = 1000 iterations)

Metric	Δ Median (MC, 2.5–97.5%)	MC Median	MC $p < 0.05$
Aboveground Biomass Density (Mg ha^{-1})	44.67–66.46	54.37	1.00
Canopy Height (m)	5.19–5.32	5.26	1.00
Plant Area Index (m^2 m^{-2})	0.64–0.67	0.65	1.00
Total Vegetation Volume (m^3 m^{-2})	0.12–0.13	0.12	1.00

390 **Table 1. Summary of structural and biomass differences between bamboo- and tree-**
391 **dominated evergreen broadleaf forests as observed by GEDI LiDAR**. Raw footprint-level
392 medians (Panel A) are contrasted with misclassification-aware median differences obtained from
393 posterior Monte Carlo resampling (Panel B; 1,000 iterations). All differences are expressed as
394 Trees − Bamboo, and the fraction of Monte Carlo iterations with $p < 0.05$ quantifies the
395 robustness of each contrast to classification uncertainty.

Despite being processed using the same EBT biomass model, bamboo- and tree-dominated GEDI footprints exhibited markedly different Level 4A aboveground biomass density patterns (Fig. 3b). We use residual tree cover as a diagnostic of structural inconsistency: for each footprint, we calculate GEDI-reported tree cover minus the tree cover expected at the same AGBD from the unstratified AGBD–tree-cover relationship (i.e., treating all EBT as structurally equivalent). In tree-dominated footprints, AGBD increased strongly with GEDI-estimated tree cover, which was predominantly high (> 50%), consistent with the model's underlying assumption that canopy cover and vertical extent scale with woody biomass. In contrast, bamboo-dominated footprints showed substantially lower AGBD (median 42% lower) and a weak cover-biomass coupling: similar AGBD values occurred across both low (< 25%) and high (> 50%) tree cover. Accordingly, residual tree cover was predominantly negative for bamboo footprints and positive for tree footprints, indicating a systematic decoupling between cover and biomass in bamboo that arises from GEDI observations alone, prior to any comparison with external biomass benchmarks. Altogether, GEDI tree cover is not an equivalent proxy for biomass in bamboo-dominated EBT forests as it is in tree-dominated EBT forests.

GEDI Level 2A metrics further indicate systematic differences in vertical canopy structure between bamboo- and tree-dominated forests (Fig. 3c; Table 1). Bamboo footprints exhibited lower canopy height, with median canopy height ~23% lower than tree footprints. Relative height (RH) profiles show that bamboo canopies concentrate vegetation material at lower heights, resulting in a more compressed vertical distribution, whereas tree canopies exhibit greater vertical extension, and more plant material allocated to upper canopy layers. Consistent with these patterns, bamboo footprints showed lower total PAI and total vegetation volume, and distinct PAVD profiles across height bins. Together, these GEDI-observed structural contrasts

provide a mechanistic explanation for the divergent biomass patterns in Fig. 3b and underscore that tree-based structural predictors embedded in the EBT model do not adequately represent bamboo-dominated forests.

To assess the sensitivity of these structural contrasts and biomass differences to land-cover misclassification, we propagated classification uncertainty using posterior resampling and Monte Carlo simulations. Across all four GEDI-derived metrics, median differences between tree- and bamboo-dominated footprints remained consistently positive and statistically significant in all Monte Carlo iterations (proportion of iterations with $p < 0.05 = 1.0$; Table 2). Accounting for misclassification reduced the magnitude of the estimated differences relative to naïve estimates—for example, the median tree–bamboo difference in aboveground biomass density decreased from 64.1 Mg ha^{-1} to 54.4 Mg ha^{-1}—but did not alter their direction or overall effect size. Similar patterns were observed for canopy height, plant area index, and total vegetation volume (Fig C.2). These results indicate that the observed structural differences are robust to plausible classification errors and are not driven by misclassification alone.

3.3 Systematic carbon overestimation in bamboo forests by GEDI EBT models

Building on the GEDI-observed structural differences between bamboo- and tree-dominated forests (Fig. 3), we next quantified the magnitude of biomass and carbon bias arising from applying the EBT model to bamboo-dominated systems using empirical benchmarks. Bamboo-dominated footprints were stratified into forest and fallow contexts based on local neighborhood composition, reflecting the heterogeneous agroforestry landscape of Xishuangbanna (Fig. 4a).

Panel A. Class-specific RH50–RH98 regressions

Class	Model	β0	β1	n	R^2	RMSE (m)
Trees	RH50 = β0 + β1·RH98	-2.316	0.628	123,142	0.675	3.74
Bamboo	RH50 = β0 + β1·RH98	-3.872	0.649	31,067	0.745	3.35

Panel B. Discrepancy between RH-only EBT re-estimation and published GEDI L4A carbon at overlapping footprints

Residual (Mg C ha^{-1}): $\Delta C_{EBT} = C_{est} - C_{L4A}$; $C_{est} = 0.5 \times AGBD_{EBT}(RH50, RH98)$

Class	n	Mean residual	SD
Bamboo	81	5.31	51.52
Trees	121	3.97	44.49

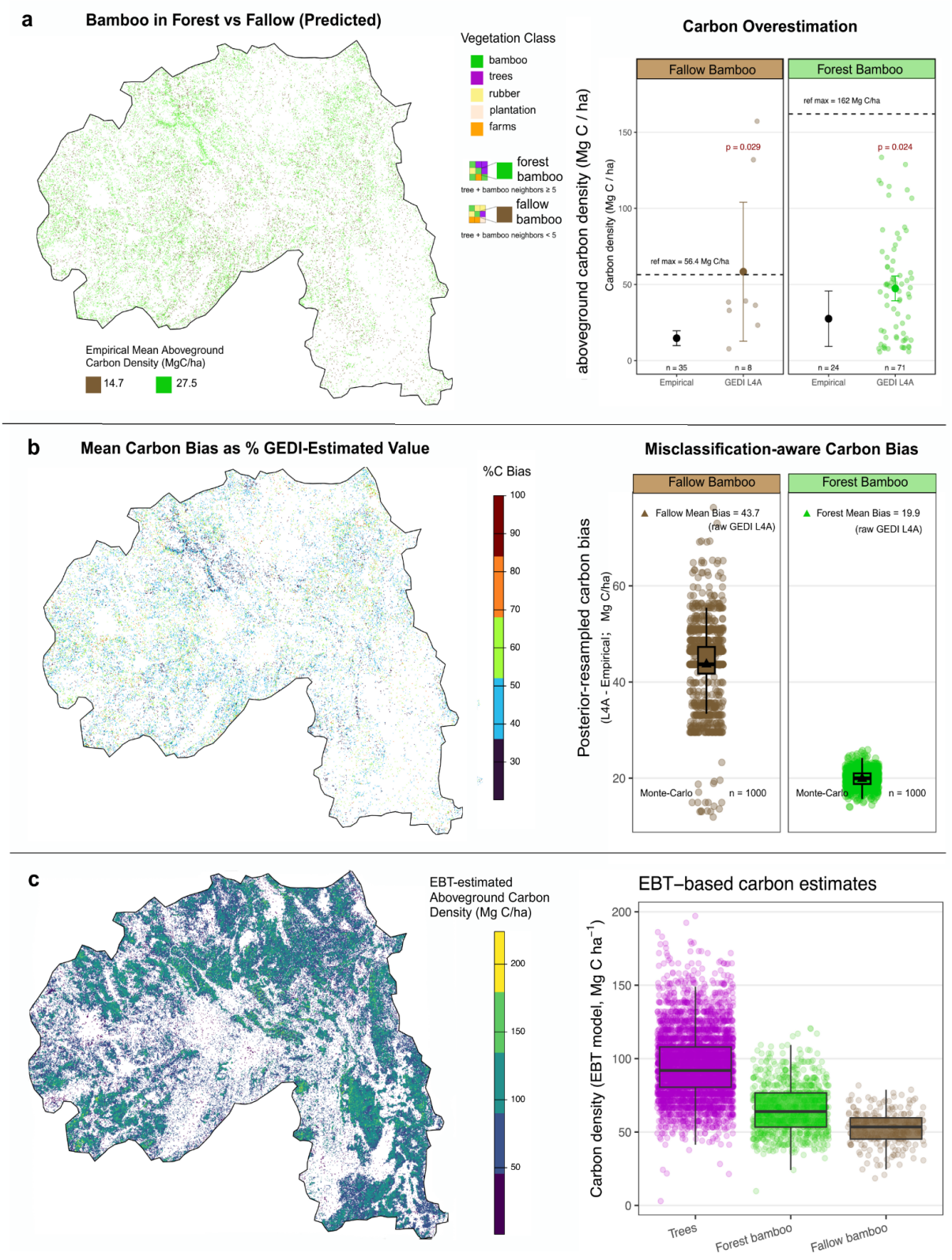
Panel C. Misclassification-aware bamboo carbon bias relative to empirical benchmarks

