ABSTRACT

Title of Dissertation: ADVANCES IN MAPPING FOREST BIOMASS
AND OLD-GROWTH CONDITIONS
USING WAVEFORM LIDAR

Jamis Michael Bruening, Doctor of Philosophy
2023

Directed by: Professor Ralph Dubayah
Department of Geographical Sciences

The Global Ecosystem Dynamics Investigation (GEDI) is a spaceborne waveform lidar system that has transformed scientific understanding of the world's forests through billions of precise measurements of ecosystem structure. Relative to forest processes that operate on decadal to millennial timescales, the four year period during which GEDI collected these measurements is short, and GEDI's ability to analyze how forest structure changes over time is mostly unproven. However, fusion efforts that integrate GEDI data with forest inventory measurements and ecosystem models hold immense potential for discovery. In this dissertation, I explore the limitations and capabilities of GEDI data for inference into structural and successional dynamics within eastern US forests. First, I used a forest gap model to quantify uncertainty in biomass predictions for individual GEDI waveforms, and discovered a relationship between biomass uncertainty and successional stage. Next, I investigated uncertainties and errors in large-scale GEDI biomass estimates relative to unbiased estimates from the US forest inventory. I developed a novel modeling framework based on fusion between GEDI and the US forest inventory data that corrected these errors, and I produced unbiased and precise maps of forest biomass for the continental US.

Lastly, I assessed GEDI's ability to identify and map different types of old-growth forests, and discovered that GEDI can detect some old forests more effectively than others. This research identified key limitations associated with using GEDI to study forest dynamics, and I leveraged these discoveries to develop new ways of using GEDI data for ecological and successional inference. These discoveries will inform new uses of GEDI data and its integration with inventory data and ecosystem modeling to better characterize changes within forest ecosystems.

ADVANCES IN MAPPING FOREST BIOMASS AND OLD-GROWTH CONDITIONS USING WAVEFORM LIDAR

by

Jamis Michael Bruening

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Dr. Ralph Dubayah, Chair
Dr. John Armston
Dr. Taylor Oshan
Dr. Benjamin Poulter
Dr. Joseph Sullivan

Dedication

To my children, Magnolia and Cooper. May you too discover the joys of science.

And to my dad, who loves old trees even more than I do.

Acknowledgments

To Ralph, thank you for sharing your knowledge and wisdom, for helping me become the scientist I am today, for pushing me and believing in me. Thank you for making me a better writer and giving me the freedom to pursue my own questions. I will be forever Grateful for my time in your lab and as a member of the GEDI team.

Thank you to my committee (John, Taylor, Ben, Joe) for your thoughtful comments and time spent improving my work. This dissertation is better because of your involvement, which I sincerely appreciate.

To John, thank you for years of mentoring and guidance regarding my work with GEDI. Thanks also for your patience and willingness to explain things to me.

To Claire, thank you for your love, support, and encouragement over the past 5 years. To Magnolia and Cooper, thank you for letting me see the world again through young eyes and reminding me of the value of curiosity. Thanks for asking “Why?”. To my parents, thank you for your love, interest, and support for my career as a scientist. None of this would have been possible without you.

To Donal, thank you for almost a decade of friendship, unwavering enthusiasm for my work, and for always being willing to brainstorm or read a draft.

To the rest of the GEDI team, Dubayah et al. lab, GEOG department, and my other collaborators, thank you for welcoming me and taking an interest in my personal and professional development.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AGBD	Above Ground Biomass Density
ALS	Airborne Laser Scanning
CI	Confidence Interval
CONUS	Continental United States
DEM	Digital Elevation Model
DBT	Deciduous Broadleaf Tree
EBLUE	Empirical Best Linear Unbiased Estimate
EPA	Environmental Protection Agency
EMAP	Environmental Monitoring and Assessment Program
FIA	Forest Inventory and Analysis Program
FF	Forest Factory
FH	Fay-Herriot
FHD	Foliage Hieght Diversity
GEDI	Global Ecosystem Dynamics Investigation
H3	Uber's Hexagon Gridding System
ha	hectare
HT	Horvitz-Thompson estimator
ISS	International Space Station
L2A	GEDI Level 2A data product
L2B	GEDI Level 2B data product
L4A	GEDI Level 4A data product
L4B	GEDI Level 4B data product
LIDAR	Light Detection and Ranging
Mg	Megagram
MIX	Mixed DBT and NT PFT
NASA	National Aeronautics and Space Administration
NLCD	National Landcover Database
nRMSE	normalized Root Mean Square Error
NT	Needleleaf Tree
OFP	Old Forest Proportion
OLS	Ordinary Least Squares
PAI	Plant Area Index
PAVD	Plant Area Volume Density
PFT	Plant Functional Type
R^2	Coefficient of determination
RH	Relative Height
RMSE	Root Mean Squared Difference
RSE	Residual Squared Error
SISA	Scale Invariant Small Area (model)
US	United States of America
USDA	United States Department of Agriculture
USFS	United States Forest Service
WFM	Waveform Matching
WSCI	Waveform Structural Complexity Index

List of Publications

The research within this dissertation has resulted in several peer-reviewed publications, and conference presentations and posters, which are listed below, ordered by the chapter in which the are presented in this dissertation.

Chapter 2

Bruening, J.M., Fischer, R., Bohn, F. J., Armston J. D., Armstrong, A, H., Knapp, N., Tang, H., Huth, A., and Dubayah R. O. (2021) Challenges to aboveground biomass prediction from waveform lidar. Environmental Research Letters, 16(12):125013. https://doi.org/10.1088/1748-9326/ac3cec

Bruening, J.M., Fischer, R., Bohn, F. J., Armston J. D., Armstrong, A, H., Knapp, N., Tang, H., Silva, C. E., Huth, A., and Dubayah R. O. (2021) Integrating GEDI observations with a forest model predicts stand-level attributes. AGU Fall Meeting

Chapter 3

Bruening, J. M., May, P. B., Armston, J., D., and Dubayah R., O. (2023). Precise and unbiased biomass estimation from GEDI data and the US forest inventory. Frontiers in Forests and Global Change, 6:1149153 https://doi.org/10.3389/ffgc.2023.1149153

Jamis M Bruening, Paul B May, John D Armston, and Ralph O Dubayah. (2023) GEDI-FIA Fusion Enables Precise and Unbiased Biomass Estimation. AGU Fall Meeting

Chapter 4

Bruening, J. M., Dubayah, R. O., Pederson, N., Poulter, B., Calle, L. (In review). Definition criteria determine the success of old-growth forest mapping. Ecological Indicators

Chapter 1: Introduction

Some of my best childhood memories are of formative experiences exploring the forests and landscapes throughout the Adirondack Mountains and New England. It was not until much later that I came to understand how ecologically valuable the region is within the global climate and carbon cycles, and the potential it holds for learning about how forests regrow and develop after widespread disturbance. This dissertation is inspired by the need to better understand structural and successional dynamics throughout eastern US forests, bounded by the Canadian border to the north, the Atlantic Ocean and Gulf of Mexico to the east and south, and desert and prairie to the west. The research presented here aims to expand our understanding of forest processes in this region, and especially how biomass changes and develops over time. A central theme to this work is identifying the capabilities and limitations of waveform lidar data from NASA's Global Ecosystem Dynamics Investigation (GEDI) for inference into forest dynamics.

In this chapter I explain why the dynamics of eastern US forests are ecologically significant given the region's history, and why we need a better understanding of these processes in the future. I then present three areas of forest research—lidar remote sensing, forest inventory, and ecosystem modeling. I outline the capabilities of GEDI lidar data to accurately and precisely characterize forest structure, and argue that GEDI has the potential to revolutionize our understanding of structural and successional dynamics through integration with forest inventory and ecosystem modeling. I conclude by presenting this dissertation's research goal, the objectives required to achieve that goal, and the scientific questions I will address.

1.1 Background and Motivation

1.1.1 Structural and successional dynamics of eastern forests

The land carbon flux is a critical and uncertain component of the global carbon budget (Friedlingstein et al., 2022). Evidence suggests that carbon uptake in forests globally has outpaced land-use emissions for decades, making the land surface a recent carbon sink (Pan et al., 2011; Houghton et al., 2018). This is due in large part to the regrowth of secondary forests throughout the eastern United States (Kondo et al., 2018; Pugh et al., 2019). As Europeans arrived throughout the 17th and 18th centuries, populations grew and most of the forested landscape was logged (sometimes repeatedly) or cleared for agriculture or pasture (Cogbill et al., 2002). Some estimates suggest that of the 381 million acres of eastern US forests that existed in 1996, only 2 million acres were undisturbed by Euro-Americans (Davis, 1993, 1996). As agriculture intensified and relocated to the mid-west throughout the late 19th and 20th centuries, many fields and pastures in the east were abandoned (Fuller et al., 1998; Foster et al., 2017). This trend initiated land use transitions back to forest that occurred sporadically over time and space, resulting in forest carbon accumulation (Canadell et al., 2007). However, regrowth-driven forest carbon sinks are inherently transient (Körner, 2017), and the future of the regrowth carbon sink in the east is not well understood (Hurtt et al., 2002; Rhemtulla et al., 2009; Keeton et al., 2011). Thus, a better understanding of successional biomass development throughout the region is necessary to more precisely characterize the future trajectory of forest carbon accumulation.

Today, the legacy of forest regrowth from logging and agricultural abandonment has produced a mosaicked forested landscape across the eastern US (Fuller et al., 1998). The matrix of secondary eastern forests is structurally and compositionally more homogeneous than pre-European settlement forests in several ways. Historical forests were a diverse mix of successional and structural stages (Wirth et al., 2009). However, today's landscape is almost entirely comprised of early or mid-successional forests. Expansive tracts of late successional and old-growth forests with very large and very old trees that were once common are now rare (Davis, 1996; Barton and Keeton, 2018). Regrowth dynamics have also altered species distributions, resulting in a more homogeneous forest composition (Thompson et al., 2013). Furthermore, there have been widespread changes to disturbance regimes and environmental stressors, such as increased drought stress from climate change (Pederson et al., 2014), more frequent and diverse pathogens (Haavik, 2019) and fire suppression and management (Barton and Keeton, 2018, Chapter 3,7). These and other changes have impacted both forest composition and structure by interacting with natural demographic and competitive processes that drive stand dynamics and secondary succession. Evidence suggests the effects are leading to forests that are less diverse, shorter in stature and longevity, and lower in potential biomass storage (McDowell et al., 2020).

Now more than ever we need to understand how eastern US forests will change in the near future. This information will further constrain the global carbon balance, and help prioritize conservation efforts, develop management plans that improve overall forest health and ecosystem function, and enact policies to ensure forest longevity and coexistence with humans.

1.1.2 Modes of inference regarding forest dynamics

There are three primary methods used to study forests: ground-based forest inventory and observation, remote sensing of forest attributes, and theory-based forest growth modeling. In the eastern US, two of these methods can be feasibly used to investigate forest dynamics at large spatial scales.

The first is forest inventory. The United States Forest Service maintains a uniformly random network of over 300,000 forest inventory plots throughout the United States, called the Forest Inventory and Analysis (FIA) program (Gray et al., 2012). The FIA program conducts systematic and repeated sampling of forest attributes across the plot network (Bechtold and Patterson, 2005). This enables unbiased estimation of forest variables (such as total biomass or biomass per unit land area) within spatial estimation units, as well as analysis of how attributes change over time at the scale of individual trees, forest stands, and landscapes.

The second is forest modeling. Gap models are computer-based models of forest ecosystems that implement ecological theory to simulate and track demographic processes of recruitment, establishment, growth, and mortality at the level of individual trees, as well as physical structure, growing space, resource availability and light environment within a stand (Shugart et al., 2020). These models simulate dynamic structural and successional processes across spatial scales. Models must be empirically calibrated for a specific site, region, or forest type, and should be validated with inventory data to ensure adequate representation of forest dynamics. This enables experiments and simulations to better understand successional processes that shape forest composition and structure (Maréchaux et al., 2021). Recent advancements in modeling capabilities have resulted in successful integration with remote sensing data and the ability to generate remote sensing-like observations from the forest simulations, significantly increasing inferential abilities and opportunities for synergy (Shugart et al., 2015; Knapp et al., 2018).

Presently, remote sensing technology is not well suited for direct analysis of forest dynamics. While optical measurements enable time-series analysis of forest attributes, the time-series are only several decades long and optical measurements are not as effective for quantifying forest structure as lidar measurements. Lidar is widely accepted as the most effective remote sensing technology to measure forest structure (Dubayah and Drake, 2000). However, repeated measurements are costly and not feasible over large spatial scales. Still, recent advances in lidar technology have transformed our capacity for ecological inference through an abundance of high-quality measurements of ecosystem structure (Dubayah et al., 2022b). NASA's Global Ecosystem Dynamics Investigation (GEDI) was launched in 2018 as the first spaceborne lidar mission designed specifically to measure forest structure (Dubayah et al., 2020a). GEDI was a multi-beam waveform lidar mounted on the International Space Station that directly measured forest structure within footprints 25 m wide along eight reference ground tracks. GEDI's sole observable is a returned waveform, from which a suite of biophysical forest attributes are derived.

GEDI's database of forest structure measurements is unprecedented due to its size and sampling density. However, there are distinct limitations associated with using GEDI data to investigate forest dynamics. GEDI observations were only collected during a four year window, and one can reasonably assume the GEDI database represents forest structure at a single snapshot in time relative to the lengthy time-scales at which forest processes operate. Capabilities of lidar can be greatly expanded through integration with of forest inventory (e.g White et al., 2016) and ecosystem modeling (Knapp et al., 2018; Rödig et al., 2019). The capabilities of GEDI data to study how forest structure changes over time are still developing (e.g. Spracklen and Spracklen, 2021; Bauer et al., 2021; Ma et al., 2022), and there is potential for these data to revolutionize scientific understanding of how forests are composted and structured and how composition and structure changes over time.

1.2 Research goal, objectives, and science questions

The goal of this dissertation is to explore the capabilities of GEDI data for inference into forest dynamics and biomass development in the eastern US. The following three research objectives help achieve this goal:

1. Quantify the uncertainty in biomass values associated with individual GEDI waveforms
2. Correct errors in large scale GEDI biomass estimates through fusion with forest inventory data
3. Characterize the structural properties of different types of late successional forests

To address each objective I ask the following scientific questions. In chapter 2, I explore the relationship between GEDI waveform shape and aboveground forest biomass using a forest gap model. I ask: *Are GEDI waveform shapes associated with larger ranges of potential biomass values than previously assumed? If so, why?*

Chapter 2 quantifies the uncertainty in possible biomass values for individual waveforms, and Chapter 3 progresses this theme by investigating the errors and uncertainties in landscape-scale maps of GEDI biomass produced for the continental United States. Here I ask: *What causes systematic differences between GEDI- and FIA-based estimates of aboveground biomass? To what extent can we leverage FIA inventory data to create an accurate and precise GEDI-based* *biomass map?*

In chapter 4 I assess the structural properties of late-successional forests using unbiased GEDI biomass prediction models developed in chapter 3. I conduct an experiment to identify how much information GEDI can provide about the structure of old-growth forests. In doing so I ask: *Can different types of old-growth be identified and mapped via a structural signature measured by GEDI? What do the structural characteristics of old-growth tell us about biomass development in old forests?*

In chapter 5 I synthesize the discoveries from chapters 2-4 and explain how this research collectively pushes the boundaries of ecological inference. I conclude with a discussion of how my work should inform future inquiries.

Chapter 2: Challenges to aboveground biomass prediction from waveform lidar

2.1 Abstract

Accurate accounting of aboveground biomass density (AGBD) is crucial for carbon cycle, biodiversity, and climate change science. The Global Ecosystem Dynamics Investigation (GEDI), which maps global AGBD from waveform lidar, is the first of a new generation of Earth observation missions designed to improve carbon accounting. This paper explores the possibility that lidar waveforms may not be unique to AGBDthat forest stands with different AGBD may produce highly similar waveformsand we hypothesize that non-uniqueness may contribute to the large uncertainties in AGBD predictions. Our analysis integrates simulated GEDI waveforms from 428 in situ stem maps with output from an individual-based forest gap model, which we use to generate a database of potential forest stands and simulate GEDI waveforms from those stands. We use this database to predict the AGBD of the 428 in situ stem maps via two different methods: a linear regression from waveform metrics, and a waveform-matching approach that accounts for waveform-AGBD non-uniqueness. We find that some in situ waveforms are more unique to AGBD than others, which notably impacts AGBD prediction uncertainty (7411 Mg ha^{-1}, average of 167 Mg ha^{-1}). We also find that forest structure complexity may influence the non-uniqueness effect; stands with low structural complexity are more unique to AGBD than more mature stands with multiple cohorts and canopy layers. These findings suggest that the

non-uniqueness phenomena may be introduced by the measuring characteristics of waveform lidar in combination with how forest structure manifests at small scales, and we discuss how this complexity may complicate uncertainty estimation in AGBD prediction. This analysis suggests a limit to the accuracy and precision of AGBD predictions from lidar waveforms seen in empirical studies, and underscores the need for further exploration of the relationships between lidar remote sensing measurements, forest structure, and AGBD.

2.2 Introduction

One of the most pressing and open questions in climate science is the extent to which forests will act as a net sink or source of carbon in the short- and mid-term future (McDowell et al., 2020; Maréchaux et al., 2021). Accurate baselines of aboveground biomass density (AGBD) within forests are crucial to answering this question, making high resolution mapping of AGBD an immediate need. The Global Ecosystem Dynamics Investigation (GEDI) is the first spaceborne lidar mission specifically intended to map global carbon stocks (Dubayah et al., 2020a). GEDI was launched to the International Space Station (ISS) in December 2018 and provides waveform lidar measurements of forest structure within footprints 25 meters in diameter that are used to predict aboveground biomass. Over the course of its prime mission (April 2019 - April 2021), GEDI recorded over 10 billion land surface observations within the ISS's orbital extent between 51.5 degrees N and S latitude.

The GEDI instrument transmits pulses of light energy towards the Earth's surface and records the intensity of returned energy over time to produce a vertical waveform. Photons reflected by the top of the forest canopy surface are returned to the sensor sooner than others that are reflected lower down in the canopy or by the ground, and more vegetation matter at a given canopy height will yield a larger waveform amplitude at that height. Relative height (RH) metrics are variables derived from the waveform that give the height above the ground at which a certain quantile of returned energy is reached. These metrics are correlated with AGBD (Drake et al., 2002) and are used as predictors in GEDI's biomass models (Dubayah et al., 2020a, 2021b).

Although waveform lidar has proven to be effective for estimating biomass, there remains uncertainty about the accuracies achievable at sub-hectare resolutions, and normalized calibration errors (nRMSE) between 40-50% are common (Zolkos et al., 2013). Others have suggested various sources of this error, such as geolocation errors (Frazer et al., 2011; Rejou-Mechain et al., 2014), tree crowns that overhang plot boundaries (Knapp et al., 2021), errors and uncertainty in allometric equations (Zhao et al., 2012; Ahmed et al., 2013; Vorster et al., 2020) and differences in environmental and edaphic conditions (Shao et al., 2018).

Entirely separate from these is the issue of waveform uniqueness with respect to AGBD. Here, we define "waveform non-uniqueness" as the possibility that a specific waveform shape may be associated with substantially different AGBD values, and "waveform-AGBD uncertainty" as the likely range of AGBD associated with a specific waveform shape. In other words, the extent of a waveform's non-uniqueness impacts the magnitude of its AGBD uncertainty. In light of this possibility a fundamental question arises: can different configurations of trees and their arrangement spatially and vertically yield highly similar waveforms, yet have very different AGBD?

It is therefore of considerable interest to better understand how uniquely waveforms relate to AGBD. To what extent does non-uniqueness occur and under what conditions? What are the implications for instruments such as GEDI, which rely on the assumption that calibration equations convert waveform metrics to AGBD in an unbiased and accurate fashion? To answer these questions we apply a model-data fusion concept using FORMIND, an individual-based gap model (Fischer et al., 2016) and the integration of simulated GEDI waveforms with the FORMIND simulations. Individual-based gap models, such as FORMIND, are a powerful tool to interpret remote sensing observations ecologically, as they allow a direct link between a patch of forest (real or simulated) and how it may appear to a remote sensing instrument (Shugart et al., 2015). By using a large number of simulations in conjunction with *in situ* plots, this link can be studied across a wide range of real world conditions. As such, our modeling framework bridges the gap between spaceborne and ground-based estimates of AGBD, and explores the relationship between AGBD and lidar waveforms in a controlled and systematic manner.

In this paper we examine the issue of waveform non-uniqueness with respect to temperate forest AGBD within the Northeast USA. Our objectives are to; 1) quantify the extent to which waveforms can be non-unique with respect to AGBD within 400m^2 plots; 2) explain possible causes of waveform-AGBD non-uniqueness; and 3) assess the implications of our findings on efforts to predict AGBD from waveform lidar. Through GEDI-FORMIND fusion we employ two distinct methods to estimate AGBD from GEDI waveforms and in doing so characterize the uniqueness of GEDI waveforms with respect to AGBD. We compare the predictions and uncertainties from both methods, and relate patterns in derived AGBD uncertainty to forest stand attributes. Finally, we discuss the relevance of these results on AGBD prediction from lidar waveforms and the possible causes of waveform-AGBD non-uniqueness.

2.3 Methods and data

Our methodological approach (fig. 2.1) is a fusion between GEDI waveforms and the FORMIND model. It uses forest simulations to gain new insights into the relationship between GEDI waveforms and AGBD across a network of 428 field sites throughout the northeast US. We calibrated FORMIND to simulate a database of 896,620 potential forest stands that could exist throughout the Northeast US, using an implementation mode called the Forest Factory (FF) (Bohn and Huth, 2017). We then generated lidar point clouds for every simulated FF stand, and simulated GEDI waveforms from each point cloud using a waveform simulator (Hancock et al., 2019). The result was a database of 896,620 GEDI waveforms from which we developed two different methods of predicting AGBD within *in situ* forest plots. The first is an ordinary least squares (OLS) linear regression model to predict AGBD from lidar waveform RH metrics. The second is a lookup table approach called waveform matching (WFM). For each *in situ* observation, the WFM algorithm identifies the set of the 100 most similar waveforms from the FF database, and infers the AGBD of the *in situ* observation from the distribution of AGBD from those 100 FF stands. We applied both methods (OLS and WFM) to predict the AGBD of 428 *in situ* stem maps from those stands' GEDI waveforms, compared the predictions to one another and to the observed AGBD calculated directly from the inventory data, and contrasted the uncertainties from these two methods. Finally, we performed a regression tree analysis to explore the relationship between the WFM-derived AGBD prediction uncertainty and forest structure.

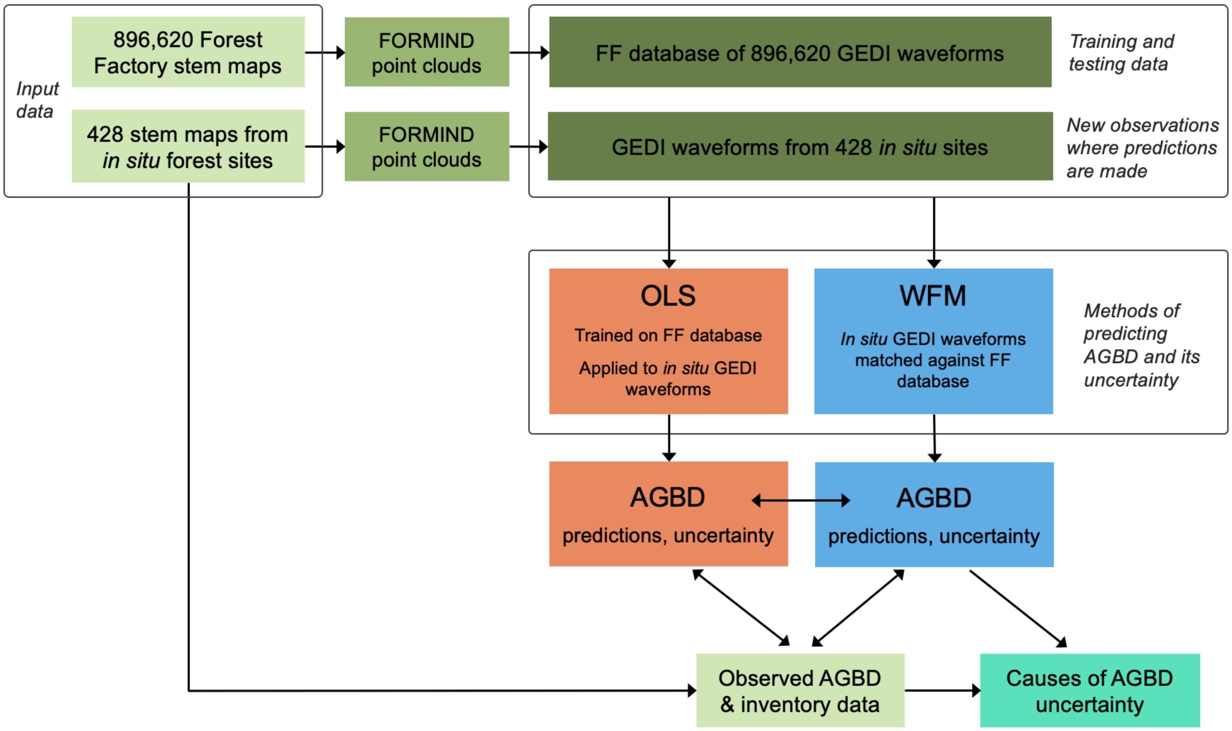

Figure 2.1: This workflow integrates GEDI waveforms and FORMIND using a waveform simulator to enable fusion between waveforms from *in situ* forest stands and those simulated by FORMIND's Forest Factory (FF). The two rows of green boxes at the top represent the *in situ* and FORMIND-simulated stem maps, point clouds, and GEDI waveforms that serve as the input data for the analysis. The FF database of 896,620 GEDI waveforms is used to predict the AGBD of 428 *in situ* stem maps using two different methods, an OLS regression (orange) and the WFM lookup table approach (blue). We quantify the uncertainty associated with each set of predictions, and compare both sets of predictions and uncertainties to each other and to the observed AGBD calculated from the *in situ* inventory data. Lastly, we identify the potential causes of WFM-derived AGBD prediction uncertainty as a function of stand dynamics and structural complexity within temperate forests.

The following sections provide an overview of our workflow, and more details are provided in appendix A.

2.3.1 Field sites

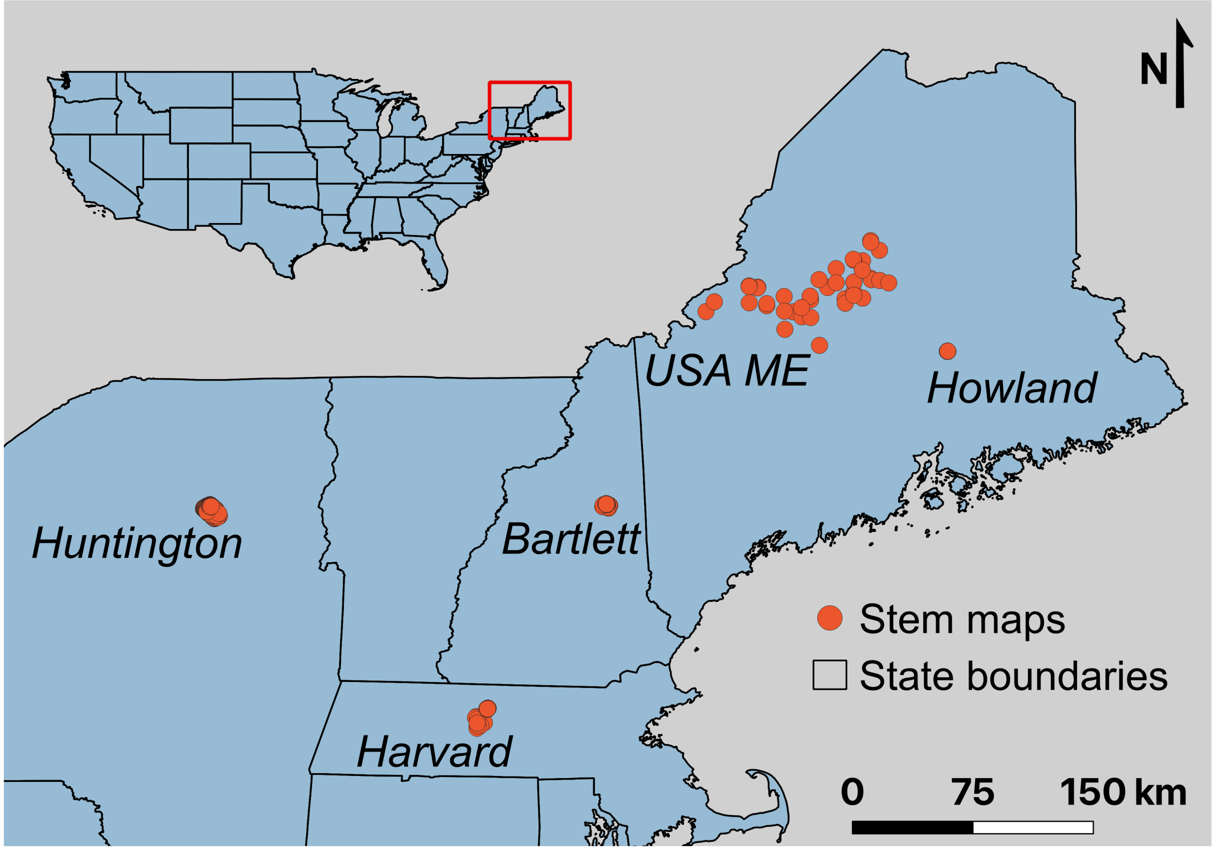

Figure 2.2: Our study uses five forest research sites and 428 stem map locations distributed throughout the northeast US.

