remote sensing

MDPI

Article

Upscaling Frameworks Drive Prediction Accuracy and Uncertainty When Mapping Aboveground Biomass Density from the Synergism of Spaceborne LiDAR, SAR, and Passive Optical Data

Inacio T. Bueno [1,*], Carlos A. Silva [1], Monique B. Schlickmann [1], Victoria M. Donovan [1], Jeff W. Atkins [2], Kody M. Brock [1], Jinyi Xia [1], Denis R. Valle [1], Jiangxiao Qiu [1], Jason Vogel [1], Andres Susaeta [3], Ajay Sharma [4], Carine Klauberg [1], Midhun Mohan [5] and Ana Paula Dalla Corte [6]

1 School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, FL 32611, USA; c.silva@ufl.edu (C.A.S.); moniquebohorasch@ufl.edu (M.B.S.); victoria.donovan@ufl.edu (V.M.D.); kody.brock@ufl.edu (K.M.B.); jinyixia@ufl.edu (J.X.); drvalle@ufl.edu (D.R.V.); qiuj@ufl.edu (J.Q.); jvogel@ufl.edu (J.V.); carine.klaubergs@ufl.edu (C.K.)
2 USDA Forest Service, Southern Research Station, New Ellenton, SC 29809, USA; jeffrey.atkins@usda.gov
3 Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, OR 97331, USA; andres.susaeta@oregonstate.edu
4 College of Forestry, Wildlife and Environment, Auburn University, Auburn, AL 36849, USA; azs0365@auburn.edu
5 Ecoresolve, San Francisco, CA 94105, USA; mikey@ecoresolve.eco
6 Department of Forest Science, Federal University of Paraná, Curitiba 80060-000, PR, Brazil; anacorte@ufpr.br
* Correspondence: ithomazbueno@ufl.edu

Abstract

Accurate mapping of aboveground biomass density (AGBD) is vital for ecological research and carbon cycle monitoring. Integrating multi-source remote sensing data offers significant potential to enhance the accuracy and coverage of AGBD estimates. This study evaluated three upscaling frameworks for integrating GEDI LiDAR, SAR, and optical satellite data to create wall-to-wall AGBD maps. The frameworks tested in this paper were: (1) a single-step approach using optical imagery, (2) a two-stage approach with GEDI-derived variables, and (3) a three-stage approach combining imagery and in situ-derived allometries. Internal validation showed that framework 1 achieved the lowest root mean square difference (%RMSD) of 53.3% and highest coefficient of determination (R^2) of 0.53. An independent external validation of the AGBD map was performed using in situ observations, also revealing that framework 1 was the most accurate (%RMSD = 39.3% and R^2 = 0.93), while frameworks 2 and 3 were less accurate (%RMSD = 54.7, 44.7 and R^2 = 0.95, 0.90, respectively). Herein, we show that upscaling frameworks significantly impacted AGBD map uncertainty and the magnitude of estimate differences. Our findings suggest that upscaling framework 1 based on a single step approach was the most effective for capturing detailed AGBD variations, while careful consideration of model sensitivity and map uncertainties is essential for reliable AGBD estimation. This study provides valuable insights for advancing forest AGBD monitoring and highlights the potential for further enhancements in remote sensing methodologies.

Keywords: GEDI; Random Forest; Harmonized Landsat Sentinel; principal component analysis; wall-to-wall mapping

check for updates

Academic Editor: Huaqiang Du

Received: 9 May 2025
Revised: 3 July 2025
Accepted: 7 July 2025
Published: 8 July 2025

Citation: Bueno, I.T.; Silva, C.A.; Schlickmann, M.B.; Donovan, V.M.; Atkins, J.W.; Brock, K.M.; Xia, J.; Valle, D.R.; Qiu, J.; Vogel, J.; et al. Upscaling Frameworks Drive Prediction Accuracy and Uncertainty When Mapping Aboveground Biomass Density from the Synergism of Spaceborne LiDAR, SAR, and Passive Optical Data. *Remote Sens.* **2025**, *17*, 2340. https://doi.org/10.3390/rs17142340

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Forest aboveground biomass density (AGBD) mapping is crucial in ecological research, climate change studies, and carbon cycle monitoring. AGBD describes the organic material in trees above the soil within a specific area and provides essential insights into the health and productivity of forest ecosystems, such as degradation and recovery [1]. It can be used to inform forest management practices, conservation efforts, and restoration activities, as it enables policymakers to assess carbon stocks and fluxes, formulate strategies to mitigate climate change, and allocate resources effectively for forest conservation [2]. Moreover, forests absorb significant amounts of atmospheric CO_2, playing a pivotal role in regulating the Earth's climate [3]. AGBD is a critical biophysical variable in global climate monitoring frameworks like the Global Climate Observing System [4], where it is used to support evaluations of forest carbon storage capabilities.

Accurate measurement, mapping, and monitoring of AGBD are vital for reducing uncertainties associated with quantifying carbon dynamics and their impacts on climate, habitat, and biodiversity [5]. This is essential not only for scientific understanding but also for policy formulation and implementation aimed at climate change mitigation. Moreover, as climate change impacts become more pronounced, the ability to monitor changes in AGBD accurately is critical for adapting forest management practices to ensure the resilience and sustainability of forest ecosystems [6,7].

Remote sensing technologies are a cornerstone for mapping AGBD at large spatial extents. Integrating diverse data types from remote sensing sources, including passive optical data, synthetic aperture radar (SAR), and spaceborne Light Detection and Ranging (LiDAR), captures detailed information on forest structure and dynamics and enhances the accuracy and precision of AGBD estimates [8–11]. In this context, the Global Ecosystem Dynamics Investigation (GEDI) mission represents a significant leap forward in measuring forest biomass and carbon stocks from space [12]. Mounted on the International Space Station, GEDI employs high-resolution LiDAR to map the three-dimensional structure of the Earth's forests. The waveform data from GEDI footprints are transformed into estimates of AGBD through the Level 4 (L4) product calibration models [12]. These models leverage collocated field plot measurements and airborne laser scanning data [13], which can provide unmatched insights into the vertical structure of forests.

Despite rapid advancements in our remote sensing capabilities and capacity, no single sensor or platform is currently able to generate wall-to-wall, representative maps of AGBD at sufficient spatial or temporal resolution, necessitating synergistic approaches to combine data from multiple sensors. For example, GEDI faces limitations in generating continuous, wall-to-wall maps due to its sampling strategy, which focuses on collecting footprint waveforms along eight parallel ground transects, with a 60 m footprint spacing center-to-center along the transect and 600 m between transects [14]. Therefore, Sentinel-2 and Sentinel-1 have been used in upscaling GEDI AGBD at national [15,16], state [17], and local scales [18–20], augmenting GEDI's detailed observations with their wide-area coverage and complementary data types. Landsat data has also significantly complemented GEDI's capabilities, especially in upscaling footprint-level forest canopy height measurements [21] and modeling woodland biomass [22]. SAR data, such as that from the TanDEM-X mission, have also strengthened the precision of AGBD estimates derived from GEDI [23]. Other studies have been improving AGBD mapping through data integration approaches; for example, GEDI has been combined with corrected field data and Sentinel-1/2 in a two-stage upscaling framework, significantly improving accuracy in arid, low-biomass regions [24]. Terrain and hydrological metrics were integrated with GEDI and Landsat data in stratified models, highlighting the benefits of biome-specific deep learning approaches for improving AGBD estimates [25]. Another recent advance is an automated model optimization

approach that aggregates GEDI at a block level and integrates Sentinel-2 features to map multiple forest structure parameters, improving accuracy and transferability in tropical forests [26].

While there is a considerable body of literature supporting the effectiveness of combining data from sources like GEDI and satellite imagery for biomass estimation, there remain persistent doubts regarding the accuracy and reliability of these integrated models. For example, early studies found that combining LiDAR with satellite imagery does not always lead to improvements in accuracy, suggesting that the integration of these data sources may depend heavily on specific conditions and contexts [27,28]. As model complexity increases, so does the potential for compounding errors, necessitating rigorous validation and calibration techniques to assess model accuracy and reliability [29]. Regression models such as Random Forest (RF; [30]) are common approaches to fuse satellite imagery with GEDI data to estimate biomass accurately based on the structural characteristics of the forest provided by GEDI [17,31]. Recent studies have also shown that deep learning models can achieve higher accuracy than traditional machine learning approaches [25]. These models often exhibit stronger generalization capacity and reduced spatial bias but require extensive training data and are computationally demanding [32]. Therefore, existing studies have not thoroughly investigated the comparative strengths and limitations of various models that combine GEDI with satellite imagery, nor have they delved deeply into the modeling details to identify what contributes to successful biomass estimation and uncertainty propagation.

Here, we conduct a comprehensive evaluation of the integration of GEDI data with other remote sensing sources. First, we assess the efficacy of merging GEDI data with data from sources such as SAR and optical satellite imagery, aiming to understand how well these combined datasets can enhance forest AGBD mapping across different landscapes. Second, we evaluate three distinct upscaling frameworks to derive continuous, or wall-to-wall, AGBD maps, comparing their effectiveness in capturing detailed biomass variations. Third, we analyze the sensitivity of these models to the input data and their associated uncertainties. By quantifying how minor changes in data input can affect biomass estimates and assessing the uncertainties inherent in each model, we identify robust methodologies that provide reliable and repeatable results, thus offering valuable insights into the precision and reliability of biomass estimation models.

2. Materials and Methods

2.1. Study Area

We conducted our assessment using four areas across the state of Florida in the southeastern US (Figure 1), each chosen for its ecological characteristics and representative AGBD range. The first two study areas were the University of Florida's Austin Cary Forest and Fisheries and Aquatic Sciences (FAS) Millhopper Unit, located near Gainesville in north-central Florida. Forest management in these areas is centered on educational and research initiatives, emphasizing sustainable forest management and conservation. Both areas are primarily composed of managed plantations of loblolly pine (*Pinus taeda* L.), longleaf pine (*Pinus palustris* Mill.), and slash pine (*Pinus elliottii* Engelm. var. elliottii) alongside naturally occurring mixed hardwoods. These plantations undergo regular thinning and harvesting practices, which not only help in managing forest health and productivity but also increase the range of biomass values present. This variability enriches the dataset, providing a broader spectrum of biomass conditions that enhance the robustness of the models used in the study. Our third study area was the Myakka State Forest, located in Southwest Florida. This area encompasses a vast expanse of protected land, featuring a mix of dry prairie, pine flatwoods, and wetland ecosystems. These varied habitats host a rich biodiversity, including many plant and animal species that contribute to the area's

complex ecological dynamics. The forest is characterized by its longleaf pine and slash pine stands, interspersed with palmetto underbrush and seasonally flooded marshes. Lastly, our fourth study area was the Okaloacoochee Slough State Forest, which encompasses a diverse array of ecosystems, including cypress swamps, pine flatwoods, and hardwood hammocks, reflecting the ecological variety of the region. The area is particularly notable for its sloughs, swamp-like waterways that provide important wildlife habitat and play a key role in local hydrology.

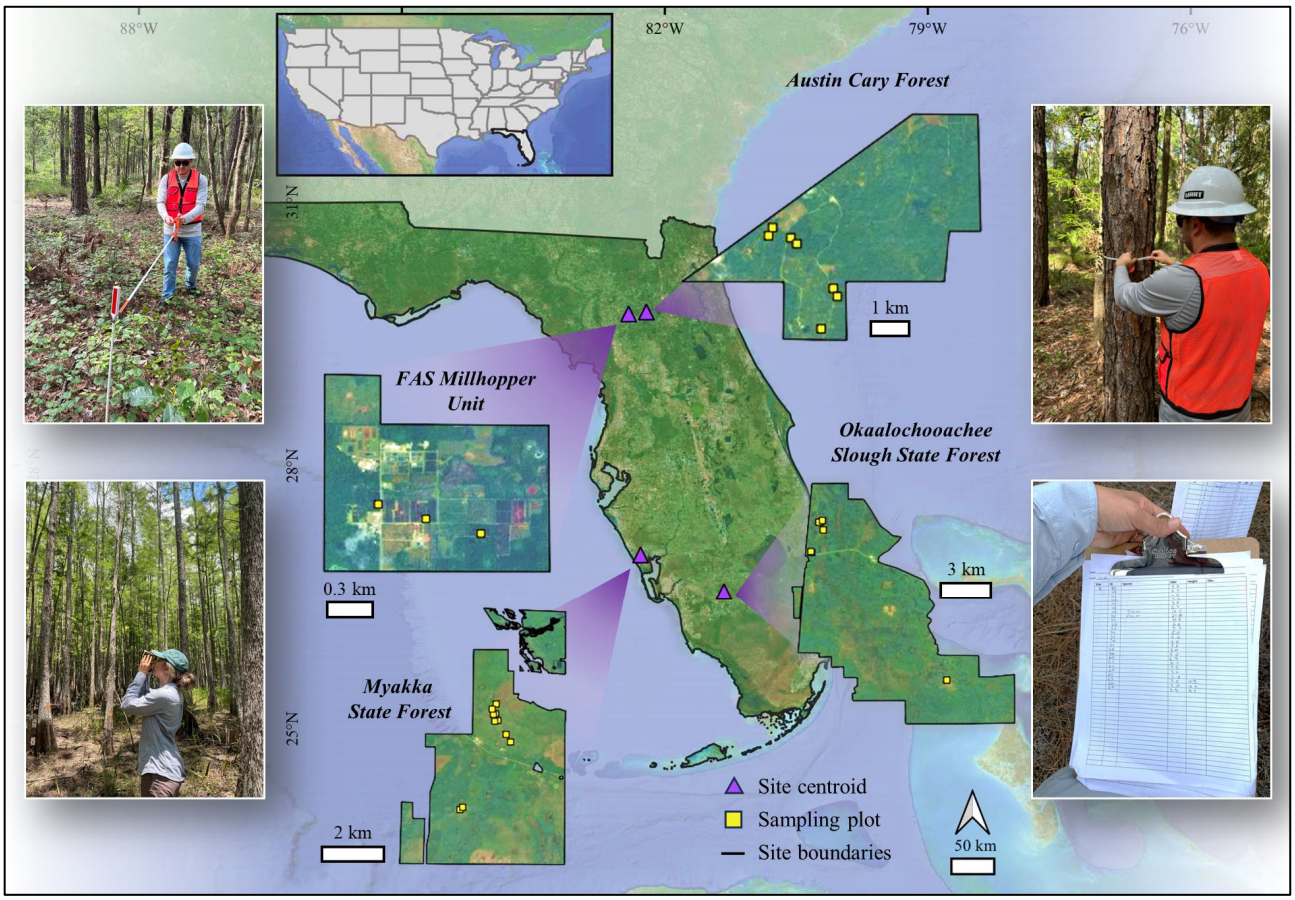


Figure 1. The location of our four study areas within the state of Florida, USA (purple triangle denotes area centroid), along with corresponding insets detailing the location of sampling plots (yellow squares) within each of the four study areas. Accompanying photographs demonstrate in situ data collection within sampling plots. The top inset highlights the location of the state of Florida (outlined in dark black) within the United States of America.

All study areas have a humid subtropical climate based on Köppen climate classifications (Cfa), with long, hot summers and mild, wet winters. Austin Cary Forest and FAS Millhopper Unit typically have annual precipitation averaging about 1170 mm, with temperatures ranging from a winter low of around 5.6 °C to a summer high of approximately 33.0 °C. Myakka State Forest, closer to the Gulf Coast, sees a slightly higher average annual precipitation of around 1350 mm and similar temperature ranges, albeit with slightly milder winters and warmer summers due to its coastal proximity. Lastly, the Okaloacoochee Slough State Forest annual precipitation levels are similar to Myakka but can experience slightly higher temperatures during the summer months.

2.2. Field AGBD Data Collection and Processing

Field data collection was conducted in each of the four study areas. We distributed 25 m × 25 m sample plots, representing the diverse range of forest types and conditions

present in each study area, from densely wooded regions to shrublands with sparse trees. A total of 27 plots were sampled, with the number of plots varying per area based on the accessibility of the area and the visual AGBD range presented. This approach involved initial data collection followed by evaluating the mapping accuracy; if the accuracy met our standards, we concluded that the sampling frequency was satisfactory. Within each plot, we identified tree species, measured diameter at breast height (DBH), and height for all the trees with a DBH greater than 10 cm.