Naive (no-correction) mean bias (Mg C/ha): Bamboo Fallow: 43.7; Bamboo Forest: 19.9

Bamboo context	Component	Mean bias (Mg C/ha)	Median	2.5%	97.5%
Fallow bamboo	posterior-resampled bamboo bias (MC)	44.1	43.7	29.6	59.0
Forest bamboo	posterior-resampled bamboo bias (MC)	19.8	19.8	16.4	23.3
Fallow bamboo	Total (MC + EBT discrepancy)	49.4	49.0	34.8	64.2
Forest bamboo	Total (MC + EBT discrepancy)	25.1	25.1	21.7	28.6

Table 2. Key models and uncertainty components used to extrapolate bamboo carbon bias across Xishuangbanna. (A) Class-specific RH50–RH98 regressions fitted to GEDI Level 2 footprints (bamboo vs. tree) and used to predict RH50 from a 10-m canopy height layer (treated operationally as RH98) for regional application of the EBT biomass model. (B) Systematic discrepancy between biomass/carbon estimated from the simplified RH-only EBT formulation (Eq. 1) and the published GEDI Level 4A carbon at footprints where Level 2 and Level 4

observations overlap, summarized as residual carbon ($C_{est,i} - C_{L4A,i}$). (C) Misclassification-aware GEDI estimates of bamboo carbon bias relative to empirical benchmarks, obtained by posterior Monte Carlo resampling that propagates uncertainty in whether mapped-bamboo footprints are truly bamboo (conditional on the mapped label) and recomputes GEDI–empirical mean differences for forest- and fallow-bamboo contexts. The "Total" component combines the misclassification-aware GEDI–empirical bias with the EBT-formulation discrepancy in Panel B to produce the bias estimates used for regional mapping. Panel C reports bias estimates conditional on footprints mapped as bamboo (false positives are removed); omission errors (false negatives) are not propagated (see Discussion).



454

Fig 4. Quantifying carbon overestimation when structurally distinct evergreen broadleaf forests are treated with GEDI's single PFT-based EBT biomass model in Xishuangbanna. **(a)** Predicted bamboo pixels were separated into forest versus fallow contexts using a 3×3 neighborhood rule (forest bamboo: ≥ 5 bamboo/tree neighbors; fallow bamboo: < 5). GEDI L4A aboveground carbon density for each bamboo type is compared with regional empirical reference values from Yuen et al. (2017). Error bars show t-based 95% confidence intervals for the group mean carbon density (mean ± t·SD/√n); p-values are from one-sided Welch's t-tests (H_1: GEDI > empirical). A re-derivation of the literature summary statistics is provided in Appendix D. **(b)** Uncertainty arising from bamboo misclassification was propagated using Monte Carlo resampling, in which footprint-level "true" land-cover classes were repeatedly drawn from posterior probabilities derived from the random-forest classifier. These misclassification-aware bias distributions were combined with bias arising from GEDI's simplified EBT biomass formulation based on RH98 and RH50 (Eq.1 applied in GEDI L4A for Xishuangbanna; Kellner et al., 2023) to estimate mean percent carbon overestimation across the region. Fallow bamboo exhibits larger absolute carbon bias. **(c)** Region-wide EBT-based aboveground carbon estimates derived from the simplified EBT RH model. The boxplots summarize the resulting EBT-based carbon densities for pixels classified as trees, forest bamboo, and fallow bamboo. Carbon uncertainties are summarized in Table 2.

Comparisons between GEDI Level 4A estimates and empirical bamboo benchmarks reveal systematic overestimation of aboveground carbon density in both bamboo contexts (Fig. 4a). GEDI-derived carbon densities for fallow bamboo were significantly higher than empirical values (one-sided Welch's *t*-test, H1: GEDI > empirical, $p = 0.029$), and forest bamboo exhibited an even larger positive deviation from its corresponding benchmark ($p = 0.024$). To limit the disproportionate influence of extreme GEDI footprint values on summary statistics, we excluded outliers in GEDI L4A bamboo carbon density (forest vs fallow) using the Tukey rule (values outside 1.5 × IQR) prior to visualization and statistical comparison.

The magnitude of overestimation also differed between bamboo contexts: fallow bamboo showed higher absolute carbon bias (43.7 Mg ha^{-1} versus 19.9 Mg ha^{-1}) than forest bamboo with greater carbon uncertainty, consistent with its lower canopy height and higher landscape complexity observed by GEDI (Fig. 3). In contrast, tree-dominated forests exhibited substantially higher carbon densities overall, as expected for woody evergreen broadleaf systems. Together, these results indicate that applying a single EBT model to bamboo-dominated forests inflates carbon estimates and obscures meaningful variation between bamboo forest and fallow systems.

Accounting for land-cover misclassification using posterior Monte Carlo resampling did not eliminate the inferred carbon bias in bamboo-dominated systems (Fig. 4b). Across 1,000 posterior resampling iterations, mean aboveground carbon bias remained consistently positive for both bamboo contexts. Fallow bamboo exhibited a mean overestimation of 44.1Mg C ha^{-1}, whereas forest bamboo showed a larger mean bias of 19.8 Mg C ha^{-1}, closely matching the naïve footprint-level estimates (triangles representing mean values). The narrow spread of the Monte Carlo distributions for forest bamboo indicates that plausible misclassification errors are insufficient to explain the observed magnitude of overestimation, while fallow bamboo with

496 wider distributions suggests higher uncertainty in AGBD estimates likely due to mixed
497 influences from rubber plantations and agroforestry mosaics. These results confirm that the
498 positive carbon bias arises from structural mismatch between bamboo and the EBT biomass
499 model rather than from classification uncertainty alone.