Table 2.1: Site information for the five forest research sites which contribute stem maps to this analysis. 'N' refers to the number of stem maps at each site, 'GEDI' indicates whether the site is part of the GEDI Forest Structure and Biomass Database (Dubayah et al., 2020a), and 'YEAR' is the plot inventory year.

Abbr.	Site name	Approximate Lat Lon	N	Source	GEDI	Year
usme	USA ME	-70.02, 45.58	42	NASA CMS	Yes	2015
harv	Harvard	-72.18, 42.53	35	NEON	Yes	2015-2017
bart	Bartlett	-71.29, 44.05	25	NEON	Yes	2016-2018
howl	Howland	-68.73, 45.20	70	Univ. of Maine	Yes	2015
hunt	Huntington	-74.22, 43.97	256	SUNY ESF	No	2011

The Northeast US (fig. 2.2) was chosen for this study due to the high availability of inventory data (table 2.1), the large range of potential AGBD values (Keeton et al., 2011), and importance to the global carbon budget through secondary forest regrowth (Pugh et al., 2019). We use 428 stem mapped forest plots that come from five different research areas, four of which are projects within the GEDI Forest Structure and Biomass Database (Dubayah et al., 2020a). Plot size and shape varied across the five areas, which necessitated standardization to a common square plot shape of 20 m x 20 m (such as clipping larger plots to conform to the smaller square shape), however the relative spatial position of all trees within the plot was preserved. Trees greater than 12.7 cm (5 in) in diameter at breast height were measured and their locations recorded within the plot.

2.3.2 FORMIND and the Forest Factory

FORMIND is a forest gap model that simulates growth dynamics at the level of individual trees (Fischer et al., 2016). It allows simulation of forest dynamics and structure, including gap formation (falling down of large trees) and succession. FORMIND simulates all physiological processes (photosynthesis, respiration, tree growth, mortality, regeneration, competition) at the tree level. Growth of a single tree is calculated on the basis of a carbon balance and depends on the tree size, climate conditions and the shading of surrounding trees. Forests are represented as a collection of square patches (stands) which may vary in successional and structural stage. The size of these stands correspond with the light competition range of trees, which in this case is 400m^2 (Fischer et al., 2016).

We calibrated a regional version of FORMIND to represent forest composition and structure characteristic of the Northeast US, using all tree-level data from the US Forest Service's Forest Inventory and Analysis (FIA) program's most recent survey of the Northeast USA. We segmented the region's 27 most abundant species into nine different plant function types (PFT) based on each species' maximum size (height and stem diameter), growth rate, and shade tolerance, and then calibrated a set of tree geometry equations for each PFT so that forest stands simulated in FORMIND represent generalized structural characteristics of forests in this region (table A.1). We then validated these parameters against the inventory data from the *in situ* field sites.

FORMIND has an implementation mode called the Forest Factory (FF), which simulates the structure (not growth) of many unique, 400m^2 forest stands (Bohn et al., 2014; Bohn and Huth, 2017). The purpose of the FF is to simulate the diversity in forest structure within a region, based on varying the PFT compositions and stem-size distributions across a large number of simulated forest stands. We implemented the FF using the Northeast US calibration of FORMIND to simulate a structural diversity database containing 896,620 unique forest stands that could exist throughout the region. This number of simulations ensured the database covered a wide range of structural and compositional diversity, given the potential occurrence of up to nine different PFTs within a stand, and various potential stem size distributions that have been observed across successional and structural gradients in temperate forests (Lorimer and Halpin, 2014; Bohn and Huth, 2017). This database serves as the basis for the OLS and WFM AGBD prediction methods, as it encompasses the myriad of forest stand structural configurations and AGBDs that are possible throughout the Northeast US.

2.3.3 Lidar simulations

This study is based on GEDI waveform comparisons between real and simulated forests, so it was necessary to standardize these data sources. This section explains our standardization approach to simulate GEDI waveforms for the 428 *in situ* stem maps and all 896,620 stem maps in the FF database.

First we input each *in situ* stem map into FORMIND using its initialization feature, and had the model construct the stand's structure according to the stem map and previously calibrated allometric relationships. We implemented FORMIND's lidar simulator to generate a point cloud representation of each stem map, according to Knapp et al. (2018), and subsequently simulated a GEDI waveform from each stem map point cloud using the GEDI waveform simulator (Hancock et al., 2019). The waveform simulations mirror GEDI's calibration process (see Dubayah et al., 2020a), except here the waveform footprint width was set to 20 meters to match the size of the 428 *in situ* and 896,620 FF stem maps. The resultant 428 GEDI waveforms represent FORMIND's rendering of forest structure within 428 real forest stands throughout the Northeast US.

We used almost the same process as above to simulate a GEDI waveform from every stand in the FF database. The only difference was an added step to eliminate edge effects between FF stands. The FF outputs 100 stands at once in a 10x10 grid, so we separated the stands when generating the lidar point clouds to eliminate the influence of tree crowns that overhang plot boundaries (see Knapp et al. (2021) and appendix A for details).

The end result is that all 896,620 stem maps in the FF database have an AGBD value and a corresponding GEDI waveform, calculated in the same exact manner as the 428 field sites. If one of the FF stands were to actually exist somewhere in a forest and was inventoried, the AGBD value calculated from the inventory data would be same as the value in the FF database, and ingesting the stem map into FORMIND (as was done for the 428 *in situ* stands) would produce a point cloud and simulated GEDI waveform identical to its counterpart in the FF database. This is crucial to our analysis because it ensures standardization between all data from the *in situ* stands and FF database. This approach removes allometric variability between tree species and across sites, because all trees in the stem maps (the 428 *in situ* and 896,620 FF database) were constructed in our Northeast US regional parameterization of FORMIND.

2.3.4 Deriving AGBD and its uncertainty from lidar waveforms

We use two methods to predict forest stand AGBD and estimate prediction uncertainty from GEDI waveforms: ordinary least squares regression (OLS), and a lookup table approach we call waveform matching (WFM). Both methods rely on the 896,620 GEDI waveforms from the FF database to predict the AGBD and uncertainty of each of the 428 *in situ* forest stands from their respective GEDI waveforms.

2.3.4.1 OLS regression modeling

We used an ordinary least squares (OLS) linear regression model to predict AGBD from GEDI waveform RH metrics in 10% increments from RH10 to RH90, with the addition of RH98 (a more stable indicator of top of canopy height than RH100). We developed a set of 18 candidate models based on relevant literature, using a square root transform on the response (Duncanson et al., 2020; Kellner et al., 2023). We trained each model on half of the FF database (n=448,310) so that the AGBD models were derived from the same simulated forest stands used in the WFM approach (section 2.3.4.2). To assess model performance, we randomly split the remaining half of the FF database not used for training into 500 different testing sets (each with n=897), and applied all 18 candidate models to all 500 testing sets. We selected the final model (fig. A.1 and table A.2) based on the lowest average RMSE across the 500 tests, and applied it to the 428 *in situ* stem map GEDI waveforms to generate OLS-based predictions of AGBD and the associated 90% prediction interval for each site.

2.3.4.2 Waveform matching

WFM is a process that quantifies the similarity in shape between two waveforms (Inman and Bradley Jr, 1989; Rödig et al., 2019). Similarity is defined as the relative overlapping area between two waveforms, expressed as the ratio of area shared by both waveforms to the entire area encompassed by either waveform, as follows

$$r = \sum_{h=0}^{mch} min(E_x(h), E_y(h)) \left(\sum_{h=0}^{mch} max(E_x(h), E_y(h)) \right)^{-1} \tag{2.1}$$

where r is the relative overlap, h is height above ground in meters, mch is the maximum canopy height between the waveforms, and E_x and E_y are the returned waveform energies at a given value of h from the stem map and FF-stand waveforms, respectively. We used the same height step of 0.15 m as the waveform simulator (Hancock et al., 2019).

We applied the WFM algorithm individually to the 428 stem map waveforms, and identified the 100 FF-stand waveforms with the largest r for each. When an *in situ* stand did not have 100 FF matches with $r > 0.75$, we removed it from the rest of the analysis to guarantee a high degree of similarity between each *in situ* waveform and its FF matches. The result was a set of 100 best matching FF stands for each *in situ* stem mapped stand, based on waveform shape.

To derive the *in situ* stand's WFM-predicted AGBD value, we use the median AGBD value from the set of 100 FF matches. We represent the WFM AGBD prediction uncertainty as the range in AGBD that encompasses the middle 90% of the 100 FF AGBD values. We define this range as the magnitude of uncertainty in AGBD associated with a stem map's waveform, which represents the extent of non-uniqueness with respect to AGBD for each stem map waveform.

2.3.5 Explaining AGBD uncertainty

WFM yields predicted values of AGBD for each stem map, as well as the uncertainty around each estimate. We performed a regression tree analysis, using the *rpart* package in R (Therneau et al., 2010), to identify the extent to which forest structure variables explain patterns in the WFM-derived AGBD uncertainty across the *in situ* stem maps. The explanatory variables used were waveform entropy and skewness, the standard deviation in tree heights, standard deviation of tree diameters, and stand basal area. To ensure a simple and interpretable model, we set the maximum tree depth to three, and only allowed a split from nodes with at least 15% of the total sample. This means the final model could have had a maximum of seven possible splits, and eight possible classes, and that a split was not allowed if a node had less than 15% of the total observations.

2.4 Results

2.4.1 Stem map waveform matches

Waveform matching (WFM) revealed 100 best matches with $r > 0.75$ for 380 of the 428 stem maps (fig. A.2). The distribution of biomass across these 380 *in situ* stands (fig. A.4A) is very similar to the distribution of biomass across all 38,000 FF stands that were identified as a match to the 380 *in situ* stands (fig. A.4B). Across these 380 sites, the WFM-derived uncertainty in AGBD ranged from 7 - 411 Mg ha^{-1}(mean of 167 Mg ha^{-1}), and from 0.56 - 5.57 (mean of 1.29) as a ratio relative to the WFM-predicted AGBD. There was considerable variability in the WFM-derived distributions of AGBD associated with different waveform shapes (e.g. figure 3), as some waveforms had a larger range of possible AGBD values than others (fig. A.3).

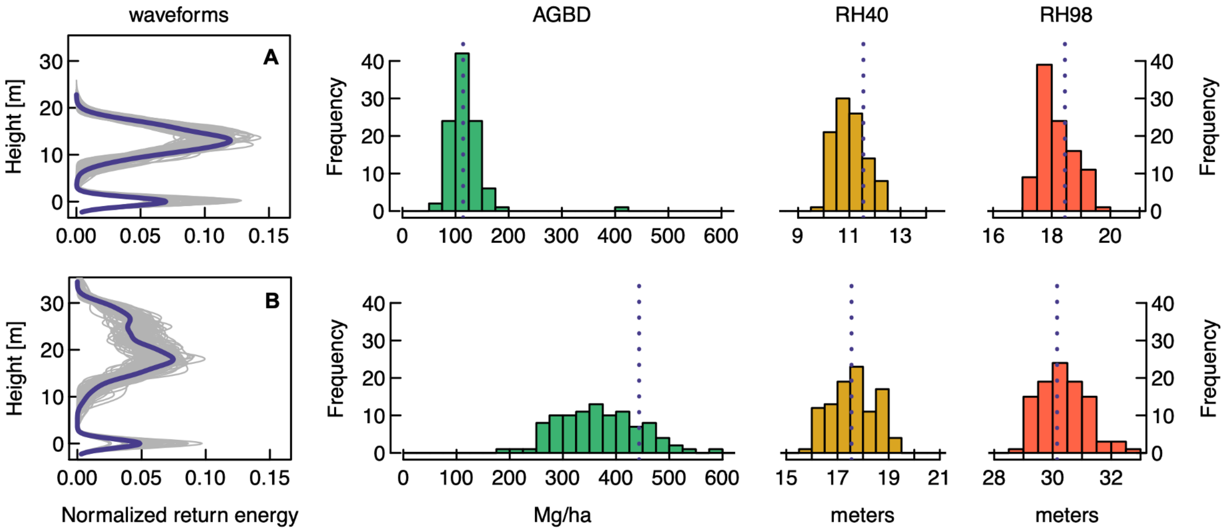

Figure 2.3: The distributions of AGBD (green) from the 100 matching waveforms (grey lines) for two example stem map waveforms (blue lines) are highly variable. Waveform A has less variability in AGBD across its 100 matches than waveform B. Waveform B has a larger uncertainty in WFM-derived AGBD than waveform A, despite similar variability in RH40 and RH98 across the 100 matches for these two examples. According to the WFM method, waveform A is more unique to AGBD than waveform B. The vertical dotted lines represent the *in situ* stem mapped stand's value for each variable.

2.4.2 AGBD predictions and uncertainty

The OLS and WFM methods performed similarly with respect to prediction accuracy (R^2 of 0.64 and 0.66 respectively) and overall error (nRMSE of 32.1% and 31.6% respectively) (fig. 2.4A-B). The agreement between each set of predictions was higher than between either set and the observed values, and there was substantial agreement between the residuals (fig. 2.4C-D). However, the uncertainty associated with WFM- and OLS-predicted AGBD were different (vertical lines in fig. 2.4A-B). Across these sites, AGBD uncertainty was lower from WFM than from OLS, despite a larger degree of variability and some extremely large values in the WFM prediction uncertainty (fig. 2.5). When the AGBD prediction uncertainty from each method was compared directly, and the magnitude of uncertainty from WFM was lower than from OLS at 313 of the 380 sites with 100 FF matches.

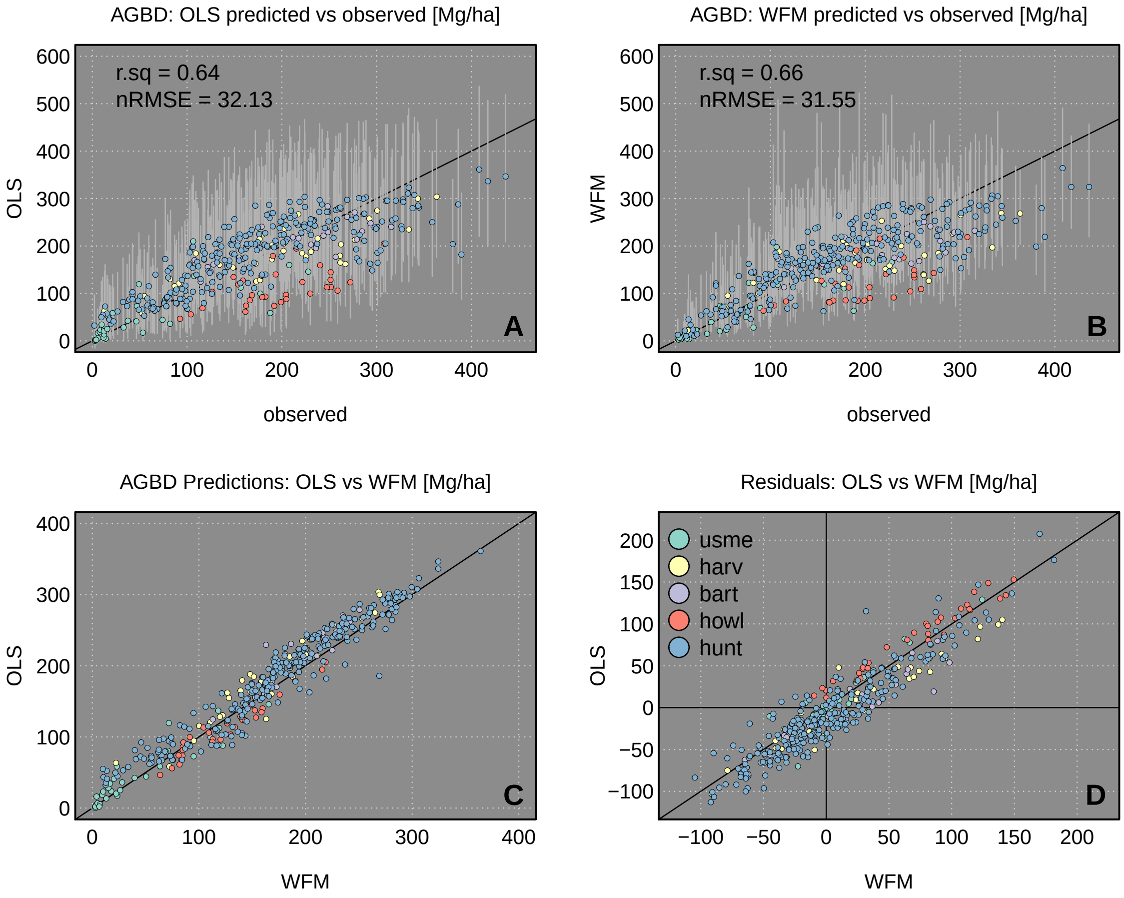

Figure 2.4: Ordinary least squares (OLS) (A) and waveform matching (WFM) (B) perform similarly in AGBD prediction compared to the observed values. There is strong agreement between the predicted values from both methods (C) and the residuals (D), despite differences in reported uncertainty (grey lines in A and B).

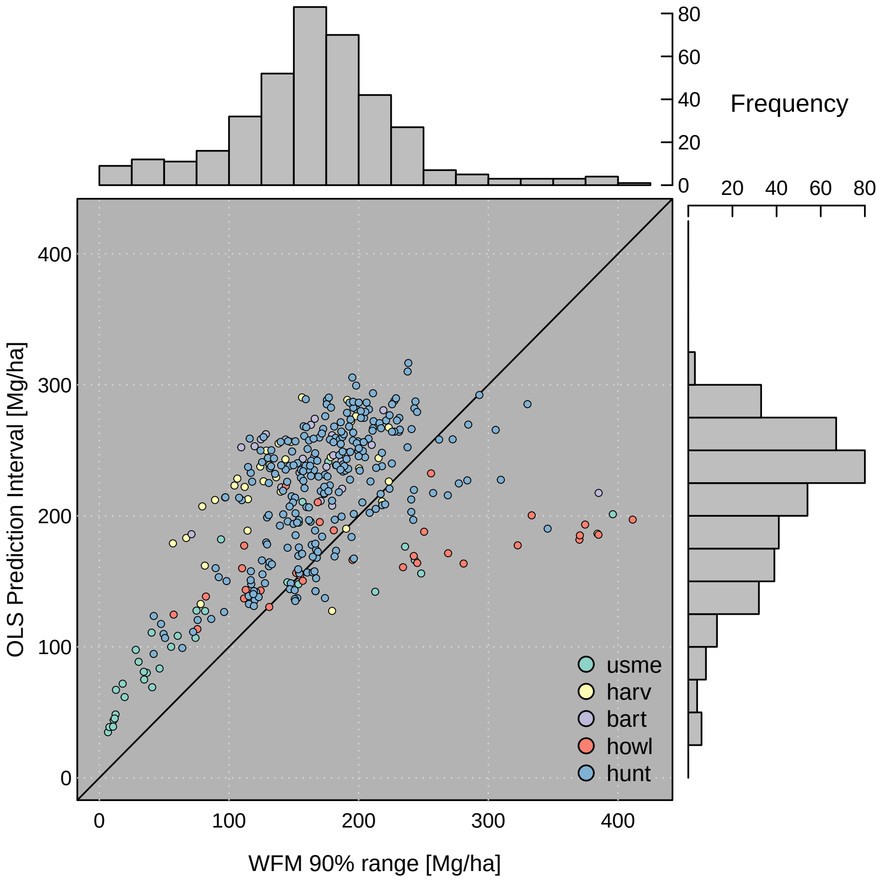

Figure 2.5: The WFM prediction uncertainty is typically lower than the OLS prediction uncertainty. The WFM prediction uncertainty is the range of AGBD that encompases 90% of the 100 FF matches AGBD values for a given *in situ* waveform, and the OLS prediction uncertainty is the width of the 90% prediction interval for each prediction. The solid line shows the 1:1 relationship, and the bin width for both histograms is 25 Mg ha^{-1}.

2.4.3 WFM AGBD uncertainty as a function of forest stand attributes

The regression tree analysis (R^2=0.4) partitioned the stem maps into five groups of increasing AGBD uncertainty (derived from WFM) (fig. 2.6). Total stand basal area had the most explanatory power with respect to predicting the AGBD uncertainty, as stands with low basal area (< 10 m^2 ha^{-1}) tended to have a smaller uncertainty than stands with a larger basal area. Among stands with basal area > 10 m^2 ha^{-1}, those with more variation in tree height (height standard deviation > 4 m) tended to have larger AGBD uncertainty than stands with less variation in tree height. Lastly, within each resulting group (both above and below height standard deviation of four meters), stands with more variation in tree diameter tended to have larger AGBD uncertainty than stands with less variation in tree diameter. Forest attributes differed substantially across the five groups (fig. 2.7).

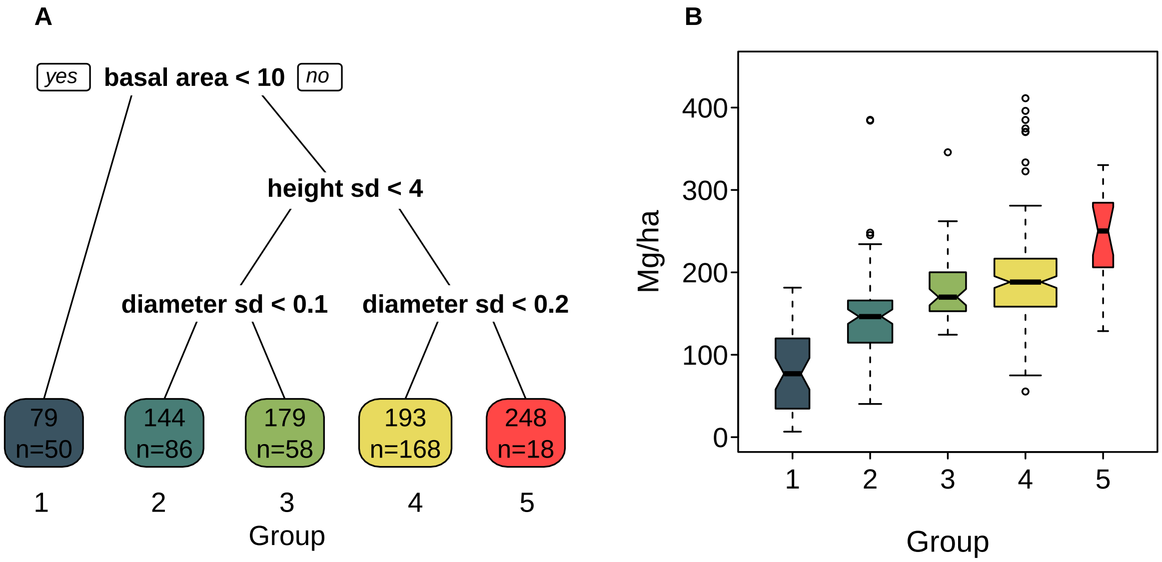

Figure 2.6: The AGBD prediction uncertainty derived from WFM is related to a stand's basal area [m^2ha^{-1}] and heterogeneity in tree height, represented here as the standard deviation in tree height [m] and diameter [m]. The regression tree (A) partitions the stem maps into groups based on the influence of various forest stand attributes on WFM-derived AGBD uncertainty (B). The top value in each terminal node of the tree is the group's mean AGBD uncertainty, and the bottom number is the number of stands in the group. The boxplot widths are weighted according to the square root of the number of observations in each group.

2.5 Discussion

The outcome of the WFM analysis indicates that lidar waveforms are not unique to AGBD, but instead are associated with a distribution of possible AGBDs (e.g. fig. 2.3), and the median

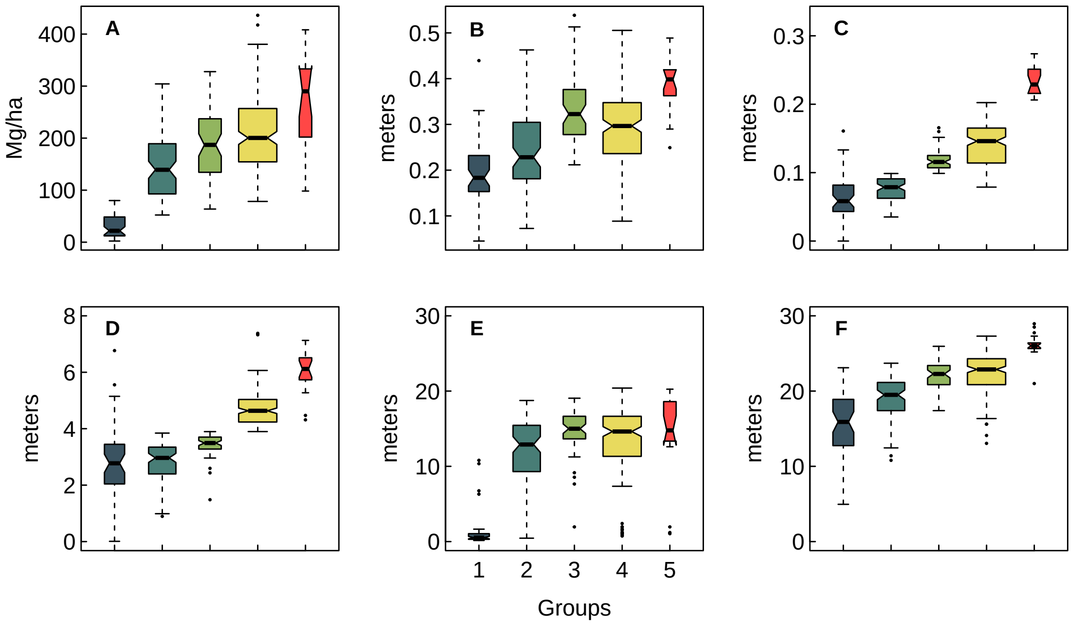

Figure 2.7: Forest stand attributes differ across the groups of increasing AGBD uncertainty. Panels are: (A) AGBD, (B) mean tree diameter, standard deviation of tree diameters (C) and heights (D), (E) RH50, (F) RH95. The boxes are colored according to group, to match figure 6.

of this distribution tends to be a reliable predictor of the observed AGBD within the waveform footprint (fig. 2.4B) when compared to the OLS predictions. Additionally, the WFM-derived distributions of AGBD associated with different waveform shapes are themselves variable (e.g. fig. 2.3). Thus some waveforms are associated with a wider range of possible AGBD than others (fig. 2.5). Our results also support that a specific value of AGBD can be associated with multiple different waveform shapes. Two stands with different tree configurations may have the same AGBD, yet these stands would produce considerably different waveforms (fig. 2.8). This is evidenced by the overlapping AGBD uncertainties for different waveforms (fig. 2.4A-B). AGBD is not unique to a specific waveform shape, perhaps because multiple pathways of forest stand development can lead to a given AGBD. In the following sections we discuss how waveform-AGBD non-uniqueness may present limitations in AGBD prediction from lidar, and we provide a possible explanation for the observed pattern in WFM-derived AGBD uncertainty across the stem map field sites used in this analysis.

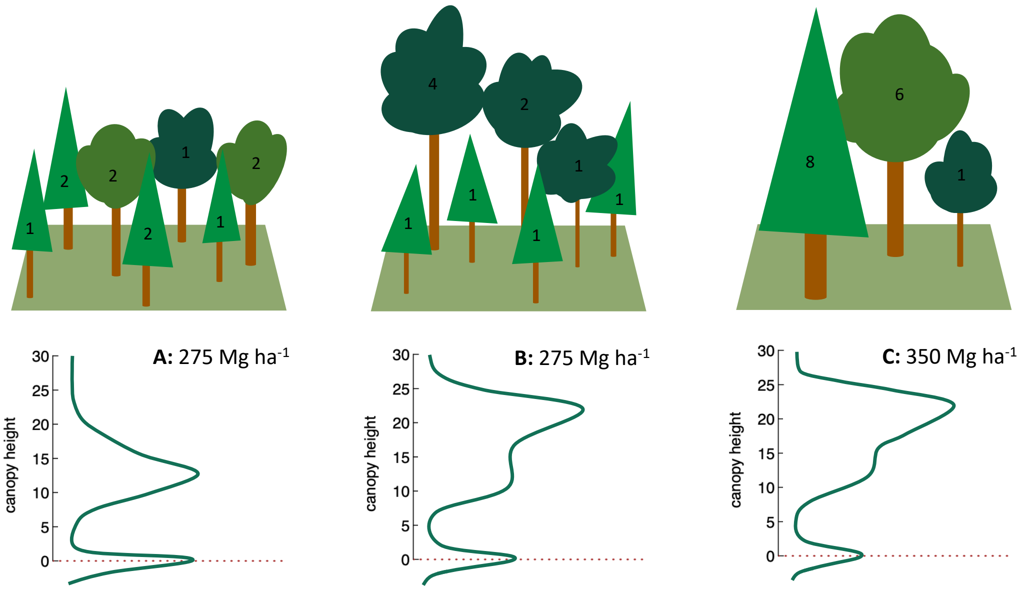

Figure 2.8: Forest stands (400m^2) with unique tree configurations can have similar AGBD yet different waveforms (A and B), or different AGBD yet similar waveforms (B and C). The numbers in each tree represent its AGB Mg, and the reported densities here have been scaled to 1 ha. These stands are only conceptual, used to illustrate the non-uniqueness phenomenon, and not actual stands within our *in situ* or simulated databases.