The data from sample plots were used to estimate plot-level AGBD through allometric equations. Tree species within sample plots consisted of slash pine, loblolly pine, longleaf pine, and cypress (*Taxodium* spp.). We also found hardwood species mixed with these pines, predominantly oaks from the Fagaceae family. For each species, we employed a specific aboveground biomass (AGB) allometric equation found in the literature, tailored to the unique growth patterns and physical characteristics of each tree. AGB equations for slash pine, loblolly pine, longleaf pine, and cypress basically consisted of the sum of their three components: stem, branches, and foliage (Table 1). Additionally, we applied an equation specifically developed for the Fagaceae family when mixed with pines, referring to these as woodland [33]. The sum of individual tree AGB provided the total AGB in kilograms per plot. This value was then converted to megagrams and scaled to a per-hectare basis (Mg ha^{-1}), resulting in AGBD.

Table 1. Allometric equations used for estimating the AGB across different species measured within our sampling plots.

Species	AGB Equations	Reference
Taxodium spp.	$AGB_{stem} = 0.008 \times (\sqrt{dbh \times Ht})^{3.10}$ $AGB_{branches} = 0.0004 \times (\sqrt{dbh \times Ht})^{3.20}$ $AGB_{foliage} = 0.0716 \times (\sqrt{dbh \times Ht})^{1.95}$	[34]
Pinus taeda	$AGB_{stem} = 0.0126 \times (dbh^{1.8167}) \times (Ht^{1.2609})$ $AGB_{branches} = 0.0660 \times (dbh^{1.7827}) \times (e^{0.0413 \times dbh}) \times (Ht^{-0.2300})$ $AGB_{foliage} = 0.9920 \times (dbh^{0.9076}) \times (e^{0.0707 \times dbh}) \times (H^{-0.8351})$	[35]
Pinus palustris	$AGB_{stem} = 0.0273 \times (dbh^{1.9745}) \times (Ht^{0.9163})$ $AGB_{branches} = 0.0070 \times (dbh^{3.6735}) \times (Ht^{-1.1735})$ $AGB_{foliage} = 0.0697 \times (dbh^{2.1631}) \times (Ht^{-0.5569})$	[36]
Pinus elliottii	$AGB_{stem} = 0.0049 \times (dbh^{1.8272}) \times (Ht^{1.6306})$ $AGB_{branches} =$ $0.0008 \times (dbh^{4.9276}) \times (e^{-0.0503 \times dbh}) \times (Ht^{-1.3343})$ $AGB_{foliage} = 0.6361 \times (dbh^{0.6790}) \times (e^{-0.4075 \times dbh}) \times (Ht^{0.0842})$	[35]
Fagaceae	$ln\ (AGB) = -3.030 + 2.498 \times ln(dbh)$	[33]

2.3. Data Acquisition and Processing

2.3.1. GEDI Aboveground Biomass Density

The GEDI Level 4A data product provides near-global AGBD (Mg ha^{-1}) estimates for each footprint collected along 25 m footprints spaced 60 m apart along-track, based on a three-dimensional measured structure. AGBD footprints were estimated from parametric linear models that correlate GEDI L2A waveform relative height metrics with AGBD [13]. A critical aspect of GEDI Level 4A data aimed to account for inter-annual variability in biomass and to enhance the robustness of our models against year-to-year changes in environmental conditions and vegetation dynamics.

We also implemented a series of filters to refine the data quality and ensure the accuracy of AGBD estimates. Initially, we selected GEDI L4A footprints where the "l4_quality_flag" equals 1 and the "degrade_flag" is set to 0, removing data not meeting the Level 4A product quality requirements. We also filtered footprints with beams lower than 4, based on the understanding that lower beam numbers correspond to higher data reliability for our specific research needs. We selected only data where "solar_elevation" was less than or equal to 0. This step ensured that all selected data points were collected at nighttime, eliminating the influence of sunlight on the GEDI L4A data. Finally, we focused on specific vegetative features of the Google Dynamic World dataset [37], selecting footprints that overlap with acquisition involving temporal selection to optimize the reliability and applicability of our models. The final set of footprints was resampled to a 30 m pixel size to ensure spatial alignment with the resolution of the remote sensing predictors.

2.3.2. Remote Sensing Predictors of AGBD

To enhance our analysis and improve the accuracy of upscaling GEDI L4A data, we expanded our dataset by incorporating additional satellite imagery sources, significantly enriching our suite of predictor variables. Utilizing the computational power of Google Earth Engine (GEE; [38]), we integrated data from the Harmonized Landsat Sentinel (HLS; [39]) and Sentinel 1C [40], along with ancillary imagery. This approach allowed us to generate an extensive collection of over 260 active and passive remote sensing predictor layers. We set a spatial resolution of 30 m to match the coarser resolution of HLS data in the stack, ensuring consistency across all layers.

We also defined a specific temporal window for the remote sensing predictors to ensure the consistency and relevance of our analysis. We selected data spanning from 1 March to 31 May for the years 2019, 2020, 2021, and 2022. This selection was informed by pre-modeling tests that highlighted this period as optimal for capturing stable and representative vegetation dynamics across the studied regions. Accordingly, this period corresponds to the dry season, characterized by slow vegetation growth and minimal biomass dynamics, which helps ensure canopy stability. Thus, rather than capturing biomass seasonality directly, we focused on minimizing seasonal variation in vegetation reflectance, which can introduce noise unrelated to actual AGBD and weaken model reliability [21,41]. Additionally, the data used here matched the earlier selection in GEDI in terms of the temporal window, ensuring consistency and compatibility in the analysis. We also conducted an extensive time-series analysis to ensure that vegetation dynamics remained stable and comparable across the selected years within this time frame. This analysis was crucial for assessing inter-annual variability and confirming that the selected period did indeed offer a window where vegetation characteristics, such as greenness, canopy structure, and biomass, showed minimal fluctuations from year to year.

1. Harmonized Landsat Sentinel (HLS)

We first acquired image data at the blue, green, red, near-infrared (NIR), shortwave infrared 1 (SWIR1), and 2 (SWIR2) bands from the HLS. The HLS project provides a consistent surface reflectance time series observed by Landsat 8 and 9 Operational Land Imager and the Sentinel-2A and 2B Multi-Spectral Instrument satellites. Consequently, the combined measurement of both sensors enables global observations of the land every 2–3 days at 30 m spatial resolution. The final data is achieved through a set of algorithms, encompassing atmospheric correction, cloud and cloud-shadow masking, spatial co-registration, common gridding, bidirectional reflectance distribution function normalization, and spectral bandpass adjustment. In our study, we accessed the Landsat HLS image catalog (HLSL30) through the GEE platform. All available HLSL30 across the study area with less than 30% of cloud cover were used.

2. Sentinel 1C

For acquiring and processing Sentinel 1C data, we used the interferometric-wide mode, a choice driven by its ability to provide high-resolution images suitable for detailed vegetation and land cover studies. In addition, we selected data with specific polarizations: vertical transmit/vertical receive (VV), and vertical transmit/horizontal receive (VH). These polarizations were chosen for their utility in capturing distinct aspects of surface features and vegetation.

From Sentinel 1C data, we further enhanced our dataset by calculating four radar indices. These indices, as detailed in Table S1, were specifically chosen to deepen our understanding of vegetation and surface characteristics.

3. Ancillary imagery

The acquisition of ancillary data complemented the primary datasets from GEDI L4A, Sentinel 1C, and HLS. A key component of this ancillary data came from NASADEM, which represents a significant enhancement over the original Shuttle Radar Topography Mission data. We extracted three topographic variables: elevation, slope, and aspect. Elevation data is fundamental in understanding the vertical distribution of vegetation and its correlation with biomass. Slope and aspect assess topography's influence on local climate conditions, soil moisture, and sunlight exposure, all of which are significant determinants of vegetation type and density.

We further augmented our analytical approach by incorporating the geographical coordinates of each footprint as critical predictor variables. The latitude and longitude information allowed us to contextualize biomass and vegetation data within a specific geographical context and enabled the modeling algorithm to explore spatial patterns and ecological dynamics across different forest patches.

2.4. Covariates Metrics and Imagery Stacking

We generated a comprehensive range of covariate metrics from the Sentinel 1C, HLS, and ancillary datasets. Spatial transformations, such as kernel window analyses [42] and the Gray Level Co-occurrence Matrix (GLCM; [43]), were calculated to capture the texture and structure of the landscape at a finer scale. For the kernel window analysis, we created 3×3 kernel covariates for variables obtained from HLS, Sentinel 1C, and DEM datasets. This involved calculating the mean, standard deviation, maximum, and minimum values of the variables within these 3×3 spatial windows. The smallest spatial window allowed us to quantify the immediate spatial variability and contextual information surrounding each pixel, minimizing the inclusion of surrounding heterogeneous land cover types, while offering insights into local heterogeneity in vegetation cover, biomass density, and terrain features.

GLCM texture analysis was applied exclusively to the HLS bands. The GLCM is a second-order statistical method that quantifies texture by measuring how often pairs of pixel values occur within a defined spatial relationship, in this case, a 3×3 moving window. It captures spatial patterns such as contrast, homogeneity, entropy, and correlation using 8-bit grayscale inputs to generate 18 standard texture indices [43]. These texture metrics are particularly relevant for interpreting variations in vegetation structure and density, enhancing the predictive power of our models by incorporating detailed information on the physical arrangement of vegetation.

A total of 260 covariates were generated and stacked into a consolidated dataset for further AGBD modeling and analysis (Table S1).

2.5. AGBD Modeling and Wall-to-Wall Mapping

In our study, we tested three distinct upscaling frameworks to estimate AGBD, each leveraging different methodologies (Figures 2 and 3). The first framework utilizes image-derived data, capitalizing on remote sensing technologies to upscale AGBD measurements from satellite imagery. The second is a sequential modeling strategy that first models variables derived from GEDI L2A and L2B before using these parameters to model biomass. The third framework relies on in situ data allometry, incorporating direct measurements from field plots such as tree height and diameter into allometric equations to calculate biomass. We used the GEE Python (version 3.11.13) application programming interface and the geemap package to access, process, and analyze the image collections and ancillary datasets in the Google Colab environment [44].

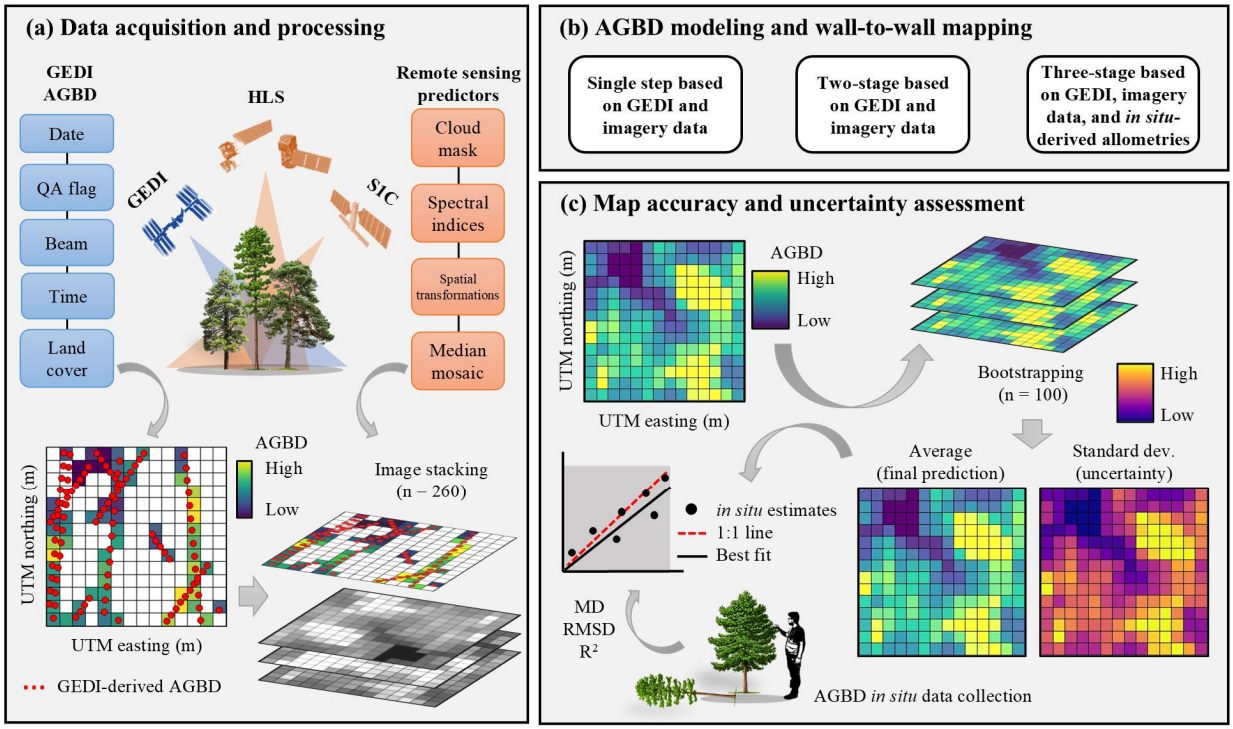


Figure 2. Flowchart of the proposed wall-to-wall ABGD upscaling frameworks used in this study. Panel (**a**) illustrates the steps for acquiring and processing GEDI AGBD data (light blue boxes) and other remote sensing predictors of AGBD (light orange boxes), resulting in 260 covariates. Panel (**b**) describes the three AGBD modeling and wall-to-wall upscaling frameworks used in the study, detailed in Figure 3. Panel (**c**) explains the map accuracy and uncertainty assessment process, showing how the final AGBD predictions and uncertainties were derived from bootstrapping (n = 100) and validated with in situ observations.

2.5.1. Upscaling Framework 1: Single Step Based on GEDI and Imagery Data

We modeled AGBD as a function of image-derived data (Figure 3a). This method began with a sampling phase where 1000 GEDI L4A footprints were selected from the image stack, ensuring that footprints from all years available were included. We employed semivariogram analysis to account for the spatial autocorrelation of the samples [45]. Initially, a semivariogram was fitted for all the footprints, revealing an effective range of 1.3 km. Based on this finding, we applied this distance as a limit between two sampled footprints to ensure that samples selected for analysis are statistically independent and not biased by spatial autocorrelation. Subsequently, the dataset underwent a split into training and validation subsets, adhering to a proportion of 70:30. This distribution was strategically chosen to optimize the learning process, allowing the model to train on a substantial portion of the data while reserving a significant subset for validation purposes.