500 Having established that bamboo-specific carbon overestimation persists after explicitly
501 accounting for classification uncertainty at the footprint level, we next examined how this bias
502 propagates spatially across Xishuangbanna. To support regional extrapolation, we quantified
503 class-specific relationships between RH50 and RH98 from GEDI Level 2 observations and
504 evaluated the discrepancy between AGBD predicted by a simplified, RH-only EBT formulation
505 and published GEDI Level 4A estimates. These components were combined to propagate
506 uncertainty into regional estimates of absolute and percent carbon bias. Inputs to this propagation
507 include (1) class-specific RH50–RH98 regressions, (2) the EBT–L4A biomass discrepancy, and
508 (3) land-cover misclassification adjustments; the height dependence of the EBT structural
509 discrepancy is diagnosed in Appendix E. Table 2 summarizes the regression equations and the
510 full bias decomposition.

511 Mapping the percentage carbon bias revealed pronounced spatial heterogeneity across
512 bamboo-dominated landscapes (Fig. 4b). Relative overestimation frequently exceeded 40% in
513 bamboo-rich areas, with systematic differences between fallow and forest bamboo (Fig. 4c).
514 Because percentage bias is defined relative to EBT-estimated carbon density (Eq. 3), spatial
515 variation reflects the joint effects of context-specific absolute bias (ΔC_i) and variation in the
516 underlying EBT-estimated carbon baseline ($AGC_{EBT,i}$). Overall, these spatial patterns reinforce
517 the footprint-level inference that GEDI's EBT model systematically overestimates bamboo

518 aboveground carbon density, while the magnitude of overestimation varies predictably with
519 bamboo stand context.

520 4. Discussion

521 4.1 Grassy trees drive carbon bias in tree-calibrated biomass models

522 Our results demonstrate systematic biomass and carbon bias in GEDI Level 4A products over
523 evergreen broadleaf forests in Xishuangbanna, tropical China. This bias persists after Bayesian
524 correction for bamboo–tree classification uncertainty, indicating that it arises primarily from
525 structural mismatch between bamboo-dominated canopies and GEDI's global EBT biomass
526 model, rather than from misclassification alone. Understanding this bias therefore requires
527 examining how the operational EBT model was trained and applied.

528 The GEDI EBT biomass model is trained on globally uneven data, with continental and
529 Southeast Asia among the most underrepresented regions (Kellner et al. 2023). Although
530 calibration relies on simulated GEDI waveforms linked to evergreen broadleaf forest field plots
531 from 21 countries, only 326 simulated waveforms support the EBT stratum in South Asia,
532 compared with 3,441 in South America and 834 in Africa (Kellner et al. 2023, Table 1). As a
533 result, statistical support for EBT biomass estimation in tropical Asia is substantially weaker than
534 in other regions, a limitation particularly relevant for Xishuangbanna at the northern margin of
535 Southeast Asia.

536 Consistent with this limitation, the CEOS Land Product Validation framework recognizes
537 that, where training data are sparse, biomass estimates must rely on models trained in other
538 regions or plant functional types, implicitly assuming geographic and structural transferability
539 (Duncanson et al. 2022; Kay 2021). While necessary for global coverage, this assumption
540 introduces systematic bias when canopy structure and species composition deviate from those

541 represented in training data, such as bamboo-dominated secondary forests relative to tree-
542 dominated evergreen broadleaf forests.

543 Ecological studies consistently show that bamboo dominance produces pronounced
544 departures from woody evergreen broadleaf forest structure, including reduced tree basal area,
545 altered gap dynamics, suppressed or redirected tree regeneration, and demographic processes
546 driven by clonal expansion and episodic flowering–dieback cycles (Tabarelli and Mantovani
547 2000; Lima et al. 2012; Fadrique et al. 2021; Zheng and Lv 2023). These stands therefore
548 represent fundamentally different canopy architectures, consistent with the structural contrasts
549 observed in GEDI footprint-level LiDAR data (Fig. 3).

550 Mechanistically, bamboo differs from trees in ways that directly violate woody biomass
551 allometries. Bamboo lacks secondary growth, forms hollow culms, and expands clonally through
552 rhizomes (Makita 1998). Whereas tree biomass accumulates through continued radial growth of
553 woody stems—supporting power-law relationships between DBH, height, and biomass (Enquist
554 et al. 1999; Chave et al. 2014)—individual bamboo culms reach near-final height and diameter
555 within a single growing season. Consequently, stand-level biomass increases primarily through
556 culm recruitment and turnover rather than through growth of existing individuals (Yen 2016).
557 Although culm-level DBH–height allometries exist (Sharma et al. 2025), estimating bamboo
558 stand biomass requires explicit consideration of colony structure and culm age composition,
559 complicating direct scaling from tree-based models (Zhang et al. 2014).

560 These biological differences intersect directly with GEDI's biomass retrieval logic. GEDI
561 Level 4A relies heavily on relative height (RH) metrics (e.g., RH50, RH98), implicitly assuming
562 that waveform-derived vertical structure provides a consistent proxy for woody biomass across
563 vegetation types (Kellner et al. 2023). In bamboo-dominated canopies, however, RH metrics may

reflect dense foliage and hollow culm packing rather than woody stem mass, yielding an RH–AGBD relationship that deviates from the tree-calibrated EBT relationship assumed by GEDI L4A. Our height-stratified diagnostics (Fig. E.1; Method E) confirm that residual bias varies systematically with canopy height, indicating structural bias intrinsic to applying the EBT formulation to bamboo-dominated canopies.

Taken together, these considerations indicate that systematic bias is expected when GEDI's global, RH-based EBT biomass model is applied to bamboo-dominated forests. Our analyses in Fig. 3 and Fig. 4 were designed to diagnose and quantify this bias, rather than to challenge the validity of GEDI products per se. Instead, our results align with the CEOS validation framework by emphasizing transparency in model assumptions and responsible interpretation of satellite-derived biomass estimates (Kay 2021). In bamboo-dominated forests, botanical traits such as clonal growth, hollow culms, and rapid culm maturation fundamentally shape canopy structure and biomass accumulation and must be considered when interpreting RH-based biomass products.

4.2 Scope, assumptions, and diagnostic interpretation

Diagnostic scope of the case study

This analysis is based on a single regional case study and is designed to develop a workflow for diagnosing and quantifying systematic bias in GEDI biomass estimates, rather than to provide definitive estimates of absolute biomass or carbon density differences between bamboo- and tree-dominated forests. Accordingly, the magnitude of bias we report should be interpreted as an order-of-magnitude reference rather than a universal correction factor. Our results are derived from one ecosystem and one broad contrast (bamboo versus evergreen broadleaf trees), and the magnitude of bias is therefore expected to vary across regions, species compositions, and

successional contexts. Classification relies on single-year Sentinel-2 composites because of limited GEDI coverage in Xishuangbanna; incorporating time-series features would likely improve robustness by leveraging temporal patterns of phenological signals.

Benchmarking strategy and data constraints

A second limitation is the absence of site-specific ground-based biomass measurements in Xishuangbanna. While direct field measurements would provide the strongest benchmark, existing literature values are themselves uncertain. Our analysis therefore does not aim to establish a precise "true" biomass baseline. Instead, our conclusions rest on internally consistent comparisons among GEDI-derived structural metrics, GEDI's operational EBT biomass model, and published empirical ranges (Yuen et al. 2017). The convergence of structural, ecological, and modeling evidence supports our central conclusion that applying tree-centric, RH-based biomass models to bamboo-dominated canopies introduces systematic bias.