2.5.1 Limits to AGBD prediction and uncertainty estimation from lidar

A consistent and perplexing issue associated with modeling AGBD is the heteroscedastic predictions and high nRMSEs in AGBD calibration equations for small plot sizes (Zolkos et al., 2013). While the community has searched for solutions to this problem through the inclusion of more and more complex metrics (Ni-Meister et al., 2010; Knapp et al., 2018) and fusion with other remote sensing data (Lu et al., 2012; Qi et al., 2019), the problem persists. Our results suggest the heteroscedasticity may not be solvable; rather it may be introduced by the measurement properties of waveform lidar at small scales.

Waveforms represent an aggregate measure of vertical structure throughout the footprint, reducing the information from multiple trees into a single observation. There is no horizontal differentiation within a waveform, as it represents the amount of combined plant matter at a given canopy elevation across the entire footprint area. Further, the laser pulse follows a Gaussian distribution within the footprint, so vegetation matter at the center of the footprint influences the waveform more than matter near the edges. We argue this aggregation across space may introduce the heteroscedasticity between waveform metrics and AGBD, and is more likely for small plot sizes. It is generally known that variability in AGBD increases with decreasing plot size, and that smaller area plots may also have much larger AGBDs than larger area plots. The scale-dependent variability in forest structure and biomass may explain why, at small scales, a specific waveform shape can be produced by stands with substantially different AGBDs. A decrease in AGBD variability in larger scales would decrease this effect, resulting in smaller ranges of AGBD associated with a given waveform shape. This would help explain the large decrease in nRMSEs observed in biomass models as calibration plot size increases (Zolkos et al., 2013). Indeed, Knapp et al. (Knapp et al., 2018) demonstrate that as plot sizes increase, the ability of lidar structural metrics to predict AGBD increases concurrently with a decrease in the variability in both AGBD and structural metrics. In effect, the uniqueness between LiDAR metrics and AGBD may increase with plot size.

The non-uniqueness effect may complicate efforts to estimate the uncertainty of AGBD predictions from statistical models. Uncertainty in regression models is often represented as a prediction interval, defined as the range in which the predicted value from a new observation will fall given what has been observed in the model training sample, for any desired probability. In the case of OLS, the size of the prediction interval for a new observation depends on the training sample mean and variance, and how far from the mean that new observation is. While statistically valid, prediction intervals do not recognize that some lidar waveforms may be more unique to AGBD than others. As such, the OLS prediction interval measures something inherently different than the WFM-derived AGBD uncertainty (the 90% range of possible AGBD associated with the 100 matches to a given stem map's waveform). The motivation for WFM is to quantify the variability in AGBD associated with a given waveform shape from a specific forest stand, absent all other influences, especially other forest stands' waveforms and AGBDs. In doing so this method accounts for the possibility that some waveforms may be associated with much larger ranges of AGBD than others, something not accounted for in OLS prediction intervals. This would help explain the differences between the OLS prediction intervals and WMF-derived uncertainty (fig. 2.5).

In both WFM and OLS, a given waveform shape will always result in the same predicted value of AGBD—there is no stochasticity in either prediction method. However, the knowledge that multiple forest stands with different structures and waveform shapes can have the same AGBD (e.g. fig. 2.8) highlights another important difference between OLS and WFM. In OLS, a predicted value of AGBD is unique to the set of predictor values used in the regression equation to make that prediction. In other words, the number of waveform shapes that could result in a specific predicted AGBD is directly related to the number of terms in the model. In a model that only uses RH98 to predict AGBD, a specific value of RH98 can only result in one predicted AGBD value, and that prediction is associated with a single prediction interval. However this is not the case in WFM, in which the prediction uncertainty is a function of the variability in AGBD associated with the matched set of waveforms.

2.5.2 Drivers of waveform-AGBD uncertainty

A conceptual model may characterize forest structure according to the interaction between three attributes: 1) the number and spatial position of trees within a stand 2) the sizes of the trees (e.g. maximum height, crown shape, stem diameter), and 3) the variation in tree sizes within the stand. A forest stand's structural stage can then be classified according to these attributes (Oliver et al., 1996; Keeton et al., 2011; Lorimer and Halpin, 2014), and it has been shown that discrete airborne lidar observations can effectively discriminate between stands in various structural classes (Falkowski et al., 2009; Kane et al., 2010). Our results indicate that the extent to which lidar waveforms are unique to AGBD depends on the structural stage of the forest stand.

The regression tree (fig. 2.6) segmented the *in situ* stands into in five groups of increasing WFM-derived AGBD uncertainty, and the magnitude of that uncertainty is related to a stands basal area and variation in tree sizes. Stands with larger basal area and more variation in tree height and diameter tend to produce waveforms that are less unique to biomass than stands with smaller basal area and more homogeneous tree sizes. Together, the five groups of forest stands represent a progression in structural complexity driven by dynamic forest processes (Oliver et al., 1996).

Investigations of temperate forest stand dynamics have resulted in multiple classifications of forest stand structural development (e.g. (Oliver et al., 1996; Spies, 1997; Palik et al., 2020)), however the general trends in structural development over time are consistent. Following a stand replacing disturbance or at the start of old-field succession, the stand-initiation phase is characteristic of open canopies with a single stratum of saplings, rapid growth, small basal area, and low biomass density (Oliver et al., 1996). The group with the lowest WFM-derived AGBD uncertainty embodies these characteristics (figs. 2.6 and 2.7).

Upon stand initiation, most available growing space is quickly occupied and trees start to directly compete for resources during the stem exclusion phase (Oliver et al., 1996). Boles become larger and a dense canopy shades the forest floor, precluding new seedlings and perpetuating a single stratum of homogeneous tree sizes. As individuals in the overstory start to die, growing space becomes available and access to resources lower down in the canopy enables a transition to the understory reinitiation stage. Overtime, the canopy may stratify into various layers, and successive mortality events sustain a multi-layered canopy composed of different cohorts, stems of various sizes, and the potential for high biomass density in the multi-strata stage (Barton and Keeton, 2018). Groups 2-5 embody a general transition from young initiated stands through the various stages of structural development to more mature, mosaiced stands with established understories and multiple canopy strata.

The characteristics of these five groups illustrate how patterns of structural development within a forest may impact the uniqueness of waveforms with respect to AGBD. It appears that the ability of lidar waveforms to uniquely represent biomass within the footprint decreases as the structural complexity inside the footprint increases. A waveform that represents a forest stand with a single, clearly defined canopy layer and moderate ground return tends to map to a small range of AGBD because the forest stands that could produce such a waveform must be composed of trees of similar sizes that form a single canopy layer, and those stands are similar in structure and AGBD (fig. 2.3A). Conversely, a waveform that indicates multiple canopy strata and more canopy cover could theoretically come from a set of stands with larger compositional and structural variety, resulting in a relatively wider range of AGBD (fig. 2.3B). Waveforms capture a finite amount of structural information about a stand, which is only a fraction of the stand's total structural information (e.g. maximum height, crown shape, and stem diameter of every tree in the footprint). We argue that this fraction likely varies based on the structural complexity of the observed stand, and in the context of AGBD prediction, the higher the fraction of structural information captured, the more unique a waveform may be to AGBD.

2.5.3 Forest simulations

The heavy reliance on simulated data in this analysis is intentional, as this experiment would not be possible using field data alone, and each stage of the workflow has been extensively tested and validated (Fischer et al., 2016; Bohn and Huth, 2017; Knapp et al., 2018; Hancock et al., 2019; Rödig et al., 2019). The use of FORMIND and the Forest Factory allows us to systematically explore the relationship between forest structure and AGBD across a wide range of structural conditions that exist in real forests, and to make novel inferences about AGBD prediction uncertainty. FORMIND captures general trends in forest structure across time and space, and it is not intended to predict the exact AGBD value of a single stand. Instead, we use it to assess the likely variability in AGBD predictions based on a given forest structure, rather than reporting our predicted AGBD values as truth. Additionally, FORMIND does not represent stochastic variability in tree allometry, or other differences due to environmental gradients or other factors known to influence forest structure (past disturbances, land use transitions, etc.). As such the point cloud representation of the *in situ* forest stands do not exactly reproduce the structural complexity of these stands in the real world, as shrubs, downed trees and debris, and small trees are not represented in the stem maps or FORMIND-generated point clouds, nor are the potential impacts of topography represented in the point clouds or waveforms. However, all of these aspects are absent from both the *in situ* and FF stem maps and waveforms, ensuring a like-to-like comparison.

We have attempted to reproduce results common to empirical studies (increasing variation in predictions as AGBD increases and large calibration nRMSEs) in a modeling framework, and in doing so expose waveform-AGBD non-uniqueness as a contributing factor. Our experimental design allowed for a controlled setting to explore this possibility, while eliminating various proposed causes of heteroscedastic predictions and large calibration nRMSEs, as follows. An unrepresentative training sample is not likely a factor, as 896,620 FF stands were used to train and test the OLS model, and expanding the training sample would not act to reduce the non-uniqueness effect. There is no geolocation error between the FORMIND-generated point clouds and the *in situ* or FF-derived stem maps, nor do any of FF point clouds have tree crowns that overhang the plot boundaries, both of which have been suggested to add to calibration equation error and prediction scatter (Frazer et al., 2011; Rejou-Mechain et al., 2014; Knapp et al., 2021). All the FF and *in situ* stands are constructed according to the same set of PFT-specific allometric equations, so there is no allometric variability within a PFT or between stands, and the same equations used to estimate the field biomass values of the stem maps were used to obtain the biomass of the simulated FF stands. We do not claim these factors have no influence on AGBD predictions from LiDAR waveforms in general, however by controlling for their influence in this analysis, we conclude that the non-uniqueness effect may account for a substantial amount of AGBD prediction uncertainty and error, and should not be overlooked in future studies.

2.6 Conclusions

In this paper we used GEDI-FORMIND fusion to explore the possibility that lidar waveforms are not unique to AGBD. Our results support that lidar waveforms may instead be associated with a range of potential AGBD values, and that this range varies among waveform shapes. We have demonstrated that within the study extent, the range of AGBD associated with specific a waveform may be a function of the stand's structural characteristics. Forest stands in early structural development tend to be relatively homogeneous and similar to one another, resulting in low biomass variation for a given structural signature (waveform shape). Over time, stand dynamics drive changes in forest composition and structure. This process of structural development over time may result in substantial differences in biomass between structurally mature stands, yet due to the measurement properties of waveform lidar, some stands may still produce similar waveform shapes. The result is that some waveform shapes are likely associated with a small range of possible AGBD, while others may be associated with a greater range of possible AGBD.

The phenomena of waveform non-uniqueness with respect to AGBD presents challenges in the context of traditional approaches to modeling AGBD from lidar waveforms. However, these challenges are not new and are well documented (Zolkos et al., 2013). Acknowledging this phenomena may help explain the heteroscadasticity and large calibration errors present in empirical studies, although further investigation into waveform-AGBD non-uniqueness across scales is necessary. The non-uniqueness effect is perhaps intuitive to some extent, however we have attempted to isolate its influence from other factors known to impact AGBD predictions. In doing so, this work highlights limitations to AGBD prediction from waveform lidar at fine scales.

2.7 Acknowledgements

This work was primarily supported by funding from the Global Ecosystem Dynamics Investigation and NASA FINESST Grant (80NSSC21K1626). The authors would like to thank Donal O'Leary and Steve Hancock for thoughtful discussions that improved the analysis, and Shawn Fraver for sharing the inventory data from Howland, ME. The authors are also appreciative of the comments from two anonymous reviewers that greatly contributed to the quality of this manuscript.

Chapter 3: Precise and Unbiased Biomass Estimation from GEDI data and the US Forest Inventory

3.1 Abstract

Atmospheric CO_2 concentrations are dependent on land-atmosphere carbon fluxes resultant from forest dynamics and land-use changes. These fluxes are not well-constrained, in part because reliable baseline estimates of forest carbon stocks and the associated uncertainties are lacking. NASA's Global Ecosystem Dynamics Investigation (GEDI) produces estimates of aboveground biomass density (AGBD) that are unique because GEDI's hybrid estimation framework enables formal uncertainty calculations that accompany the biomass estimates. However, GEDI's estimates are not without issue; a recent validation using design-based AGBD estimates from the US Forest Inventory and Analysis (FIA) program revealed systematic differences between GEDI and FIA estimates within a hexagon tessellation of the continental United States. Here, we explored these differences and identified two issues impacting GEDI's estimation process: incomplete filtering of low quality GEDI observations and regional biases in GEDI's footprint-level biomass models. We developed a solution to each, in the form of improved data filtering and GEDI-FIA fusion AGBD models, developed in a scale-invariant small area estimation framework, that were compatible with hybrid estimation. We calibrated 10 regional Fay-Herriot models at the hexagon scale for application at the unit scale of GEDI footprints, for which we provide a mathematical justification and empirical testing of the models' scale-invariance. These models predicted realistic distributions of unit level AGBD, with equal or improved performance relative to GEDI's L4A models for all regions. We then produced GEDI-FIA fusion estimates that were more precise than the FIA estimates and resulted in a bias reduction of 86.7% relative to the original GEDI estimates: 19.3% due to improved data filtering and 67.5% due to the new AGBD models. Our findings indicate that (1) small area estimation models trained in a scale-invariant framework can produce realistic predictions of AGBD, and (2) there is substantial spatial variability in the relationship between GEDI forest structure metrics and AGBD. This work is a step toward achieving reliable baseline forest carbon stocks, provides a viable methodology for training remote sensing biomass models, and may serve as a reference for other investigations of GEDI AGBD estimates.

3.2 Introduction

NASA's Global Ecosystem Dynamics Investigation (GEDI) is the first spaceborne mission designed specifically to estimate aboveground biomass (Dubayah et al., 2020a). GEDI is a multi-beam waveform lidar mounted on the International Space Station that directly measures forest structure within footprints 25 meters wide along eight reference ground tracks. GEDI's sole observable is a reflected waveform, from which forest structure metrics are calculated. These waveforms enable above ground biomass density (AGBD) mapping within a regular spatial grid or jurisdictional boundaries, as follows. First, GEDI's level 2A (L2A) signal processing algorithms are applied to each waveform to calculate metrics representative of forest structure attributes

(Hofton et al., 2019). Second, GEDI's level 4A (L4A) footprint AGBD models use the metrics to predict AGBD within the waveform footprint (Kellner et al., 2023). Third, hybrid inference estimators are applied to the footprint level AGBD predictions produce estimates of mean AGBD and uncertainty in the form of a standard error of the mean estimate (Healey et al., 2022). Throughout this paper we refer to these steps collectively as the GEDI's "estimation process", to individual GEDI waveforms as "observations", and to the derived mean estimates of AGBD in step three as "estimates".

Hybrid inference (hereafter "hybrid") is an established method to estimate AGBD and its uncertainty from lidar remote sensing (Healey et al., 2012; Margolis et al., 2015; Nelson et al., 2017). Hybrid is valuable because it provides a formal estimate of the mean estimate's uncertainty, which improves the utility of the resultant map. GEDI's AGBD maps are unique relative to most other global biomass maps (e.g. Saatchi et al., 2011; Avitabile et al., 2016; Santoro and Cartus, 2021) because the hybrid variance estimators account for both sampling and model uncertainty in a transparent manner, compared to ad hoc methods of uncertainty characterization that do not rely on a statistical framework. GEDI's AGBD estimation process was designed to satisfy the assumptions necessary for hybrid inference; 1) the GEDI sample approximates a simple random cluster sample within each spatial estimation unit, in which each laser ground track is a cluster sample, and there are at least two cluster samples in each spatial unit; 2) the L4A models are properly specified and unbiased throughout the entire area in which they are applied; and 3) over large-enough areas the residual error from the L4A models sums to zero (Patterson et al., 2019).

While the first and third assumptions were addressed by Patterson et al. (2019) during GEDI's pre-launch phase, the extent to which the L4A models are unbiased everywhere is not well-established (Dubayah et al., 2022b; Duncanson et al., 2022). This possibility is acknowledged by Duncanson et al. (2022), and represents a substantial challenge in developing globally representative prediction models due to geographic gaps in training data used to calibrate the L4A models that must be transferable to entire continents. Such an assessment requires validation of GEDI's AGBD estimates with independent reference data, however standardized reference data does not exist to validate a global biomass map at 1 km resolution (Duncanson et al., 2021). Instead, McRoberts et al. (2019) outline a strategy for local validation of global biomass maps using design-based estimates of biomass from a national forest inventory at the same spatial scale of the global map. Since the true AGBD for an estimation unit is unknown, the authors suggest validation by comparing population estimates, characterizing the difference between the reference estimate and the remote sensing estimate in a probabilistic manner.

Such a validation of GEDI's AGBD estimates was recently performed. The United States Forest Inventory and Analysis (FIA) program maintains a random sample of national forest inventory plots, from which Menlove and Healey (2020) produced post-stratified design-based reference estimates of AGBD within an equal area hexagon tessellation ($\sim$640 km^2, n = 12,550) covering the continental United States. Generated specifically for validating remote sensing biomass maps, the FIA hexagon reference estimates were used by Dubayah et al. (2022b) to validate an analogous set of GEDI hexagon AGBD estimates. The analysis exhibited strong spatial patterns of both positive and negative systematic differences between the GEDI and FIA reference estimates that were unlikely due to chance. Here, we assume the FIA reference estimates are unbiased, and that systematic differences between the GEDI and FIA estimates are caused by issues biasing GEDI's estimation process. If maps of biomass from remote sensing are to be used more widely, it is important that differences relative to well-designed national forest inventory networks be reconciled.

The goal of this paper is to mitigate the primary issues that adversely affect GEDI's biomass estimation process. We hypothesize these issues to be 1) incomplete filtering of low-quality observations not suited for AGBD estimation, and 2) regional misspecification of the L4A footprint level AGBD prediction models. Specifically, our objectives are:

1. To identify and remove GEDI observations not suited for biomass estimation, that are not caught by GEDI's standard quality filtering.
2. Develop unbiased footprint level AGBD models that may be used within GEDI's hybrid inference framework.
3. Quantify the overall improvement in GEDI AGBD estimation, relative to the FIA estimates, brought by these advancements to GEDI's estimation process.

While our approach is aimed at improving GEDI's AGBD estimation process, the changes we make in the form of improved quality filtering and new footprint level AGBD models only affect the inputs to the population estimation method of hybrid inference, and not the actual hybrid mean and variance estimators put forth by Patterson et al. (2019).

3.3 Materials and Methods

First, we repeated Dubayah et al.'s (2022b) validation of GEDI hexagon estimates to highlight regions of systematic difference relative to the FIA estimates, which informed our hypotheses about the two factors adversely impacting GEDI's original AGBD hexagon estimates; incomplete quality filtering and footprint AGBD model misspecification (fig. 3.1). We designed and applied additional quality filters to produce a set of GEDI hexagon AGBD estimates using hybrid inference that excluded the observations flagged by our additional filters, which we call ‘filtered’ GEDI estimates. Next, we developed new footprint AGBD models using GEDI-FIA fusion in a scale-invariant small area (SISA) estimation framework, which we refer to as the SISA models. We calibrated Fay-Herriot small area estimation models at an aggregated scale, regressing the FIA hexagon estimates of AGBD against hexagon aggregations of GEDI metrics in such a way that the resultant equations were largely unbiased and applicable at the footprint level. We also developed a theoretical justification and validation of these models given the available data. We then applied the SISA models to the GEDI observations resulting in new footprint level AGBD predictions, from which we generated a new set of GEDI hexagon AGBD estimates again using hybrid inference, which we refer to as the ‘fusion’ estimates. We also generated ‘fusion’ estimates at the 1 km scale, based on the flexible nature of GEDI’s hybrid estimation framework. Lastly, we compared the ‘filtered’ and ‘fusion’ GEDI-based estimates with the estimates from Dubayah et al. (2022b) (hereafter the ‘original’ estimates) to quantify the improvements that our filtering and modeling methods had on GEDI’s AGBD estimates, relative to the design-based FIA reference estimates. In the following sections we first outline the GEDI and FIA data sources used in our work, and our method of comparing the GEDI and FIA estimates to identify non-random systematic differences. We then catalog the steps taken to develop, implement, and assess our solutions to both the data filtering and model misspecification issues.

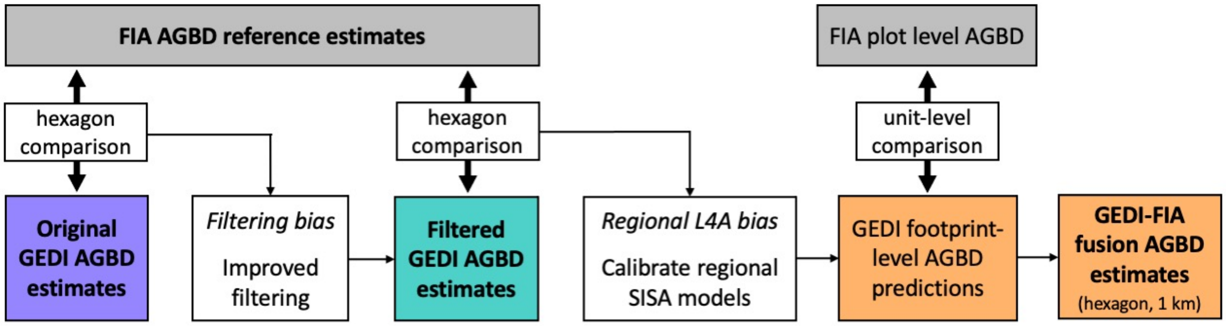

Figure 3.1: Methodological overview. We began by comparing the original GEDI and FIA hexagon estimates, and hypothesized that inclusion of GEDI observations not suited for biomass estimation were causing systematic differences relative to the FIA estimates. We developed methods to identify those observations, removed them from GEDI's biomass estimation process, and generated an updated set of 'filtered' GEDI estimates. A comparison of the filtered GEDI and FIA hexagon estimates revealed a second source of difference between GEDI and FIA estimates, which we hypothesized were due to regional L4A model biases. In response, we developed a scale-invariant, small area (SISA) estimation framework to calibrate new GEDI footprint level AGBD models using GEDI-FIA fusion. When applied to the GEDI observations, the SISA models resulted in mostly unbiased predictions that were more accurate and realistic than the L4A models. We then used the SISA predictions to generate updated 'fusion' GEDI estimates, at both the hexagon and 1 km scale.

3.3.1 GEDI data and algorithms

The GEDI data used were the L2A footprint level waveform metrics (Dubayah et al., 2020b) and the L4A footprint level AGBD predictions (Dubayah et al., 2021a), collected between April 18th 2019 and May 11 2022. We applied the quality filtering criteria put forth in Dubayah et al. (2022b) to ensure we used only the highest quality observations for biomass estimation (see appendix B), and generated the set of 'original' GEDI hexagons estimates using the exact same hybrid inference methods as in Dubayah et al. (2022b). After developing and implementing both our additional filtering criteria and the SISA footprint level AGBD models, we again used hybrid inference to generate the 'filtered' and 'fusion' GEDI hexagon level biomass estimates. This was done using the same code base and hybrid estimators used to produce GEDI's L4B data products (Dubayah et al., 2022a,b), with slight modifications to accommodate our changes to the estimation process. We also used the 1 km GEDI L4B gridded biomass product in comparisons with our 1 km gridded fusion estimates (Dubayah et al., 2022a).

3.3.2 FIA data

The FIA program conducts systematic sampling of forest attributes across a network of field plots evenly distributed throughout the US (Bechtold and Patterson, 2005). The sampling design has approximately 27 evenly distributed plots per hexagon, and unbiased estimates of population parameters can be produced through design-based statistical techniques in the form of a total (i.e. total biomass) or ratio (i.e. total biomass per unit of forested land), along with the associated uncertainty in the form of a percent sampling error. The FIA program estimates forest attributes for forest land only, and assumes forest attributes to be zero on sampled plots that do not meet its definition of forested land—land that is at least 10 percent covered by trees, at least 1 acre (0.405 ha) in size, and at least 120 feet (0.305 meters) wide (Bechtold and Patterson, 2005). Ratio estimates may be adjusted to reflect the entire land area within a spatial estimation unit by changing the denominator to reflect total land area rather than just forested land. In this way, Menlove and Healey (2020) produced estimates of mean AGBD for the total land area within the hexagon grid, and we converted the reported percent sampling errors into standard errors of the mean, according to Bechtold and Patterson (2005). These mean and standard error estimates were the independent reference data we used to validate the GEDI estimates, as well as the response variable that we used in calibrating the SISA models.

3.3.3 Comparison of GEDI and FIA estimates

The comparison of GEDI and FIA hexagon AGBD estimates underpinned our analysis, aiding our diagnosis of the issues potentially biasing GEDI's estimation process, as well as our evaluation of the impact that our solutions had on the updated GEDI estimates. Consider the following difference between two population estimates

$$dB_{j_i} = \hat{\mu}_{j_{FIA}} - \hat{\mu}_{j_{GEDI_i}} \tag{3.1}$$

in which j represents a specific hexagon, $\hat{\mu}_{FIA}$ is the FIA mean AGBD hexagon estimate derived from the FIA's design-based statistical estimators, (Bechtold and Patterson, 2005; Pugh et al., 2018; Menlove and Healey, 2020), $\hat{\mu}_{GEDI_i}$ is the GEDI mean AGBD estimate (eq. 4 from Patterson et al., 2019), and i denotes the version of GEDI estimate being evaluated (original, filtered, or fusion). To characterize the difference between the FIA and GEDI mean estimates we calculated the following test statistic from McRoberts et al. (2019),

$$t_{j_i} = \frac{dB_{j_i}}{\sqrt{M\hat{S}E(\hat{\mu}_{j_{FIA}}) + M\hat{S}E(\hat{\mu}_{j_{GEDI}})}} \tag{3.2}$$

in which $M\hat{S}E(\hat{\mu}_{j_{FIA}})$ and $M\hat{S}E(\hat{\mu}_{jGEDI_i})$ are the respective mean squared errors associated with $\hat{\mu}_{j_{FIA}}$ and $\hat{\mu}_{j_{GEDI_i}}$. Instead of formal hypothesis testing to determine significance based on an arbitrary confidence level, we take t_{j_i} as a heuristic to assess the difference between the GEDI and FIA mean estimates relative to the respective estimate uncertainties (Dubayah et al., 2022b). Our goal was to focus on regions with non-random systematic differences in the estimates, relative to the magnitude and precision of those estimates, as a way to identify potential issues somewhere in GEDI's estimation process.

The GEDI-FIA comparison analysis was performed sequentially on the three sets of GEDI-derived mean AGBD estimates ($j =$ original, filtered, fusion) outlined in section 3.3.1.

3.3.4 Addressing deficiencies in GEDI observation filtering

GEDI observations in which part of the laser pulse is reflected from something other than flat bare ground or vegetation may result in biased predictions of footprint level AGBD, as these conditions are not represented in the data used to train GEDI's L4A footprint AGBD models. Examples include waveforms that intersect buildings, low clouds, steep slopes, rough terrain or other topographic features, both with and without vegetation. The presence of a steep slope (vegetated or not) or non-vegetated object within the waveform footprint alters the relative height metrics and may also cause ground finding errors. If many such observations are used in hybrid estimation, the resultant estimates may differ substantially from unbiased independent reference data. While there are quality flags built into both the L2A and L4A algorithms that trigger when it is obvious that a waveform does not represent the ground surface conditions (such as a cloud high above the land surface), the GEDI algorithms cannot differentiate between waveforms from forest canopies and those that contain a building, low cloud, or non-vegetated topographic feature such as a rock outcropping, canyon wall, or steep slopes.

Ancillary information is necessary to identify such observations and remove them from the set used in hybrid AGBD estimation. For the case of buildings, GEDI's level 4B (L4B) algorithm uses a custom urban mask to identify observations that are likely to have intersected a human-built structure (Healey et al., 2022). The L4B algorithm also identifies likely cloud affected observations—those with a maximum height larger than 150 meters, and also segments of GEDI orbits where the deviations between GEDI canopy top and TanDEM-X DEM elevations are systematically larger than those from other nearby GEDI orbits—although there are instances in which localized improvements to the cloud filtering process can be made, such as isolated cloud affected observations. However, instances of topography affected waveforms are not identified, in part because removal of all such GEDI observations may violate the hybrid assumption that the GEDI sample approximates a random cluster sample within each spatial estimation unit. These waveforms are not suited for biomass estimation because a steep slope or other three dimensional topographic features can result in a waveform with large relative height metrics that look like trees, even when there is no vegetation within the footprint. GEDI has not yet implemented a global method to identify and remove these observations, and instead assumes that the L4A predictions for these observations are suitable for inclusion in hybrid estimation in current data products.