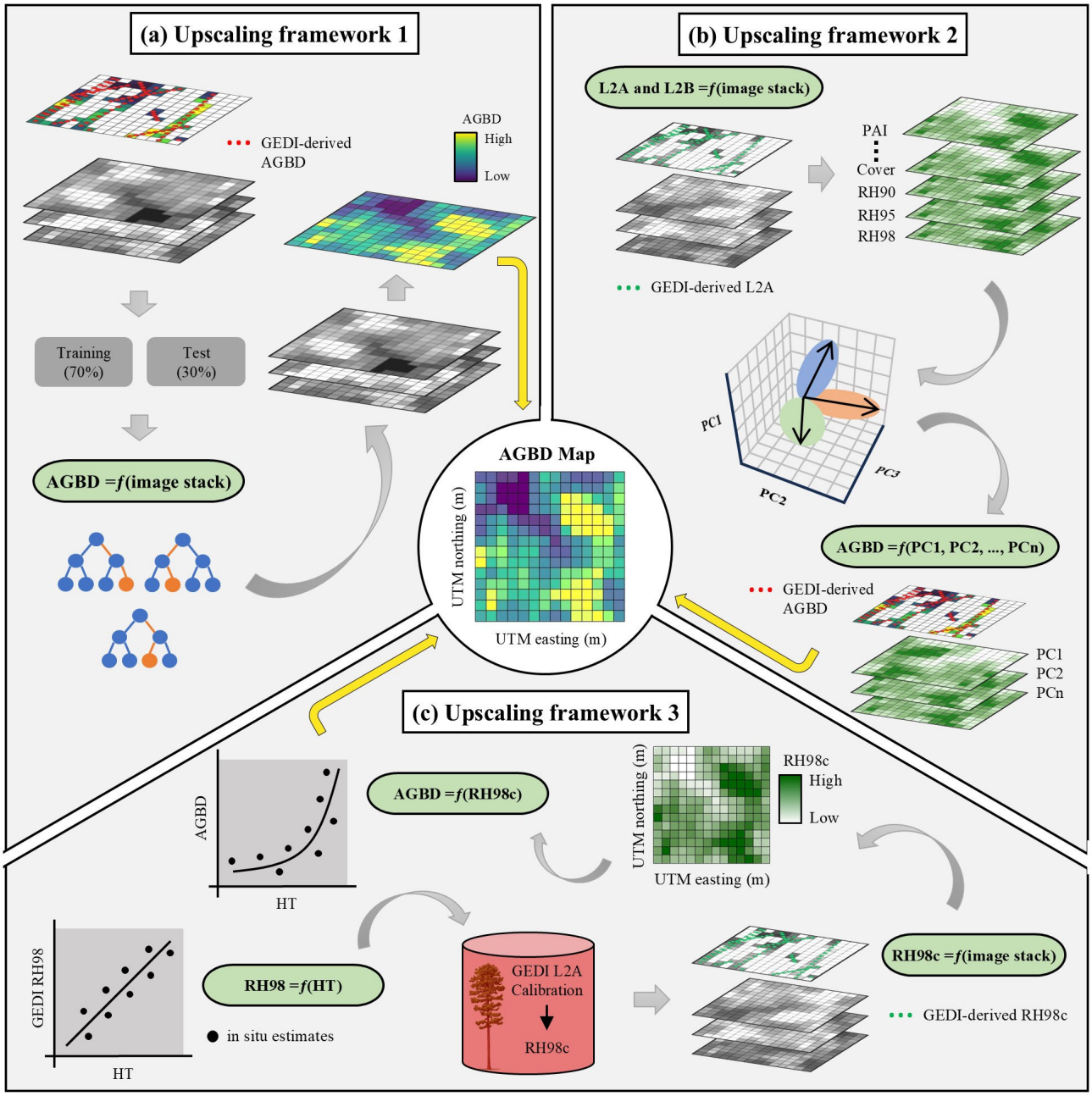


Figure 3. Illustration of overall steps used for (**a**) upscaling framework 1: single step based on GEDI and imagery data, (**b**) upscaling framework 2: two-stage based on GEDI and imagery data, and (**c**) upscaling framework 3: three-stage based on GEDI, imagery data, and in situ-derived allometries.

We employed an RF regressor using the Scikit-learn (sklearn) package [46] in Python (version 3.11.13) to estimate AGBD. We configured the RF with 250 trees to balance model complexity and computational efficiency while ensuring robust predictive performance. The sklearn SelectFromModel selected the most informative features, reducing dimensionality and enhancing its accuracy and generalizability. This process involves fitting an RF estimator to the entire dataset, which then computes a model-specific importance score for each feature based on the mean decrease in impurity. After these importance scores are determined, only features greater than or equal to the mean score across all features are retained for further model training. It is important to note that the RF's robustness to multicollinearity, due to its randomized feature selection and aggregation of multiple decision trees, minimizes the risk of inflating the influence of correlated predictors during the feature selection process. We kept the importance scores of the remaining variables for

further ranking and analysis. Finally, the trained model was applied to the comprehensive image stack, allowing us to predict the AGBD across the study area, transforming localized model outputs into a continuous, spatially explicit AGBD map.

We used a bootstrap procedure of 100 iterations to ensure robustness and reliability in our outcomes. During each bootstrap iteration, a random sample of the data was selected with replacement to train the RF model, and an AGBD map was generated based on the model's predictions for that iteration. This approach allowed us to assess the stability and consistency of the model across different subsets of data. The final AGBD prediction was derived by taking the average of the 100 AGBD maps generated from each bootstrap iteration, providing a consensus estimate of biomass density. Additionally, the uncertainty associated with this prediction was quantified using the standard deviation of the values across the 100 maps.

Variable importance analysis was conducted after all bootstrap iterations. During each run, variables were selected based on their contribution to model accuracy, as quantified by the mean decrease in impurity. After completing all bootstrapping iterations, we evaluated the frequency with which each variable was selected across these runs and calculated the cumulative mean decrease in impurity for each variable. To determine the final ranking of variable importance, we summed the mean decrease in impurity for each variable across the bootstrap procedure. This strategy, combined with RF's robustness to multicollinearity, helped mitigate the influence of highly correlated predictors.

2.5.2. Upscaling Framework 2: Two-Stage Based on GEDI and Imagery Data

In the second framework of our study, we modeled AGBD by focusing on the structural attributes of the forest: variables derived from GEDI L2A and L2B data (Figure 3b). The sampling design and RF settings were consistent with those used in our previous method, ensuring a uniform method for comparison and integration across different modeling strategies.

We began by sampling 1000 GEDI footprints, then proceeded with a detailed modeling process. Relative height metrics at 10% (RH10), 15% (RH15), 20% (RH20), 25% (RH25), 30% (RH30), 35% (RH35), 40% (RH40), 45% (RH45), 50% (RH50), 55% (RH55), 60% (RH60), 70% (RH70), 75% (RH75), 80% (RH80), 85% (RH85), 90% (RH90), 95% (RH95), and 98% (RH98), were modeled using the GEDI L2A data as functions of the image stack, resulting in individual maps for each metric. Canopy cover, foliage height diversity (FHD), and plant area index (PAI) were modeled using the GEDI L2B data, also as functions of the image stack, to produce respective maps. Since GEDI L4A is derived from the same LiDAR pulses as GEDI L2A and L2B, these footprints are naturally co-located.

From GEDI L2A and L2B-derived maps, we applied a Principal Component Analysis (PCA) to minimize multicollinearity issues among predictors and enhance model interpretability. By transforming the data, PCA ensures that AGBD is not directly modeled as a function of height metrics, such as those derived from GEDI L2A data, which is essential because GEDI AGBD itself is derived from height. This approach avoids potential biases and redundancy, enabling a more robust and interpretable modeling framework. PCA works by transforming the input L2A and L2B maps into a set of orthogonal vectors or principal components (PCs), which preserve the information contained in the original data. Each PC accounts for a portion of the total variance in the input data. In this study, we retained the number of PCs that collectively explained 99.9% of the total data variance, which allowed us to preserve meaningful information while minimizing uncorrelated noise.

Maps of each selected PC were then generated, providing a transformed representation of the original L2A and L2B variables. These maps were then used as inputs to model AGBD using GEDI L4A data, allowing us to upscale the AGBD predictions across the study

area. We also performed a bootstrap of 100 estimates to ensure robustness. The final AGBD prediction consisted of the average of these 100 maps, with uncertainty quantified by the standard deviation of the map values.

2.5.3. Upscaling Framework 3: Three-Stage Based on GEDI, Imagery Data, and In Situ-Derived Allometries

The third framework extracted AGBD biomass estimates across the study area from in situ allometric equations (Figure 3c). To integrate this method with remote sensing data, we began by simulating GEDI waveforms using the rGEDI package [47], which allowed us to generate GEDI-like data from Airborne Laser Scanning (ALS) data. ALS data was acquired from the United States Geological Survey (USGS) from November 2018 to January 2020. This dataset presents a point density of approximately 10 points per square meter, 2.5 m ground sample distance, and vertical accuracies of 11.13 cm in non-vegetated areas and 18.01 cm in vegetated areas [48]. The gediWFSimulator function was configured with specific geographic coordinates for the center of each field plot, ensuring that the simulated waveforms accurately represented the spatial location of the ground measurements. Further refinement of the waveform data was achieved using the gediWFMetrics function. Noise linkage was set to values typical for GEDI operations (3.0103 mean and 0.95 multiplier), to mimic the noise characteristics of actual GEDI data. The maxDN parameter was set to 4096, specifying the maximum digital number, and the sWidth was set to 0.5, indicating the width of the Gaussian smoothing applied to the waveform. The varScale, set to 3, adjusted the scaling of variance components within the waveform metrics calculation. We then fit a linear model between the simulated GEDI RH98 values and the maximum tree height measured inside each plot. Using this regression model, we calibrated the GEDI L2A footprints to reflect more accurately the real-world tree height measurements observed in the field, thereby improving the accuracy of our AGBD estimates.

We proceeded to model the calibrated height (RH98c) as a function of the image stack. The method's consistency was also considered by maintaining the sampling design, RF settings, and bootstrapping strategy with those used in the previous methods. We applied an allometric equation that relates aboveground biomass to canopy height using a power-law regression function [49]. This relationship was expressed through a power function, capturing the non-linear relationship typically observed between tree height and biomass accumulation. By leveraging this power function, we could estimate the AGBD and derive a continuous wall-to-wall map based on the heights indicated in the RH98c map.

2.5.4. Modeling and Internal Validation

The model performance was evaluated by calculating key statistical metrics. Specifically, we computed the mean difference (MD; Equation (1)) and relative mean difference (%MD; Equation (2)) to measure the average deviation of the predicted AGBD from the observed values, which allowed us to understand the model's overestimation or underestimation. In addition, we calculated the Root Mean Square Difference (RMSD; Equation (3)) and relative RMSD (Equation (4)), which provided insight into the overall error magnitude of the predictions by measuring the square root of the average squared differences between predicted and actual values. Lastly, the R-squared (R^2; Equation (5)) value was determined to assess the proportion of variance in the dependent variable that is predictable from the independent variables, indicating the model's explanatory power.

$$MD(Mg\ ha^{-1}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \quad (1)$$

$$\%MD = \frac{MD}{\bar{y}} \times 100 \quad (2)$$

$$RMSD\left(Mg\ ha^{-1}\right) = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad (3)$$

$$\%RMSD = \frac{RMSD}{\overline{y}} \times 100 \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})} \quad (5)$$

where y_i is the reference AGBD in the GEDI footprint i, $\hat{y}_i$ is our estimated AGBD in the GEDI footprint i, and $\overline{y}$ is the mean of the reference AGBD sample used for testing. %RMSD (Equation (4)) and %MD (Equation (2)) are calculated by dividing the respective absolute values (Equations (1) and (3)) by the mean of GEDI AGBD observations used for model testing.

2.6. Map Accuracy and Uncertainty Assessment

2.6.1. Error Propagation and Map Uncertainty Assessment

For each upscaling framework in our study, we generated a final AGBD prediction map along with an accompanying uncertainty map to evaluate spatial mapping accuracy and predictive performance across areas and frameworks. Statistical metrics were utilized to capture and quantify spatial uncertainty in the predictions, providing insights into how variability and precision change across the mapped area. The AGBD estimated mean value $\widehat{E(\mu)}$, calculated as the average of final predicted values within each area, serves as the central tendency of the predictions. Accompanying this mean, the 95% confidence interval provides a range within which the true mean AGBD is expected to fall. To assess spatial variability, the total variance $\widehat{VE(\mu)}$ was calculated as the sum of variances of pixel-level bootstrap estimates, which quantifies the prediction variability and reflects the degree of uncertainty or consistency in the model outputs. From this variance, the standard error ($\widehat{SE}$) is computed, which measures the average amount by which the estimated means are expected to deviate from the actual population mean. Lastly, the relative standard error ($\%\widehat{SE}$) relates this standard error to the mean AGBD, providing a relative measure of uncertainty, which is particularly useful for comparing precision across different models or datasets.

We conducted a Kruskal–Wallis [50] test at a 95% confidence level to ensure the robustness of our comparisons and to determine if there were significant differences among the final predictions and uncertainties. The null hypothesis (H0) for the Kruskal–Wallis test is that the median values of the groups are equal. The alternative hypothesis (H1) is that at least one group's median is different, suggesting a significant difference among the frameworks. By using this test, we assessed the differences without assuming normality in the data. If the test indicated significant differences, we conducted post hoc pairwise comparisons using the Wilcoxon rank-sum test [51], a non-parametric method for non-normally distributed data, to identify which specific framework differed from each other.

$$\widehat{E(\mu)}(Mg\ ha^{-1}) = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{1}{N}\sum_{i=1}^{N} m_{ij}\right) \quad (6)$$

$$\widehat{VE(\mu)} = \sum_{j=1}^{k}\left(\frac{1}{N}\sum_{i=1}^{N}(m_{ij} - \hat{m}_j)^2\right) \quad (7)$$

$$\widehat{SE}\left(Mg\ ha^{-1}\right) = \sqrt{\widehat{VE(\mu)}} \quad (8)$$

$$\%\widehat{SE} = \frac{\widehat{SE}}{\widehat{E(\mu)}} \times 100 \quad (9)$$

where k is the number of pixels in each map, N is the number of bootstrap runs, m_{ij} is the value at position j in the i-th bootstrap map, $\hat{m}_j$ is the estimated mean value at position j across all bootstrap maps.

We also compared uncertainties from bootstrapping to evaluate whether the AGBD estimates for a pixel, when comparing paired frameworks, were statistically different. For each AGBD pixel, we collected 100 values from the bootstrapping process of each framework. We then paired these 100 values from one upscaling framework with the corresponding 100 values from another for the same pixel. We applied the Wilcoxon Signed-Rank test to evaluate whether the differences between pairs were significant, determining if the distributions of the two sets of bootstrap results were significantly different, based on a p-value threshold set at 0.05. Pixels were classified as significantly different (supporting the alternative hypothesis H1) or not significantly different (supporting the null hypothesis H0). We visualized these comparisons by plotting the pixels, allowing us to spatially identify where the upscaling frameworks yielded equivalent results and where they diverged. Additionally, we analyzed the significance of the pixels in relation to the AGBD values, examining the value ranges where the methods most differed or agreed.

2.6.2. Map External Validation with In Situ Observations

To assess the accuracy and reliability of the AGBD maps, we conducted a thorough external validation using in situ data collected from the 27 designated plots. This validation process involved comparing predicted AGBD values from the maps to actual measurements obtained on the ground. It is important to note that these plots were not used for training or internal validation of the models. Key statistical metrics were calculated to quantify the accuracy and precision of the AGBD maps, including MD, %MD, RMSD, %RMSD, and R^2.

To further ensure the statistical robustness of our comparisons between in situ data and model estimates for each upscaling framework, we conducted a series of statistical tests. Initially, we performed a Shapiro–Wilk test [52] to check for the normality of the distribution of differences between predicted and observed values. Given that the residuals did not necessarily follow a normal distribution, as indicated by the Shapiro–Wilk test results, we proceeded with the Wilcoxon Signed-Rank test. This test compared in situ and model estimates to determine if their population mean ranks differed significantly. We applied this test at a 95% confidence level, with a p-value threshold of less than 0.05 indicating statistically significant differences. This step confirmed whether the deviations observed were systematic across the dataset or due to random variation, thereby reinforcing the validity of our model evaluation process.

3. Results

3.1. AGBD Modeling Performance and Internal Validation

The upscaling framework 1 demonstrated reasonable modeling prediction accuracy, evidenced by modest MD ranging from −4.8 to 9.2 Mg ha^{-1} and %MD from −5.8% to 12.7% (Figure 4a,b). This suggests that it offers reliable estimates with a moderate level of variability in biomass estimation based on the internal validation dataset. Framework 2 displayed similar MD, ranging from −6.9 to 8.1 Mg ha^{-1}, and %MD from −11.5% to 15.6%, indicating slightly more variability in its predictions (Figure 4a,b). It could still be viewed as balancing bias and precision, although its predictive accuracy appears slightly reduced compared to the first framework. Conversely, framework 3 showed significantly higher deviations in MD, ranging from −15.9 to 2.1 Mg ha^{-1}, and %MD from −18.6% to 2.7%, highlighting

relatively poorer modeling performance (Figure 4a,b). The higher values suggest greater errors and a less reliable model for predicting biomass compared to the others.