Modeling assumptions and uncertainty treatment

Our uncertainty propagation relied on standard but simplifying assumptions. We treated the visually interpreted confusion matrix as an observation model and assumed that misclassification rates were stationary and that errors were conditionally independent given the mapped label. We used the same posterior resampling framework throughout the analysis: for footprint-level structural contrasts (Fig. 3; Table 1), we propagated uncertainty across all mapped classes to assess how tree–bamboo differences changed under plausible label reassignment; for the empirical-benchmark carbon comparison (Fig. 4; Table 2), we propagated uncertainty conditional on the mapped bamboo subset to maintain alignment with bamboo-specific benchmarks and context stratification (forest vs fallow). A fully landscape-level correction that also recovered omission errors (false negatives) would require reconstructing latent true-class

610 composition across all land-cover classes and re-deriving bamboo context within each
611 resampling iteration, which was beyond the scope of this workflow-focused case study.
612 Taken together, these limitations do not weaken our conclusions, because our objective is
613 diagnostic rather than predictive: to identify the presence, direction, and structural origin of
614 systematic bias in GEDI biomass estimates. Importantly, our contribution extends beyond bias
615 identification. Using bamboo as a test case, we show that explicitly stratifying structurally
616 distinct canopies using texture and structural features transforms the same remote-sensing data
617 from a source of bias into an opportunity for refinement. This logic generalizes beyond bamboo
618 to other vegetation types whose structure departs from global-model assumptions, including
619 other grassy trees such as palms and bananas, and to fragmented agroforestry systems. We next
620 outline applications that demonstrate how this workflow supports more responsible, bias-aware
621 use of GEDI and related LiDAR products.

622 **4.3 From regional case study to scalable global applications**

623 Our workflow highlights the complementarity between horizontal information from optical
624 imagery (spatial distribution and land-use context) and vertical information from LiDAR
625 (canopy structure), both of which are necessary to interpret ecosystem function and avoid
626 misrepresenting structurally distinct canopy types in biomass estimation. More broadly,
627 stratification is a problem of scale selection under a coverage–accuracy trade-off: the features
628 that best separate target vegetation classes often emerge at intermediate spatial or temporal
629 scales, where signals can be detected consistently across heterogeneous landscapes. These results
630 further emphasize that remote-sensing–derived products must be interpreted with biological
631 awareness, because trait-driven deviations that are negligible at global scales can dominate local
632 error budgets where those traits are prevalent.

One immediate application of this framework is place-based monitoring in structurally complex landscapes. By combining Sentinel-2 optical imagery with GEDI LiDAR structure, the workflow enables explicit identification and mapping of locally relevant canopy strata or species of interest, such as grassy trees (e.g., bamboo, palms, and bananas) that are fast-growing yet frequently conflated with trees in global products (Zheng and Lu 2025). This capability is particularly important in fragmented forests and agroforestry mosaics, where compositional changes occur over spatial scales too fine for coarse plant functional type classifications. Rather than relying solely on globally defined PFTs, the workflow allows locally meaningful strata to be detected using region-specific training data and local knowledge, supporting community-led monitoring in heterogeneous landscapes.

A second application lies in improving measurement, reporting, and verification (MRV) of forest carbon. Once structurally distinct vegetation types are explicitly identified, the same framework enables more targeted carbon accounting by reducing systematic bias that arises when biomass models are applied beyond their domain of transferability. This is especially relevant for REDD+ and other carbon-crediting or monitoring programs operating in secondary forests, forest–agriculture transition zones, and agroforestry systems, where grassy trees are often promoted or maintained through human management and can contribute substantially to landscape-level biomass yet remain poorly represented in global allometric frameworks (Tang et al. 2025). By separating bias driven by structural mismatch from true spatial variation in biomass, the workflow supports more transparent MRV.

Beyond carbon accounting, stratified structure–biomass estimates can guide targeted regeneration, restoration, and management interventions. In multilayered systems such as cocoa- and coffee-based agroforestry, LiDAR-enabled separation of canopy layers—combined with

spectral or phenological indicators—can help identify vegetation components that contribute most to productivity, microclimate buffering, or climate resilience (Atalaya-Marin et al. 2025; Pippi et al. 2025). More broadly, our stratify–quantify–scale approach can be extended to track successional trajectories by separating shifts in canopy species composition from disturbance-driven changes in canopy structure, enabling interventions tailored to different successional stages.

Conclusion

As remote sensing technologies advance and global products like GEDI become more accessible, they create new opportunities while also increasing the risk of misinterpretation. Responsible use therefore requires bias-aware interpretation grounded in ecology, organismal biology, land-use history, and local knowledge. Using bamboo as an example, we demonstrate that biologically informed canopy stratification can diagnose and quantify systematic bias in GEDI's tree-calibrated biomass estimates and provide a practical workflow for doing so. Looking ahead, conceptual and technical advances can further improve the decision value of global products by specifying when stratification is needed, which signals reliably distinguish canopy types, and what validation is sufficient for targeted applications.

Data Availability Statement

All remote-sensing datasets used in this study are publicly available. Data sources and processing details, together with all analysis scripts, are provided at https://github.com/aiyuz/GEDI_Carbon_Workflow.

Author Contributions

AZ, YY, and ML jointly developed the conceptual framework of carbon discrepancy from structural heterogeneity. AZ implemented the workflow, analyzed the data, and prepared all

679 figures and tables. AZ wrote the manuscript and compiled the scripts, with revisions from ML
680 and YY.

681 **Declaration of interests**

682 The authors declare no competing interests.

683 **Acknowledgements**

684 We thank Thaise Emilio, Ximena Londoño, Juan Carlos Camargo García, Junwei Luan, Cuiju
685 Liu, Wenjun Zhou, and Jiayue Zhang for their collaboration in establishing the grassy tree
686 research network. We also thank members of the Lu Lab and the Yin Lab in the Department of
687 Environmental Studies at New York University for helpful discussions.

688 **Declaration of generative AI and AI-assisted technologies**

689 During the preparation of this work, the authors used ChatGPT (version 5.2) to help organize and
690 annotate scripts for publication and to check the language and grammar of the manuscript. After
691 using this service, the authors reviewed and edited the content and take full responsibility for the
692 content of the published article.

693 **References**

694 Atalaya-Marin, Nilton, Teiser Sanchez-Fuentes, Malluri Goñas, et al. 2025. "Estimation of
695 Aboveground Biomass and Carbon Sequestration in a Cocoa Agroforestry System Using
696 UAV-LiDAR in Northwestern Peru." *Remote Sensing Applications Society and*
697 *Environment* 40 (101750): 101750.

698 Avalos, Gerardo, Milena Cambronero, and Carolina Alvarez-Vergnani. 2022. "Allometric
699 Models to Estimate Carbon Content in Arecaceae Based on Seven Species of Neotropical
700 Palms." *Frontiers in Forests and Global Change* 5 (July).
701 https://doi.org/10.3389/ffgc.2022.867912.