To identify GEDI observations impacted by low clouds and fog, we determined the maximum observed tree height within each hexagon from the FIA tree-level attribute tables. Since the FIA samples a small fraction of the forested area within each hexagon, we assumed the true maximum to be substantially larger than the observed maximum. We then multiplied the observed height by an expansion factor to decrease the likelihood of removing GEDI observations from trees taller than the observed maximum height. We applied different height expansion factors and visually compared the GEDI observations with larger maximum heights to their neighboring observations, as well as satellite imagery depicting the land cover. We arrived at a final expansion factor of 1.75 to remove as many cloud affected observations as possible while maintaining a low probability of removing valid forested observations.

Our second filtering procedure identifies a small set of GEDI observations that are highly impacted by steep slopes and rough terrain, the details of which are provided in the appendix B. We focus on only the most impacted observations because removing every observation in which the waveform relative height metrics are impacted by topography would violate the assumptions of hybrid inference. Slope and topography can impact waveforms in a variety of ways (Yang et al., 2011; Chen et al., 2014; Park et al., 2014), and here we are focused on removing those with the most inflated relative height metrics, as they are the easiest to identify and have the largest impact on AGBD estimates. To do so, we compute the 99th percentile of canopy heights from GEDI waveforms returned from flat ground within five predefined ranges of canopy cover (table B.1). If a GEDI observation from slopped terrain had a canopy height larger than the 99th percentile of heights from the flat ground observations in its same range of canopy cover, we deemed it as too impacted by topography for the L4A model to be applicable and disqualified it for use within hybrid estimation.

3.3.5 Addressing L4A model bias

In this section we present our methods for developing unbiased footprint level AGBD models. First, we discuss why GEDI's L4A models may be biased in some regions of the US. Second, we explain how we addressed this issue using a scale-invariant small area estimation framework to train new footprint level AGBD models, and the assumptions required to do so. Third, we demonstrate how the models were scale-invariant and produced relatively unbiased predictions at the footprint level.

3.3.5.1 Potential L4A model misspecification

The L4A models were calibrated using one of the most extensive databases of forest inventory plots (“calibration plots”) coincident with airborne laser scanning (ALS) retrievals yet compiled (Duncanson et al., 2022). A GEDI waveform was simulated from ALS data for each plot, and the inventory-based AGBD values were then related to the simulated waveform metrics to calibrate footprint-level AGBD prediction models. Operationally, at least 50 plots were required to train a model, and there were not enough plots for localized partitioning below the continental scale (Kellner et al., 2023). Instead, a combination of continental region and MODIS plant functional type (PFT) (Friedl et al., 2019) was used to delineate large regions that approximate biomes, and a unique L4A model was calibrated for each region (henceforth called GEDI “prediction strata”) using the calibration plots located within. In North America there are three L4A models and associated prediction strata; deciduous broadleaf (DBT, n=873 calibration plots), needleleaf (NT, n=1391 calibration plots), and grasslands, shrublands, and woodlands (GSW, n=89 calibration plots). Evidence suggests that relationships between lidar-derived forest structure metrics (from both airborne lidar and GEDI) and AGBD vary spatially within PFT, as models that use these metrics to predict AGBD over large areas explain more overall variation in AGBD when a spatial component is included, compared to similar aspatial models (Babcock et al., 2015; May et al., 2023). However, the L4A models are not sensitive to within-strata variations in the structure-biomass relationship, which may result in locally varying L4A model bias.

3.3.5.2 SISA model development

To ensure a footprint level AGBD model that is unbiased for its entire prediction stratum requires 1) calibration data that represent local conditions equally throughout the stratum, and 2) an ability to reflect the continuous spatial processes that result in spatial variations in the structure-biomass relationship. An ideal model would be calibrated on representative data evenly distributed throughout the prediction stratum, and would incorporate a spatially varying component (Babcock et al., 2015, 2016, 2018; Taylor-Rodriguez et al., 2019). However, such methods require near-continuous, spatially representative calibration data to adequately capture the spatial processes impacting the relationship between physically-based predictors and the biophysical response, which is not available in this case.

Accordingly, we developed a modeling framework that satisfied the first of these requirements completely, but the second requirement only partially. Instead of a spatially varying model, we developed various regional models for different areas in an attempt to isolate local relationships between the predictor and response variables and capture spatial impacts on the relationship between forest structure and biomass. While methodologically different than spatially varying models, this regional stratification was a pragmatic solution that also allowed our models to integrate with GEDI's hybrid estimation framework. To ensure spatially balanced and representative calibration data, we used the FIA-based hexagon AGBD estimates of Menlove and Healey (2020) as our response variable, and hexagon level averages of various GEDI waveform-derived canopy height metrics as the predictor variables (see appendix B). The hexagons are of equal area and cover the entire continental US, so every part of the prediction region was equally represented in the calibration data.

There were two motivations for training models at the hexagon scale using averaged predictor variables, instead of at the native resolution of the GEDI footprint. First, exact overlays of a GEDI observation and an FIA plot were not possible because the size and spatial configuration of FIA plots does not align with the GEDI footprint diameter, and training models on misaligned data 1) introduces additional, unwanted uncertainty and 2) decreases the signal captured in the training data, and is discouraged (Duncanson et al., 2021). Furthermore, the combined geolocation uncertainties of GEDI observations (approximately 10 horizontal meters) and FIA plots (approximately 10 horizontal meters) (Hoppus and Lister, 2007) would substantially increase the variability in actual overlap between GEDI shots and FIA plots, thus adding more noise to the relationship of interest.

The second motivation for training on aggregated data is the inherent scalability of linear relationships. A theoretical linear relationship at the level of a single GEDI waveform takes the form

$$y_{ij} = \mu + x_{ij}\beta + \epsilon_{ij} \tag{3.3}$$

in which y_{ij} represents the true AGBD within the waveform footprint; x_{ij} represents a GEDI waveform derived height metric or a linear transformation thereof; index $i = 1...n$ and represents a continuous partitioning of the hexagon into discrete GEDI footprints; $j = 1...N$ and represents all the hexagons within a given region of the US for which we assume there is a single, constant linear relationship between y and x. Parameters μ and β are true regression parameters and ϵ_{ij} is a randomly distributed error term for which we assume an expected value of zero and a constant variance. Averaging y_{ij} across all footprints within hexagon j yields

$$\frac{1}{n}\sum_{i=1}^{n} y_{ij} = \mu + \left(\frac{1}{n}\sum_{i=1}^{n} x_{ij}\right)\beta + \frac{1}{n}\sum_{i=1}^{n} \epsilon_{ij}, \tag{3.4}$$

which can be simplified to

$$\overline{y}_j = \mu + \overline{x}_j\beta + \overline{\epsilon}_j. \tag{3.5}$$

Therefore, assuming a linear regression at the footprint level directly implies a linear regression at the hexagon (or any area aggregate) level with the same regression parameters. Importantly, x_{ij} in eq. (3.3) could be a non-linear transformation of a GEDI metric, accounting for a non-linear relationship, but enabling a model that is linear with respect to parameters β and ϵ_{ij}. In this case, $\overline{x}_j$ in eq. (3.5) would represent the same non-linear transformation, applied at the unit level before averaging to calculate $\overline{x}_j$. If the true values of $\overline{y}_j$ and $\overline{x}_j$ were known, we could fit a linear model accordingly. However, we do not know the true values, and in turn must rely on estimates. Menlove and Healey (2020) provide estimates of both AGBD and the associated sampling error, from which we can estimate $\overline{y}_j$:

$$\hat{y}_j = \overline{y}_j + \delta_j \tag{3.6}$$

in which $\hat{y}_j$ is the estimate of AGBD within the hexagon, and δ_j is the associated sampling error. Similarly, we can use the GEDI sample within a hexagon to estimate $\overline{x}_j$, as follows:

$$\hat{x}_j = \overline{x}_j + \zeta_j \tag{3.7}$$

in which $\hat{x}_j$ is the average value of x_{ij} across all GEDI shots within the hexagon and ζ_j is the GEDI sampling error. If the variance of ζ_j is non-trivial in magnitude, it could be accounted for with a measurement error model (Fuller, 2009). However, we assume the GEDI sample is large enough within each hexagon (tens of thousands of observations per intact hexagon) for ζ_j to be negligible and thus eq. (3.7) becomes

$$\hat{x}_j \approx \overline{x}_j \tag{3.8}$$

Rearranging eq. (3.6), we can rewrite our aggregate model from eq. (3.5) as

$$\hat{y}_j = \mu + \hat{x}_j \beta + (\overline{\epsilon}_j + \delta_j) \tag{3.9}$$

This is known as the Fay-Herriot (FH) model, which yields reliable inference on β by accounting for the sampling error δ_j associated with $\hat{y}_j$ (Fay and Herriot, 1979). The FH model belongs to the family of small area estimation models used to increase estimation precision, relative to design-based estimates, especially for small sample sizes (Rao and Molina, 2015).

For implementation within GEDI's hybrid estimation framework, the required unknown quantities are estimates of the regression parameters μ, β, and the associated variance of these estimates. Let $\boldsymbol{b} = [\mu \ \beta]^T$ be a vector of the regression parameters. Let $\bar{\boldsymbol{X}}$ be a $N \times p$ matrix of the N hexagon aggregates of the p different GEDI predictors, let $\boldsymbol{Z} = [\boldsymbol{1} \ \bar{\boldsymbol{X}}]$, appending a column of ones to $\bar{\boldsymbol{X}}$, and let $\hat{\boldsymbol{y}}$ be the vector of N direct estimates. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_\epsilon + \boldsymbol{D}$, where $\boldsymbol{\Sigma}_\epsilon$ is a $N \times N$ covariance matrix such that $[\boldsymbol{\Sigma}_\epsilon]_{jk} = \text{Cov}[\overline{\epsilon}_j, \overline{\epsilon}_k]$, and $\boldsymbol{D}$ is a diagonal matrix of the sampling variances, $[\boldsymbol{D}]_{jj} = \text{Var}[\delta_j]$. The FH estimate of $\boldsymbol{b}$ is

$$\hat{\boldsymbol{b}} = \left(\boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{y}}. \tag{3.10}$$

The FH estimate, $\hat{\boldsymbol{b}}$, is the best linear unbiased estimate (BLUE) for $\boldsymbol{b}$ by the Gauss-Markov theorem, meaning out of all linear unbiased estimates, $\hat{\boldsymbol{b}}$ has minimum variance:

$$\text{Var}[\hat{\boldsymbol{b}}] = \left(\boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\right)^{-1}. \tag{3.11}$$

However, in practice, matrix $\boldsymbol{\Sigma}_\epsilon$ is unknown and must be estimated from the data, call this $\hat{\boldsymbol{\Sigma}}_\epsilon$. Substituting this estimate for $\boldsymbol{\Sigma}_\epsilon$ yields the empirical best linear unbiased estimate (EBLUE) for $\boldsymbol{b}$. A common assumption is $\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2\boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix, implying identical and independently distributed errors $\bar{\epsilon}_j$. This ignores potential spatial correlation in $\bar{\epsilon}_j$, which may lead to over-confident estimates of $\boldsymbol{b}$, i.e. estimates of $\text{Var}[\hat{\boldsymbol{b}}]$ that are too small. We instead use a simultaneous autoregressive (SAR) model for $\bar{\epsilon}_j$ (Pratesi and Salvati, 2008) to account for dependence between nearby hexagons, so that

$$\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2\left[(\boldsymbol{I} - \rho\boldsymbol{W})(\boldsymbol{I} - \rho\boldsymbol{W}^T)\right]^{-1}, \tag{3.12}$$

where $\boldsymbol{W}$ is a proximity matrix such that $[\boldsymbol{W}]_{jk} = 1$ if hexagons j and k are neighbors and $[\boldsymbol{W}]_{jk} = 0$ otherwise, $\rho \in (-1, 1)$ is a correlation parameter, and σ_ϵ^2 is the variance parameter. The unknown parameters are σ_ϵ^2, ρ, so that

$$\hat{\boldsymbol{\Sigma}}_\epsilon = \hat{\sigma}_\epsilon^2\left[(\boldsymbol{I} - \hat{\rho}\boldsymbol{W})(\boldsymbol{I} - \hat{\rho}\boldsymbol{W}^T)\right]^{-1}. \tag{3.13}$$

We use 'R' package 'sae' (Molina and Marhuenda, 2015), which uses restricted maximum likelihood to estimate σ_ϵ^2, ρ in order to compute EBLUE $\hat{\boldsymbol{b}}$ and its variance.

Given the assumptions that the relationship between x_{ij} and y_{ij} is consistent and linear throughout all N hexagons, the above is justification that β is the same at both the unit scale of GEDI footprints (eq. (3.3)) and aggregate scale of hexagons (eq. (3.9)). If these assumptions are satisfied the resultant model form in eq. (3.9) is applicable at the unit level (GEDI footprint). This is what we refer to as the scale-invariant small area AGBD model. In this context, 'scale-invariant' refers to the difference in scale between SISA model calibration (aggregation at the hexagon scale) and prediction (unit level of GEDI footprint). This is different from other scale-invariant remote sensing analyses in which larger scale variables are not aggregations of unit level variables.

3.3.5.3 SISA model calibration

In this section we summarize the SISA model calibration methodology, a complete description of which can be found in appendix B. First we delineated ten SISA calibration and prediction strata at the hexagon level, based on the predominant forest type and climate (fig. 3.2). Next, we used the GEDI calibration plots to determine which GEDI waveform variables (combinations of waveform relative height metrics with and without transformations) had a linear relationship with AGBD. We then calculated hexagon level averages of the variables that were linear with respect to AGBD. We fit candidate SISA models (66 for the DBT strata and 91 for the MIX and NT strata) within each of the ten prediction strata using the hexagon averaged GEDI variables as predictors and the post-stratified FIA hexagon AGBD estimates from Menlove and Healey

(2020) as the response. From the candidate models tested for each strata, we selected six models with the lowest training RSE for assessment and validation at the unit level. We applied the final six candidate models for each strata to all GEDI observations within the strata, and selected each stratum's final model (table B.2) based on how well the distribution of footprint-level predicted AGBD values matched the distribution of AGBD from the FIA plots. Each region's final SISA model was the one that produced the most similar unit level distribution of biomass when compared to that regions' distribution of FIA plot-level biomass, based on quantile-quantile plots, side by side comparisons, and a Kolmogorov-Smirnov test. The predictions and parameter estimate covariance matrices from the final SISA models were then used to generate hexagon level AGBD estimates using GEDI's hybrid estimation framework. We refer to these estimates as the fusion estimates.

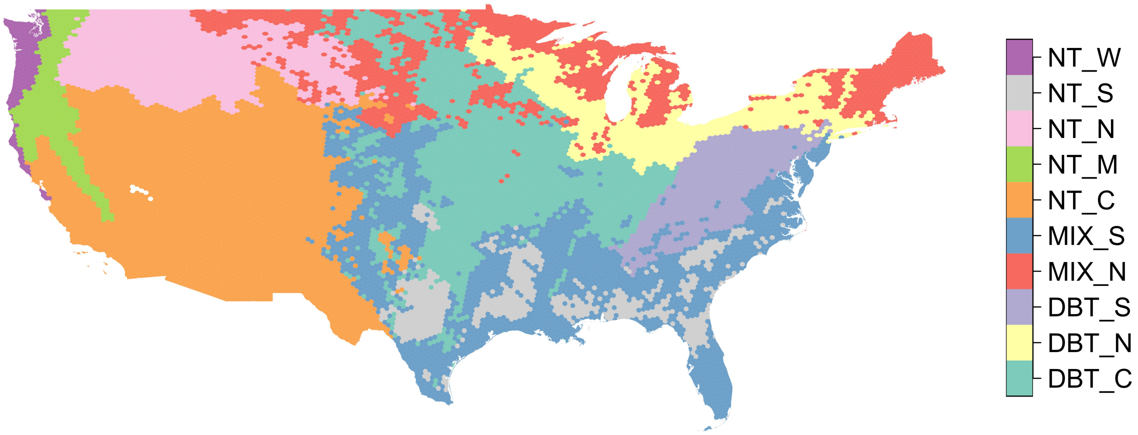

Figure 3.2: Map of the ten forest strata regions used in this analysis. For each hexagon we calculated the most abundant forest PFT according to the NLCD 2019 classification map (DBT, NT, MIX), and then we further segmented these PFTs into regional strata based primarily on the level two EPA ecoregion classification map.

3.3.5.4 SISA model validation

Here we explain our approach to validate the SISA models, in which validation primarily hinges on whether or not the models produced realistic footprint level distributions of AGBD. Implicit in the usage of the models at the unit level for AGBD prediction is the assumption that the models are spatially invariant—the models are unbiased at the unit level and realistically predict AGBD despite being trained at an aggregate scale. Our validation approach assesses the extent to which this assumption of scale invariance is met based on the data available. We first examined the applicability of the β parameter estimates at the unit level by comparing distributions of SISA predicted AGBD with the distributions of FIA plot AGBD. Second, we assessed the internal consistency of the β parameter estimates and their uncertainty between the hexagon (aggregate) and footprint (unit) levels. If 1) the unit level distribution of SISA AGBD predictions matched the unit level distribution of FIA plot AGBD, and 2) $\hat{\beta}$ appeared consistent across different spatial scales of aggregation, we determined there was not sufficient evidence to invalidate the assumption of spatial invariance, thus rendering unit level AGBD prediction via the SISA models reliable for hybrid estimation of AGBD.

The final SISA model for each forest stratum was selected to maximize the similarity in unit level AGBD distributions between the FIA plots and SISA predictions. We assumed the ability of a given SISA model to reproduce its stratum's unit level distribution of FIA plot biomass was a reliable indicator of model performance at the unit level. Therefore, if a stratum's distribution of SISA predictions was similar to the distribution of FIA plot level AGBD, we determined the corresponding SISA model was unbiased and the predictions were realistic. We also included the L4A predictions in these distribution comparisons, allowing us to characterize each SISA model's performance, relative to that of L4A, within its stratum. For example, if the SISA distribution had obvious differences relative to the FIA distribution, but less so than the L4A distribution, we deemed the SISA model as preferable to the L4A model. We also applied the appropriate SISA model to each GEDI calibration plot based on its location, to compare model performance (L4A and SISA) directly, relative to the field estimate of AGBD (figs. B.7 and B.8).

To assess the internal consistency of the β parameter estimates, and by extension whether the SISA predictions were reliable at the unit level, we asked the following question: If we recalibrated the SISA models at a different spatial scale of aggregation, would the various $\hat{\beta}$ confidence intervals overlap? In other words, does the spatial scale of calibration impact the statistical consistency of $\hat{\beta}$? To answer this question, we recalibrated each stratum's final SISA model (table B.2) at three different hexagon grid resolutions, with approximate areas of 12,400 km^2, 1770 km^2, and 250 km^2 (Uber, 2018). This required generating versions of our SISA prediction strata and the response and predictor variables within each new hexagon grid resolution, as follows. We mapped the prediction strata into each new hexagon grid based on which strata was most abundant within each hexagon in the new grids. For the predictor variables we aggregated the GEDI metrics within each new hexagon grid as we did for the original FIA hexagon grid. For the response variable, we recalculated the FIA-based estimates within each new grid using the Horvitz-Thompson estimator, which produced similar estimates to the post-stratified estimates of Menlove and Healey (2020), and were easier to implement and reproduce (May et al., 2023). We then refit the final SISA models for each new resolution, the result of which was a new β parameter estimate associated with the model fit for each stratum and resolution combination. The statistical consistency of each stratum's SISA model form was assessed by comparing the resultant β parameter estimates and the associated 95% confidence intervals across the different spatial scales of model calibration.

3.4 Results

In the following section we present the collective SISA model performance, quantify the impact of our filtering and SISA modeling procedures on GEDI's AGBD estimates, and compare each set of GEDI AGBD hexagon estimates (original, filtered, and fusion) to the FIA AGBD hexagon estimates. Additional figures can be found in appendix B.

3.4.1 SISA model validation

The SISA models produced distributions of footprint level AGBD that were similar to the FIA plot level distributions for all ten forest strata, but to varying degrees (fig. 3.3). Relative to the L4A unit level distributions, the SISA unit level distributions were equally or more similar to the FIA unit level distributions in all regions, based on a side-by-side visual comparison and quantile-quantile plots. When applied to GEDI calibration plots, the L4A models explained a larger fraction of the overall variance in the field-based estimate of AGBD and had a lower root mean squared error (RMSE) than the SISA models (fig. B.7). This is expected, as the L4A models were fit on these data using a least squares approach which mathematically guarantees a lower squared error compared to the SISA models that were not calibrated on these data. However, the SISA models appeared less biased than the L4A models for low (<60 Mg ha^{-1}) and high (>500 Mg ha^{-1}) field-based AGBD. The SISA models were biased high in the 150 – 250 Mg ha^{-1}range of field-based AGBD.

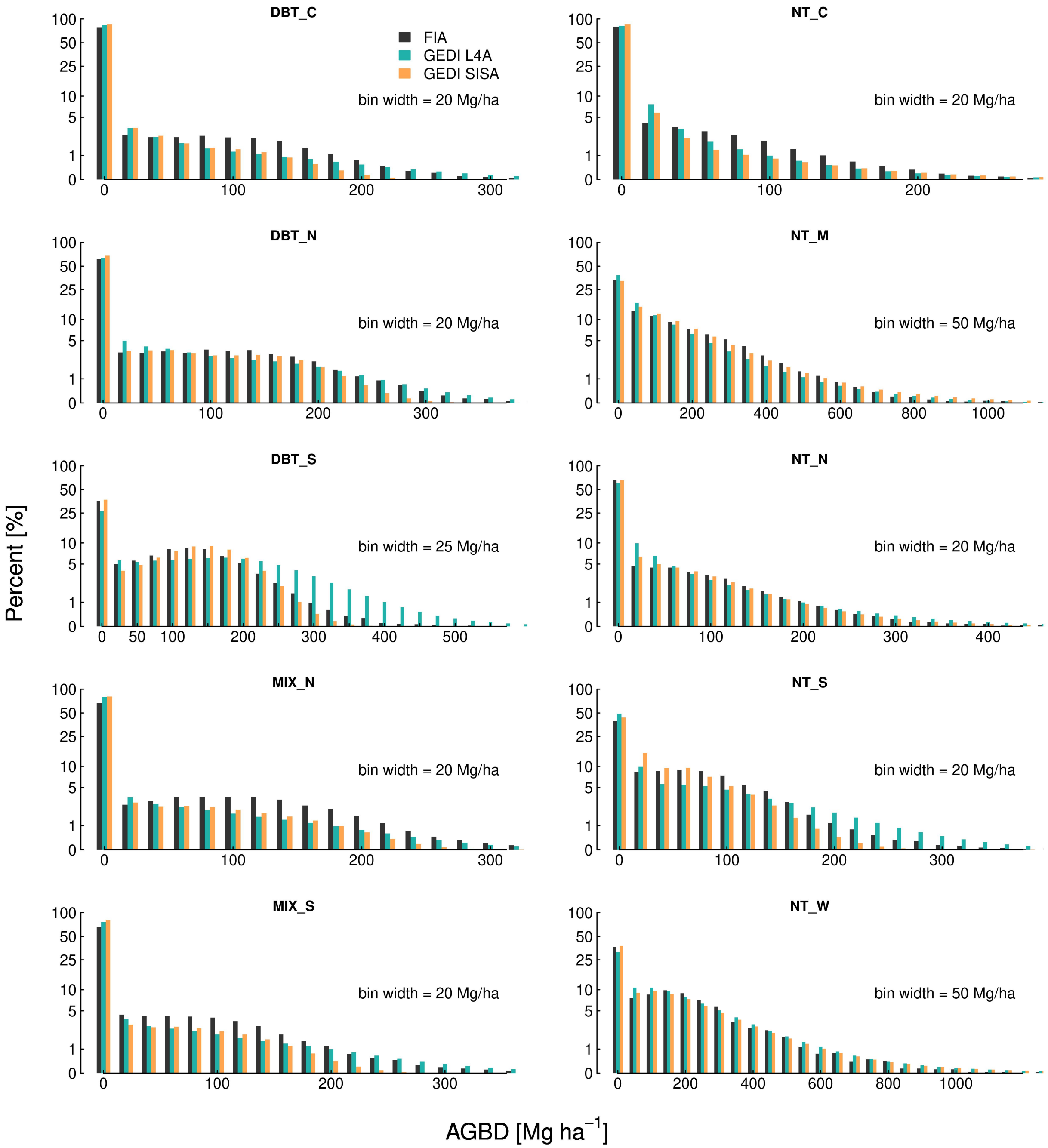

Figure 3.3: Unit level AGBD histograms demonstrate the SISA models generally produce realistic distributions that align with the FIA distributions, more so than the L4A distributions.

As a whole, there was moderate overlap in the SISA β parameter estimate confidence intervals across spatial scales, with some strata displaying a higher degree of consistency than others (fig. 3.4). The DBT_S, DBT_N, NT_N, NT_W strata exhibited relative consistency across spatial scales; these strata were more homogeneous and spatially compact relative to the other strata, individually had above average proportions of forest (table 3.1), and collectively were 44.4% forested. Conversely, the parameter estimates for the DBT_C, MIX_N, NT_C strata varied across spatial scales (fig. 3.4), and in some instances (e.g. NT_C) the variation was large relative to the point estimates. Relative to the four strata with more stable parameter estimates, these three strata were more disjointed and ecotonal, and collectively were only 25.5% forested. The MIX_S, NT_S, NT_W strata exhibited mixed consistency in β parameter estimates across scales; these strata had characteristics in between those with consistent and inconsistent β parameter estimates and collectively were 42.8% forested.

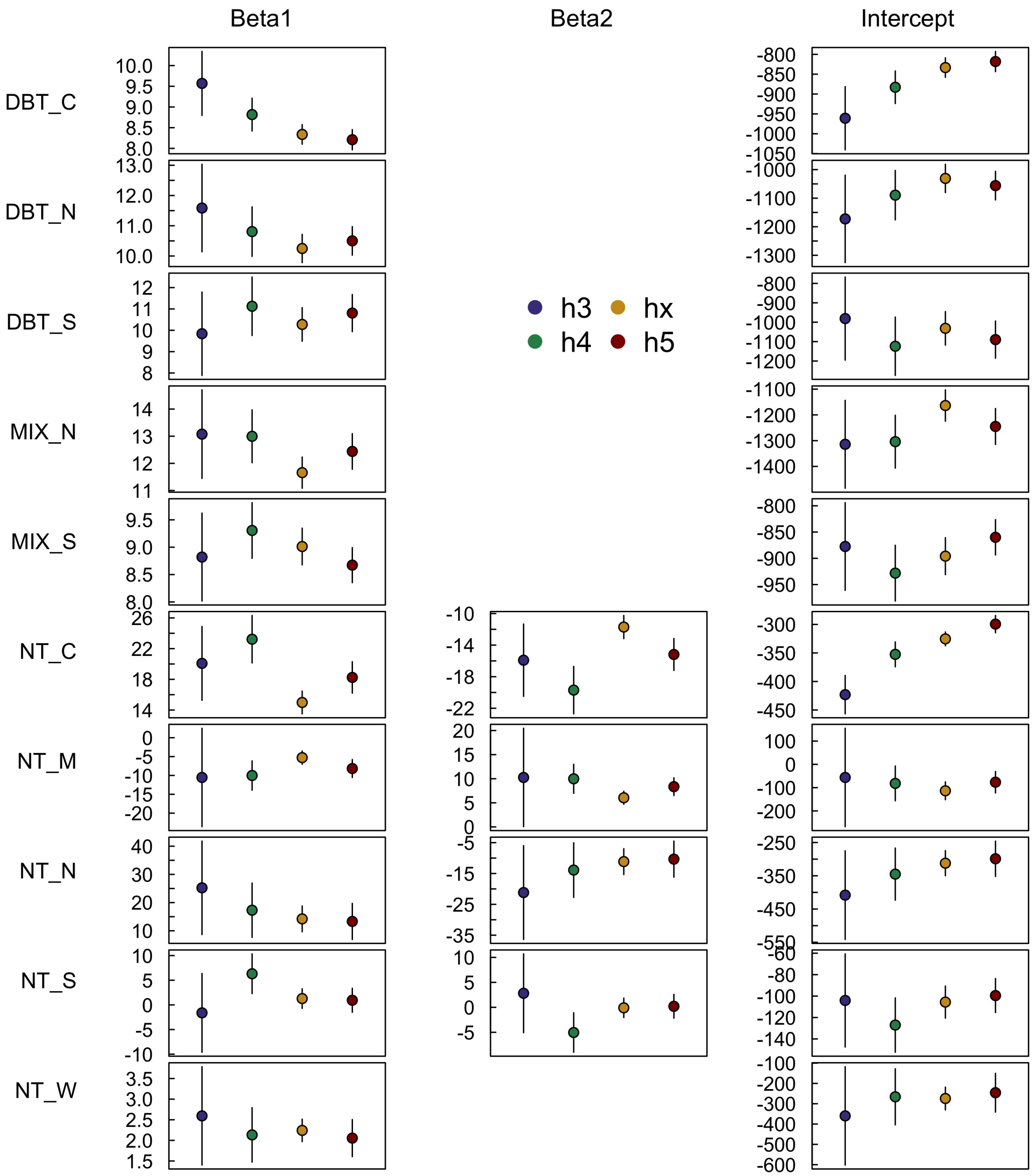

Figure 3.4: SISA model β parameter estimates consistent to varying degrees when calibrated at four different spatial resolutions. The h3 (approximately 12,393 km^2), h4 (approximately 1,770 km^2), and h5 (approximately 253 km^2) resolutions are different hexagon tesselations from the H3 hierarchical spatial indexing system, and the hx resolution (approximately 640 km) is the FIA's hexagon grid used throughout this analysis. Broadly, the more homogeneous strata with high proportions of forest cover exhibited a greater degree of stability in the parameter estimates across spatial scales than strata with lower forested proportions and more forest type heterogeneity.