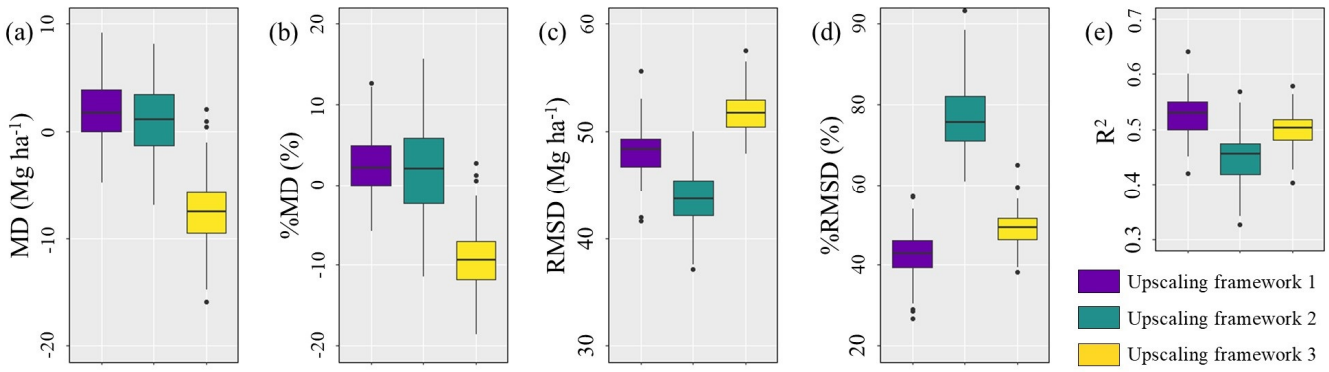


Figure 4. A comparison of model performance across 3 upscaling frameworks in terms of (**a**) mean difference (MD); (**b**) relative mean difference percentage (%MD); (**c**) root mean squared difference (RMSD); (**d**) relative root mean squared difference (%RMSD); and (**e**) R^2 for the testing datasets across upscaling frameworks within the study areas.

Framework 2 returned higher precision in the biomass predictions, indicated by low RMSD (median of 43.7 Mg ha^{-1}, compared to 48.3 and 51.7 Mg ha^{-1} for 1 and 3, respectively; Figure 4c). However, framework 1 showed the lowest median %RMSD at 53.3%, outperforming framework 2 and 3 (77.8 and 64.6%, respectively; Figure 4d). The higher %RMSD observed in framework 2 can be attributed to its lower mean AGBD value, which amplifies relative error despite its lower absolute RMSD. These internal validation metrics indicate that while framework 2 offers more precise predictions in absolute terms, framework 1 achieves better relative accuracy and lower bias. Furthermore, framework 1 also accounted for a greater extent of the variability in AGBD, with a median R^2 of 0.53 compared to 0.46 and 0.50 for frameworks 2 and 3, respectively (Figure 4e).

Framework 1 used sklearn's SelectFromModel to determine the most relevant covariates for modeling AGBD from a set of 260 potential variables. On average, it identified 34 ± 5 covariates as significant contributors to the model. Based on the average results from SelectFromModel, the red band was the most crucial spectral variable, with its kernel covariates R_MIN, R_MAX, and R_MEAN, and the GLCM sum average R_SAVG ranking among the top ten of all covariates (Figure 5). Similarly, the green band also shows importance, with G_MIN, G_MAX, and G_AVG also making it to the top ten (Figure 5). The inclusion of Longitude and Latitude highlights the influence of spatial location in determining biomass distribution. Additional covariates from passive optical imagery such as CMR, GVMI, CVI, SWIR2, and FSOIL also feature prominently (Figure 5). Notably, GVMI and CMR, which incorporate SWIR1 and SWIR2 bands, underline the significance of moisture content in vegetation and soil properties in biomass estimation. Radar-based covariates were also important in the model, indicating the utility of radar data in penetrating cloud cover and providing reliable data irrespective of weather conditions. Finally, ancillary information such as Elevation and Slope, though ranking lower, provided context to the model, aiding in the understanding of topographical influences on biomass distribution.

3.2. AGBD Wall-to-Wall Maps Predictions and Uncertainties

The AGBD wall-to-wall maps generated using the three different upscaling frameworks across various areas visually underscore the quantitative findings from our analysis (Figure 6). Additionally, the distribution of AGBD modeled reveals varied outcomes for each study area (Figure 7). For FAS Millhopper Unit and Austin Cary Forest areas, all frameworks returned similar distributions overall. However, framework 2 detected patches

of low AGBD values and lacked the high AGBD patches presented in frameworks 1 and 3, suggesting it may underestimate biomass in certain areas. For the Austin Cary Forest, framework 1 predicted a higher estimated AGBD mean value $\widehat{E(\mu)}$ of 151.1 Mg ha^{-1}, followed by framework 3 with a slightly lower value of 140.9 Mg ha^{-1}, and framework 2 with the lowest at 115.1 Mg ha^{-1} (Figure 7a). Similar trends were observed in the FAS Millhopper Unit, where framework 1 again yielded the highest mean AGBD of 131.0 Mg ha^{-1}, followed by framework 3 at 125.1 Mg ha^{-1} and framework 2 at 92.6 Mg ha^{-1} (Figure 7b).

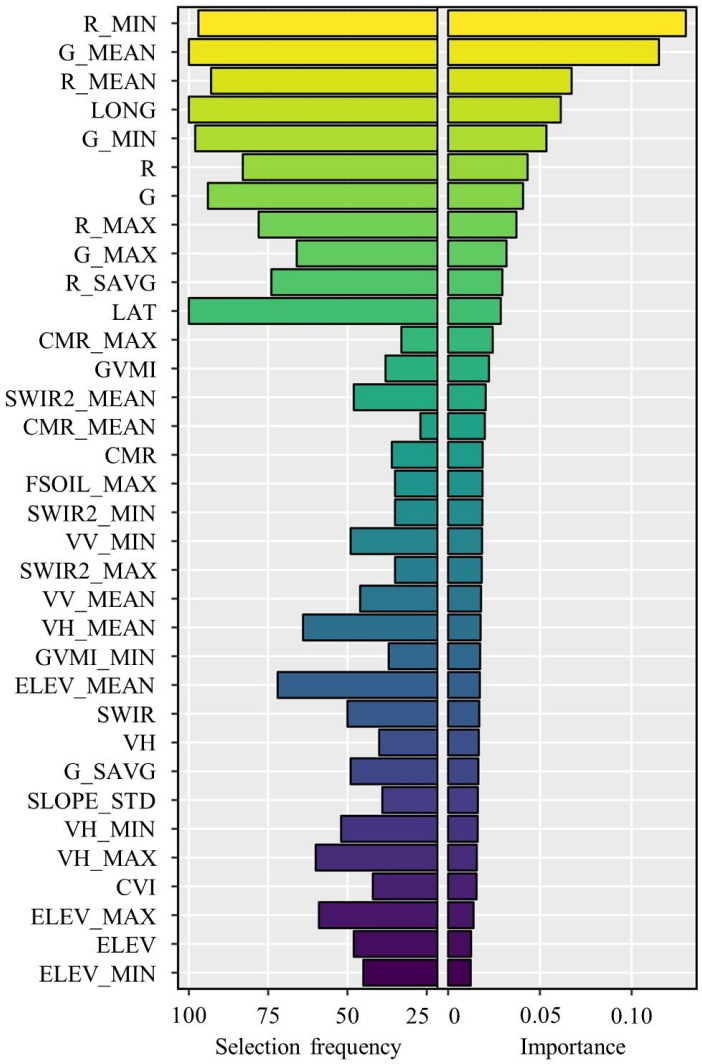


Figure 5. The most important covariates in bootstrapping analysis of upscaling framework 1: on the left, the frequency with which each covariate was selected across multiple iterations, and on the right, the overall importance of each covariate in influencing the final AGBD prediction.

The Myakka State Forest showed a sharp contrast in the AGBD estimates. Framework 1 predicted an estimated AGBD mean value of 36.4 Mg ha^{-1}, framework 2 predicted a lower value of 18.5 Mg ha^{-1}, and framework 3 predicted 39.7 Mg ha^{-1}. The distribution of framework 2 differed from the others, showing values centered on lower AGBD and broader distributions (Figure 7c). Lastly, for the Okaloacoochee Slough State Forest, framework 1 predicted higher biomass values at 65.4 Mg ha^{-1} across the area, while frameworks 2 and 3 presented more moderate estimates at 48.4 Mg ha^{-1} and 44.4 Mg ha^{-1}, respectively. The distributions of AGBD values across the site were similar (Figure 7d).

Uncertainty maps (Figure 8) and uncertainty distributions (Figure 9) associated with each upscaling framework across different forest areas provided insights into the reliability

of each method. Uncertainty values for framework 1 were high in Austin Cary Forest, averaging around 81.5 Mg ha^{-1}, reflecting substantial variability in the estimates. Framework 2 showed lower uncertainties around 64.5 Mg ha^{-1} indicating a marginally more stable output. Framework 3 returned the highest average value of 91.0 Mg ha^{-1} (Figure 9a). FAS Millhopper Unit also presented higher uncertainty values across frameworks, indicating lower confidence in the biomass estimates in this area, with averages at 75.1 Mg ha^{-1} in framework 1, 54.9 Mg ha^{-1} in framework 2, and 82.5 Mg ha^{-1} in framework 3 (Figure 9b). For areas in south Florida, Myakka State Forest was markedly lower, especially for framework 2, with values at 17.4 Mg ha^{-1}, reflecting more consistent and reliable predictions. In contrast, framework 1 and 3 show slightly higher uncertainties at 29.8 and 35.1 Mg ha^{-1}, respectively (Figure 9c). Okaloacoochee Slough State Forest shows an uncertainty of 62.1 Mg ha^{-1} for framework 1, which is higher than 2 at 36.6 Mg ha^{-1} and 3 at 42.7 Mg ha^{-1} (Figure 9d).

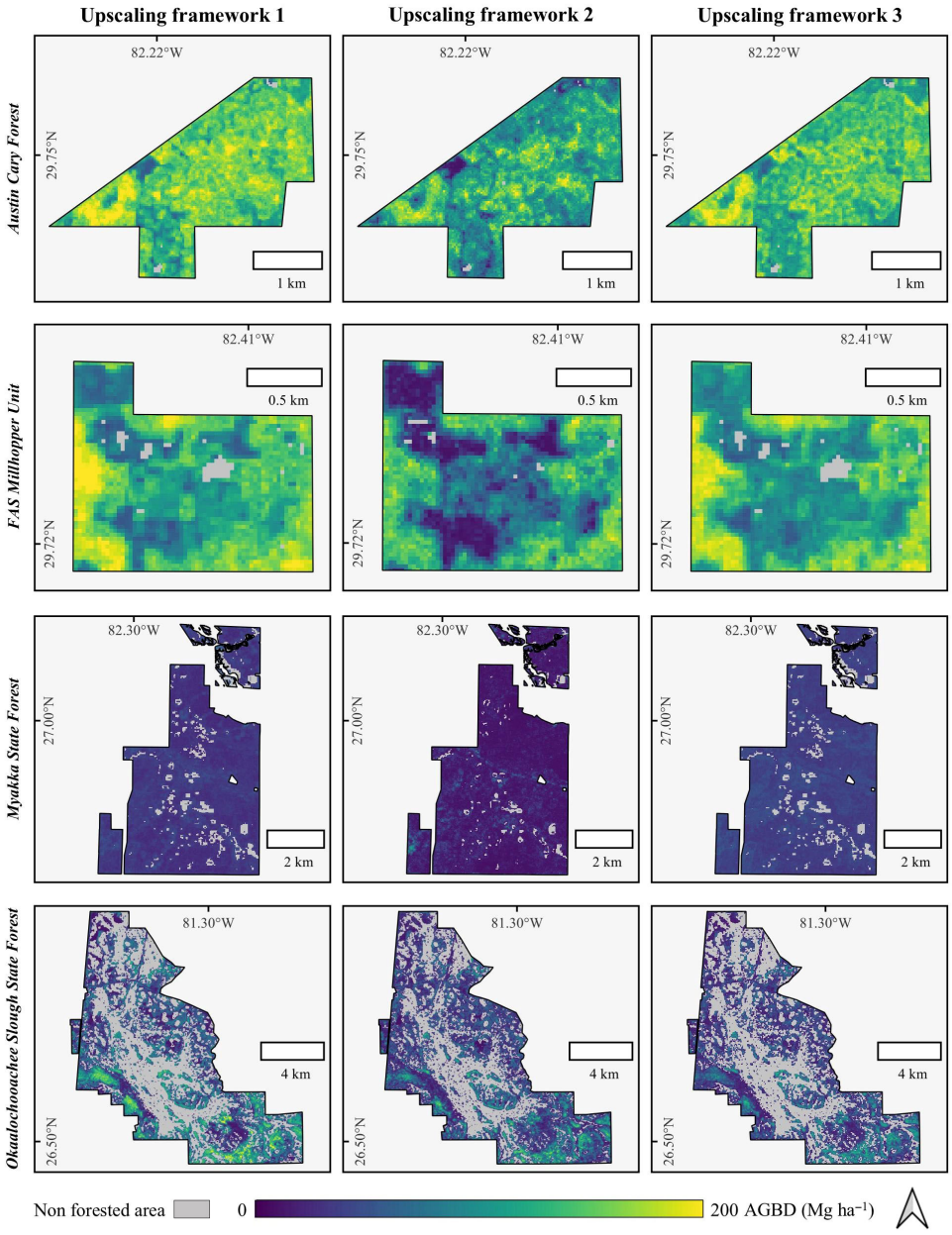


Figure 6. Spatial distribution of AGBD across each of our 4 study areas relative to our three upscaling frameworks.

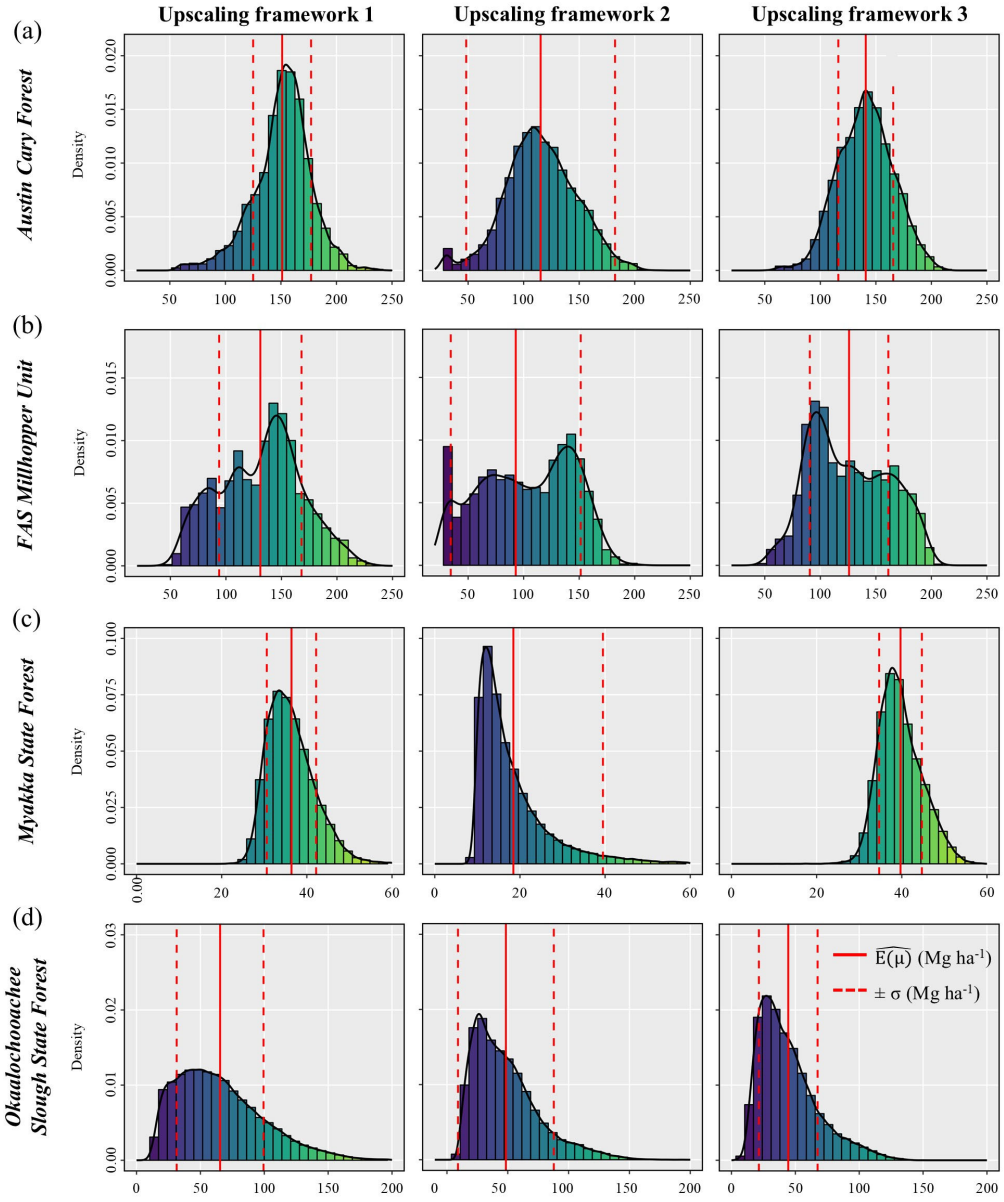


Figure 7. Distribution of AGBD estimates across (**a**) Austin Cary Forest, (**b**) FAS Millhopper Unit, (**c**) Myakka State Forest, and (**d**) Okaloacoochee Slough State Forest, relative to our three upscaling frameworks. Solid red lines represent the estimated mean values $\widehat{E(\mu)}$ and dashed red lines represent the standard deviation of the estimated AGBD values (σ).