702 Bruening, Jamis, Paul May, John Armston, and Ralph Dubayah. 2023. "Precise and Unbiased
703 Biomass Estimation from GEDI Data and the US Forest Inventory." *Frontiers in Forests*
704 *and Global Change* 6 (1149153). https://doi.org/10.3389/ffgc.2023.1149153.

705 Camargo García, Juan Carlos, Angela Maria Arango Arango, and Long Trinh. 2023. "The
706 Potential of Bamboo Forests as a Carbon Sink and Allometric Equations for Estimating
707 Their Aboveground Biomass." *Environment Development and Sustainability* 26 (8): 20159–
708 20187.

709 Carvalho, Anelena L. de, Bruce W. Nelson, Milton C. Bianchini, Daniela Plagnol, Tatiana M.
710 Kuplich, and Douglas C. Daly. 2013. "Bamboo-Dominated Forests of the Southwest
711 Amazon: Detection, Spatial Extent, Life Cycle Length and Flowering Waves." *PloS One* 8
712 (1): e54852.

713 Chave, Jérôme, Maxime Réjou-Méchain, Alberto Búrquez, et al. 2014. "Improved Allometric
714 Models to Estimate the Aboveground Biomass of Tropical Trees." *Global Change Biology*
715 20 (10): 3177–3190.

716 Descals, Adrià. 2023. "High-Resolution Global Map of Closed-Canopy Coconut Palm." *Earth*
717 *System Science Data* 15: 3991–4010.

718 Drusch, M., U. Del Bello, S. Carlier, et al. 2012. "Sentinel-2: ESA's Optical High-Resolution
719 Mission for GMES Operational Services." *Remote Sensing of Environment* 120 (May): 25–
720 36.

721 Dubayah, Ralph, John Armston, Sean P. Healey, et al. 2022. "GEDI Launches a New Era of
722 Biomass Inference from Space." *Environmental Research Letters* 17 (9): 095001.

723 Duncanson, Laura, James R. Kellner, John Armston, et al. 2022. "Aboveground Biomass
724 Density Models for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar
725 Mission." *Remote Sensing of Environment* 270 (112845): 112845.

726 Enquist, Brian J., Geoffrey B. West, Eric L. Charnov, and James H. Brown. 1999. "Allometric
727 Scaling of Production and Life-History Variation in Vascular Plants." *Nature* 401 (6756):
728 907–911.

729 Fadrique, Belen, Daniel Gann, Bruce W. Nelson, Sassan Saatchi, and Kenneth J. Feeley. 2021.
730 "Bamboo Phenology and Life Cycle Drive Seasonal and Long-term Functioning of
731 Amazonian Bamboo-dominated Forests." *The Journal of Ecology* 109 (2): 860–876.

732 Fadrique, Belén, Joseph W. Veldman, James W. Dalling, et al. 2020. "Guidelines for Including
733 Bamboos in Tropical Ecosystem Monitoring." *Biotropica* 52 (3): 427–443.

734 Heinrich, Viola H. A., Ricardo Dalagnol, Henrique L. G. Cassol, et al. 2021. "Large Carbon Sink
735 Potential of Secondary Forests in the Brazilian Amazon to Mitigate Climate Change."
736 *Nature Communications* 12 (1): 1785.

737 Huang, Xuying, Weimin Ju, Zhanghua Xu, and Jing Li. 2024. "A Novel Method for Mapping
738 Moso Bamboo Forests Using Remote Sensing Data with the Consideration of Phenological
739 Status." *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE*
740 *Geoscience and Remote Sensing Society* 62: 1–18.

741 Humeau-Heurtier, Anne. 2019. "Texture Feature Extraction Methods: A Survey." *IEEE Access:*
742 *Practical Innovations, Open Solutions* 7: 8975–9000.

743 Huy, Bao, Giang Thi Thanh, Krishna P. Poudel, and Hailemariam Temesgen. 2019. "Individual
744 Plant Allometric Equations for Estimating Aboveground Biomass and Its Components for a
745 Common Bamboo Species (Bambusa Procera A. Chev. And A. Camus) in Tropical Forests."
746 *Forests* 10 (4): 316.

747 IPCC. 2006. *Intergovernmental Panel on Climate Change Guidelines for National Greenhouse*
748 *Gas Inventories. Chapter 4: Forest Lands*. https://www.ipcc-
749 nggip.iges.or.jp/public/mtdocs/pdfiles/LUCF_EGPM_Meeting_Report_Final.pdf.

750 Jia, Xiaowei, Ankush Khandelwal, Kimberly M. Carlson, et al. 2020. "Automated Plantation
751 Mapping in Southeast Asia Using MODIS Data and Imperfect Visual Annotations." *Remote*
752 *Sensing* 12 (4): 636.

753 Kay, Heather. 2021. "Aboveground Woody Biomass Product Validation Good Practices
754 Protocol." Preprint, Land Product Validation Subgroup (Working Group on Calibration and
755 Validation, Committee on Earth Observation Satellites).
756 https://doi.org/10.5067/DOC/CEOSWGCV/LPV/AGB.001.

757 Kellner, James R., John Armston, and Laura Duncanson. 2023. "Algorithm Theoretical Basis
758 Document for GEDI Footprint Aboveground Biomass Density." *Earth and Space Science*
759 *(Hoboken, N.J.)* 10 (4). https://doi.org/10.1029/2022ea002516.

760 Kotowska, Martyna M. 2015. "Quantifying above-and Belowground Biomass Carbon Loss with
761 Forest Conversion in Tropical Lowlands of S Umatra (I Ndonesia)." *Global Change Biology*
762 21: 3620–3634.

763 Lang, Nico, Nikolai Kalischek, John Armston, Konrad Schindler, Ralph Dubayah, and Jan Dirk
764 Wegner. 2022. "Global Canopy Height Regression and Uncertainty Estimation from GEDI
765 LIDAR Waveforms with Deep Ensembles." *Remote Sensing of Environment* 268 (112760):

112760.

Li, Hongmei, Youxin Ma, T. Mitchell Aide, and Wenjun Liu. 2008. "Past, Present and Future Land-Use in Xishuangbanna, China and the Implications for Carbon Dynamics." *Forest Ecology and Management* 255 (1): 16–24.

Li, Longwei, Nan Li, Dengsheng Lu, and Yuyun Chen. 2019. "Mapping Moso Bamboo Forest and Its on-Year and off-Year Distribution in a Subtropical Region Using Time-Series Sentinel-2 and Landsat 8 Data." *Remote Sensing of Environment* 231 (111265): 111265.

Lima, Renato A. F., Débora C. Rother, Ana E. Muler, Igo F. Lepsch, and Ricardo R. Rodrigues. 2012. "Bamboo Overabundance Alters Forest Structure and Dynamics in the Atlantic Forest Hotspot." *Biological Conservation* 147 (1): 32–39.

Lin, Di, Mario Elia, Onofrio Cappelluti, et al. 2025. "From Spaceborne LiDAR to Local Calibration: GEDI's Role in Forest Biomass Estimation." *Remote Sensing* 17 (16): 2849.

Makita, Akifumi. 1998. "The Significance of the Mode of Clonal Growth in the Life History of Bamboos." *Plant Species Biology* 13 (2-3): 85–92.

Mohammadpour, Pegah, Domingos Xavier Viegas, and Carlos Viegas. 2022. "Vegetation Mapping with Random Forest Using Sentinel 2 and GLCM Texture feature—A Case Study for Lousã Region, Portugal." *Remote Sensing* 14 (18): 4585.