3.4.2 Changes to the GEDI hexagon biomass estimates

The additional filtering criteria that we implemented had substantial impacts on GEDI's biomass estimates in certain areas, and no impact in others (fig. B.10A). The spatial pattern of differences between the original and filtered GEDI estimates shows that filtering was most helpful in the western third of the country, which is almost entirely due to the removal of topography impacted observations. The impact of cloud filtering was not nearly as concentrated nor as large, and was mostly randomly distributed throughout eastern deciduous forests. Cloud filtering resulted in the removal of 618,615 GEDI observations, or 0.11% of all observations that made it through GEDI's original quality filters.

The original GEDI estimates contained an average of 43,516 observations per intact hexagon (excludes coastal and partial hexagons with a land area less than 600 km^2, N=11,631). Our filtering procedures removed an average of 59 observations per hexagon, resulting in an average decrease of 1.91 Mg ha^{-1}between the original and filtered GEDI estimates. However, of the 12,550 total hexagons, our filtering techniques did not remove any observations in 3752 hexagons. Further, 14.1% of hexagons had more than 100 observations removed (resulting in a mean decrease of 8.8 Mg ha^{-1}between the original and filtered estimates for these 1,768 hexagons), and 0.58% of hexagons had more than 1,000 shots removed (resulting in a mean decrease of 21.94 Mg ha^{-1}between the original and filtered estimates for these 73 hexagons).

The SISA models resulted in an average decrease of 9.85 Mg ha^{-1}between the filtered and fusion estimates (table 3.1), although the magnitude and direction of change varied considerably by region (fig. B.10B). The largest absolute mean change at the region level occurred in the Appalachian Mountians (DBT_S), where the SISA models resulted in an average decrease in

Table 3.1: FIA and GEDI estimates and GEDI change statistics by forest stratum and for all the continental US (COUNS). Units are Mg ha^{-1}for all columns except N (counts) and t which is unitless. All reported values are the mean of each metric, expect the t-values, for which the median is reported.

Region	Area %		FIA		GEDI original (v0)			GEDI filtered (v1)			GEDI-FIA fusion (v2)			Change in GEDI estimates		
	CONUS	FIA	% forest	mean	mean	dB	t	mean	dB	t	mean	dB	t	v0 - v1	v1 - v2	v0 - v2
DBT_C	16.43	78.13	21.06	25.01	35.54	-10.53	-1.11	35.20	-10.19	-1.07	22.15	2.86	0.37	-0.34	-13.05	-13.39
DBT_N	6.94	96.27	36.46	57.02	72.69	-15.67	-1.32	72.23	-15.21	-1.29	55.89	1.13	0.05	-0.46	-16.34	-16.80
DBT_S	5.32	100.00	61.08	100.73	153.36	-52.63	-3.06	151.69	-50.96	-2.98	98.01	2.72	0.06	-1.67	-53.68	-55.35
MIX_N	11.16	63.62	33.08	44.60	49.77	-5.17	-0.87	49.27	-4.67	-0.83	45.01	-0.41	-0.16	-0.50	-4.25	-4.76
MIX_S	17.12	80.26	37.48	36.34	51.48	-15.14	-1.61	51.24	-14.90	-1.58	34.65	1.69	0.13	-0.24	-16.59	-16.82
NT_C	24.60	70.00	24.31	14.50	17.31	-2.80	-0.49	12.35	2.15	0.36	13.71	0.80	0.09	-4.95	1.35	-3.60
NT_M	3.02	100.00	78.44	171.35	141.45	29.90	1.00	138.32	33.03	1.08	164.76	6.58	0.04	-3.13	26.45	23.32
NT_N	9.12	79.34	39.06	39.89	45.37	-5.48	-0.75	41.33	-1.44	-0.40	37.14	2.75	0.05	-4.04	-4.20	-8.23
NT_S	5.14	99.58	66.94	55.62	72.85	-17.23	-1.59	72.39	-16.77	-1.59	53.68	1.93	0.12	-0.46	-18.70	-19.16
NT_W	1.15	99.79	63.52	148.72	187.35	-38.63	-1.08	182.07	-33.35	-1.07	165.94	-17.22	-0.49	-5.28	-16.12	-21.40
CONUS	100.00	79.42	35.43	41.92	52.40	-10.48	-1.13	50.38	-8.46	-0.87	40.53	1.39	0.11	-2.02	-9.85	-11.87

the mean estimate of 53.68 Mg ha^{-1}per hexagon. The other region with a notably large change included the Cascade and Sierra Mountain ranges (NT_M), with an average increase of 26.45 Mg ha^{-1}per hexagon.

3.4.3 GEDI and FIA hexagon estimate comparisons

There were widespread and systematic differences between the original GEDI and FIA estimates (fig. 3.5). The GEDI map overestimated relative to FIA throughout eastern deciduous forests, and underestimated in the PFT-mixed or predominantly conifer forests in Maine, southwest, and Sierra and Cascade mountain ranges. The mean difference (FIA - GEDI) across all hexagons was -10.48 Mg ha^{-1}(table 3.1), and the RMSD was 27.99 Mg ha^{-1}. The median of test statistic $t_{original}$ from eq. (3.2) was -1.13, with first and third quartiles of -2.81 and 0.24.

The spatial pattern of differences between the filtered GEDI and FIA estimates was mostly similar to that of the differences between the original GEDI and FIA estimates (fig. 3.6), with clustered areas of improvement in the west. The mean difference improved by 2.02 Mg ha^{-1}to -8.46 Mg ha^{-1}(table 3.1), and the RMSD improved slightly to 27.04 Mg ha^{-1}. The median of test statistic $t_{filtered}$ from eq. (3.2) was -0.87, with first and third quartiles of -2.47 and 0.48.

The SISA models resulted in considerable improvement in the fusion estimates relative to the FIA estimates (table 3.1). The spatial pattern of the FIA-GEDI differences changed substantially and the general pattern of overestimation in eastern deciduous forests and underestimation of western conifer forests was eliminated (fig. 3.7). The mean difference across all hexagons was 1.39 Mg ha^{-1}, and the RMSD was 18.10 Mg ha^{-1}. The median of test statistic t_{fusion} from eq. (3.2) was 0.11, with first and third quartiles of -0.88 and 0.98.

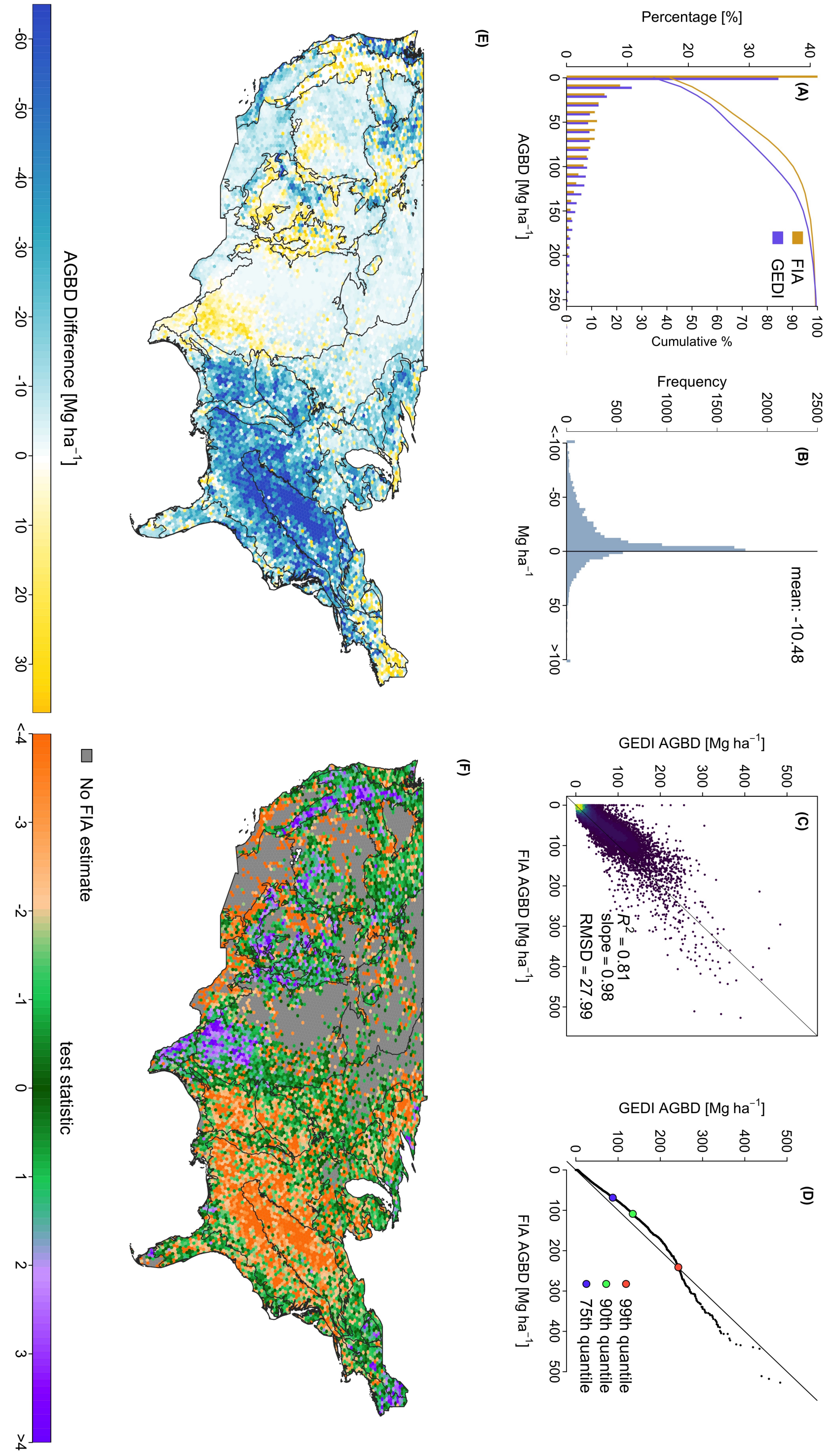

Figure 3.5: Comparison of the original GEDI and FIA estimates. a) Side-by-side histogram comparison truncated to 300 Mg ha^{-1}, b) Difference histogram (FIA-GEDI), c) Scatter plot of FIA vs GEDI estimates, d) Quantile-quantile plot of hexagon estimates, and maps of e) estimate differences (FIA-GEDI) and f) $t_{original}$, with EPA level II ecoregions overlaid.

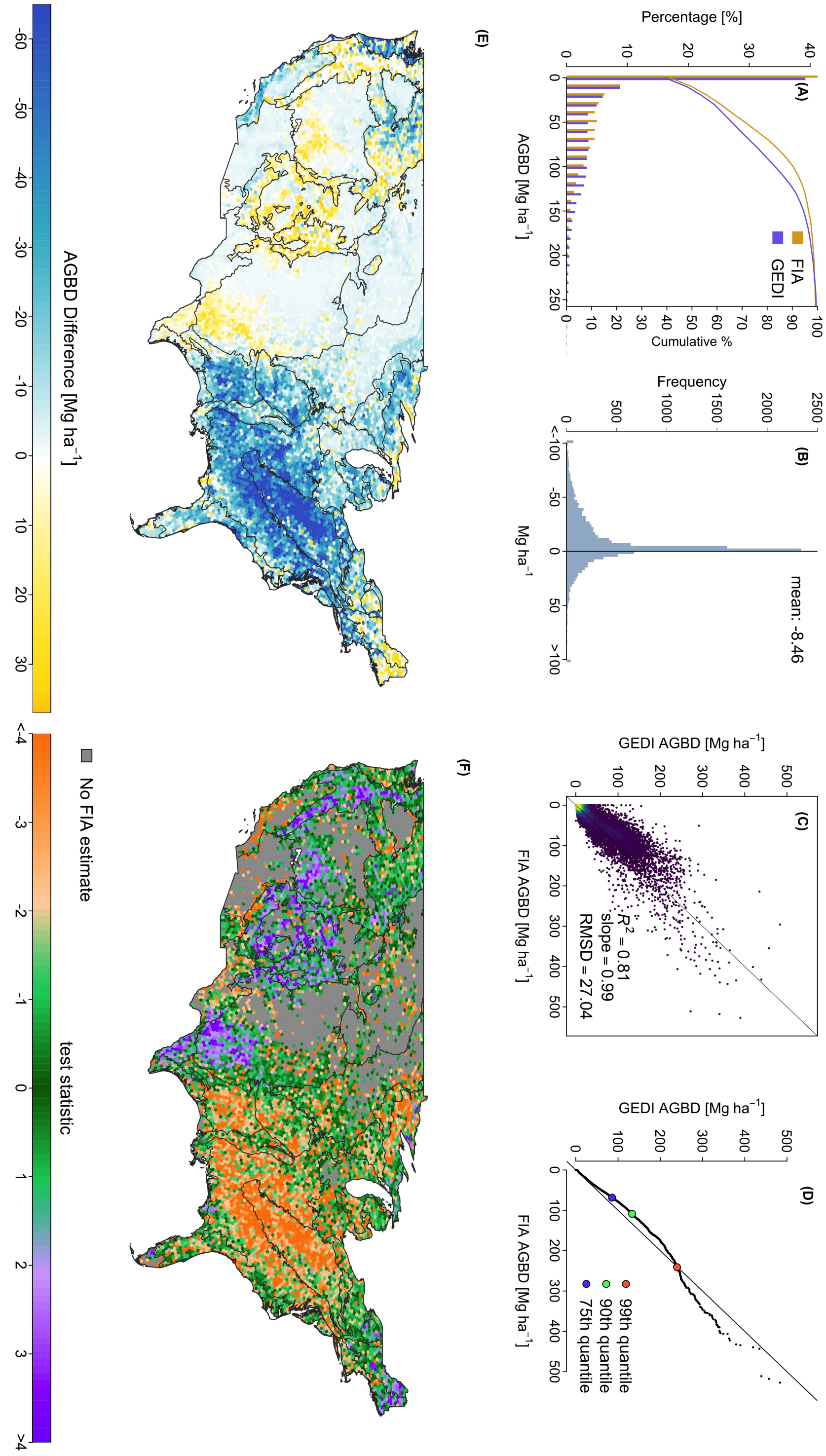

Figure 3.6: Comparison of the filtered GEDI and FIA estimates. a) Side-by-side histogram comparison truncated to 300 Mg ha^{-1}, b) Difference histogram (FIA-GEDI), c) Scatter plot of FIA vs GEDI estimates, d) Quantile-quantile plot of hexagon estimates, and maps of e) estimate differences (FIA-GEDI) and f) $t_{filtered}$, with EPA level II ecoregions overlaid.

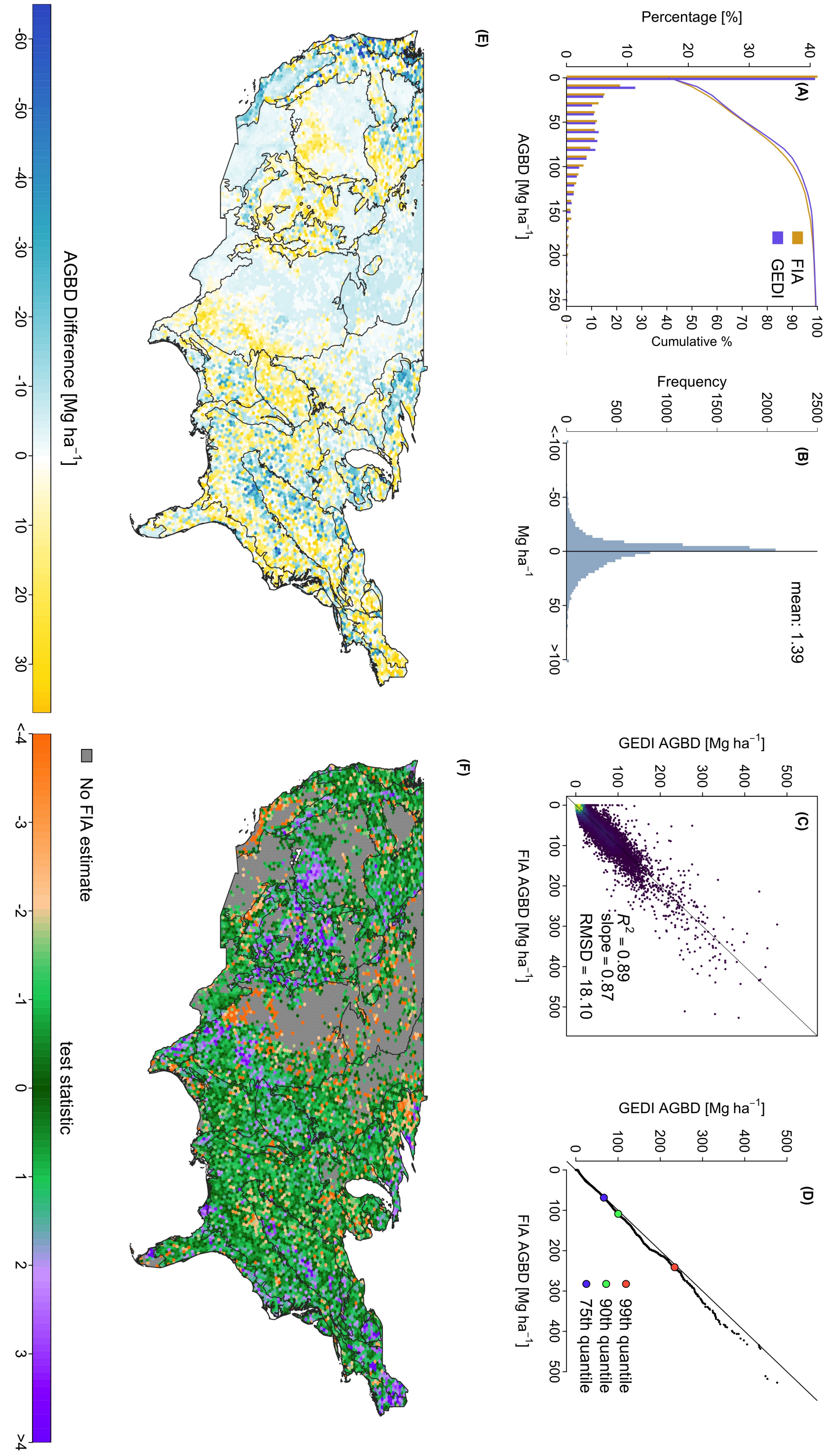

Figure 3.7: Comparison of the fusion GEDI and FIA estimates a) Side-by-side histogram comparison truncated to 300 Mg ha^{-1}, b) Difference histogram (FIA-GEDI), c) Scatter plot of FIA vs GEDI estimates, d) Quantile-quantile plot of hexagon estimates, and maps of e) estimate differences (FIA-GEDI) and f) t_{fusion}, with EPA level II ecoregions overlaid.

In total, our improvements to GEDI's AGBD estimation process resulted in a bias reduction of 86.7%; 19.3% due to improved filtering, and the remaining 67.5% due to the SISA models. Here, we define estimate bias as the percent change between mean absolute differences in the FIA and GEDI estimates. In every strata the distribution of t_{fusion} was more centered on 0 than for $t_{filtered}$ or $t_{original}$ (fig. 3.8). In all strata the interquartile range t_{fusion} was smaller than for $t_{filtered}$ or $t_{original}$ with the exception of NT_W, where it was comparable. This is not a surprising result, because the SISA models were calibrated using the FIA hexagon estimates (see discussion).

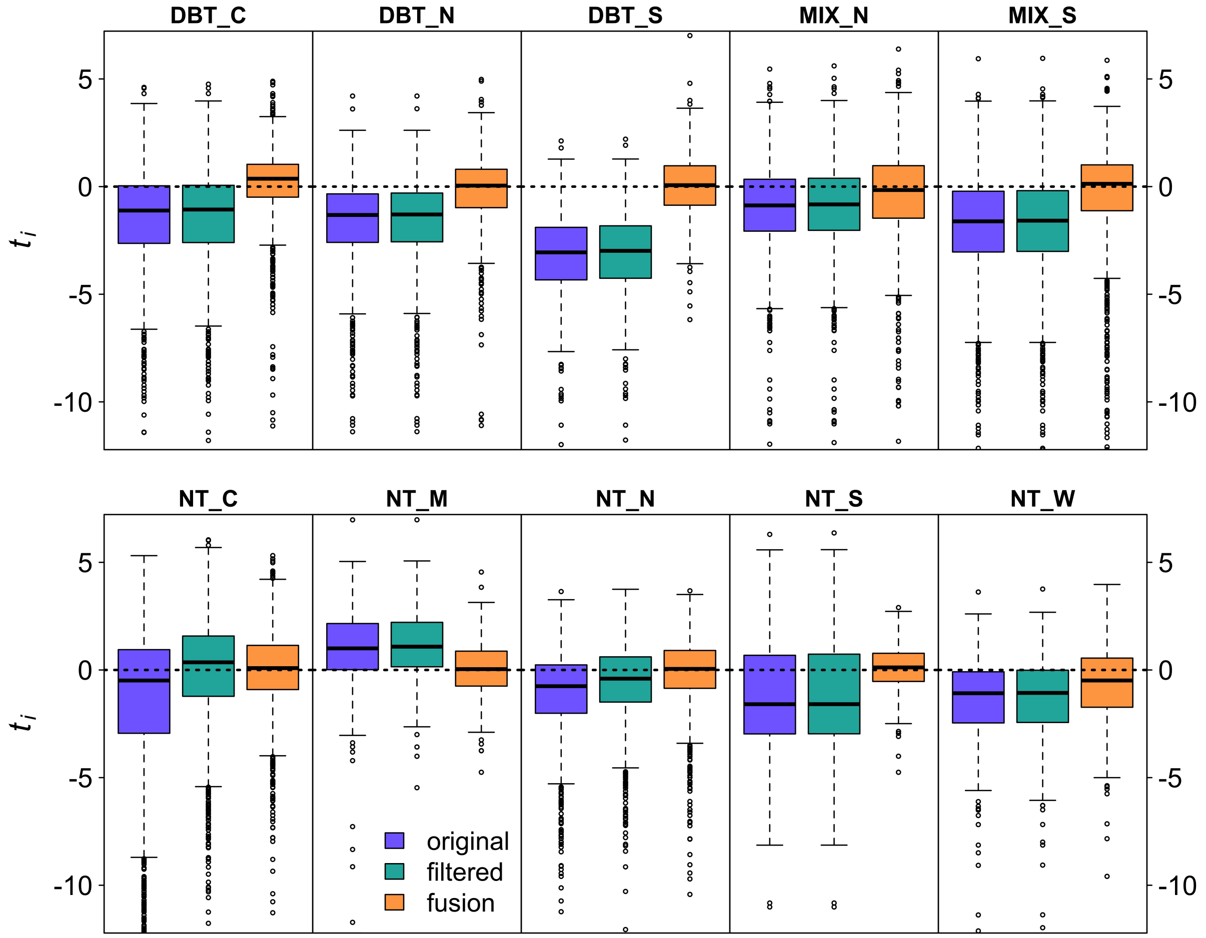

Figure 3.8: Boxplots of t_i from the original, filtered, and fusion estimates show that for all regions, the fusion estimates are in better agreement with the FIA estimates than the original or filtered estimates, with smaller interquartile ranges that are more centered on 0.

The standard error associated with the original and filtered estimates were similar (fig. 3.9), with respective mean values of 1.94 Mg ha^{-1}and 1.85 Mg ha^{-1}, compared to a mean standard error of 2.57 Mg ha^{-1}for the fusion estimates, and 10.88 Mg ha^{-1}for the FIA estimates. The respective mean standard errors as a percentage of the mean estimate were 6.0%, 7.1%, 12.9%, and 38.1%. These reported values represent hexagons with two or more forested FIA plots (N=9851), because it is not possible to produce a valid standard error estimate for hexagons with fewer than 2 plots.

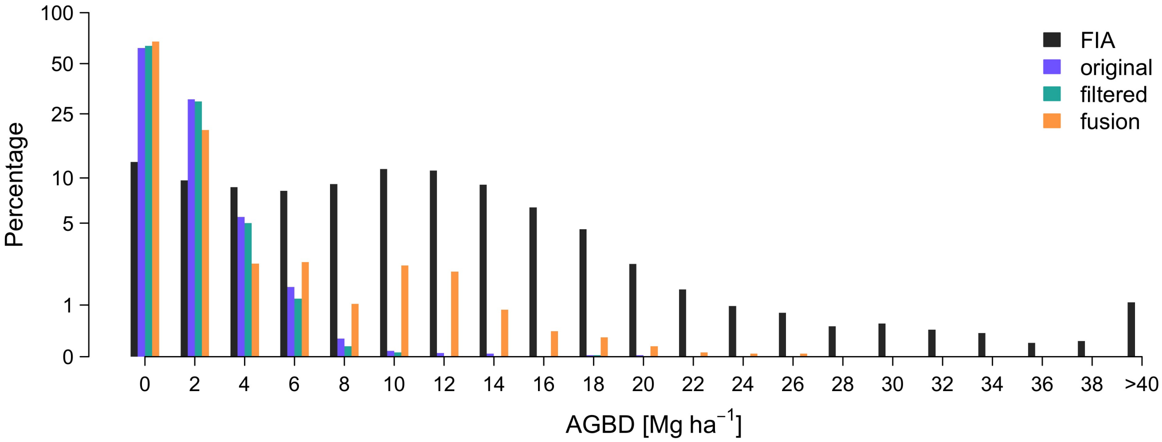

Figure 3.9: FIA's design-based hexagon standard error estimates are substantially larger than GEDI's hybrid standard error estimates from GEDI's original, filtered, and fusion mean estimates. The fusion hexagon standard error estimates are more likely than the filtered or original estimates to be less than 2 Mg ha^{-1}, while also displaying a multimodal response similar to that of the FIA standard errors, but to a lesser degree.

3.4.4 Gridded 1km resolution fusion estimates

At the 1 km scale, mean AGBD was 45.6 Mg ha^{-1}from the original GEDI estimates, and 37.2 Mg ha^{-1}from the fusion estimates (fig. 3.10). For 1 km cells with mean estimates less than 100 Mg ha^{-1}, the respective mean standard errors were 4.5 Mg ha^{-1}and 4.8 Mg ha^{-1}, and for 1 km cells with mean estimates larger than 100 Mg ha^{-1}, the average standard error as a percentage of the mean estimate were 11.4% and 10.9%. Here the difference in uncertainty reporting (Mg ha^{-1}vs. percentage of the mean) is to coincide with GEDI's specific precision requirements; estimate precision for grid cells with a mean estimate below 100 Mg ha^{-1}should be reported in units of Mg ha^{-1}, while estimate precision for grid cells with a mean estimate greater than or equal to 100 Mg ha^{-1}should be reported as a percentage of the mean estimate.