Besides the estimated mean values previously described, error measures were also important statistics to compare our upscaling frameworks' outcomes (Table 2). Additionally, they were statistically different across frameworks, highlighting variations in their performance. In Austin Cary Forest and FAS Millhopper Unit, framework 1 showed moderate standard error and precision, while framework 3 had slightly higher errors and less precision. Framework 2's precision fell between frameworks 1 and 3. For Myakka State Forest, Framework 3 had the highest standard error while framework 1 kept the lowest error. Finally, Okaloacoochee Slough State Forest presented high standard error for frameworks 1 and 3, with framework 2 showing better precision.

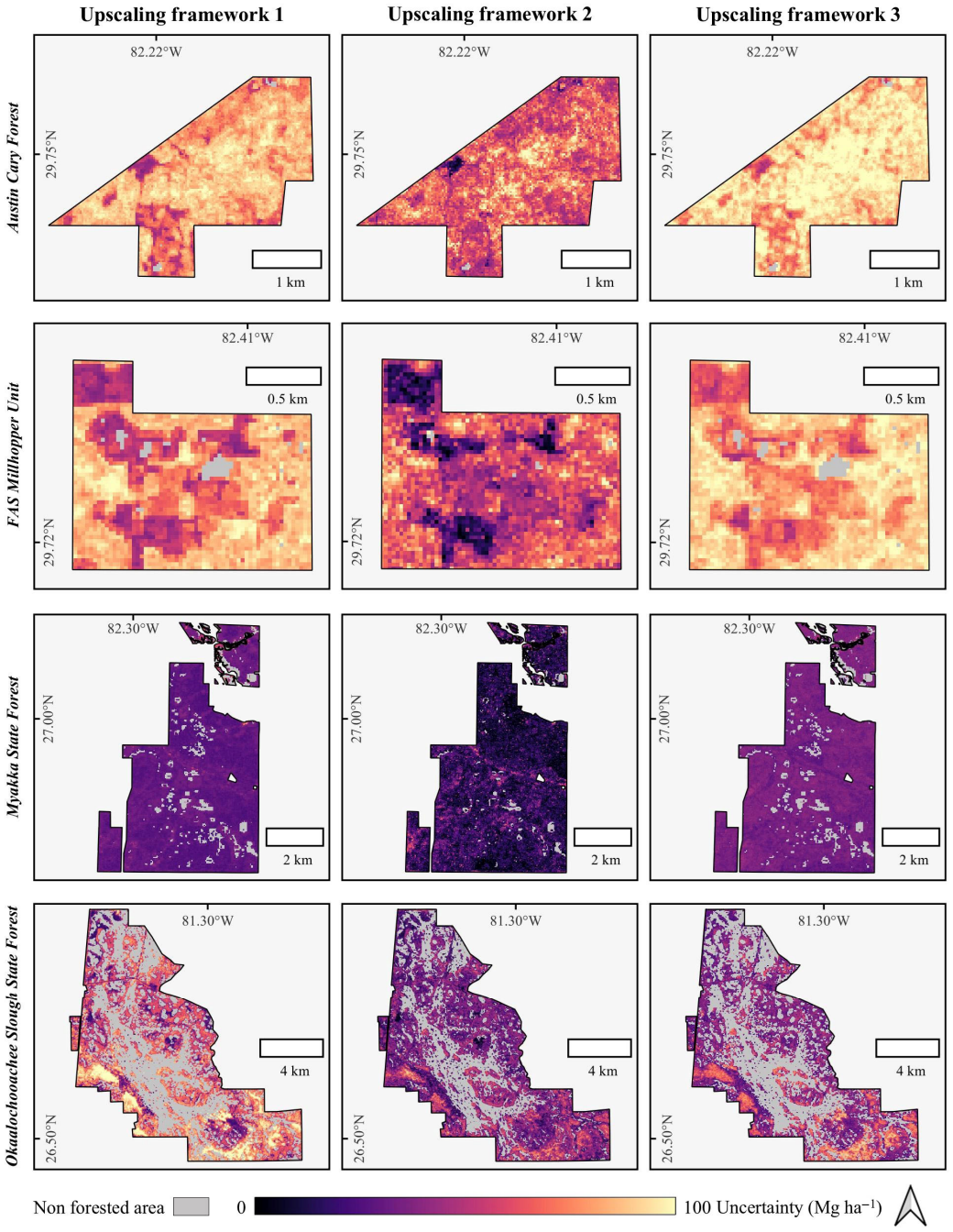


Figure 8. Uncertainty estimates of AGBD across our 4 study areas relative to our 3 modeling upscaling frameworks. Warmer colors indicate higher levels of uncertainty.

Table 2. Summary of the AGBD and variance estimates for each area and upscaling framework. The estimated mean value $\widehat{E(\mu)}$: estimated AGBD mean value, $V\widehat{E(\mu)}$: total variance, standard error ($\widehat{SE}$), and relative standard error (%$\widehat{SE}$). Different letters indicate a significant difference ($p < 0.05$).

Site	Upscaling Framework	$\widehat{E(\mu)}$(Mg ha^{-1})	$V[\widehat{E(\mu)}]$	$\widehat{SE}$(Mg ha^{-1})	%$\widehat{SE}$
Austin Cary Forest	1	151.1 ± 164.7 ^{a}	6784.4	82.4 ^{a}	54.5
	2	115.1 ± 134.0 ^{b}	4489.2	67.0 ^{b}	58.2
	3	140.9 ± 183.2 ^{c}	8394.2	91.6 ^{c}	65.0
FAS Millhopper Unit	1	131.0 ± 153.7 ^{a}	5907.5	76.9 ^{a}	58.7
	2	92.6 ± 117.0 ^{b}	3422.7	58.5 ^{b}	63.2
	3	125.1 ± 167.1 ^{c}	6977.0	83.5 ^{c}	66.3

Table 2. *Cont.*

Site	Upscaling Framework	$\widehat{E(\mu)}$ (Mg ha^{-1})	$V[\widehat{E(\mu)}]$	$\widehat{SE}$ (Mg ha^{-1})	$\%\widehat{SE}$
Myakka State Forest	1	36.4 ± 60.4 [a]	911.0	30.2 [a]	83.0
	2	18.5 ± 42.1 [b]	442.3	21.0 [b]	113.5
	3	39.7 ± 70.5 [c]	1242.1	35.2 [c]	88.7
Okaloacoochee Slough State Forest	1	65.4 ± 130.9 [a]	4281.0	65.4 [a]	100.0
	2	48.4 ± 78.8 [b]	1551.6	39.4 [b]	81.4
	3	44.4 ± 90.1 [c]	2031.2	45.2 [c]	101.6

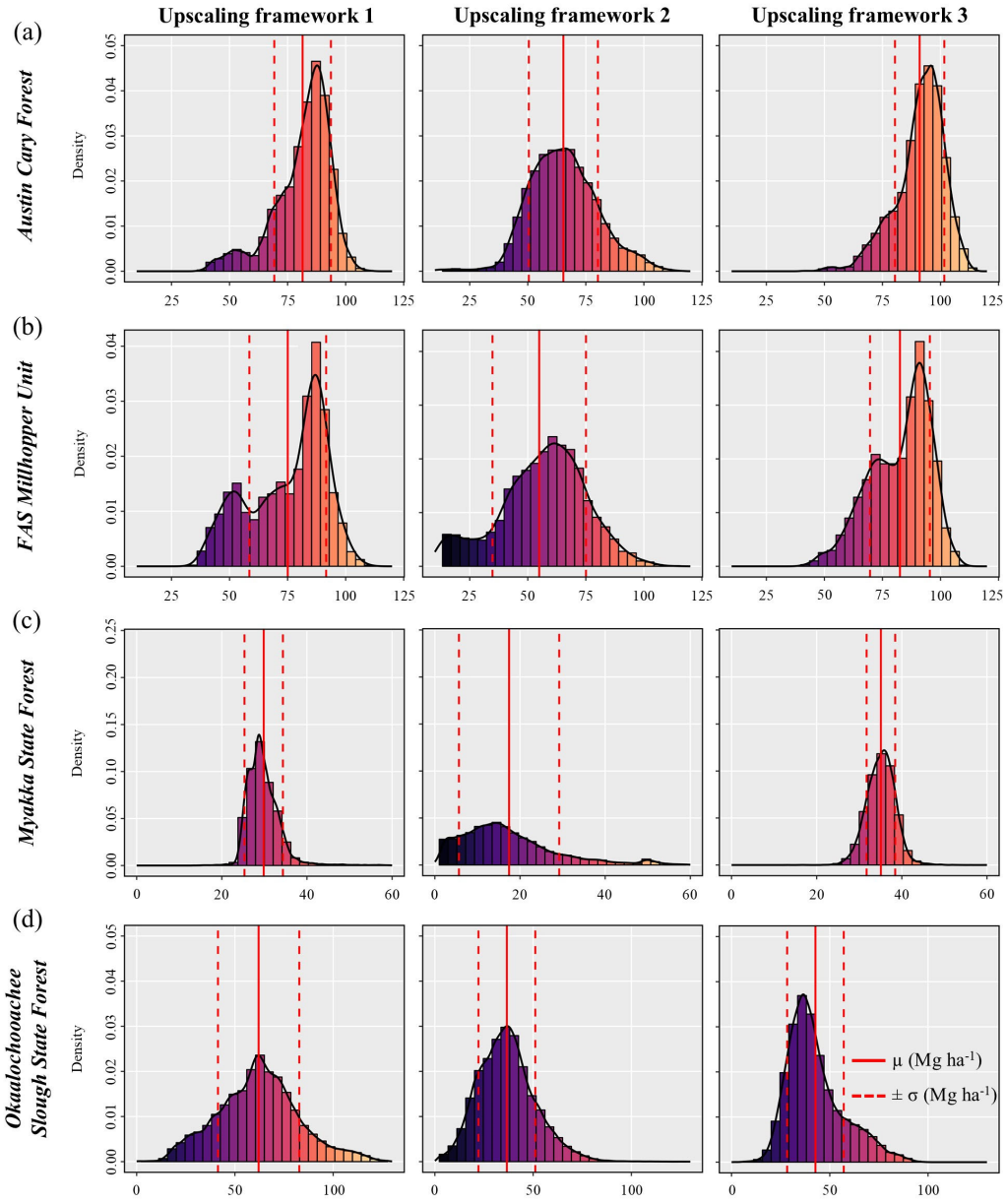


Figure 9. Distribution of uncertainty estimates across in (**a**) Austin Cary Forest, (**b**) FAS Millhopper Unit, (**c**) Myakka State Forest, and (**d**) Okaloacoochee Slough State Forest, relative to our three upscaling frameworks. Solid red lines indicate the uncertainty average for each site (μ) and dashed red lines indicate its standard deviation (σ).

There was high variability in the uncertainties from bootstrapping when comparing paired frameworks across different AGBD values and study area (Figure 10). Frameworks 1 and 2 demonstrated similar predictions for AGBD values up to 60 Mg ha^{-1},

beyond which significant differences began to emerge, indicating that these frameworks perform comparably under lower to moderate biomass conditions but diverge in higher biomass settings. In contrast, frameworks 2 and 3 were almost entirely significantly different, particularly for lower AGBD values up to 25 Mg ha^{-1}. Similarly, Frameworks 2 and 1 also showed significant differences in lower AGBD values, mainly up to 60 Mg ha^{-1}, suggesting divergent performance in low to moderate biomass settings. Discrepancies were more pronounced in regions like Myakka State Park and FAS Millhopper Unit, which exhibit lower AGBD values and thus more pronounced differences.

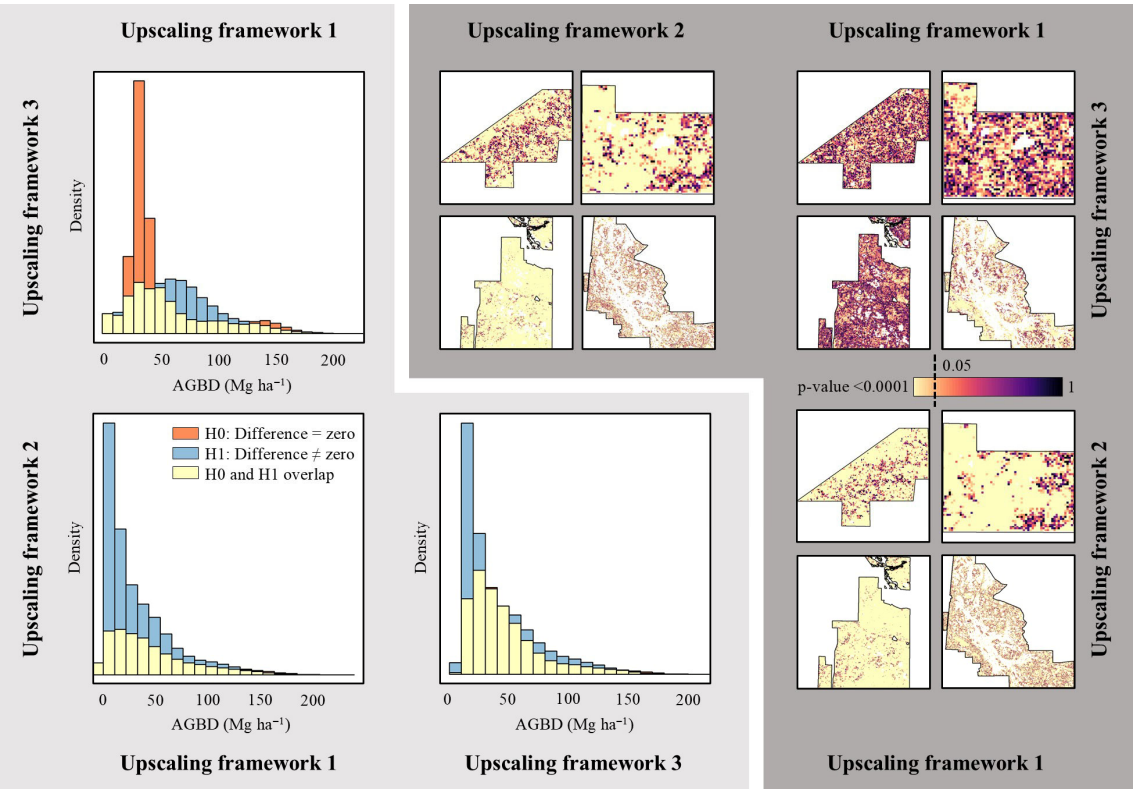


Figure 10. Comparison analysis between upscaling frameworks. Frequency histograms in light gray background display the significance of the pixels in relation to the AGBD values, examining the value ranges where the methods most differed or agreed. Study area maps in the dark gray background spatially illustrate the distribution of *p*-values from paired upscaling frameworks, where warmer colors indicate a lower *p*-value.