Nath, Arun Jyoti, Rattan Lal, and Ashesh Kumar Das. 2015. "Managing Woody Bamboos for Carbon Farming and Carbon Trading." *Global Ecology and Conservation* 3 (January): 654–663.

Navarro, M. N. V., C. Jourdan, T. Sileye, et al. 2008. "Fruit Development, Not GPP, Drives Seasonal Variation in NPP in a Tropical Palm Plantation." *Tree Physiology* 28 (11): 1661–1674.

Pan, Yude, Richard A. Birdsey, Jingyun Fang, et al. 2011. "A Large and Persistent Carbon Sink in the World's Forests." *Science (New York, N.Y.)* 333 (6045): 988–993.

Pan, Yude, Richard A. Birdsey, Oliver L. Phillips, et al. 2024. "The Enduring World Forest Carbon Sink." *Nature* 631 (8021): 563–569.

Pippi, Lorenzo, Michael Alibani, Daniele Antichi, et al. 2025. "Use of Digital Technologies into Agroforestry Systems: A Review." *Agronomy (Basel, Switzerland)* 15 (12): 2671.

Qi, Shuhua, Bin Song, Chong Liu, et al. 2022. "Bamboo Forest Mapping in China Using the Dense Landsat 8 Image Archive and Google Earth Engine." *Remote Sensing* 14 (3): 762.

Senf, Cornelius, Dirk Pflugmacher, Sebastian Van der Linden, and Patrick Hostert. 2013. "Mapping Rubber Plantations and Natural Forests in Xishuangbanna (southwest China) Using Multi-Spectral Phenological Metrics from MODIS Time Series." *Remote Sensing* 5 (6): 2795–2812.

Sharma, Aastha, Santosh Ayer, Keshav Ayer, et al. 2025. "Empirical Allometric Models for Estimating Aboveground Biomass of Bambusa Teres and Bambusa Tulda in Non-Forest Areas of Nepal." *Agroforestry Systems* 99 (5). https://doi.org/10.1007/s10457-025-01221-9.

Song, Xinzhang, Xiaofeng Chen, Guomo Zhou, Hong Jiang, and Changhui Peng. 2017. "Observed High and Persistent Carbon Uptake by Moso Bamboo Forests and Its Response to Environmental Drivers." *Agricultural and Forest Meteorology* 247 (December): 467–475.

Tabarelli, Marcelo, and Waldir Mantovani. 2000. "Gap-Phase Regeneration in a Tropical Montane Forest: The Effects of Gap Structure and Bamboo Species." *Plant Ecology* 148 (2): 149–155.

Tang, Yuzhi, Chao Yang, Haishan Wu, et al. 2025. "Tropical Forest Carbon Offsets Deliver Partial Gains amid Persistent over-Crediting." *Science (New York, N.Y.)* 390 (6769): 182–187.

Venkatappa, Manjunatha, Sutee Anantsuksomsri, Jose Alan Castillo, Benjamin Smith, and Nophea Sasaki. 2020. "Mapping the Natural Distribution of Bamboo and Related Carbon Stocks in the Tropics Using Google Earth Engine, Phenological Behavior, Landsat 8, and Sentinel-2." *Remote Sensing* 12 (18): 3109.

Wang, Kanglin, Ramanatha Rao, and Hong. 2002. "Diversity and Indigenous Utilization of Bamboo in Xishuangbanna, Yunnan Province, Southwest China." *Journal of Bamboo and Rattan* 1 (3): 263–273.

Wang, Yao, Hongliang Fang, Yu Li, Sijia Li, and Hao Tang. 2025. "Validation of the Vertical Plant Area Index Profile Product Derived from GEDI over Global Forest Sites." *Agricultural and Forest Meteorology* 371 (110612): 110612.

Xayalath, Singkone, Isao Hirota, Shinsuke Tomita, and Michiko Nakagawa. 2019. "Allometric Equations for Estimating the Aboveground Biomass of Bamboos in Northern Laos." *Journal of Forest Research* 24 (2): 115–119.

Yen, Tian-Ming. 2016. "Culm Height Development, Biomass Accumulation and Carbon Storage in an Initial Growth Stage for a Fast-Growing Moso Bamboo (Phyllostachy Pubescens)." *Botanical Studies (Taipei, Taiwan)* 57 (1): 10.

Yang, Qing, Zhu-Biao Duan, Zheng-Liang Wang, Kai-Hong He, Qi-Xiang Sun, and Zhen-Hua Peng. 2008. "Bamboo Resources, Utilization and Ex-Situ Conservation in Xishuangbanna, South-Eastern China." *Journal of Forestry Research* 19 (1): 79–83.

Yen, Tian-Ming. 2016. "Culm Height Development, Biomass Accumulation and Carbon Storage in an Initial Growth Stage for a Fast-Growing Moso Bamboo (Phyllostachy Pubescens)." *Botanical Studies (Taipei, Taiwan)* 57 (1): 10.

Yuen, Jia Qi, Tak Fung, and Alan D. Ziegler. 2017. "Carbon Stocks in Bamboo Ecosystems Worldwide: Estimates and Uncertainties." *Forest Ecology and Management* 393 (June): 113–138.

838 Yusof, Nurasmalaily, Helmi Zulhaidi Mohd Shafri, and Nur Shafira Nisa Shaharum. 2021. "The
839 Use of Landsat-8 and Sentinel-2 Imageries in Detecting and Mapping Rubber Trees."
840 *Journal of Rubber Research (Kuala Lumpur, Malaysia)* 24 (1): 121–135.

841 Zhang, Houxi, Shunyao Zhuang, Bo Sun, Haibao Ji, Changming Li, and Sai Zhou. 2014.
842 "Estimation of Biomass and Carbon Storage of Moso Bamboo (Phyllostachys Pubescens
843 Mazel Ex Houz.) in Southern China Using a Diameter–age Bivariate Distribution Model."
844 *Forestry (London, England)* 87 (5): 674–682.

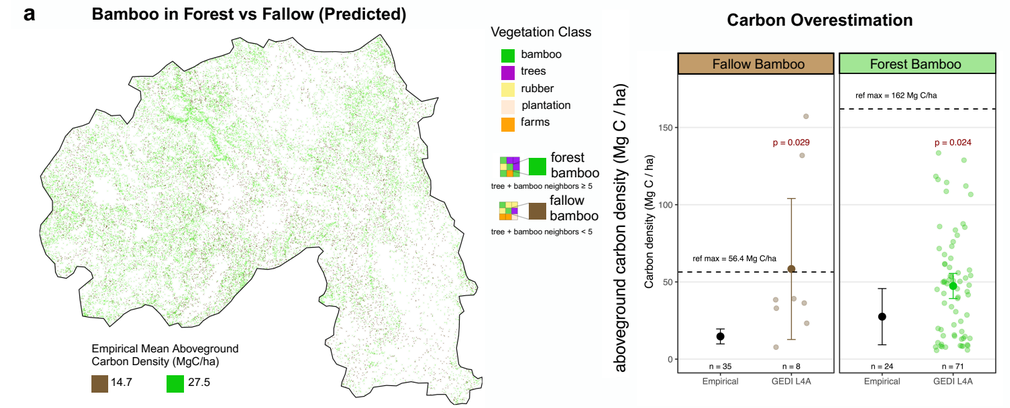
845 Zhang, Jianhou, and Min Cao. 1995. "Tropical Forest Vegetation of Xishuangbanna, SW China
846 and Its Secondary Changes, with Special Reference to Some Problems in Local Nature
847 Conservation." *Biological Conservation* 73 (3): 229–238.