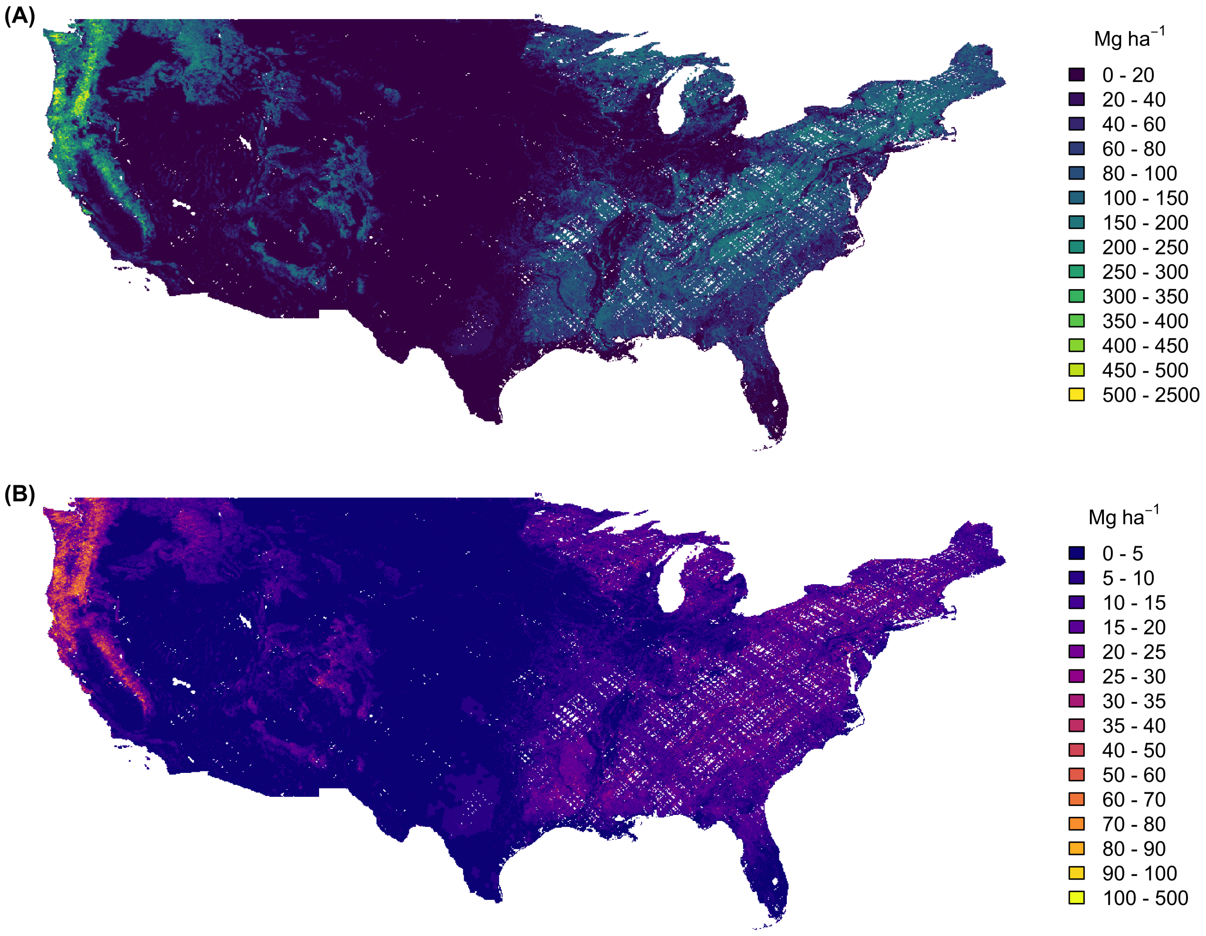

Figure 3.10: Mean 1 km estimates of AGBD (A) and its uncertainty (B). The white areas in both maps are gaps in the GEDI sample at 1km scale.

3.5 Discussion

The GEDI mission represents an advance in global biomass mapping, because the ability to characterize estimate uncertainty and flag deviations from reference estimates that are unlikely due to chance allows for iterative improvements to GEDI's overall estimation process (Dubayah et al., 2022b). We leveraged this ability to identify and mitigate the filtering and L4A model misspecification issues that adversely impacted GEDI's estimation process. In the discussion that follows we first summarize why stricter data filtering is important and should be implemented with care. We present a possible cause of L4A model bias in certain regions within the US, and explain how the SISA framework accommodates this phenomena. We then acknowledge the shortcomings in our SISA modeling approach, specifically how modeling assumptions may be violated to varying degrees in certain areas, and the implications for the resultant predictions and estimates. We conclude with comments on the circularity of comparing the fusion GEDI and FIA references estimates in a probabilistic manner.

GEDI's data filtering procedure applies a set of rules to identify waveforms not suited for biomass estimation (Healey et al., 2022). Our findings suggest that additional filtering of a small number of observations impacted by topography or low clouds can improve GEDI's biomass estimates at the hexagon scale, and by extension, other spatial scales. However, caution must be exercised when identifying such observations so as not to violate the hybrid assumption of a random cluster sample. We did not implement any additions to the AGBD estimation process to deal with an increased number of non-response samples relative to GEDI's original estimates, for two reasons. First, we needed to implement the exact estimation method used to produce the original GEDI estimates to ensure a like-comparison of results. Second, the number of observations designated as non-response by our added filtering practices was very small relative to the number of GEDI observations already removed from the sample by GEDI's current filtering procedures. Our topographic filtering approach removes a small number of GEDI observations heavily impacted by topography, deceasing AGBD estimation bias without substantially impacting the GEDI sample. Reducing the impact of clouds and fog within 150 meters of the land surface globally is more challenging, especially when clouds and fog can occur below the height of the tallest tree in an area. Our cloud filtering approach is straightforward and only removes affected observations with heights larger than the tallest trees. In future studies we recommend development of algorithms to improve detection of low clouds and fog impacting GEDI waveforms.

Availability of calibration data is a limiting factor to unbiased, continental scale biomass mapping with GEDI. The spatial pattern of differences between the filtered GEDI and FIA estimates (fig. 3.6e,f) implies a systematic issue somewhere else in the AGBD estimation process that is not related to data filtering. Despite the considerable effort spent to ensure a representative L4A training sample and unbiased models (Duncanson et al., 2022), the attenuation of this spatial pattern in the differences between the fusion GEDI and FIA estimates (fig. 3.7e,f) suggests that L4A model bias is the primary cause.

The results for MIX_N and NT_N are especially informative, given these strata respectively contained 85.3% and 67.9% of the calibration plots used to train the L4A DBT and NT models (table B.3). In both strata the L4A models appear unbiased, as the unit level distributions are similar to those from the FIA (fig. 3.3), and there are not widespread systematic differences between the filtered GEDI and FIA hexagon level AGBD estimates (fig. 3.7e,f; fig. 3.8). Conversely in other strata, there is substantial disagreement in between the L4A and FIA unit level distributions and the GEDI filtered and FIA hexagon estimates. Two such strata are DBT_S and NT_M, which respectively contain 0.4% and 4.5% of the GEDI calibration plots for the L4A DBT and NT models (table B.3). As geographic transferability was prioritized during the L4A model calibration process, these specific L4A models had the best performance when applied to validation data outside of the geographic extent of training data, relative to all the candidate models tested during L4A calibration (Duncanson et al., 2022). In other words, the L4A models are as unbiased as possible given the available training data. Yet, substantial discrepancies in observed L4A performance exist across space; the observation that the L4A models appear better specified in MIX_N and NT_N than in DBT_S and NT_M allows for the possibility that the relationship between forest structure and AGBD varies spatially. Variation in the relationship between forest structure and AGBD between the DBT_S and MIX_N strata, and between the NT_M and NT_N strata, could explain the difference in L4A model performance because the training data are more representative of the MIX_N and NT_N strata than of the DBT_S and NT_M strata. That the L4A and FIA unit level distributions and the filtered GEDI and FIA estimates match at all in some regions despite spatially varying relationships and limited training data is a testament to the L4A calibration and validation process.

The SISA modeling framework accounts for spatial variation in the relationship between GEDI derived forest structure metrics and AGBD in two ways. First, model development at the aggregate (hexagon) scale ensured training data with uniform spatial coverage throughout each of the ten prediction strata, appropriately capturing within-strata variability in the structure to biomass relationship. Second, the large number of hexagons within the US allowed for more localized and homogeneous prediction strata relative to L4A's two continental scale prediction strata. The SISA prediction strata contain presumably less variation in the local structure to biomass relationships than the continental scale DBT and NT L4A strata, thus decreasing the likelihood of localized model bias within a SISA stratum. Our comparisons of the SISA, L4A, and FIA unit level distributions of AGBD demonstrate the importance of accounting for spatial variability in forest structure and biomass. For all strata, the SISA distributions were equally or more realistic than the L4A distributions, which leads us to conclude that as a whole the SISA models produced more accurate and realistic predictions than the L4A models. This was most apparent in the DBT_S and NT_M strata, where the application of the SISA models resulted in marked improvement over the L4A models, for both the unit level distributions (fig. 3.3) and the hexagon estimate comparisons (fig. 3.7). The SISA models for DBT_N, DBT_S, NT_N, NT_M, NT_W appeared definitively unbiased. In DBT_C and NT_C the SISA and FIA distributions were marginally different above approximately 50 Mg ha^{-1}, with SISA biased low relative to FIA, implying a low to moderate level of bias in these strata. In contrast, larger differences between the SISA and FIA unit level distributions in MIX_N, MIX_S, and NT_S suggested the SISA models are likely biased to a greater extent in these strata.

Instances of SISA model bias presupposes that the assumption of SISA model spatial invariance has not been met in these areas. The SISA model derivation in section 3.3.5.2 provides a theoretical justification for the assumption of scale invariance necessary for prediction at the unit level, which we then evaluated using the available *in situ* data following our validation scheme in section 3.3.5.4. The unit level distributions, used here as a method of evaluating the extent to which the SISA models may be biased or not, suggest that the SISA models are mostly unbiased, but that some areas of moderate to substantial model bias are probable. This interpretation is supported by the analysis of SISA model parameter estimates across spatial scales. While an overlap of confidence intervals does not guarantee equality in the parameter estimates by itself, it is another manner by which we evaluate the assumption of spatial invariance, in combination with unit level distribution comparisons. The overlap of confidence intervals in the DBT_N, DBT_S, NT_M, NT_N, NT_W strata, along with the matching FIA and SISA unit level distributions, suggests a lack evidence sufficient for invalidation of the assumption of scale invariance in these strata. However, parameter estimates for which the confidence intervals do not overlap across spatial scales within a strata is sufficient evidence of invalidation of this assumption, especially when sample sizes are large. Thus the observed scale-dependent variability in the parameter estimates for DBT_C, NT_C, MIX_N, MIX_S supports the results from the unit level distributions, that the assumption of scale invariance in these strata is not met to the same degree as in other strata.

Degradation of scale invariance leading to SISA model bias in certain regions leads us to believe that, despite our best efforts, one or both of the assumptions underpinning our SISA model derivation were violated to some degree in certain areas. There are several mechanisms by which these assumptions—a consistent relationship between predictor and response variables that is linear at both the aggregate and unit level—could be impacted. The first is that within strata variations to the relationship between FIA AGBD estimates and forest structure quantified by GEDI violates the assumption of a consistent relationship. This is likely true to some extent within several strata, as seen in the remaining differences between the FIA and fusion estimates within the MIX_N region (fig. 3.7). Relative to the FIA estimates, the fusion estimates are systematically larger in the northern mid-west, and systematically smaller throughout much of the northeast. The relationship between FIA biomass and GEDI forest structure variables may vary longitudinally within this stratum, and as a result positive model bias in the northern midwest leads to overestimation of AGBD, while negative model bias in the northeast leads to underestimation. The decision of ten strata was somewhat arbitrary, and future work could employ a more data driven stratification, such as Scarth et al. (2019), aimed at maximizing within-strata structural and biomass characteristics.

Similarly, a non-linear relationship between predictor and response variables at the unit level would also violate the SISA model assumption. We used all available plot level data to ensure only linear relationships qualified as candidate SISA models during calibration (section 3.3.5.3, appendix B.3, figs. B.2 to B.4). However, as the GEDI calibration plots are spatially clustered and represent some SISA strata far better than others, we may have chosen SISA relationships for which the linearity constraint breaks down in areas that are not well represented by the GEDI calibration sites.

A third mechanism by which the SISA model assumptions could be violated relates to a small but important difference in the population of interest between the FIA and GEDI. While GEDI does not make a distinction between forested and non-forested areas in biomass estimation, the FIA only estimates biomass within forested lands and explicitly ignores vegetation located on lands that do not meet its definition of forest. A mismatch between GEDI predictor variables used in SISA calibration that reflect non-forested vegetation and the FIA AGBD response variable that ignores such vegetation would at best add noise to a calibration data set that would not otherwise be present, and at worse could violate both the linearity and consistency assumptions within a strata or part of a strata. The strata where the SISA models appear partially biased (DBT_C, NT_C, MIX_N, MIX_S) contain some of the lowest proportions of FIA estimated forest cover of all strata (table 3.1) and thus have more opportunity for non-forested vegetation to be captured by GEDI.

The comparison of the fusion GEDI hexagon estimates and the FIA hexagon estimates is circular in that the FIA reference estimates were used to calibrate the SISA models, which underpin the fusion estimates. Thus it is not surprising that the fusion estimates were in far better agreement with the FIA estimates than either the original or filtered GEDI estimates. The more interesting result was that the SISA models produced mostly realistic distributions of unit level AGBD, matching the FIA unit level distributions of AGBD. Since the SISA models were mostly unbiased and yielded realistic predictions, it follows that the fusion and reference estimates are in relative agreement. The hexagon level comparison of the fusion and reference estimates was only necessary to demonstrate that the SISA models result in mostly unbiased estimates when applied within hybrid inference. The ability to calibrate unit level models at an aggregated scale is important because it helps overcome challenges in obtaining representative training data for remote sensing based models of AGBD. Further, the fusion hexagon estimates are more precise than the FIA reference estimates (fig. 3.9), which reduces overall uncertainty in the US forest carbon stock and may help constrain future fluxes. This work represents not only a refinement of GEDI's biomass estimation process, but also an advance in remote sensing based methods of biophysical inference.

3.6 Acknowledgements

The authors thank Zhiqiang Yang for assistance with the FIA database.

This work was supported by funding from the Global Ecosystem Dynamics Investigation (NASA contract NNL15AA03C) and NASA FINESST Grant (80NSSC21K1626).

Chapter 4: Dimensions of Eastern Old-Growth Appear Different From Space

4.1 Abstract

Old-growth forests have been widely studied for decades. The extreme diversity of old forest characteristics has inspired an equally diverse set of old-growth definitions. This lack of consensus makes mapping old-growth difficult across space and varied forest types. While the use of remote sensing in old-growth research is not new, there is a growing need for large scale mapping to improve understanding of old forest processes and to support old-growth conservation. Old-growth mapping requires definitions that are ecologically relevant to old forests while also transferable to remote sensing data. In this paper we develop a conceptual framework to evaluate three dimensions of old-growth—a temporal dimension related to tree ages, a physical dimension related to tree sizes, and a functional dimension related to forest processes. In the first part of our analysis, we classify forests throughout the eastern US as old or not with respect to each old-growth dimension using existing old-growth definitions and data from the US Forest Inventory and Analysis (FIA) program. We estimate the proportion of forest classified as old within a hexagon grid, resulting in a unique map of old forest proportion (OFP) for each dimension. Subsequently, we use spaceborne lidar data from NASA's Global Ecosystem Dynamics Investigation (GEDI) to reproduce each OFP map in a modeling framework designed to 1) assess the extent to which each dimension of forest oldness can be mapped at large spatial scales, and

2) identify biophysical GEDI variables related to each dimension of forest oldness. We found substantial spatial variation in the mapped OFP estimates across the three dimensions, highlighting how definition criteria impacts old-growth maps. We also found that physically old forests were more effectively mapped than functionally or temporally old forests, and that physically old forests were more structurally similar to one another than temporally or functionally old forests. Our modeling results indicate that while lidar remote sensing may be best suited to mapping physical old-growth characteristics, definitions that rely solely on physical characteristics do not adequately represent old forests throughout the eastern US. We propose that remote sensing efforts to map old-growth with spaceborne data may maximize utility through collaboration with western and indigenous old-growth experts to determine ecologically relevant and precise old-growth mapping definitions, appropriate spatial resolutions, and flexible quantitative frameworks that account for the complexities and heterogeneity of old forests.

4.2 Introduction

Considerable ecological research has focused on forests that are generally referred to as old-growth—forests with relatively old trees and characteristics that require a long time to develop (Frelich and Reich, 2003; Spies, 2004). Early old-growth studies by Western scientists were reliant on ground-based observations to analyze a single or small collection of sites within a localized region or specific forest type (Davis, 1996; Gaines et al., 1997; Tyrrell, 1998). Initial attempts to map old-growth conditions were also local, utilizing a combination of ground-based and remotely sensed data (e.g. Helmer et al., 2000; Falkowski et al., 2009; Hansen et al., 2014). As mapping efforts expand in geographic extent, there is a growing need for data sources and methods that can identify old-growth across a variety of forest types and diverse environmental conditions at scale.

The use of spaceborne remote sensing in old-growth mapping is limited but evolving (e.g. Spracklen and Spracklen, 2019, 2021; Davis et al., 2022; DellaSala et al., 2022). High resolution optical time series are not long enough to appropriately characterize stand level disturbances and forest longevity on the temporal scale of old-growth processes, and mapping efforts have instead used forest structure as a proxy for longevity and to identify biophysical conditions within known old forests. Presently, forest structure is most effectively measured by lidar systems such as airborne laser scanning (ALS) or spaceborne waveform lidars (ex. the Global Ecosystem Dynamics Investigation (GEDI)), or synthetic aperture radar (SAR) systems (ex. TanDEM-X or NISAR) (Krieger et al., 2007; Kellogg et al., 2020). The future of large scale old-growth mapping will likely involve multi-sensor fusion that merges different types of information related to forest structure and functioning derived from lidar, radar, or multispectral, hyperspectral, and stereo optical imagery at high spatial resolutions, and the integration of these data sets with forest inventory and dendroecology observations (Hirschmugl et al., 2023).

The evolution in old-growth mapping efforts has renewed debate over old-growth definitions. In this new context the question becomes: what exactly is being mapped? Old-growth cannot be detected directly, and instead must be inferred through modeled relationships between remotely sensed biophysical predictor variables and a response variable related to old-growth. This method requires the response variable to be defined precisely and consistently throughout the area being mapped. However, forests age in many different ways (Wirth et al., 2009). Old forests are highly complex and dynamic, composed of gradients and interconnected processes with diverse manifestations based on forest type, disturbance regimes, site quality, topographic position, and climate (Pesklevits et al., 2011). The difficulty in appropriately representing old-growth diversity with a consistently defined response variable is perhaps the biggest challenge to old-growth mapping at large spatial scales (Hirschmugl et al., 2023).

The diversity of old forests has lead to agreement within the scientific community that a single unified old-growth definition is not possible nor preferable. There are many ideas regarding how old-growth should be defined and what specific criteria should be used, based on both Western forest science and Traditional Ecological Knowledge from indigenous communities (Hilbert and Wiensczyk, 2007; Wirth et al., 2009). Recently, different old-growth definitions (Pelz et al., 2023; Barnett et al., 2023) were developed for the US Forest Inventory and Analysis (FIA) network that resulted in divergent estimates of how much old-growth exists throughout North America, despite using the same inventory data and estimation methods. The results suggest old-growth estimates are highly sensitive to definition criteria, the extent to which was previously unknown.

Pelz et al. (2023) and Barnett et al. (2023) have advanced old-growth discourse in the US by highlighting the sensitivity in amount estimates to the definition criteria, which is especially important when informing forest management plans, resource extraction, or conservation efforts with spatially explicit old-growth information. Different old-growth definitions are helpful in understanding various aspects of old forests and their spatial patterns, as long as the respective differences are understood and recognized. A systematic comparison of the Pelz et al. (2023) and Barnett et al. (2023) definitions is thus necessary to contextualize the results, map the spatial pattern of old forests according to these definitions, and inform future old-growth investigations within the US.

We propose the following conceptual framework to evaluate forest oldness definitions within the FIA network. Forest attributes related to old-growth can be grouped into three categories, which we refer to as old-growth dimensions:

1. **Temporal**: tree ages and overall stand age structure, the number and age of cohorts (if applicable)
2. **Physical**: the size and shape of trees, stand stem density, basal area, biomass, canopy cover and vertical profile, structural complexity
3. **Functional**: biogeochemical processes such as net primary or ecosystem production (NPP, NEP), nutrient cycling,

These forest attributes change constantly over time, dependent on demographic processes (recruitment, growth, mortality) and disturbance regimes that shape these processes. Our framework simplifies changes in forest attributes by assuming development over time, such as from less to more biomass, or from a younger to an older maximum tree age. Forest oldness is then defined as the progression of a given attribute's value over time, and can be independently assessed with respect to each dimension (fig. 4.1A). *In situ* forest stands within the FIA network can be classified as "old" or not along a given dimension by 1) choosing a specific attribute representative of the dimension (ex. mean stand age), 2) setting a threshold value along its developmental gradient to designate the onset of oldness (ex. 120 years), 3) and comparing this value to those from inventoried forest stands (fig. 4.1B). Importantly, classifications are neither mutually inclusive or exclusive across the dimensions. While this method conflicts with views of old-growth as a dynamic process instead of a developmental state that can be classified (e.g. Spies, 2004; Pesklevits et al., 2011; Barton and Keeton, 2018), it is useful in contrasting definition criteria, estimating the

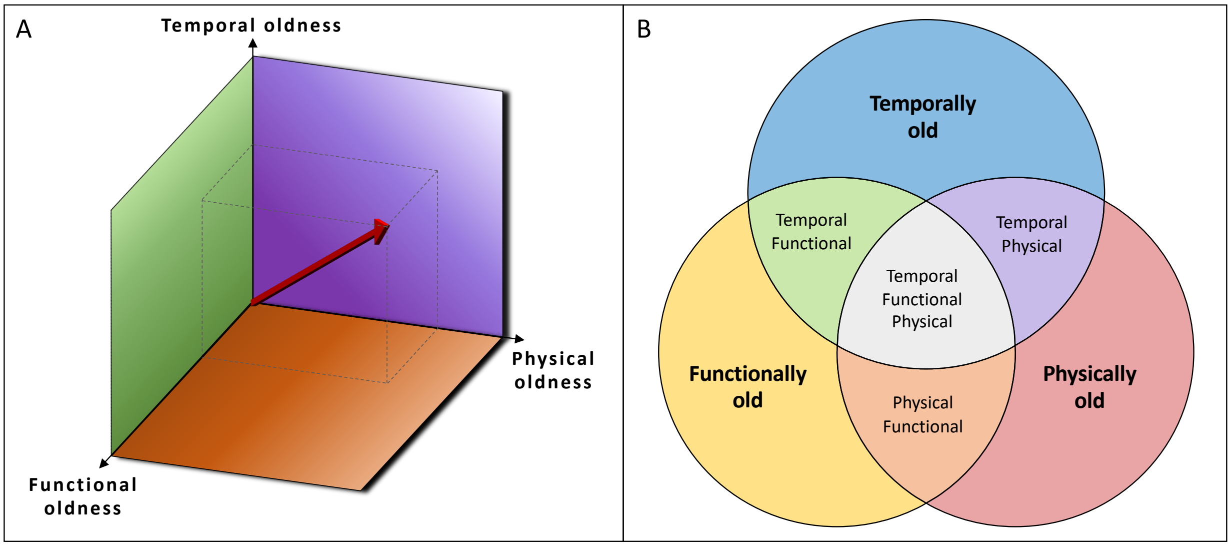

Figure 4.1: (A) A theoretical framework for quantifying forest oldness along three dimensions: temporal, physical and functional. The red arrow represents how a forest stand's attributes could be mapped onto each dimension, and demonstrates that a stand's oldness in each dimension may not be equal. Threshold values can be set along each dimension (not shown here) to delineate the onset of old-growth characteristics for that dimension. Evaluating data from *in situ* stands against the threshold values results in a binary classification of forest oldness along each dimension, which are neither mutually inclusive nor exclusive. (B) Possible combinations of stand-level old-growth classifications when considering all three dimensions. A stand's classification in one dimension is independent from the other dimensions, in that physical characteristics are not be considered when classifying temporal oldness, and so on. There may or may not be multiple old-growth classifications across the dimensions for a given forest stand.

spatial patterns of old forests, and quantifying the extent to which various old-growth definitions can be mapped with remote sensing.

In this paper we contrast old-growth definition criteria from Pelz et al. (2023) and Barnett et al. (2023) and evaluate the definitions with respect to old-growth mapping throughout the Eastern US, using data from the GEDI mission as a case study. Our goal is to systematically assess the impact of old-growth definition criteria on the amount of old forest estimated throughout the Eastern US, and the ability to map old forests using GEDI data. In achieving this goal we seek to answer the following science questions:

1. What are the spatial patterns associated with eastern old forests and how do they vary by dimension?
2. To what extent can each dimension of old-growth be mapped within our framework using GEDI lidar data?
3. Are there structural characteristics unique to each dimension of old-growth?

We hypothesize the physical dimension of old-growth will be more effectively mapped with GEDI data than the temporal or function dimension, and will evaluate modeling success using normalized prediction error. While physical old-growth should have a structural signature that is easily identified using GEDI data, we expect that temporally and functionally old forests may also poses structural signatures that can be detected and leveraged for mapping. We intend the results of this study to inform the definition development and theoretical underpinnings of future old-growth mapping efforts.

4.3 Methods and Data

We begin our analysis (fig. 4.2) by integrating the old-growth definition criteria from Pelz et al. (2023) and Barnett et al. (2023) into our forest oldness conceptual framework, and derived unique temporal, physical, and functional old-growth definition criteria by forest type for the eastern US. We then classified every forest stand sampled by the FIA network as old or not according to each dimension's oldness definition. From the classifications we produce areal estimates of old forest proportion (OFP) for each dimension, which represent the ratio of old forest area to total forest area within an equal-area hexagonal tessellation of the land surface. We then assessed the extent to which each dimension's OFP estimates can be modeled using biophysical forest attributes inferred from GEDI lidar forest structure data. We use regionally calibrated regression trees to model the OFP estimates for each old-growth dimension as a function of GEDI variables sampled from forested land and aggregated to the spatial resolution of the OFP estimates. Lastly, we determine which types of GEDI forest variables were most important in predicting each set of OFP estimates.

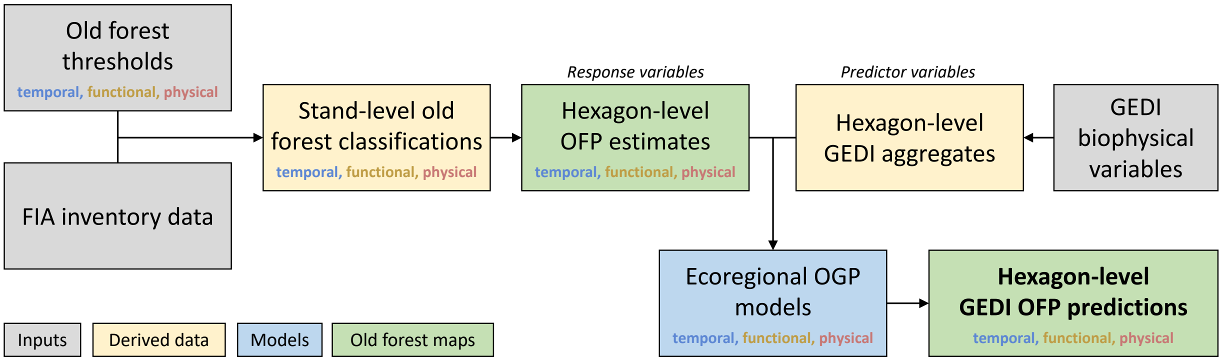

Figure 4.2: Methodological overview: Old forest definitions from Pelz et al. (2023) from Barnett et al. (2023) were standardized within our conceptual oldness framework (fig. 4.1) and applied to FIA inventory data resulting in binary old forest classifications at the stand-level across the temporal, functional, and physical old-growth dimensions. From these classifications, we estimated the old forest proportion as a ratio of old forest area to total forest area for each dimension, using ratio estimators adapted from Bechtold and Patterson (2005) applied within a hexagonal tessellation covering the eastern US. The result was three separate maps of OFP for eastern US forests. Footprint-level GEDI variables were then aggregated within the hexagonal grid and used as predictor variables in modeling the OFP estimates. Regression tree models were calibrated at the ecoregion level for each dimension's OFP estimates, resulting in GEDI-based prediction maps of OFP for each dimension. The variable importance from these models were compared to gain ecological inference about the structural characteristics and mapability of each old-growth dimension.

4.3.1 National forest inventory data

The FIA data used in this analysis were obtained from the plot and condition tables in the FIA database, for plots sampled between 2010 and 2022 (Gray et al., 2012). If an individual plot was sampled twice during this period, we used the most recent inventory information. Data were obtained and analyzed at the stand-level using the ‘rFIA’ package in ‘R’ (Stanke et al., 2020; R Core Team, 2022)

4.3.2 Old forest definitions

On April 20th 2023, the United States Forest Service (USFS) released a national inventory of old-growth and mature forests within US federal lands (Barndt et al., 2023). Definitions were developed by USFS foresters and ecologists for predominant forest types within each of the nine Forest Health Protection regions within the US, and criteria identifying old-growth characteristics were set individually for different old-growth community types, which are defined as a collection of FIA forest type groups. The definitions were designed specifically for FIA data, and an explanation of the definition development process is provided by Pelz et al. (2023). Here, we used the eastern and southern region definitions (respectively, tables 15-17 in Barndt et al. (2023)) as the foundation for our temporal and physical old forest definitions. This required some standardization of old growth community types and definition criteria to ensure consistency in the application of old-growth definitions throughout our study region, as outlined below.