3.3. AGBD Map Accuracy and External Validation

External validation of mapping upscaling frameworks compared the predicted AGBD values to observations from the 27 designated field plots. Framework 1 demonstrated a high correlation with in situ measurements, evidenced by an R^2 of 0.93, indicating satisfactory predictive accuracy (Figure 11). However, it also showed a slight underestimation with a MD of −10.5 Mg ha^{-1} and %MD of −9.5%. The RMSD is 43.3 Mg ha^{-1}, with a percentage of 39.3%, indicating a moderate spread in the prediction errors. Framework 3 underestimated biomass more significantly than upscaling framework 1, with an MD of −19.3 Mg ha^{-1} and an %MD of −16.9%. The variability in its estimates is also higher, as shown by an RMSD of 51.0 Mg ha^{-1} and an %RMSD of 44.7%, which suggests a greater inconsistency in the predictions. Framework 2 demonstrated a very strong correlation with in situ measurements, with an R^2 of 0.95. However, it showed higher underestimation compared to the other upscaling frameworks, with an MD of −31.7 Mg ha^{-1} and an %MD of −31.8%, indicating a tendency towards underestimation. The variability in upscaling framework 2's estimates was also higher, with an RMSD of 54.6 Mg ha^{-1} and an %RMSD of 54.7%. Despite

these discrepancies, upscaling framework 2 provided robust predictive power in external validation, as reflected in its high R^2 value.

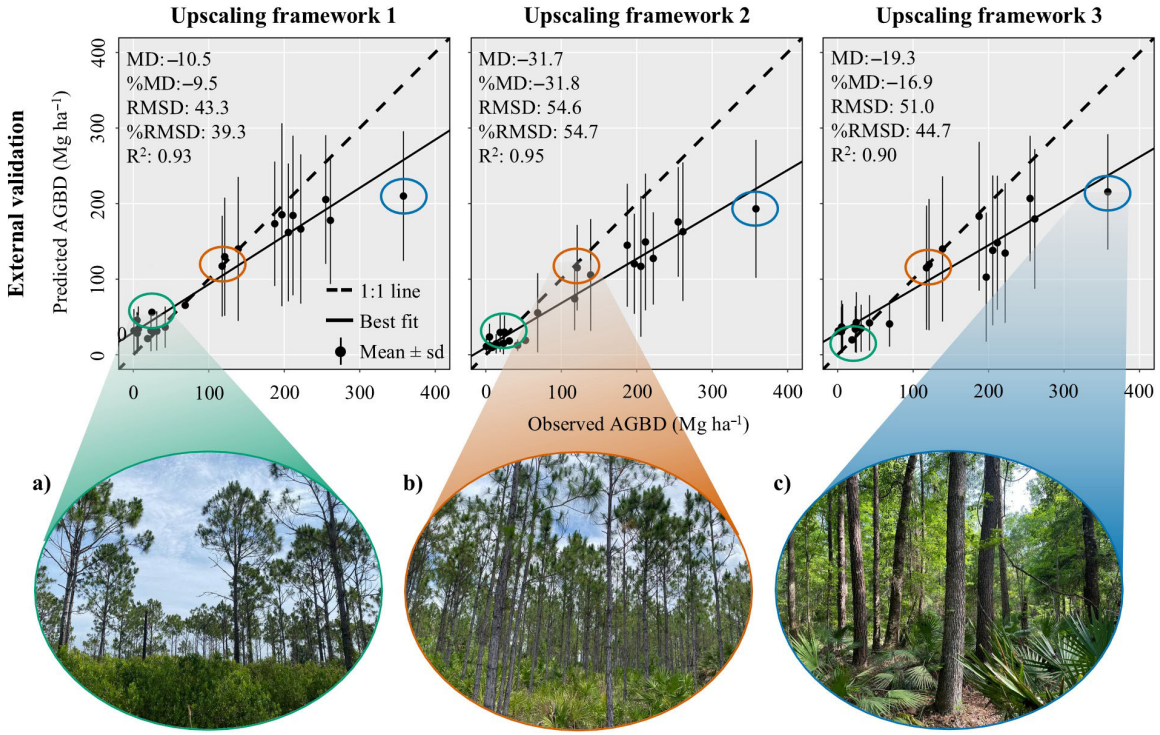


Figure 11. Comparison of in situ estimates and derived wall-to-wall AGBD estimates (average of the 100 bootstrapping runs) across the mapping upscaling frameworks. The black dashed line represents the 1:1 relationship, the solid black line is the best-fit linear model for the pairwise measurements, and vertical lines indicate the standard deviation from each estimate. Statistical metrics are computed by the linear relationship between predicted AGBD and the 27 field plots (observed AGBD). Field pictures illustrate real AGBD conditions of selected plots, highlighting (**a**) an underestimated, (**b**) a well-fitted, and (**c**) an overestimated example.

The Wilcoxon test results indicated no significant difference between the predicted and observed AGBD for all upscaling frameworks, suggesting that all three upscaling frameworks provided predictions consistent with the in situ measurements. Framework 1 yielded a V value of 155 and a *p*-value of 0.6, upscaling framework 2 had a V value of 288 and a *p*-value of <0.1, and upscaling framework 3 showed a V value of 176 and a *p*-value of 0.3.

Overestimation plots (Figure 11a) were related to fewer trees in the plot and higher understory; since we only measure trees, biomass could be overestimated in these scenarios. However, this assumption must be interpreted with caution as understory in pine flatwoods can contribute up to approximately 5 Mg ha^{-1} of biomass only [53,54]. Underestimation plots (Figure 11c) seemed to be associated with a high density of trees with high DBH, areas where most of the biomass information from satellite images (which is derived from height and canopy coverage) tends to underestimate the biomass. Good estimates, however, came from plots with a well-distributed mix of tree sizes and biomass (Figure 11b).

4. Discussion

The development of robust methods for obtaining accurate and widespread AGBD estimates has long been a significant goal within the remote sensing and ecological research communities [55]. While the integration of data from various sources such as GEDI and satellite imagery has been widely documented for its effectiveness in enhancing biomass

estimation, there exist gaps concerning the accuracy and reliability of these integrated models. Challenges persist in the scientific community, due to inconsistencies in model performance and the complexities involved in fusing heterogeneous data sources. In the following sections, we delve deeper into our research findings: discussing the integration of GEDI with other remote sensing sources (in Section 4.1), exploring how our upscaling frameworks contribute to the overall effectiveness of biomass estimation (in Section 4.2), analyzing sensitivity and uncertainty (in Section 4.3), and evaluating the overarching challenges and future directions in research (in Section 4.4).

4.1. Integration of GEDI with Other Remote Sensing Sources

The strategic integration of the GEDI with SAR and passive optical data represents a significant advancement in AGBD mapping. GEDI's high-resolution LiDAR technology offers precise structural measurements of forests, which is crucial for assessing biomass. However, GEDI's sampling strategy, which focuses on specific transects rather than continuous coverage, creates gaps that could limit its applicability for extensive, wall-to-wall forest mapping. By complementing GEDI's detailed but spatially limited LiDAR data with the broader coverage provided by SAR and passive optical sensors, we were able to produce more comprehensive AGBD maps. Studies on predicting AGBD have leveraged both SAR and optical data due to their complementary strengths in capturing different forest characteristics [56,57]. In our study, we specifically chose to use the HLS dataset, due to its data consistency and reliable temporal resolution, providing comprehensive time coverage that is essential for continuous mapping and monitoring biomass changes. Ensuring temporal consistency and image availability was a crucial element in our study since we focused on data collected during a singular month across consecutive years. Another important finding was that the most significant variables found in the models were spectral bands directly derived from the HLS dataset, underscoring its critical role in our study. These results reflect those of other studies that also found satisfactory performance of Landsat in predicting biomass [22]. Notably, throughout our analysis, we encountered no significant gaps in the HLS dataset, which underscored its robustness and suitability for our AGBD upscaling frameworks.

In our study, the integration of diverse remote sensing data sources, including SAR, passive optical data, vegetation indices, texture information, and geographical and terrain data, exemplifies a comprehensive approach to biomass estimation. By combining these varied data types, we effectively leverage the strengths of each sensor and data characteristic, significantly enhancing the accuracy and reliability of our biomass estimates. This multifaceted approach allows us to minimize the reliance on any single data source, which traditionally could introduce bias or errors into our analysis. For example, SAR data, with its all-weather capability, ensures consistent data acquisition, while passive optical data provides high-resolution spectral insights critical for detailed vegetation analysis. Additionally, vegetation indices and texture information contribute to a deeper understanding of vegetation health and structural properties. Geographical data, such as latitude and longitude, alongside terrain information, further refine our models by accounting for spatial variations and environmental conditions that influence biomass. By synergistically using these data sources, we mitigate the error margins associated with individual datasets and achieve a more robust, reliable, and comprehensive assessment of forest biomass across our study areas.

We observed several advantages of integrating GEDI with other remote sensing sources, particularly noting improvements in the temporal resolution of biomass monitoring. Other studies have also highlighted the benefits of combining GEDI with additional remote sensing data in providing more comprehensive and accurate monitoring [58]. This

integration effectively compensates for the limitations in GEDI's temporal coverage, which is constrained by its targeted transect sampling approach. By leveraging the frequent revisits of other remote sensing sources, we ensured a more continuous and comprehensive monitoring of biomass changes over time. Additionally, our upscaling frameworks involved training models on a dataset spanning multiple years. This strategic choice was crucial in enhancing the temporal generalization of our findings. By incorporating data from different years, our models could adapt to more temporal conditions, making them more reliable and applicable to different years. This methodology enabled a temporally inclusive remote sensing tool that can support ongoing environmental monitoring and decision-making processes.

It is important to emphasize the cost-effectiveness of integrating GEDI with other remote sensing sources, especially compared to airborne laser scanning. Optical satellite and space LiDAR technologies provide wide geographic coverage and frequent data collection at no cost through public data portals, unlike airborne laser scanning, which involves significant operational expenses [59].

4.2. Evaluation of AGBD Upscaling Frameworks and Methodological Design

Considering the three upscaling frameworks for modeling AGBD, some critical common features emerged that significantly enhanced the effectiveness of our methodologies. The employment of a semivariogram analysis to account for the spatial autocorrelation of GEDI footprints prevented spatial bias of our results by managing spatial dependencies effectively. Additionally, the parameters of the RF algorithm were particularly well-suited for this modeling, handling large and complex datasets robustly and preventing overfitting. Furthermore, the implementation of bootstrapping across the models ensured the robustness and reliability of our findings, allowing us to evaluate the stability of our predictions across different samples and enhancing the generalizability of our results.

In our study, upscaling framework 1 employed image-derived data, using remote sensing technologies to upscale AGBD measurements from satellite imagery, a method commonly referenced in the literature for creating wall-to-wall GEDI maps. However, we introduced innovations to enhance robustness and accuracy. First, we strategically selected multiple years of GEDI Level 4A data to ensure temporal robustness, capturing interannual variability in AGBD and enhancing the model's temporal generalization. Second, we accounted for spatial autocorrelation, mitigating the spatial dependence of the samples. This approach's modeling results were reasonable, aligning well with similar studies that use image-derived data for AGBD modeling in different regions. For example, an RMSE of 7.05 Mg ha^{-1}, %RMSE of 42.0%, and an R^2 value of 0.64 in Mozambique [60]; an RMSE of 70.8 Mg^{-1}, %RMSE of 46%, and R^2 of 0.67 in the United States [16]; and an RMSE of 42.2 Mg^{-1}, and %RMSE of 10.3%, and an R^2 of 0.69 in India [15]. Furthermore, the external validation of this method using in situ data was particularly satisfactory. The approach showed a strong correlation between predicted and observed biomass values, indicating robust model performance. While there was a slight tendency towards underestimation, overall accuracy was high.

The second upscaling framework implemented a sequential modeling strategy, beginning with the modeling of L2A and L2B variables, which were then reduced to principal components and subsequently used to estimate forest biomass. While this framework showed the poorest internal validation results compared to the other two approaches, with the lowest R^2 and highest %RMSD, it still achieved satisfactory accuracies in biomass estimation. Other studies have reported the efficacy of this method across different forest environments, achieving RMSE and R^2 of 39.49 Mg ha^{-1} and 0.65 [17] and 13.86 Mg ha^{-1} and 0.77 [31], respectively. External validation with in situ data further highlighted the

framework's mixed performance. The model demonstrated a high R^2, indicating strong predictive relationships, but also exhibited high error margins and underestimation bias (%MD = −31.8%), reflecting notable uncertainty in the AGBD estimates. This can be attributed to the nature of the relationships modeled and the propagation of errors across modeling stages. The relationships modeled can preserve patterns, resulting in a high R^2, but still showing high error if, for example, key predictors such as spectral indices saturate at high biomass, leading to underestimation [16,61]. On the other hand, errors introduced during earlier steps, such as uncertainties in previous models, can propagate through the modeling pipeline, impacting prediction error despite strong apparent correlations [62]. Canopy height and cover are considered critical forest structural covariates in determining AGBD, as taller trees carry greater biomass, making the use of accurate canopy height models increasingly common in AGBD mapping [63,64]. This correlation typically guides the expectation that increases in canopy height will proportionally increase biomass estimates. However, our findings highlight exceptions to this general rule, particularly in ecosystems like old-growth forests. In such environments, trees may exhibit high DBH relative to their height (Figure 11c). This characteristic results in large stem volumes that contribute significantly to biomass but may not be adequately reflected by height alone. Consequently, traditional models based primarily on canopy height can often underestimate biomass in these settings, as they fail to account for the substantial biomass contributions from large, mature stems.

Upscaling framework 3 exhibited the highest variability and underestimation. The allometric equation seemed to fit well in our database, returning good accuracies. Our results were similar to those reported by other studies, which achieved an RMSD of 33.56 Mg ha^{-1} and an R^2 of 0.52 using a similar approach [65]. The primary limitation of this method lies in the propagation of errors through the scaling-up chain from in situ data collection to satellite-based large-scale mapping, a challenge also highlighted in other studies [66]. A significant source of error in this approach is related to the choice of the allometric model [67]. The variability and biases introduced by different allometric equations can significantly impact the accuracy of biomass estimates, emphasizing the need for careful selection and validation of these models to improve the reliability of AGBD assessments.

The external validation revealed a trend of underestimation across the upscaling frameworks. This underestimation likely arises from the use of optical data, which tends to saturate in high-biomass areas, with saturation occurring around 200 Mg ha^{-1} in our study. This saturation reduces the model's robustness to predict AGBD in dense forests. Saturation is a well-known challenge in remote sensing-based AGBD estimation, with several studies reporting its impact on high-biomass areas. For example, Sentinel-2 data was unable to capture vertical forest stand structure, and data saturation was a key factor contributing to the relatively low performance in AGBD estimation [68]. Landsat-based indices also show high sensitivity to saturation when modeling GEDI-derived AGBD, with prominent saturation effects around 75 Mg ha^{-1} [60]. Integrating a diverse range of data sources, including SAR, Sentinel-2 multispectral data, elevation, and land cover, has been suggested as a way to potentially mitigate saturation in high-biomass areas [16], as, saturation has been reported in optical imagery when predicting AGBD from ALS data [22,69,70]. Additionally, potential factors such as scale mismatches between field plot data and larger spatial scales, as well as limitations in the allometric equations used, may contribute to the observed underestimation. These challenges collectively impact the accuracy of AGBD predictions, particularly in regions with high canopy density.

Overall, the upscaling frameworks revealed important trade-offs in model complexity, robustness, and sensitivity to uncertainty. Specific design choices, such as spatial sam-

pling strategies, ensemble modeling, or allometries, shaped each framework's strengths and limitations. These choices underscore the importance of aligning methodological approaches with study objectives and landscape conditions to ensure accurate and reliable AGBD mapping.

4.3. Map Uncertainty Analysis

The map uncertainty analysis provided important insights into the performance of the different AGBD upscaling frameworks. Internal validation showed that bootstrapping effectively captured variability for all frameworks, ensuring robust and reliable predictions. AGBD was sensitive to passive optical information, particularly the spectral bands of the HLS, highlighting the importance of this data in biomass estimation.