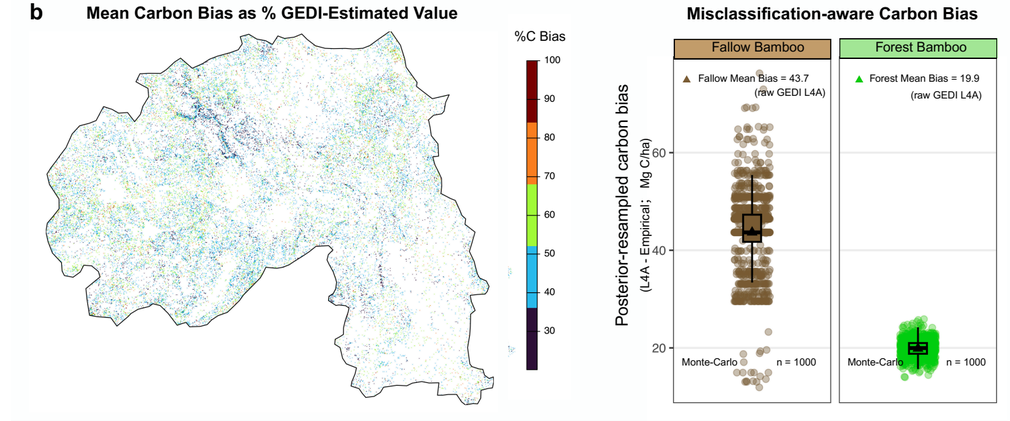
848 Zheng, Aiyu, and Mingzhen Lu. 2025. *Grassy Trees: The Neglected Hybrids for Sustainability*.
849 October 23. https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(25)00287-
850 3.

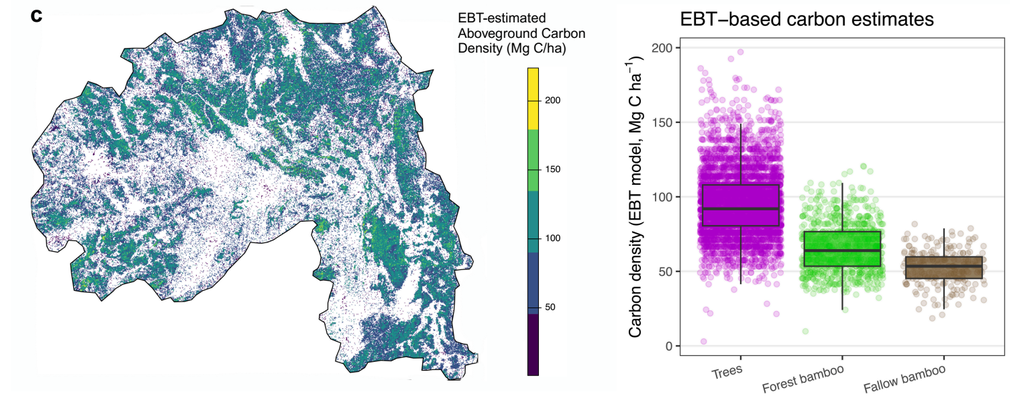
851 Zheng, Aiyu, and Jianhua Lv. 2023. "Spatial Patterns of Bamboo Expansion across Scales: How
852 Does Moso Bamboo Interact with Competing Trees?" *Landscape Ecology* 38 (12): 3925–
853 3943.

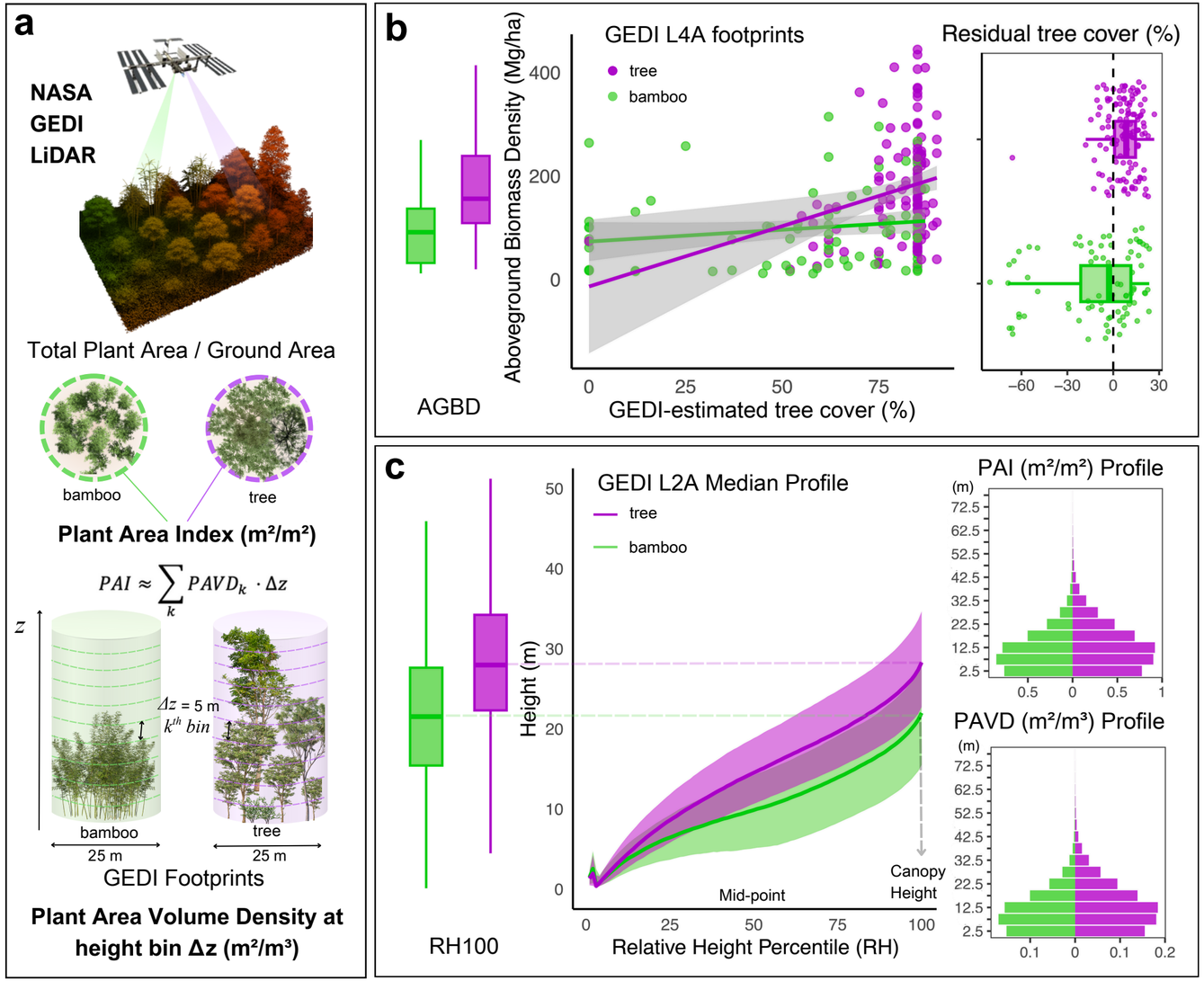
854 Zhu, Hua, Min Cao, and Huabin Hu. 2006. "Geological History, Flora, and Vegetation of
855 Xishuangbanna, Southern Yunnan, China[1]." *Biotropica* 38 (3): 310–317.