First, we standardized definition criteria for old growth community types that occurred in both regions (e.g. northern hardwoods). The standardization of definition criteria was done to ensure a single set of definition thresholds were used for community types that occurred in both regions. We also merged community types across the regions when the species compositions and definition thresholds were comparable based on our informed opinion of forest composition in eastern forests, and resulted in the merging of five community types.

Next, we harmonized the temporal and physical criteria for each community type across the regions. The southern and eastern regions used identical temporal criteria, in that stand age must exceed a threshold value, and we made no further modifications. The physical criteria however varied between regions. The eastern region used a minimum density of large trees, in which the density (ex. 10 trees per acre) and diameter of large trees (ex. 20 inches) varied independently by community type. The southern region also required a minimum density of large trees, but the density was held constant at six trees per acre and while the size threshold for large trees was allowed to vary. There were also live basal area and standing dead tree criteria included in the southern region's physically old criteria. For simplicity, we opted to use the USFS eastern-style definition of physical forest oldness, and estimated the appropriate density and large tree threshold values for southern region forest types using data from Gaines et al. (1997). The result was a set of harmonized temporal (stand age) and physical (large tree density) oldness criteria for the old-growth community types set by Pelz et al. (2023) in the eastern US (table 4.1).

Functional forest oldness criteria were taken directly from Barnett et al. (2023). The analysis used a space-for-time substitution with FIA plot data to relate carbon accumulation to stand age for a combination of FIA forest type group and productivity classes, stating that functional old-growth characteristics are reached when stand-level aboveground carbon density reaches 95% of a theoretical maximum value derived from FIA stand data. They calculated the age at which functional old-growth characteristics were reached at the stand level for each FIA forest type group (table 2 in Barnett et al. (2023)). We used the same forest types and age thresholds to classify functional oldness according to our conceptual framework. We also performed a comparison of the mean temporal and functional age thresholds for eastern forests to aid in understanding estimate differences between the temporal and functional dimensions.

4.3.3 FIA stand classification

The temporal, physical and functional old forest definitions were applied to FIA inventory data at the stand level based on the FIA forest type associated with each stand. For the temporal and physical dimensions the forest type assignment is documented in table 4.1, while for the functional dimension the forest type assignment was based on the combination of stand productivity class and FIA forest type group. An FIA stand is an identification of specific forest characteristics based on land use, reserve status, ownership, regeneration status, tree density, forest type and stand size within a plot, and there can be multiple stands located on a plot (Bechtold and Patterson, 2005). Each stand is assigned an age based on the average age (from tree cores) of two or three dominant canopy trees. This variable served as the basis for the temporal and functional old-growth dimensions. We calculated the physical old-growth dimension's large tree density for each stand using the tree-level inventory data and associated large tree size and density thresholds in table 4.1. We then classified each FIA stand as old or not by comparing its values to the temporal, physical and functional thresholds. The result was a set of three old forest classifications for each stand.

4.3.4 Areal estimation of old forest proportion

From the binary stand-level old forest classifications, we estimated the proportion of old forest land relative to all forest land for each old-growth dimension, which we refer to as the old-forest proportion (OFP) estimates because values ranged between 0 and 1, inclusive. We adapted the ratio-of-means estimator from section 4.3.4 of Bechtold and Patterson (2005), using the Horvitz-Thompson estimator to estimate both the numerator and denominator terms, as follows

$$\hat{R}_j = \frac{\hat{Y}_j}{\hat{X}_j} = \frac{\frac{1}{n_j}\sum_{i=1}^{n} Y_{ij}}{\frac{1}{n_j}\sum_{i=1}^{n} X_{ij}}, \tag{4.1}$$

in which $\hat{R}_j$ is the ratio estimate of old-growth forest area to total forest area for a given type, $\hat{Y}$ and $\hat{X}$ are the Horvitz-Thompson estimates of old-growth forest area and all forest area respectively, j is an index representing each spatial estimation unit, n_j represents the number of FIA plots within the spatial estimation unit, i is an index representing each individual FIA plot within spatial estimation unit j, Y_{ij} represents the proportion of each FIA plot that is classified as old-growth forest expressed as a unit interval [0,1], and X_{ij} represents the proportion of each FIA plot that is classified as forest expressed as a unit interval [0,1]. FIA plots frequently contain multiple condition classes, so to properly account for this possibility we calculated Y_{ij} by summing the plot-area proportions of all forested conditions classified as old-growth on each plot. Similarly, we calculated X_{ij} by summing the plot-area proportions of forested conditions on each plot.

To estimate the variance associated with $\hat{R}_j$ we first calculated the variance associated with the Horvitz-Thompson estimates of $\hat{Y}_j$ and $\hat{X}_j$ as

$$\hat{\sigma}^2_{Y_j} = V\hat{a}r[Y_j] = \frac{1}{n_j(n_j-1)}\sum_{i=1}^{n_j}(y_{ij} - \hat{Y}_j)^2 \tag{4.2}$$

$$\hat{\sigma}^2_{X_j} = V\hat{a}r[X_j] = \frac{1}{n_j(n_j-1)}\sum_{i=1}^{n_j}(x_{ij} - \hat{X}_j)^2 \tag{4.3}$$

along with the covariance of $\hat{Y_j}$ and $\hat{X_j}$ in the form of

$$\hat{\sigma}^2_{XY_j} = \hat{Cov}[XY_j] = \frac{1}{n_j(n_j - 1)} \sum_{i=1}^{n_j} (x_{ij} - \hat{X}_j)(y_{ij} - \hat{Y}_j) \tag{4.4}$$

in which y_{ij} and x_{ij} are respective plot level values of proportion old-growth forest and total forest for plot i within spatial estimate unit j. We then used the variance and covariance to calculate the variance of $\hat{R_j}$ as

$$\hat{\sigma}^2_{Rj} = \hat{Var}[R_j] = \frac{1}{\hat{X}_j^2} \left(\hat{\sigma}^2_{Yj} + \hat{R}_j^2 \times \hat{\sigma}^2_{Xj} - 2 \times \hat{R}_j \times \hat{\sigma}^2_{XY_j} \right) \tag{4.5}$$

While slightly different from standard estimation procedures with FIA data that require post-stratification and strata weights (see Bechtold and Patterson (2005); Pugh et al. (2018)), ratio estimation using a Horvitz-Thompson estimator has been shown to yield highly similar results to the post-stratified methods when estimating aboveground biomass density, and has the benefit of being easier to reproduce (May et al., 2023). The spatial estimation units were delineated by an equal-area hexagon tessellation grid covering the Eastern US (fig. 4.3). This grid informs the FIA sampling design and is the highest spatial resolution for which estimates from FIA data should be made using the FIA's standard estimation methods (Menlove and Healey, 2020).

4.3.5 Old forest proportion modeling

There were two motivations for the second phase of our study involving OFP modeling using GEDI data, ordered here by importance; 1) to determine the extent to which the old forest definitions produced OFP estimates that could be modeled using GEDI data, and 2) to identify which GEDI variables were important in predicting each dimension's OFP estimates. Producing the most accurate predictions of OFP was not an objective, and as such we prioritized interpretability and ecological inference over maximizing map accuracy. We trained regression tree models using the ‘R’ package ‘rpart’, relating each old-growth dimension’s OFP estimates to the aggregated GEDI variables (R Core Team, 2022; Therneau and Atkinson, 2022). Model fitting was performed individually within six different ecoregions, informed by the level II EPA ecoregion delineations within the Eastern US (Omernik and Griffith, 2014) mapped onto the hexagons (fig. 4.3). The OFP estimates for each dimension of old-growth were the response variables, and the predictor variables were hexagon-scale biophysical forest attributes aggregated from the GEDI data within each hexagon. For each fitted model we summed the variable importance by variable type to determine an importance hierarchy that characterized which types of GEDI forest variables were most important in the models. The following subsections explain the steps of our OFP modeling exercise.

4.3.5.1 GEDI data

NASA’s Global Ecosystem Dynamics Investigation is the first spaceborne mission designed to map forest structural attributes (Dubayah et al., 2020a). GEDI is a multi-beam waveform lidar sensor that directly measured forest structure within footprints 25 meters wide, and was operational on the International Space Station from April 2019 - March 2023. GEDI’s sole observable is a returned waveform, and throughout its first epoch GEDI is estimated to have collected 90 billion observations globally. From the waveform a suite of forest attributes are derived via signal processing and modeling, which are grouped into various data products. Here, we used GEDI footprint level 2A (L2A, canopy height) and 2B (L2B, canopy profile) data products collected

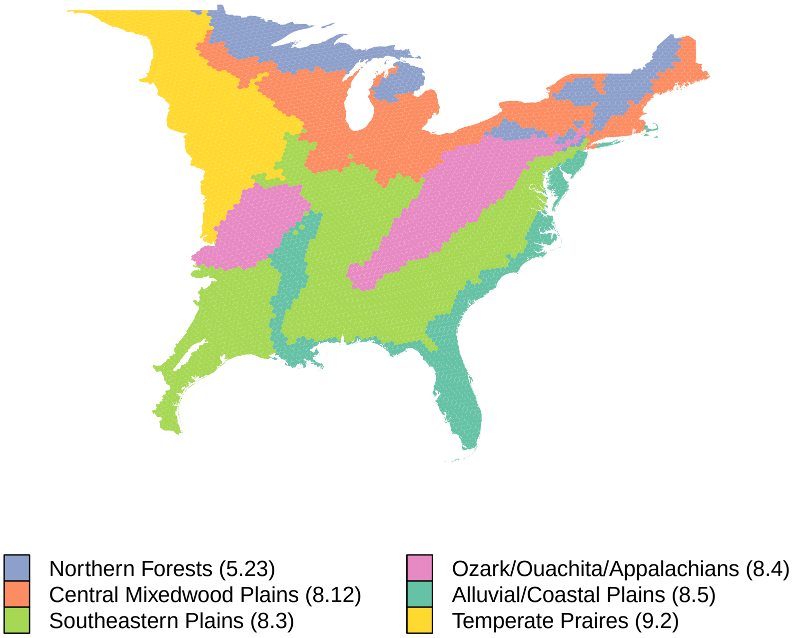

Figure 4.3: The extent of eastern forests considered in this analysis, and the ecoregion delineations used in model calibration, which are based on the EPA level II ecoregions mapped onto the hexagon grid. EPA ecoregions 5.2 and 5.3, and 8.1 and 8.2 were combined based on size and similarity.

between April 2019 and October 2022 (Dubayah et al., 2021c,d). The GEDI mission produces a footprint level 4A (L4A, aboveground biomass) data product, however a recent analysis produced updated footprint level biomass models that produced predictions that were in better agreement with estimates from the FIA network, and these models were used in place of the official GEDI 4A data product (Bruening et al., 2023). We filtered the GEDI sample so that only observations over forest land were used, selecting only multi-modal waveforms (Hofton et al., 2019), and those that intersected a 30-meter resolution forest mask (Wickham et al., 2021).

4.3.5.2 Predictor variables and dimensionality reduction

Hexagon-level aggregates of the GEDI metrics were used as the OFP model predictor variables. For each GEDI metric we calculated the mean, standard deviation, 50th, 75th, 95th and 99th percentiles at the hexagon level. We used L2A relative height metrics in 10 percentile increments from 10% to 90%, and 98%, which is a common proxy for maximum canopy height using GEDI data. The L2B canopy profile variables we used were percent canopy cover, plant area index (PAI), plant area volume density (PAVD), and foliage height diversity (FHD) (Hofton et al., 2019; Tang and Armston, 2019). We calculated partial PAI and plant area volume density (PAVD) variables specific to upper canopy foliage for each waveform. To calculate these upper-canopy metrics, we summed the PAI profiles and averaged the PAVD values for the upper half, third, quarter, fifth, and two-fifths of the canopy for each waveform, and then aggregated these values in the same manner as the other GEDI variables. Additionally, we used the number of modes in each waveform as a proxy for the number of canopy layers (Hofton et al., 2019). Lastly, we generated additional predictor variables for a select set of GEDI metrics by calculating the proportion of GEDI observations within each hexagon with values above specific threshold values (for example, the proportion of waveforms with a maximum canopy height above 35 meters). Combined, there were 205 candidate predictor variables used for model calibration. We organized the predictor variables by metric type into eight predictor variable groups for variable reduction prior to model fitting: biomass, percent canopy cover, FHD, canopy height, number of canopy layers, total canopy PAI, upper canopy PAI, and upper canopy PAVD.

GEDI metrics can be highly correlated with one another, and so we implemented a dimensionality reduction routine prior to fitting each OFP-ecoregion model. This routine minimized multicolinearity and decreased the overall number of predictor variables used to fit each model, and allowed us to assess which variable group wasmost important in predicting the OFP estimates (see section 4.3.5.4). We applied the following procedure to each of the eight variable groups separately. For each variable within a group we collected the within-group covariates with a correlation above 0.9, and selected the single predictor from that collection with the largest correlation to the OFP response variable. This process was repeated using the selected predictors within each group until all within-group correlations were below 0.9, resulting in a final set of variables for each group that minimized within-group correlations while maximizing correlations with the response variable. We then took the five variables with the strongest correlation to the response variable from each final set, to ensure each variable group had the same number of predictors allowed in model fitting. We allowed correlations above 0.9 between predictors in different groups as this information was valuable for ecological inference. Exactly 40 variables were used (five per group) in calibrating each model.

4.3.5.3 Model calibration

We calibrated 18 individual models based on the combination of old-growth dimension (three) and ecoregion (six). We used simple regression tree models as the outputs are interpretable and the variable importance scores are straightforward and comparable between models. We did not separate calibration data into testing and training sets, for several reasons. Excluding a subset of hexagons during calibration would preclude insight into which forest attributes were related to the OFP estimates in those hexagons, and generalizability was not a priority because the models were not applied to other hexagons not used in model training. We calibrated the models using hexagons with a forest proportion of at least 0.2 to reduce noise from those with little forest cover. Case weights were also assigned based on each hexagon's proportion forest, to ensure each hexagon's influence in the model was proportional to the amount of forest it contained. We used five-fold cross validation during calibration to determine a pruning length for each tree that minimized prediction error. We then allowed two additional splits in each model to improve the final predictions at the cost of some over-fitting. Each calibrated model was then applied to the GEDI aggregates to predict OFP for that dimension within each ecoregion. Final predicted OFP maps for each old-growth dimension were produced by combining the predictions from all six ecoregions.

4.3.5.4 Model interpretation and inference

The 'rpart' model's variable importance measure reflects each predictor's ability to explain variation in the response variable (Therneau and Atkinson, 2022). We scaled the raw importance scores to assess each predictor variable's importance relative to all predictors in the model and summed the scaled scores for each predictor by variable group from section 4.3.5.2. By limiting each variable group to only five predictors, we ensured a fair comparison of importance by group, otherwise groups with more than five predictors would have inflated importance measures relative to the other groups. This resulted in a measure that was comparable between old-growth dimensions for a given region, and between regions for a given oldness dimension, allowing inference into which variable groups were the best identifiers of each oldness dimension's OFP estimates. We also calculated a combined measure of variable importance for each old-growth dimension that represented all six regions by weighting each region's importance values by the amount of forested land, which allocated appropriate weight to regions with different amounts of forest and allowed inference for eastern forests as a whole.

4.4 Results

4.4.1 Old forest estimation and classification

Estimates of old forest proportion (OFP) varied by dimension throughout eastern forests (table 4.2). Physically old forests, those with a high density of large trees, were the most common at 45.2%. Functionally old forests, in which annual net biomass change is presumed to be near zero, were less prevalent at 6.5%. Temporally old forests, with a relatively old mean stand age, were least common at 4.6%. Only 0.8% of eastern forests qualified as old in all three dimensions. The spatial patterns of OFP were mostly different between dimensions. The only areas with notable overlap in old forest hotspots between the temporal and physical dimensions occurred along the spine of the Appalachian Mountains and the Adirondack Mountains. Between the physical and functional dimensions, the only overlap occurred in the eastern Cross Timbers region(fig. 4.4).

Regional trends in estimated OFP varied by dimension. Temporally old forests were most common in the Ozark-Ouachita-Appalachian region (8.7%) and the Northern Forests region (5.8%) and least common in the the Southeastern Plains region (2.2%) and the Ozark-Ouachita-Appalachian region (3.4%). The southern regions (Southeastern Plains, Ozark-Ouachita-Appalachian Mountains, and Alluvial and Coastal Plains) combined for more than five times as much functionally old forest than the northern regions (Northern Forests, Central Mixedwood Plains, and Temperate Prairies), 9.4% compared to 1.7%, despite similar temporal (4.5% compared to 4.9%)

estimates. A weaker latitudinal trend in physical estimates between the southern and northern regions was also present (49.0% compared to 38.7%).

Regarding old forest classifications, stands that were classified as physically old were unlikely to be classified as old in another dimension, whereas most stands that were classified as temporally or functionally old were likely to classified as old in at least one other dimension fig. 4.5. When considering overlapping classifications, the likelihood of old forest classification increased if a stand was classified as old in another dimension, and further increased if that stand was classified as old in both other dimensions (table 4.4). This effect was consistent across all dimensions, for eastern forests as a whole and when ecoregions were considered independently. Classification rates were highly variable across the old-growth dimensions within any given forest type, as well as across forest types within any given dimension (table 4.3). Temporal and functional old forest rates by forest type were mostly below 10%, while physically old rates ranged from 14% to 91%.

The comparison of temporal and functional oldness age thresholds yielded similar mean values for eastern forests as a whole, however there were substantial differences by forest type (table 4.5). The spatial pattern of mean age thresholds within the hexagon grid revealed a slight latitudinal trend in the temporal thresholds, and a much larger latitudinal trend in the functional thresholds (fig. 4.6).

4.4.2 Old forest prediction

Physical OFP estimates were modeled with the most success (fig. 4.7). When combining results from the regional models, the difference in physical old forest percent between the es-

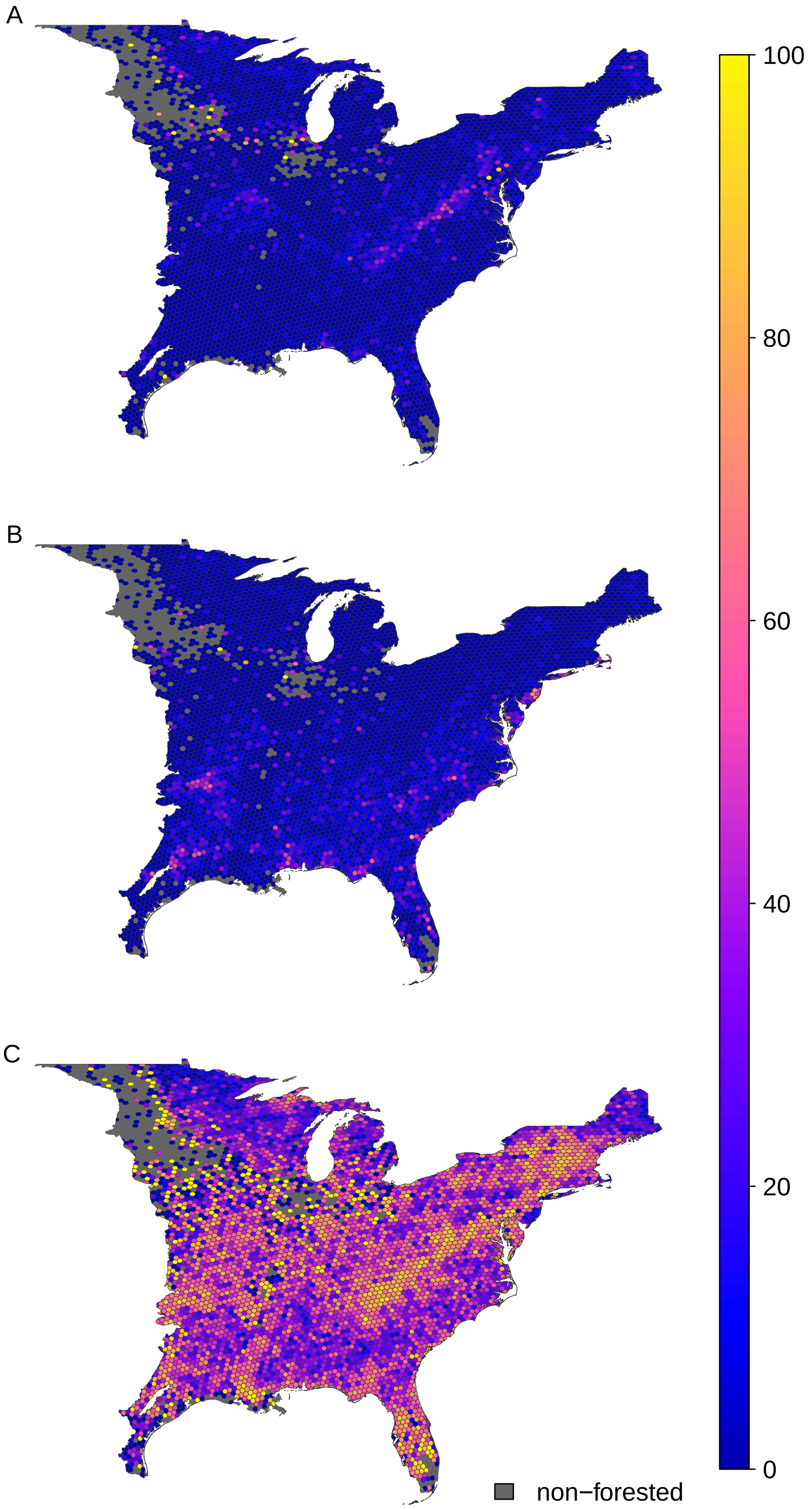

Figure 4.4: Hexagon-level estimates of temporal (A), functional (B), and physical (C) old-growth proportion [0-1] mapped as a percentage [0-100] across the eastern US.

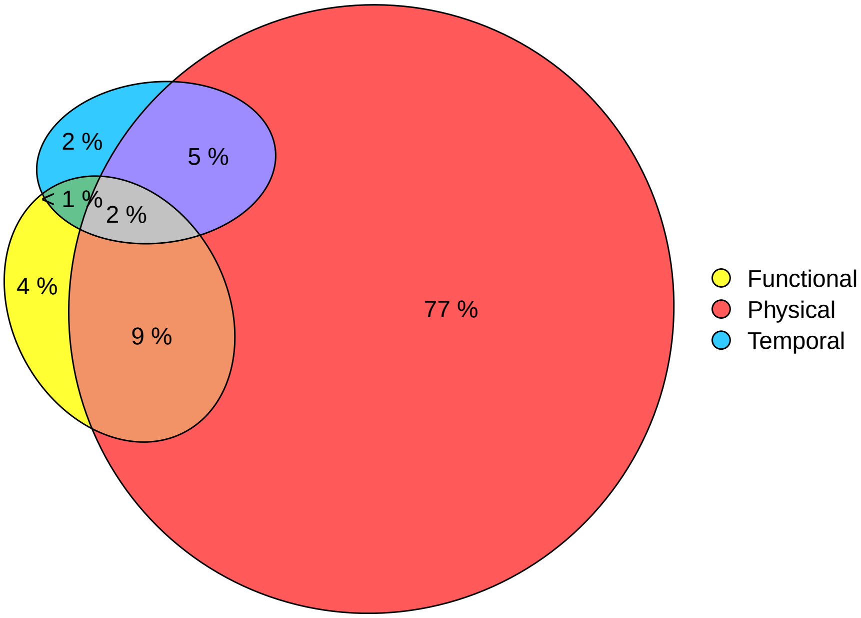

Figure 4.5: Euler diagram of *in situ* old forest classifications at the stand level for all stands in the FIA network considered in this analysis; the empirical counterpart to the theoretical fig. 4.1B. Percentages sum to 100, and thus represent the percent of stands classified relative to number of total stands with at least one old forest classification across the three dimensions.

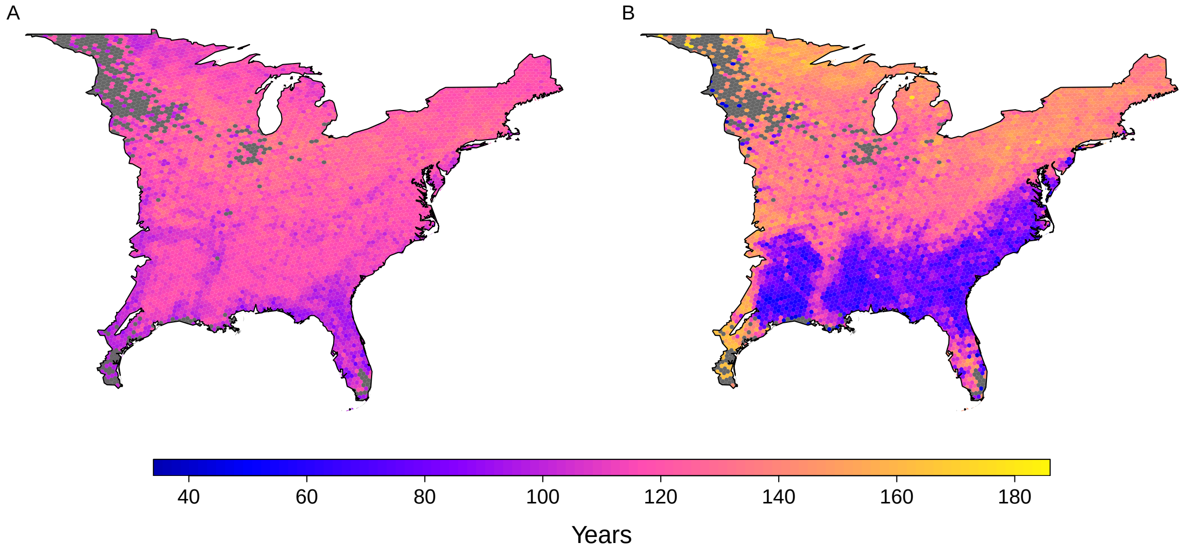

Figure 4.6: Mean stand age threshold values by hexagon for the temporal (A) and functional (B) old-growth definitions.

timated and predicted values (RMSD) was 14.9, which relative to the mean estimate (nRMSE) was 32%, and the GEDI variables explained 43% of the variation in OFP estimates. The functional OFP models were relatively less successful, with respective RMSD and nRMSD values of 7.4 and 119%, and the GEDI variables explained 45% of the functional OFP estimate variation. The temporal models had respective RMSD and nRMSD values of 6.2 and 148%, and the GEDI variables only explained 29% of the variation in OFP estimates.

Model bias, defined as the mean difference between estimated and predicted OFP, for each dimension was near zero (top row in table 4.6), and the OFP predictions represented large-scale spatial patterns of each type of old forests to varying degrees (fig. 4.7). However, attenuation bias resulted in an inability to predict the magnitude of OFP hotspots, evidenced by obvious spatial patterns in the model residuals for each dimension. The temporal models exhibited negative bias for large estimates of OFP (table 4.6) predominately throughout the Appalachian and Adirondack Mountains and Mark Twain National Forest. In contrast, the functional OFP models were able to accurately identify some but not all hotspots of functionally old forests, most notably the New Jersey Pine Barrens, Ouachita National Forest, and Choctaw Nation. The physical OFP models were most successful in reproducing large-scale spatial patterns of physically old forests. These models tended to over-predict physical OFP on throughout the Allegheny plateau and western slopes of the Appalachian Mountains, the southeastern Peidmont region, and northwestern Great Lake states, while under-predicting throughout central New England, Adirondack Mountains, and eastern cross timbers region.

The GEDI variable types that best explained physically old forests were similar across ecoregions (fig. 4.8); GEDI variables of foliage height diversity (FHD), a measure of canopy strata represented as a diversity index, aboveground biomass, and canopy height were consistently important predictors of physical OFP, while other variable types were comparatively unimportant. Functional OFP estimates were also best explained by variables related to canopy height and FHD, although there was not as much differentiation in importance across the variable groups as for the physical models. In contrast, temporal OFP model variable importance was the least differentiated by variable group for eastern forests as a whole, with substantial variation in importance across regions.