Our analysis revealed that the frameworks were statistically different, showing sensitivity to specific ranges of AGBD values, with some methods performing better at certain biomass levels. Despite these differences, the overall spatial pattern of AGBD estimate maps was similar regardless of the method used, indicating a general consistency in the spatial distribution of biomass.

Uncertainties in the internal validation were also statistically different across frameworks, indicating varying levels of precision and reliability. We noticed the effect of error propagation in our analysis. For example, frameworks 1 and 2 presented a similar method with both utilizing the same image stack initially, but the second involved sequential modeling. These multiple stages seemed to impact their external validation, resulting in decreased accuracy. Framework 3, which also involved multiple modeling stages, showed a decrease in accuracy when compared to 1, further emphasizing the impact of modeling strategies and the choice of proper equations on predictive reliability.

When it came to external validation, framework 1 emerged as the best-performing method, demonstrating the highest accuracy and lowest error rates when compared to in situ measurements. This highlights the robustness and applicability of this method for accurate AGBD estimation in diverse forest conditions. However, it is important to note that uncertainties in our AGBD predictions may arise not only from the frameworks but also from the inherent limitations of the GEDI L4A. Although widely used for biomass estimation, GEDI L4A data carries intrinsic uncertainties, as its accuracy can be influenced by factors such as vegetation type [64], terrain [71], and acquisition conditions [72]. These inherent limitations may affect the reliability of the reference AGBD values used across our frameworks.

4.4. Challenges and Future Research

Advancing forest AGBD mapping through the synergism of spaceborne LiDAR, SAR, and passive optical data presented significant challenges. One of the key challenges in our study was selecting suitable GEDI data for the analysis, particularly in determining which period would best represent the biomass without being affected by seasonal variations. GEDI data are sensitive to seasonal changes that can significantly affect biomass estimations, for example, leaf-on and leaf-off seasons can dramatically alter the LiDAR measurements of canopy structure and density. Choosing a data collection period that accurately reflects the standing biomass while minimizing the effects of phenological changes is crucial. This selection process involved analyzing historical data to identify periods of peak biomass or stability across multiple years, thereby reducing the influence of seasonal anomalies. However, the inherent variability in climate and phenological patterns across different geographical regions and forest types posed a challenge. In large regions, different areas may experience distinct seasonal dynamics, such as variations in the timing of leaf-on and leaf-off periods. These differences can dramatically alter LiDAR measurements of

canopy structure and density, making it challenging to select a single time frame that accurately captures biomass for the entire area. Additionally, spatially sparse areas may have unique environmental conditions that do not align with the broader regional patterns. Consequently, relying on a single period for data collection could lead to inaccurate biomass estimations in certain parts of the study area. Moreover, the selection of the GEDI data periods also needed to account for the availability of passive optical data, particularly avoiding high rates of cloud cover.

Furthermore, it is essential to consider several critical factors to implement our approach for advancing forest AGBD mapping. First, select GEDI data collection periods by analyzing historical data to identify times of peak biomass or stability across multiple years, minimizing the effects of seasonal anomalies. Recognizing that different regions within large study areas may have distinct seasonal dynamics, such as variations in leaf-on and leaf-off periods, can significantly impact LiDAR measurements of canopy structure and density. Therefore, it is crucial to tailor data collection periods to the specific phenological patterns of each sub-region rather than relying on a single period for the entire area.

Data consistency and compatibility posed as substantial challenges in forest AGBD estimation, particularly due to temporal misalignment. When integrating data from different sensors that collect information at various times, aligning these datasets accurately is crucial. Forests are dynamic environments where conditions can change rapidly due to factors such as seasonal growth, logging, succession, or natural disturbances. GEDI data and satellite imagery can result in significant discrepancies in the final biomass estimation if not collected within a closely aligned time frame. These temporal misalignments can lead to incorrect assessments of biomass, as the data may reflect distinct stages of forest development or recovery [73].

We also found technical challenges related to data integration in our study. Combining data from different sensors involved complex preprocessing to match spatial resolutions and to calibrate spectral responses, which can be technically challenging and time-consuming. In addition, sensors may have varying calibration standards, spectral responses, and measurement techniques, which can lead to inconsistencies and require rigorous adjustment and standardization. To address these challenges, it is important to develop standardized protocols for preprocessing and calibrating remote sensing data from different sources [74]. This includes creating unified calibration standards, developing automated tools for spectral response matching and resolution harmonization, and investing in advanced data fusion techniques [11].

Processing and analyzing large datasets from multiple sources also require substantial computational resources. In our study, we utilized the GEE Python API along with the geemap package to access, process, and analyze image collections and ancillary datasets within the Jupyter environment. This approach allowed us to leverage GEE's powerful cloud computing capabilities, which are particularly adept at handling large-scale geospatial data efficiently. However, our analysis was limited to the algorithms available within GEE, which presented certain limitations. While GEE provides a robust set of tools for broad-scale data analysis, it restricts the user to predefined algorithms. It does not accommodate custom or advanced machine learning algorithms, such as those found in deep learning frameworks. Future research should explore new alternatives that extend beyond the GEE platform to enhance the precision and accuracy of biomass estimates.

5. Conclusions

This study demonstrated that merging spaceborne LiDAR, SAR, and passive optical satellite imagery significantly enhances forest AGBD mapping across different forest-dominated landscapes. This synergistic approach proved effective in capturing the complex

variations in forest biomass, thereby confirming the efficacy of combining these datasets for improved AGBD mapping. The upscaling framework that used a single step based on GEDI and image-derived data showed reasonable accuracy and a strong correlation with in situ data, although it tended to slightly underestimate biomass. The framework that involved a two-stage modeling strategy focusing on GEDI-derived canopy height and cover appeared to be the most accurate and precise method, showing the lowest mean difference and root mean square difference. Lastly, modeling based on a three-step upscaling framework involving GEDI, imagery and in situ-derived allometries exhibited the highest degree of underestimation and variability, making it the least reliable upscaling framework. The modeling estimated mean AGBD and uncertainties were statistically different across frameworks, highlighting variations in their performance, and demonstrating that upscaling frameworks drove prediction uncertainty when mapping AGBD from the synergism of spaceborne LiDAR, SAR, and passive optical data. The analysis underscored the importance of robust data integration and advanced modeling techniques to ensure reliable AGBD estimates. Moreover, the findings contribute valuable insights into improving forest biomass mapping and carbon monitoring and highlight the potential for ongoing advancements in remote sensing methodologies, with broad applicability to forested regions worldwide. Future research can evaluate upscaling frameworks across diverse biomes, integrate new satellite data, and explore uncertainty quantification methods that explicitly propagate errors across modeling stages to improve the interpretability of AGBD estimates.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs17142340/s1, Table S1. The remote sensing covariates used in this study. References [39,42,43,75–89] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, I.T.B. and C.A.S.; Methodology, I.T.B.; Software, I.T.B.; Validation, I.T.B.; Formal analysis, I.T.B.; Investigation, I.T.B.; Resources, C.A.S.; Data curation, I.T.B., C.A.S., M.B.S. and J.X.; Writing—original draft, I.T.B.; Writing—review & editing, I.T.B., C.A.S., M.B.S., V.M.D., J.W.A., K.M.B., J.X., D.R.V., J.Q., J.V., A.S. (Andres Susaeta), A.S. (Ajay Sharma), C.K., M.M. and A.P.D.C.; Visualization, I.T.B.; Supervision, C.A.S.; Project administration, C.A.S.; Funding acquisition, C.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by NASA's Commercial Smallsat Data Scientific Analysis program (CSDSA, 80NSSC24K0055). In addition, other funding was provided by the intramural research program of the U.S. Department of Agriculture, National Institute of Food and Agriculture, Rapid Response to Extreme Weather Events Across Food and Agricultural Systems (Grant #2023-68016-39039), U.S. National Science Foundation (Grant #2409886), and NASA's grants (ICESat-2, 80NSSC23K0941) and Carbon Monitoring System (CMS, 80NSSC23K1257). The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

Data Availability Statement: The datasets presented in this article are not readily available. Requests to access the datasets should be directed to the corresponding author.

Acknowledgments: We would like to thank the School of Forest, Fisheries, and Geomatics Sciences and the University of Florida Institute of Food and Agricultural Sciences (UF/IFAS) for providing access to the FAS Millhopper Unit and Austin Cary Forest areas. We also extend our gratitude to the personnel responsible for the Myakka State Forest and Okaloacoochee Slough State Forest for their cooperation and support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cao, L.; Coops, N.C.; Innes, J.L.; Sheppard, S.R.J.; Fu, L.; Ruan, H.; She, G. Estimation of Forest Biomass Dynamics in Subtropical Forests Using Multi-Temporal Airborne LiDAR Data. *Remote Sens. Environ.* **2016**, *178*, 158–171. [CrossRef]
2. Vargas-Larreta, B.; López-Martínez, J.O.; González, E.J.; Corral-Rivas, J.J.; Hernández, F.J. Assessing Above-Ground Biomass-Functional Diversity Relationships in Temperate Forests in Northern Mexico. *For. Ecosyst.* **2021**, *8*, 8. [CrossRef]
3. Erb, K.-H.; Kastner, T.; Plutzar, C.; Bais, A.L.S.; Carvalhais, N.; Fetzel, T.; Gingrich, S.; Haberl, H.; Lauk, C.; Niedertscheider, M.; et al. Unexpectedly Large Impact of Forest Management and Grazing on Global Vegetation Biomass. *Nature* **2018**, *553*, 73–76. [CrossRef] [PubMed]
4. Liu, Y.Y.; van Dijk, A.I.J.M.; de Jeu, R.A.M.; Canadell, J.G.; McCabe, M.F.; Evans, J.P.; Wang, G. Recent Reversal in Loss of Global Terrestrial Biomass. *Nat. Clim. Change* **2015**, *5*, 470–474. [CrossRef]
5. Thum, T.; MacBean, N.; Peylin, P.; Bacour, C.; Santaren, D.; Longdoz, B.; Loustau, D.; Ciais, P. The Potential Benefit of Using Forest Biomass Data in Addition to Carbon and Water Flux Measurements to Constrain Ecosystem Model Parameters: Case Studies at Two Temperate Forest Sites. *Agric. For. Meteorol.* **2017**, *234–235*, 48–65. [CrossRef]
6. Cantarello, E.; Newton, A.C.; Martin, P.A.; Evans, P.M.; Gosal, A.; Lucash, M.S. Quantifying Resilience of Multiple Ecosystem Services and Biodiversity in a Temperate Forest Landscape. *Ecol. Evol.* **2017**, *7*, 9661–9675. [CrossRef]
7. Johnston, M.; Lindner, M.; Parrotta, J.; Giessen, L. Adaptation and Mitigation Options for Forests and Forest Management in a Changing Climate. *For. Policy Econ.* **2012**, *24*, 1–2. [CrossRef]
8. Galidaki, G.; Zianis, D.; Gitas, I.; Radoglou, K.; Karathanassi, V.; Tsakiri–Strati, M.; Woodhouse, I.; Mallinis, G. Vegetation Biomass Estimation with Remote Sensing: Focus on Forest and Other Wooded Land over the Mediterranean Ecosystem. *Int. J. Remote Sens.* **2017**, *38*, 1940–1966. [CrossRef]
9. Liu, X.; Neigh, C.S.R.; Pardini, M.; Forkel, M. Estimating Forest Height and Above-Ground Biomass in Tropical Forests Using P-Band TomoSAR and GEDI Observations. *Int. J. Remote Sens.* **2024**, *45*, 3129–3148. [CrossRef]
10. Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Li, G.; Moran, E. A Survey of Remote Sensing-Based Aboveground Biomass Estimation Methods in Forest Ecosystems. *Int. J. Digit. Earth* **2016**, *9*, 63–105. [CrossRef]
11. Silva, C.A.; Duncanson, L.; Hancock, S.; Neuenschwander, A.; Thomas, N.; Hofton, M.; Fatoyinbo, L.; Simard, M.; Marshak, C.Z.; Armston, J.; et al. Fusing Simulated GEDI, ICESat-2 and NISAR Data for Regional Aboveground Biomass Mapping. *Remote Sens. Environ.* **2021**, *253*, 112234. [CrossRef]
12. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002. [CrossRef]
13. Duncanson, L.; Kellner, J.R.; Armston, J.; Dubayah, R.; Minor, D.M.; Hancock, S.; Healey, S.P.; Patterson, P.L.; Saarela, S.; Marselis, S.; et al. Aboveground Biomass Density Models for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar Mission. *Remote Sens. Environ.* **2022**, *270*, 112845. [CrossRef]
14. Dubayah, R.; Armston, J.; Healey, S.P.; Bruening, J.M.; Patterson, P.L.; Kellner, J.R.; Duncanson, L.; Saarela, S.; Ståhl, G.; Yang, Z.; et al. GEDI Launches a New Era of Biomass Inference from Space. *Environ. Res. Lett.* **2022**, *17*, 095001. [CrossRef]
15. Mohite, J.; Sawant, S.; Pandit, A.; Sakkan, M.; Pappula, S.; Parmar, A. Forest Aboveground Biomass Estimation by GEDI and Multi-Source EO Data Fusion over Indian Forest. *Int. J. Remote Sens.* **2024**, *45*, 1304–1338. [CrossRef]
16. Shendryk, Y. Fusing GEDI with Earth Observation Data for Large Area Aboveground Biomass Mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *115*, 103108. [CrossRef]
17. Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Goulden, T. State-Wide Forest Canopy Height and Aboveground Biomass Map for New York with 10 m Resolution, Integrating GEDI, Sentinel-1, and Sentinel-2 Data. *Ecol. Inform.* **2024**, *79*, 102404. [CrossRef]
18. Kanmegne Tamga, D.; Latifi, H.; Ullmann, T.; Baumhauer, R.; Bayala, J.; Thiel, M. Estimation of Aboveground Biomass in Agroforestry Systems over Three Climatic Regions in West Africa Using Sentinel-1, Sentinel-2, ALOS, and GEDI Data. *Sensors* **2022**, *23*, 349. [CrossRef]
19. Wang, C.; Zhang, W.; Ji, Y.; Marino, A.; Li, C.; Wang, L.; Zhao, H.; Wang, M. Estimation of Aboveground Biomass for Different Forest Types Using Data from Sentinel-1, Sentinel-2, ALOS PALSAR-2, and GEDI. *Forests* **2024**, *15*, 215. [CrossRef]
20. Zhao, X.; Hu, W.; Han, J.; Wei, W.; Xu, J. Urban Above-Ground Biomass Estimation Using GEDI Laser Data and Optical Remote Sensing Images. *Remote Sens.* **2024**, *16*, 1229. [CrossRef]
21. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping Global Forest Canopy Height through Integration of GEDI and Landsat Data. *Remote Sens. Environ.* **2021**, *253*, 112165. [CrossRef]
22. Campbell, M.J.; Dennison, P.E.; Kerr, K.L.; Brewer, S.C.; Anderegg, W.R.L. Scaled Biomass Estimation in Woodland Ecosystems: Testing the Individual and Combined Capacities of Satellite Multispectral and Lidar Data. *Remote Sens. Environ.* **2021**, *262*, 112511. [CrossRef]