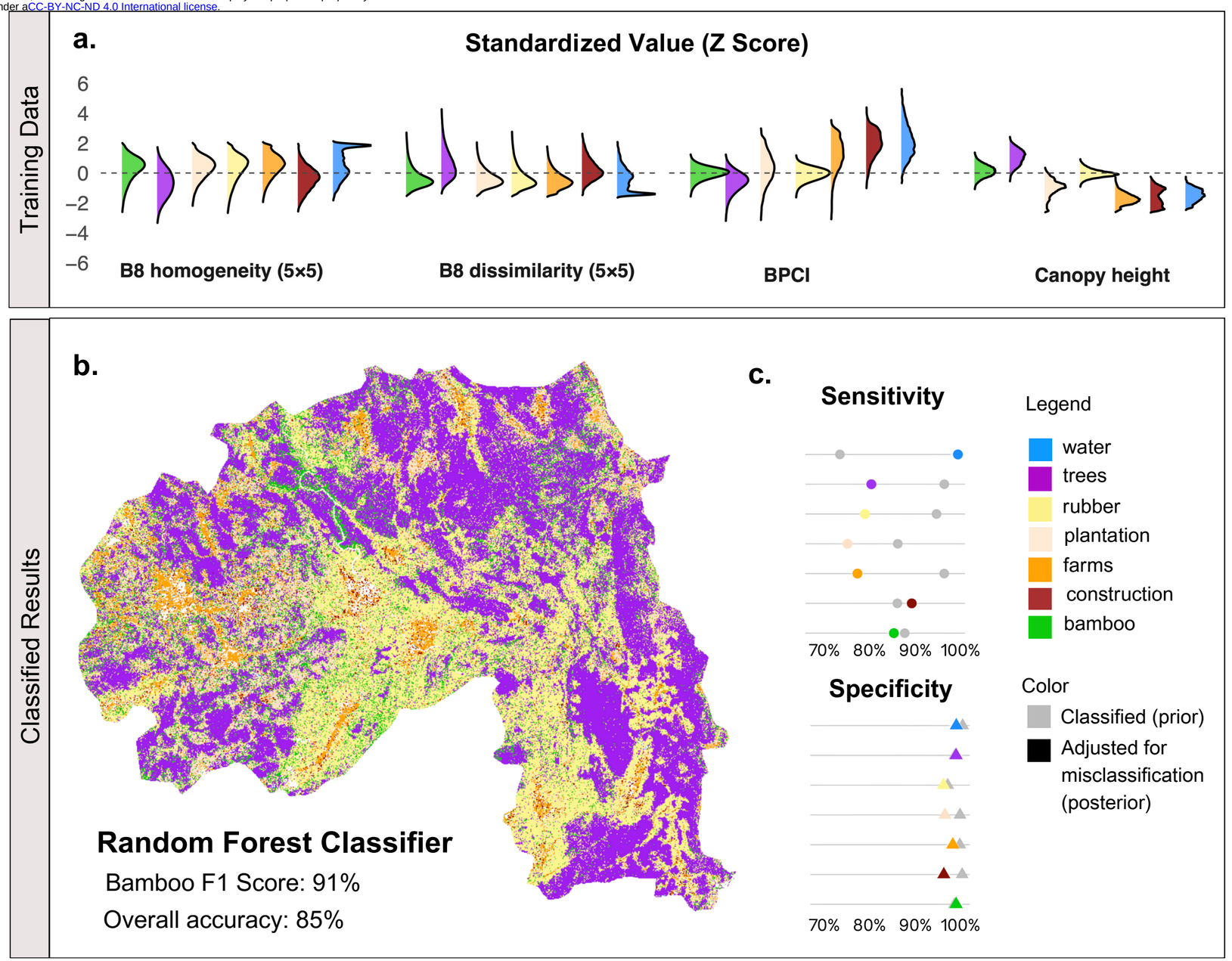
856





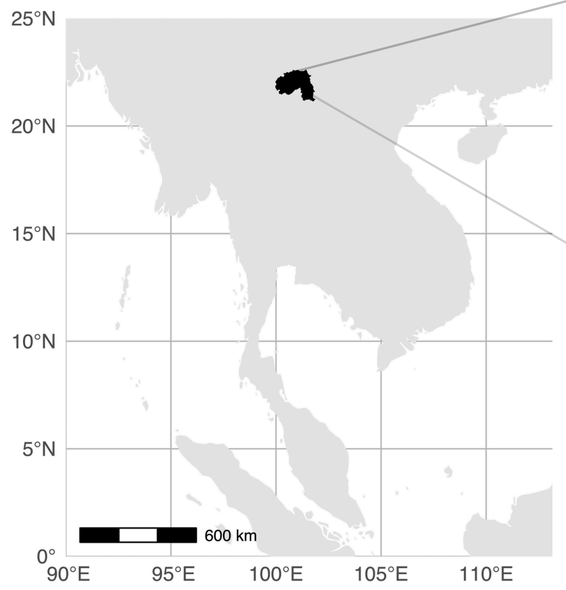






Study Region

Xishuangbanna,
Tropical China



Study Aim

Quantify carbon bias arising from applying a single PFT biomass model to structurally heterogeneous vegetation

Bamboo Map

ΔCarbon (MgC/ha)

GEDI's Source of carbon bias

Bamboo-dominant

tree-dominant

both are labelled
Evergreen Broadleaf Trees

despite significant
Structural Differences

Workflow

Objective 1: Stratify

Classify Bamboo- and Non-bamboo Evergreen Broadleaf Forests using Sentinel-2

Objective 2: Quantify

Quantify Structural Differences under Coarse PFT labeling using GEDI LiDAR

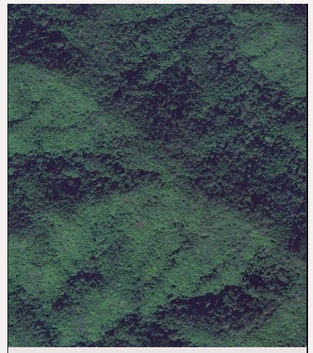
Objective 3: Scaling

Carbon Discrepancy and Other Implications

Training Data

Maxar Optical (<1m) + DEM/DSM

Observation data (Literature)



1.1 Expert Labeling

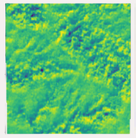
Mapping Data

Sentinel 2 Optical (10 m)

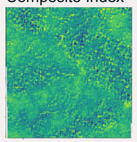
Global Canopy Height (10 m)

Texture Metrics (GLCM)

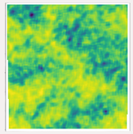
NIR



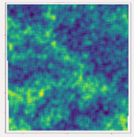
Bamboo Probability Composite Index



NIR Dissimilarity



NIR Homogeneity



1.2 Feature Extraction

Classification Algorithm

1.3 Supervised Classification

1.4 Evaluation and Validation