23. Qi, W.; Saarela, S.; Armston, J.; Ståhl, G.; Dubayah, R. Forest Biomass Estimation over Three Distinct Forest Types Using TanDEM-X InSAR Data and Simulated GEDI Lidar Data. *Remote Sens. Environ.* **2019**, *232*, 111283. [CrossRef]
24. Hernández-Martínez, L.A.; Dupuy-Rada, J.M.; Medel-Narváez, A.; Portillo-Quintero, C.; Hernández-Stefanoni, J.L. Improving Aboveground Biomass Density Mapping of Arid and Semi-Arid Vegetation by Combining GEDI LiDAR, Sentinel-1/2 Imagery and Field Data. *Sci. Remote Sens.* **2025**, *11*, 100204. [CrossRef]
25. Perpinyà-Vallès, M.; Cendagorta-Galarza, D.; Ameztegui, A.; Huertas, C.; Escorihuela, M.J.; Romero, L. High-Resolution Aboveground Biomass Mapping: The Benefits of Biome-Specific Deep Learning Models. *Remote Sens.* **2025**, *17*, 1268. [CrossRef]
26. Zhang, B.; Zhang, L.; Yan, M.; Zuo, J.; Dong, Y.; Chen, B. High-Resolution Mapping of Forest Parameters in Tropical Rainforests Through AutoML Integration of GEDI With Sentinel-1/2, Landsat 8, and ALOS-2 Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 9084–9118. [CrossRef]
27. Clark, M.L.; Roberts, D.A.; Ewel, J.J.; Clark, D.B. Estimation of Tropical Rain Forest Aboveground Biomass with Small-Footprint Lidar and Hyperspectral Sensors. *Remote Sens. Environ.* **2011**, *115*, 2931–2942. [CrossRef]
28. Feng, Y.; Lu, D.; Chen, Q.; Keller, M.; Moran, E.; Dos-Santos, M.N.; Bolfe, E.L.; Batistella, M. Examining Effective Use of Data Sources and Modeling Algorithms for Improving Biomass Estimation in a Moist Tropical Forest of the Brazilian Amazon. *Int. J. Digit. Earth* **2017**, *10*, 996–1016. [CrossRef]
29. Bruening, J.M.; Fischer, R.; Bohn, F.J.; Armston, J.; Armstrong, A.H.; Knapp, N.; Tang, H.; Huth, A.; Dubayah, R. Challenges to Aboveground Biomass Prediction from Waveform Lidar. *Environ. Res. Lett.* **2021**, *16*, 125013. [CrossRef]
30. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
31. Padalia, H.; Prakash, A.; Watham, T. Modelling Aboveground Biomass of a Multistage Managed Forest through Synergistic Use of Landsat-OLI, ALOS-2 L-Band SAR and GEDI Metrics. *Ecol. Inform.* **2023**, *77*, 102234. [CrossRef]
32. Mitsuhashi, R.; Sawada, Y.; Tsutsui, K.; Hirayama, H.; Imai, T.; Sumita, T.; Kajiwara, K.; Honda, Y. Re-Estimating GEDI Ground Elevation Using Deep Learning: Impacts on Canopy Height and Aboveground Biomass. *Remote Sens.* **2024**, *16*, 4597. [CrossRef]
33. Chojnacky, D.C.; Heath, L.S.; Jenkins, J.C. Updated Generalized Biomass Equations for North American Tree Species. *Forestry* **2014**, *87*, 129–151. [CrossRef]
34. Mitsch, W.J.; Ewel, K.C. Comparative Biomass and Growth of Cypress in Florida Wetlands. *Am. Midl. Nat.* **1979**, *101*, 417–426. [CrossRef]
35. Gonzalez-Benecke, C.A.; Gezan, S.A.; Albaugh, T.J.; Allen, H.L.; Burkhart, H.E.; Fox, T.R.; Jokela, E.J.; Maier, C.A.; Martin, T.A.; Rubilar, R.A.; et al. Local and General Above-Stump Biomass Functions for Loblolly Pine and Slash Pine Trees. *For. Ecol. Manage.* **2014**, *334*, 254–276. [CrossRef]
36. Gonzalez-Benecke, C.; Zhao, D.; Samuelson, L.; Martin, T.; Leduc, D.; Jack, S. Local and General Above-Ground Biomass Functions for Pinus Palustris Trees. *Forests* **2018**, *9*, 310. [CrossRef]
37. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near Real-Time Global 10 m Land Use Land Cover Mapping. *Sci. Data* **2022**, *9*, 251. [CrossRef]
38. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
39. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.-C.; Skakun, S.V.; Justice, C. The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set. *Remote Sens. Environ.* **2018**, *219*, 145–161. [CrossRef]
40. Mullissa, A.; Vollrath, A.; Odongo-Braun, C.; Slagter, B.; Balling, J.; Gou, Y.; Gorelick, N.; Reiche, J. Sentinel-1 SAR Backscatter Analysis Ready Data Preparation in Google Earth Engine. *Remote Sens.* **2021**, *13*, 1954. [CrossRef]
41. Ren, C.; Jiang, H.; Xi, Y.; Liu, P.; Li, H. Quantifying Temperate Forest Diversity by Integrating GEDI LiDAR and Multi-Temporal Sentinel-2 Imagery. *Remote Sens.* **2023**, *15*, 375. [CrossRef]
42. Conners, R.W.; Trivedi, M.M.; Harlow, C.A. Segmentation of a High-Resolution Urban Scene Using Texture Operators. *Comput. Vision, Graph. Image Process.* **1984**, *25*, 273–310. [CrossRef]
43. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural Features for Image Classification. *IEEE Trans. Syst. man Cybern.* **1973**, *6*, 610–621. [CrossRef]
44. Wu, Q. Geemap: A Python Package for Interactive Mapping with Google Earth Engine. *J. Open Source Softw.* **2020**, *5*, 2305. [CrossRef]
45. Curran, P.J. The Semivariogram in Remote Sensing: An Introduction. *Remote Sens. Environ.* **1988**, *24*, 493–507. [CrossRef]
46. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
47. Silva, C.A.; Hamamura, C.; Valbuena, R.; Hancock, S.; Cardil, A.; Broadbent, E.N.; Almeida, D.R.A.; Silva Junior, C.H.L.; Klauberg, C. RGEDI: An R Package for NASA's Global Ecosystem Dynamics Investigation (GEDI) Data Visualizing and Processing. Available online: https://github.com/carlos-alberto-silva/rGEDI (accessed on 15 October 2024).

48. USGS FL_Peninsular_FDEM_2018_D19_DRRA. Available online: https://rockyweb.usgs.gov/vdelivery/Datasets/Staged/Elevation/metadata/FL_Peninsular_FDEM_2018_D19_DRRA/USGS_FL_Peninsular_FDEM_2018_D19_DRRA_Project_Report.pdf (accessed on 5 May 2025).
49. Köhler, P.; Huth, A. Towards Ground-Truthing of Spaceborne Estimates of above-Ground Life Biomass and Leaf Area Index in Tropical Rain Forests. *Biogeosciences* **2010**, *7*, 2531–2543. [CrossRef]
50. Breslow, N. A Generalized Kruskal-Wallis Test for Comparing K Samples Subject to Unequal Patterns of Censorship. *Biometrika* **1970**, *57*, 579–594. [CrossRef]
51. Wilcoxon, F. Individual Comparisons of Grouped Data by Ranking Methods. *J. Econ. Entomol.* **1946**, *39*, 269–270. [CrossRef]
52. Shapiro, S.S.; Wilk, M.B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52*, 591. [CrossRef]
53. Vogel, J.G.; Bracho, R.; Akers, M.; Amateis, R.; Bacon, A.; Burkhart, H.E.; Gonzalez-Benecke, C.A.; Grunwald, S.; Jokela, E.J.; Kane, M.B.; et al. Regional Assessment of Carbon Pool Response to Intensive Silvicultural Practices in Loblolly Pine Plantations. *Forests* **2021**, *13*, 36. [CrossRef]
54. Vogel, J.G.; Suau, L.J.; Martin, T.A.; Jokela, E.J. Long-Term Effects of Weed Control and Fertilization on the Carbon and Nitrogen Pools of a Slash and Loblolly Pine Forest in North-Central Florida. *Can. J. For. Res.* **2011**, *41*, 552–567. [CrossRef]
55. Xu, C.; Morgenroth, J.; Manley, B. Integrating Data from Discrete Return Airborne LiDAR and Optical Sensors to Enhance the Accuracy of Forest Description: A Review. *Curr. For. Reports* **2015**, *1*, 206–219. [CrossRef]
56. Chen, Z.; Sun, Z.; Zhang, H.; Zhang, H.; Qiu, H. Aboveground Forest Biomass Estimation Using Tent Mapping Atom Search Optimized Backpropagation Neural Network with Landsat 8 and Sentinel-1A Data. *Remote Sens.* **2023**, *15*, 5653. [CrossRef]
57. Zhang, Y.; Wang, N.; Wang, Y.; Li, M. A New Strategy for Improving the Accuracy of Forest Aboveground Biomass Estimates in an Alpine Region Based on Multi-Source Remote Sensing. *GIScience Remote Sens.* **2023**, *60*, 2163574. [CrossRef]
58. Chen, L.; Ren, C.; Bao, G.; Zhang, B.; Wang, Z.; Liu, M.; Man, W.; Liu, J. Improved Object-Based Estimation of Forest Aboveground Biomass by Integrating LiDAR Data from GEDI and ICESat-2 with Multi-Sensor Images in a Heterogeneous Mountainous Region. *Remote Sens.* **2022**, *14*, 2743. [CrossRef]
59. Puliti, S.; Dash, J.P.; Watt, M.S.; Breidenbach, J.; Pearse, G.D. A Comparison of UAV Laser Scanning, Photogrammetry and Airborne Laser Scanning for Precision Inventory of Small-Forest Properties. *Forestry* **2020**, *93*, 150–162. [CrossRef]
60. Liang, M.; Duncanson, L.; Silva, J.A.; Sedano, F. Quantifying Aboveground Biomass Dynamics from Charcoal Degradation in Mozambique Using GEDI Lidar and Landsat. *Remote Sens. Environ.* **2023**, *284*, 113367. [CrossRef]
61. Li, X.; Wessels, K.; Armston, J.; Duncanson, L.; Urbazaev, M.; Naidoo, L.; Mathieu, R.; Main, R. Evaluation of GEDI Footprint Level Biomass Models in Southern African Savannas Using Airborne LiDAR and Field Measurements. *Sci. Remote Sens.* **2024**, *10*, 100161. [CrossRef]
62. Saarela, S.; Holm, S.; Healey, S.P.; Patterson, P.L.; Yang, Z.; Andersen, H.E.; Dubayah, R.O.; Qi, W.; Duncanson, L.I.; Armston, J.D.; et al. Comparing Frameworks for Biomass Prediction for the Global Ecosystem Dynamics Investigation. *Remote Sens. Environ.* **2022**, *278*, 113074. [CrossRef]
63. Nandy, S.; Srinet, R.; Padalia, H. Mapping Forest Height and Aboveground Biomass by Integrating ICESat-2, Sentinel-1 and Sentinel-2 Data Using Random Forest Algorithm in Northwest Himalayan Foothills of India. *Geophys. Res. Lett.* **2021**, *48*, e2021GL093799. [CrossRef]
64. Wang, C.; Jia, D.; Lei, S.; Numata, I.; Tian, L. Accuracy Assessment and Impact Factor Analysis of GEDI Leaf Area Index Product in Temperate Forest. *Remote Sens.* **2023**, *15*, 1535. [CrossRef]
65. Bullock, E.L.; Healey, S.P.; Yang, Z.; Acosta, R.; Villalba, H.; Insfrán, K.P.; Melo, J.B.; Wilson, S.; Duncanson, L.; Næsset, E.; et al. Estimating Aboveground Biomass Density Using Hybrid Statistical Inference with GEDI Lidar Data and Paraguay's National Forest Inventory. *Environ. Res. Lett.* **2023**, *18*, 085001. [CrossRef]
66. Réjou-Méchain, M.; Barbier, N.; Couteron, P.; Ploton, P.; Vincent, G.; Herold, M.; Mermoz, S.; Saatchi, S.; Chave, J.; de Boissieu, F.; et al. Upscaling Forest Biomass from Field to Satellite Measurements: Sources of Errors and Ways to Reduce Them. *Surv. Geophys.* **2019**, *40*, 881–911. [CrossRef]
67. Chave, J.; Condit, R.; Aguilar, S.; Hernandez, A.; Lao, S.; Perez, R. Error Propagation and Scaling for Tropical Forest Biomass Estimates. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* **2004**, *359*, 409–420. [CrossRef] [PubMed]
68. Indirabai, I.; Nilsson, M. Estimation of above Ground Biomass in Tropical Heterogeneous Forests in India Using GEDI. *Ecol. Inform.* **2024**, *82*, 102712. [CrossRef]
69. Fararoda, R.; Reddy, R.S.; Rajashekar, G.; Chand, T.R.K.; Jha, C.S.; Dadhwal, V.K. Improving Forest above Ground Biomass Estimates over Indian Forests Using Multi Source Data Sets with Machine Learning Algorithm. *Ecol. Inform.* **2021**, *65*, 101392. [CrossRef]
70. Urbazaev, M.; Thiel, C.; Cremer, F.; Dubayah, R.; Migliavacca, M.; Reichstein, M.; Schmullius, C. Estimation of Forest Aboveground Biomass and Uncertainties by Integration of Field Measurements, Airborne LiDAR, and SAR and Optical Satellite Data in Mexico. *Carbon Balance Manag.* **2018**, *13*, 5. [CrossRef] [PubMed]

71. Liu, A.; Cheng, X.; Chen, Z. Performance Evaluation of GEDI and ICESat-2 Laser Altimeter Data for Terrain and Canopy Height Retrievals. *Remote Sens. Environ.* **2021**, *264*, 112571. [CrossRef]
72. Wang, C.; Elmore, A.J.; Numata, I.; Cochrane, M.A.; Shaogang, L.; Huang, J.; Zhao, Y.; Li, Y. Factors Affecting Relative Height and Ground Elevation Estimations of GEDI among Forest Types across the Conterminous USA. *GIScience Remote Sens.* **2022**, *59*, 975–999. [CrossRef]
73. Babcock, C.; Finley, A.O.; Cook, B.D.; Weiskittel, A.; Woodall, C.W. Modeling Forest Biomass and Growth: Coupling Long-Term Inventory and LiDAR Data. *Remote Sens. Environ.* **2016**, *182*, 1–12. [CrossRef]
74. Su, Y.; Zhang, W.; Liu, B.; Tian, X.; Chen, S.; Wang, H.; Mao, Y. Forest Carbon Flux Simulation Using Multi-Source Data and Incorporation of Remotely Sensed Model with Process-Based Model. *Remote Sens.* **2022**, *14*, 4766. [CrossRef]
75. Vincini, M.; Frazzi, E.; D'Alessio, P. A Broad-Band Leaf Chlorophyll Vegetation Index at the Canopy Scale. *Precis. Agric.* **2008**, *9*, 303–319. [CrossRef]
76. Drury, S.A. *Image Interpretation in Geology*, 2nd ed.; Routledge: London, UK, 1987.
77. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.; Gao, X.; Ferreira, L. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
78. Ceccato, P.; Gobron, N.; Flasse, S.; Pinty, B.; Tarantola, S. Designing a Spectral Index to Estimate Vegetation Water Content from Remote Sensing Data: Part 1. *Remote Sens. Environ.* **2002**, *82*, 188–197. [CrossRef]
79. Gitelson, A.A.; Viña, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote Estimation of Canopy Chlorophyll Content in Crops. *Geophys. Res. Lett.* **2005**, *32*. [CrossRef]
80. Camps-Valls, G.; Campos-Taberner, M.; Moreno-Martínez, Á.; Walther, S.; Duveiller, G.; Cescatti, A.; Mahecha, M.D.; Muñoz-Marí, J.; García-Haro, F.J.; Guanter, L.; et al. A Unified Vegetation Index for Quantifying the Terrestrial Biosphere. *Sci. Adv.* **2021**, *7*. [CrossRef]
81. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]
82. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, Washington, DC, USA, 10–14 December 2014; NASA: Washington, DC, USA, 1974; pp. 309–317.
83. Gao, B. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]
84. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [CrossRef]
85. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]
86. Gitelson, A.A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *J. Plant Physiol.* **2004**, *161*, 165–173. [CrossRef] [PubMed]
87. Schlund, M.; Erasmi, S. Sentinel-1 Time Series Data for Monitoring the Phenology of Winter Wheat. *Remote Sens. Environ.* **2020**, *246*, 111814. [CrossRef]
88. NASA. *NASADEM Merged DEM Global 1 Arc Second V001*; NASA Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA. [CrossRef]
89. Irons, J.R.; Petersen, G.W. Texture Transforms of Remote Sensing Data. *Remote Sens. Environ.* **1981**, *11*, 359–370. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.