4.5 Discussion and Conclusions

Our analysis contrasts old forest dimensions and corresponding maps of OFP. The OFP mapping results show that definition criteria has a strong and direct effect on the total estimates and spatial patterns of old forests. This finding helps reconcile estimate differences between Pelz et al. (2023) and Barnett et al. (2023). We infer that the spatial manifestation of old forest at-

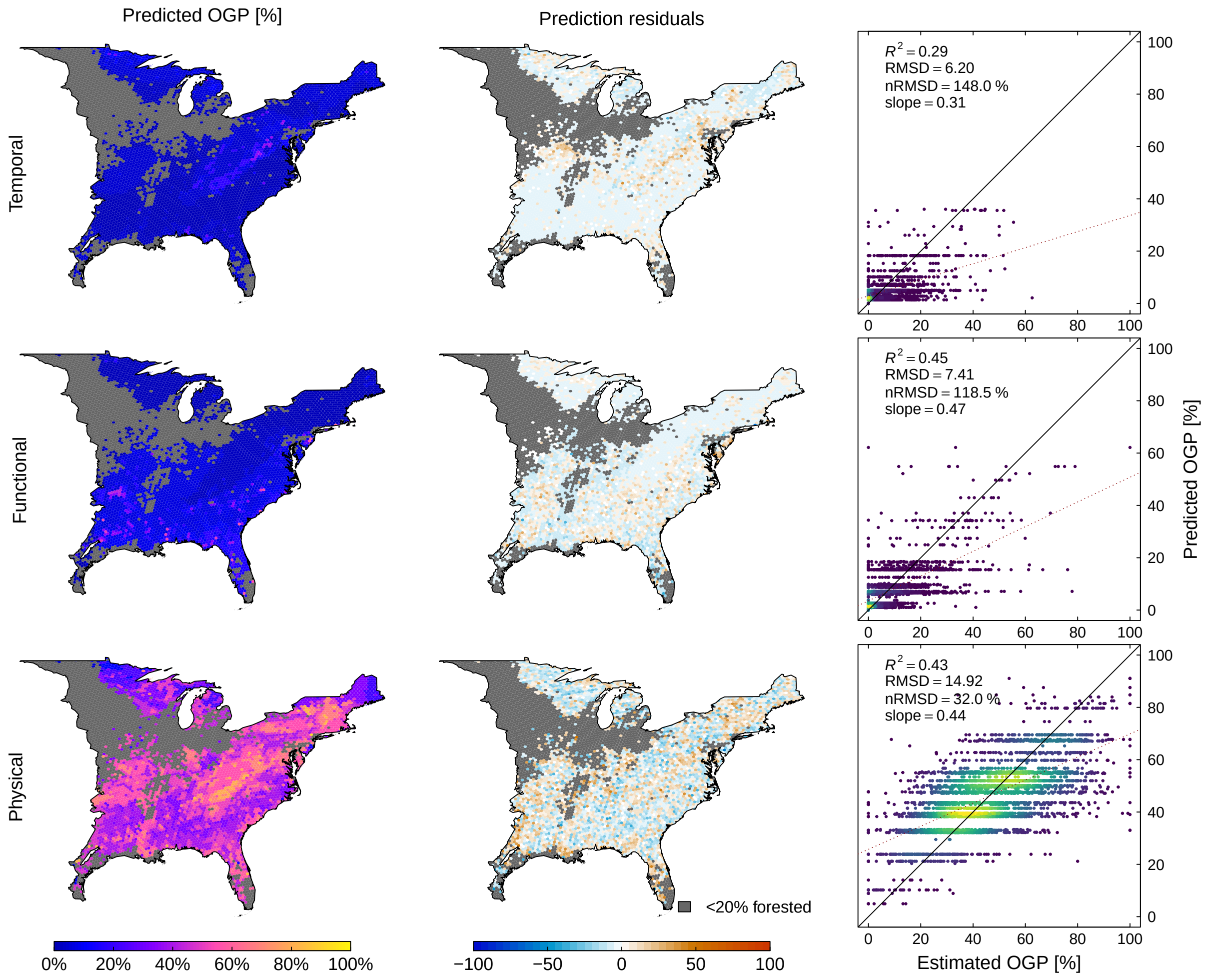

Figure 4.7: OFP model prediction and residual maps and scatter plots of predicted (from GEDI) vs. estimated (from FIA) OFP values, combined across all ecoregions for each dimension of old-growth. The OFP prediction map units are percentages, and the residual maps units are the difference in percentage points (not a percent difference) between the estimated and predicted value. The solid dashed in the scatter plots is the 1:1 line, and the dark red dotted line is the trend line between the estimated and predicted values, the slope of which is reported along with the R^2, RMSD and normalized RMSD as a percent of the mean OFP estimate.

Figure 4.8: OFP model variable importance scores aggregated by variable type and colored by biophysical attribute category, by ecoregion. The top row is a weighted measure of variable importance across all regions based on the ecoregions' relative proportions of forest.

tributes is different for each dimension; in other words, large trees are not necessarily old, and old trees are not necessarily large. This inference is supported by relatively little overlap in OFP hotspots between the dimensions (fig. 4.4) and the result that most physically old stands are not functionally or temporally old (fig. 4.5). These results demonstrate how competing definitions of old-growth are useful to characterize multifaceted old-growth conditions by mapping and contrasting the spatial patterns of old forests.

The OFP modeling exercise revealed a hierarchy with respect to old forest mapping using structure derived from remote sensing (GEDI); physical OFP maps were modeled with the most success, functional OFP maps were modeled with relatively less success, and temporal OFP maps were not modeled effectively. The mapping hierarchy supports our hypothesis that definition criteria impacts the extent to which an old-growth definition can be mapped using GEDI data within our modeling framework. We expect GEDI's height, biomass, and foliage height diversity (FHD) to be directly related to the density of large trees within a stand, so it is not surprising that the physical dimension is mapped most effectively and consistently. Relative consistency in variable importance across the regions (fig. 4.8) suggests a higher degree of structural similarity in physically old forests throughout the eastern US than for temporally or functionally old forests.

In contrast, the temporal OFP models had relatively weak predictive power and a large combined nRMSD (fig. 4.7), with considerable inter-regional variability in variable importance (fig. 4.8). From these results we infer that temporally old forests do not have a strong and consistent structural signature that is different from temporally young forests when aggregated across space and forest types, which is contrary to our expectation. We propose two related explanations for this finding. First, old trees in the eastern US come in many different shapes and sizes (Pederson, 2010). The structure of temporally old forest stands likely varies spatially within and between these ecoregions, may be specific to individual forest types, and could also be dependent on a suite of environmental covariates not considered in our analysis. Thus mixing old forest classifications across forest types within a hexagon during our estimation process may weaken the overall signal between GEDI structural attributes and old forest prevalence if the structural signatures of old forests are different across forest types. Secondly, we speculate that sensitivity to structural attributes of temporally old forests is diminished by the spatial scale of our prediction framework, as localized high resolution analyses of old-growth forests have identified numerous age-dependent relationships to remote sensing data (Falkowski et al., 2009; Kane et al., 2010; Pinto et al., 2012; Martin et al., 2021). Our estimation process produces ratio estimates of old forests that represent an expanse of forested land that is very large, and most hexagons contain very low, if any, ratios of old forest. A smaller hexagon grid would result in more variation in the OFP estimate and may help to amplify a common structural signal within temporally old forests that may by currently overwhelmed within our current framework.

The modeling results for functionally old forests were mixed in that predictive power was in between that of the physical and temporal models, but there was not consistency across ecoregions in the GEDI attributes related to functionally old forests. Despite the temporal and functional dimensions' shared usage of stand age as a definition criterion (albeit with different thresholds), GEDI variables were able to explain variations in functional OFP more than variations in temporal OFP. This suggests a structural signal is at least somewhat present within functionally old forests. An explanation for this could be that functional oldness—the age at which stand biomass accumulation is presumed to be near zero—represents the onset of a forest condition with more structural consistency than temporal oldness. A recent analysis of GEDI data colocated with FIA plot data indicated both FHD and maximum canopy height as strong indicators of carbon storage capacity on FIA plots (Crockett et al., 2023). These findings corroborate the result that FHD and canopy height are relatively effective predictors of functional oldness, given that functionally old forests should be near the maximum carbon storage capacity (Barnett et al., 2023). However, we did not find consistency in variable importance across ecoregions for functional OFP prediction, suggesting that while functional old-growth characteristics may be detectable via remote sensing, there is not same degree of structural consistency throughout functionally old forests as there is for physically old forests.

An old-growth definition characterized by near-equilibrium conditions with respect to biomass development is antithetical to views of old-growth as a dynamic and cyclical forest process that is “heterogeneously heterogeneous”, according to Pesklevits et al. (2011). Nonetheless, the functional definition is valuable for identifying forests that are approaching maximum carbon storage capacity, and we argue a diversity of viewpoints as to what constitutes old-growth may be beneficial for scientists or land managers defining old-growth for a specific use case, region, or forest type. For example, observations of late successional biomass development within temporally old northern hardwood stands show that functional oldness may not be realized until long after temporal and physical oldness (Keeton et al., 2011), while surprisingly young ages of functional old growth reported by Barnett et al. (2023) demonstrate the opposite may be true for longleaf and shortleaf pines and other predominately southern forest types (table 4.5). The spatial pattern of differences between the temporal and functional age thresholds (fig. 4.6) explains latitudinal differences between the temporal and functional OFP estimates (fig. 4.4). However, we did not directly investigate reasons for the latitudinal gradient in mean onset age of functional oldness. Yet, tree longevity and growth rates are known to be inversely related (Körner, 2017), and growth rates in lower latitude forests tend to be larger than those in northern latitude forests (Gillman et al., 2015). Together, these phenomena could explain the latitudinal gradient in mean onset of functional old-growth characteristics and differences relative to the mean temporal age thresholds (fig. 4.6).

The temporal and functional age thresholds represent fundamentally different forested conditions; the functional thresholds estimate the age at which net carbon accumulation is near zero, while the temporal thresholds seem to identify a stand age old enough to suggest the absence of recent and widespread human activity within the stand. However, the FIA's stand age variable is an imperfect measure of age structure, and stands that originated from land use transitions in the early 19th century could have a mean stand age value above the temporal oldness thresholds. The nature of the FIA stand age variable precludes a direct mapping to the age ranges which underpin the temporal thresholds, and helps explain this possibility. The comparison of temporal and functional age thresholds raises an interesting and important question; at what point should forests that long ago regenerated after human-induced land use transitions be considered within old-growth discussions? According to definitions in which widespread human activity precludes old-growth status forever, most of today's forests can never become old-growth. Conversely, Pesklevits et al. (2011) suggest that time since disturbance should matter more than the type of disturbance, as long as enough time has passed and forest complexity and heterogeneity is allowed to develop. In this context, the marriage of remote sensing data and forest simulation models that track land use transitions and forest regrowth could help identify data driven answers to this question and assist in mapping of old forest attributes (Caspersen et al., 2000; Hurtt et al., 2011; Ma, 2021).

Each dimension of forest oldness attempts to identify a forested condition that takes a long time to develop. However, physically old stands were far more common than temporally or functionally old stands, although most temporally or functionally old stands were also physically old. From this we infer that in general, physical thresholds alone are ineffective as an old-growth definition. Furthermore, an argument could be made that physically-based definition criteria actually limits the effectiveness of old-growth definitions. Old trees are not necessarily large, and definitions that reflect a Euro-American colonialist perspective may ignore other manifestations of old forests appreciated by other worldviews (Moore and Nelson, 2023). Indeed, many of today's oldest eastern forests contain trees that are relatively short and gnarled in stature, passed over for logging or clearing for agricultural due poor site quality and accessibility issues (Davis, 1996). In contrast, our modeling results demonstrate that physical oldness is most identifiable by GEDI, demonstrating old-growth definitions related to physical oldness may be more effectively mapped than definitions that do not incorporate a physically-based component. From a strictly-mapping perspective, the physical dimension of oldness as defined may provide more utility than the functional or temporal dimension. However, in the view that old-growth forests must contain very old trees, the temporal dimension of oldness is the most meaningful. In this interpretation, our results suggest a potential trade-off with respect to ecological relevance and mapability of old-growth definitions: arguably the most important part of any old-growth definition—the presence of very old trees—is also the most difficult to precisely identify across large geographic extents based on the available data and methods.

While this trade-off presents a significant challenge to old growth mapping, it also provides exciting opportunities for collaboration between ground-based forest scientists, traditional ecological knowledge experts, and remote sensing scientists. A recent review on the possibilities of mapping old-growth concluded a lack of well-defined and mappable definitions of old-growth is the largest barrier to the field (Hirschmugl et al., 2023). It is our view that the challenge of appropriately mapping old-growth requires the consideration and balance of at least four different

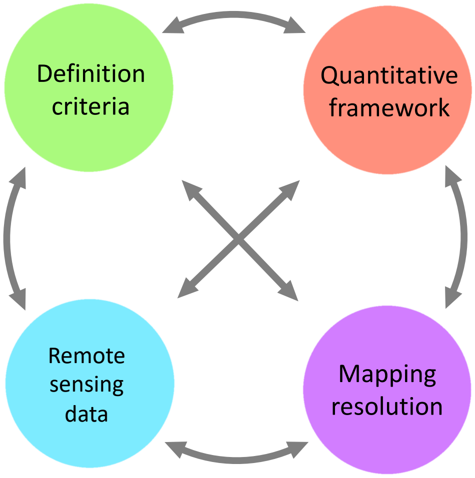

Figure 4.9: Considerations that must be balanced within old-growth mapping analyses, as each factor necessarily impacts the others. For example, old-growth definitions may be developed by blending knowledge from western forest ecologists and traditional ecological knowledge from indigenous communities, and these definitions may necessitate a specific modeling approach or spatial resolution. Alternatively, the spatial resolution of a specific old-growth definition might impact the quantitative framework used to make predictions or identify old-growth conditions, and could impact which remote sensing data types were used as predictors. These examples are not exhaustive, and are provided to demonstrate the holistic nature of old-growth mapping.

components, conceptualized in fig. 4.9: 1) clearly stated definitions, 2) remote sensing and other spatial predictor data, 3) the spatial resolution at which the old-growth conditions are mapped, and 4) the quantitative framework and statistical methods used to generate the map. In this conceptual model each component impacts the others, and thus we believe collaborative efforts may be more successful than efforts that do not appropriately account for interactions between these components. Particularly, such collaborations may be necessary to determine the appropriate spatial resolution for old-growth mapping—a factor not at all considered in our analysis. It is yet unknown what spatial resolution or resolutions are most appropriate for old-growth mapping, and whether the mapping resolution and methods should inform definition development, or if definitions should dictate the mapping resolution. Characterizing landscape-level distributions of forest attributes and establishing relationships to old-growth conditions at a meaningful spatial resolution will be critical, as the native resolution of most remote sensing instruments is almost certainly too fine to adequately characterize spatial processes in old forests that operate at the landscape scale.

4.6 Acknowledgements

We thank the USFS Mature and Old-Growth Inventory Technical Team for their willingness to discuss old-growth definition development and for a working relationship that inspired the ideas in this chapter. We also thank The Wilderness Society's Kevin Barnett and Greg Aplet for meaningful discussions related to old-growth definitions and related considerations.

This work was supported by funding from the Global Ecosystem Dynamics Investigation (NASA contract NNL15AA03C), NASA FINESST Grant (80NSSC21K1626), and the Moore Foundation.

4.7 Tables

Table 4.1: Old-growth forest types and associated temporal and physical dimension thresholds, adapted from Pelz et al. (2023) and harmonized across the eastern and southern regions for consistency.

Short Name	Long Name	Region	FORTYPCD	STDAGE	TPA	DBH
Xeric Oak (N)	Dry oak	north	162, 163, 165, 167, 182, 184, 404, 405, 501, 502, 506, 507, 509, 510, 513, 515	100	20	16
Sub-boreal Spruce/Fir	Sub-boreal spruce/fir	north	122, 125	140	10	12
N. Pine	Northern pine	north	101, 102, 103	100	20	12
Beech/Maple/Basswood	Beech maple basswood	north	805	140	10	16
Wetland hardwood	Wetland hardwood	both	600, 608, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 809, 962	120	10	18
Other	other	both		100	10	14
N. hardwood	Northern hardwood	both	517, 520, 800, 801, 802, 809	120	10	15
Montane Spruce/Fir	Montane spruce and spruce-fir	both	121, 123, 124, 128, 129, 902	120	10	15
Mesic Oak	Dry-mesic oak (S) + mesic northern oak (N)	both	500, 502, 503, 504, 505, 510, 511, 512, 515, 516, 519, 962	140	6	20
Conifer N. hardwood	Conifer northern hardwood	both	104, 105, 123, 124, 400, 401, 902	140	10	16
Xeric Pine, Pine/Oak	Xeric pine and pine-oak forest and woodland	south	162, 163, 165, 167, 171, 400, 402, 404, 405, 409, 500, 510, 514	100	10	10
Xeric Oak (S)	Dry and xeric oak forest, woodland, and savanna	south	500, 501, 502, 510, 514, 515, 519, 600, 962	90	10	8
S. wet Pine	Southern wet pine forest, woodland, and savanna	south	141, 142, 166, 400, 407	80	10	9
Coastal plain hardwood	Coastal plain upland mesic hardwood forest	south	600, 962	120	10	18
Seasonally wet hardwood	Seasonally wet oak-hardwood woodland	south	500, 504, 520, 600, 962	100	10	16
Dry-mesic Oak/Pine	Dry and dry-mesic oak-pine forest	south	161, 162, 163, 400, 404, 405, 406, 409	120	10	15
Mesophytic	Mixed mesophytic and western mesophytic forest	south	500, 506, 511, 516, 517, 800, 801, 805, 962	140	10	22
Longleaf/Slash Pine	Upland longleaf and south Florida slash pine forest	south	141, 400, 403	80	10	13
Floodplain hardwood	River floodplain hardwood forest	south	500, 508, 600, 601, 602, 605, 705, 706, 708, 962	100	10	14
Eastern riverfront	Eastern riverfront forest	south	600, 700, 702, 703, 704, 705, 709, 962	100	10	20
Cypress/Tupelo swamp	Cypress-tupelo swamp forest	south	607, 609	120	10	8

Table 4.2: OFP estimates and 95% confidence intervals for each dimension of old-growth, and all possible dimensional combinations, for all eastern forests and by ecoregion.

Ecoregion	Temporal		Functional		Physical		T&F		T&P		F&P		All	
	OFP	CI	OFP	CI	OFP	CI	OFP	CI	OFP	CI	OFP	CI	OFP	CI
all	4.64	0.000733	6.51	0.00084	45.2	0.0017	1.07	0.000359	3.54	0.000648	5.13	0.000758	0.838	0.00032
5.23	5.81	0.00302	1.73	0.00169	37.5	0.00626	1.38	0.00152	3.27	0.00232	1.05	0.00133	0.803	0.00116
8.12	3.37	0.00537	1.60	0.00376	40.2	0.0146	0.883	0.00285	2.10	0.00435	1.11	0.00318	0.632	0.00242
8.3	2.20	0.00135	10.1	0.00273	44.2	0.00449	0.788	0.000825	2.02	0.00131	8.27	0.00253	0.761	0.000816
8.4	8.71	0.00428	6.00	0.00344	56.9	0.00726	0.661	0.00122	7.59	0.00402	5.09	0.00322	0.587	0.00116
8.5	4.13	0.00797	13.7	0.0136	50.1	0.0197	2.54	0.00632	3.39	0.00727	9.88	0.0119	2.090	0.00578
9.2	4.00	0.0604	1.69	0.0387	41.4	0.150	0.59	0.0249	2.85	0.0529	0.933	0.0297	0.407	0.0210

Table 4.3: Stand-level old-growth classification rates for each dimension of old-growth and all possible combinations, grouped by all eastern forests, ecoregion, USFS old-growth community type from Pelz et al. (2023), and FIA forest type group from Barnett et al. (2023).

	group	source	Temporal	Functional	Physical	T&P	T&F	F&P	All
1	all	NA	4.2	6.44	40.02	3.02	0.95	4.71	0.72
2	5.23	ecoregion	5.43	1.55	31.92	2.84	1.23	0.92	0.7
3	8.12	ecoregion	3.3	1.48	34.81	1.82	0.76	0.95	0.53
4	8.3	ecoregion	1.96	9.74	39.29	1.71	0.69	7.32	0.63
5	8.4	ecoregion	7.76	6.43	51.52	6.57	0.62	5.02	0.52
6	8.5	ecoregion	3.89	13.46	46.13	3.03	2.38	9.25	1.88
7	9.2	ecoregion	3.84	1.76	35.75	2.28	0.48	0.84	0.28
8	Beech/Maple/Basswood	Pelz 2023	0.18	0.09	42.98	0.14	0.05	0.05	0.05
9	Conifer N. hardwood	Pelz 2023	3.97	11.17	63.26	3.67	3.16	10.21	3.01
10	Cypress/Tupelo swamp	Pelz 2023	4.41	4.53	87.34	4.3	1.39	4.41	1.28
11	Eastern riverfront	Pelz 2023	6.52	0	41.3	6.52	0	0	0
12	Floodplain hardwood	Pelz 2023	2.07	0.97	51.57	1.78	0.6	0.89	0.55
13	Longleaf/Slash Pine	Pelz 2023	11.28	1.03	40.51	9.74	1.03	1.03	1.03
14	Mesophytic	Pelz 2023	0.22	0.51	14.16	0.15	0.07	0.22	0
15	Mesic Oak	Pelz 2023	0.7	0.85	39.95	0.6	0.38	0.74	0.33
16	Dry-mesic Oak/Pine	Pelz 2023	0.06	15.53	27.77	0.03	0.05	10.41	0.02
17	Montane Spruce/Fir	Pelz 2023	2.39	1.63	16.74	1.1	1.63	0.91	0.91
18	N. hardwood	Pelz 2023	2.14	0.36	43.88	1.84	0.34	0.31	0.29
19	N. Pine	Pelz 2023	7.43	0.64	42.59	5.29	0.64	0.49	0.49
20	Other	Pelz 2023	7.69	3.7	22.93	3.56	1.78	1.82	1.01
21	Sub-boreal Spruce/Fir	Pelz 2023	1.54	1.98	14.36	0	1.54	0	0
22	Seasonally wet hardwood	Pelz 2023	20.27	0.74	62.43	14.28	0.74	0.59	0.59
23	Wetland hardwood	Pelz 2023	1.65	0.51	20.31	0.55	0.29	0.2	0.07
24	S. wet Pine	Pelz 2023	8.75	33.4	68.79	8.13	8.19	30.6	7.62
25	Xeric Oak (N)	Pelz 2023	23.55	12.7	22.91	9.73	2.37	1.13	0.76
26	Xeric Oak (S)	Pelz 2023	24.29	0.29	91.78	23.87	0.29	0.29	0.29
27	Xeric Pine, Pine/Oak	Pelz 2023	3.81	35.74	75.13	3.56	1.77	30.59	1.67
28	Aspen/Birch	Barnett 2023	0.99	0	17.88	0.74	0	0	0
29	Elm/Ash/Cottonwood	Barnett 2023	2.03	0.36	27.21	0.78	0.3	0.11	0.08
30	Fir/Spruce/Mountain Hemlock	Barnett 2023	0	0	0	0	0	0	0
31	Loblolly/Shortleaf Pine	Barnett 2023	0.65	26.93	31.32	0.43	0.65	18.1	0.43
32	Longleaf/Slash Pine	Barnett 2023	8.29	34.78	67.64	7.67	8.29	31.89	7.67
33	Maple/Beech/Birch	Barnett 2023	2.49	0.31	50.43	2.22	0.3	0.28	0.28
34	Oak/Gum/Cypress	Barnett 2023	2.45	2.08	49.35	2.2	0.99	1.85	0.89
35	Oak/Hickory	Barnett 2023	5.94	0.66	43.37	4.62	0.4	0.52	0.3
36	Oak/Pine	Barnett 2023	2.67	5.6	52.86	2.29	0.97	5.03	0.9
37	Spruce/Fir	Barnett 2023	13.45	4.25	19.18	5.16	4.13	1.96	1.96
38	Tropical Hardwoods	Barnett 2023	2.56	0	38.46	1.28	0	0	0
39	White/Red/Jack Pine	Barnett 2023	7.93	1.97	48.41	6.05	1.73	1.79	1.55

Table 4.4: Stand-level rates of overlapping old-growth classifications (columns with '&'), along with baseline rates of classification for each dimension. For example, while 4.2% of all plots were classified as temporally old, 14.8% of functionally old plots were classified as temporally old, 7.6% of physically old plots were classified as temporally old, and 15.3% of functionally and physically old plots were classified at temporally old.

Region	Temporal	T&F	T&P	T&FP	Functional	F&T	F&P	F&TP	Physical	P&T	P&F	P&TF
all	4.20	14.8	7.56	15.3	6.44	22.6	11.8	23.8	40.0	72.0	73.1	75.7
5.23	5.43	79.6	8.89	76.3	1.55	22.7	2.87	24.7	31.9	52.3	59.1	56.7
8.12	3.30	51.5	5.22	56.5	1.48	23.1	2.72	29.4	34.8	55.1	63.9	70.0
8.3	1.96	7.07	4.35	8.67	9.74	35.1	18.6	37.1	39.3	87.1	75.2	92.2
8.4	7.76	9.61	12.8	10.3	6.43	7.96	9.74	7.89	51.5	84.7	78.1	83.9
8.5	3.89	17.7	6.56	20.3	13.5	61.1	20.1	62.1	46.1	77.8	68.8	79.1
9.2	3.84	27.3	6.38	33.3	1.76	12.5	2.35	12.3	35.8	59.4	47.7	58.3

Table 4.5: Mean temporal and functional old-growth age thresholds and the mean and median difference defined as the temporal threshold minus the functional threshold, by ecoregion, and both Pelz et al. 2023 and Barnett et al. 2023 groupings. Values were rounded to the nearest year, resulting in occasional discrepancies in the mean difference by one year.

group	source	Temporal	Functional	mean difference	median difference
all	NA	117	115	1	-10
5.23	ecoregion	117	147	-30	-24
8.12	ecoregion	123	138	-15	-15
8.3	ecoregion	115	89	26	26
8.4	ecoregion	119	124	-5	-13
8.5	ecoregion	106	89	17	19
9.2	ecoregion	117	138	-21	-19
Beech/Maple/Basswood	Pelz 2023	140	146	-6	-4
Conifer N. hardwood	Pelz 2023	140	114	26	22
Cypress/Tupelo swamp	Pelz 2023	120	121	-1	-15
Eastern riverfront	Pelz 2023	100	128	-28	-39
Floodplain hardwood	Pelz 2023	100	118	-18	-17
Longleaf/Slash Pine	Pelz 2023	80	111	-31	-21
Mesophytic	Pelz 2023	140	126	14	8
Mesic Oak	Pelz 2023	140	133	7	8
Dry-mesic Oak/Pine	Pelz 2023	120	50	70	76
Montane Spruce/Fir	Pelz 2023	120	140	-20	-14
N. hardwood	Pelz 2023	120	140	-20	-24
N. Pine	Pelz 2023	100	129	-29	-38
Other	Pelz 2023	100	149	-49	-78
Sub-boreal Spruce/Fir	Pelz 2023	140	130	10	6
Seasonally wet hardwood	Pelz 2023	100	133	-33	-32
Wetland hardwood	Pelz 2023	120	132	-12	-19
S. wet Pine	Pelz 2023	80	47	33	38
Xeric Oak (N)	Pelz 2023	100	125	-25	-32
Xeric Oak (S)	Pelz 2023	90	140	-50	-42
Xeric Pine, Pine/Oak	Pelz 2023	100	78	22	26
Aspen/Birch	Barnett 2023	103	178	-75	-78
Elm/Ash/Cottonwood	Barnett 2023	116	133	-17	-19
Fir/Spruce/Mountain Hemlock	Barnett 2023	100	428	-328	-326
Loblolly/Shortleaf Pine	Barnett 2023	116	41	76	76
Longleaf/Slash Pine	Barnett 2023	80	42	38	38
Maple/Beech/Birch	Barnett 2023	124	147	-23	-24
Oak/Gum/Cypress	Barnett 2023	109	116	-6	-15
Oak/Hickory	Barnett 2023	123	133	-9	-12
Oak/Pine	Barnett 2023	112	99	13	19
Spruce/Fir	Barnett 2023	118	127	-9	3
Tropical Hardwoods	Barnett 2023	100	146	-46	-46
White/Red/Jack Pine	Barnett 2023	106	130	-23	-18

Table 4.6: OFP model bias, defined as the mean difference between estimated and predicted OFP, combined across ecoregions for each dimension. The values are in percentage points, so a value of 14.2 means the models under-predicted OFP relative to the estimated value by 14.2 points. Bias is reported for all predictions in each dimension (top row), as well as segmented by ranges of estimated OFP to assess model bias as a function of estimated OFP. The number of model data points (hexagons) within each range is represented by N for each dimension.

OFP range	Temporal		Functional		Physical	
	Bias	N	Bias	N	Bias	N
0-100%	0.0167	3941	-0.257	3941	0.313	3941
0-10%	-1.8	3347	-2.59	3059	-21	107
10-20%	7.76	427	4.48	556	-16	224
20-30%	14.2	107	11	176	-12.3	485
30-40%	18	36	12.4	86	-6.19	738
40-50%	22.3	19	18.3	36	-0.76	738
50-60%	28.2	4	20.2	16	4.92	636
60-70%	60.5	1	36.7	5	10	492
70-80%	-	-	35	6	16.5	323
80-90%	-	-	37.9	1	30.7	52
90-100%	-	-	-	-	21.5	146

Chapter 5: Conclusion

This chapter begins with a summary and synthesis of the main discoveries from chapters 2-4 (fig. 5.1). I then discuss the limitations of my research and the related challenges. I conclude with recommendations for future work based on the interpretations of my findings.

5.1 Synthesis

Understanding changes within eastern forests is of great interest to the scientific community. This dissertation explores the capabilities of GEDI waveform lidar observations for inference into structural and successional dynamics. This work has helped clarify several limitations to characterizing forest dynamics using GEDI data. More importantly, I discovered ways in which GEDI can be integrated with forest inventory data and ecosystem modeling that enhanced GEDI's capacity for ecological inference. The following is a summary of my scientific discoveries, their synthesis, and the associated implications.

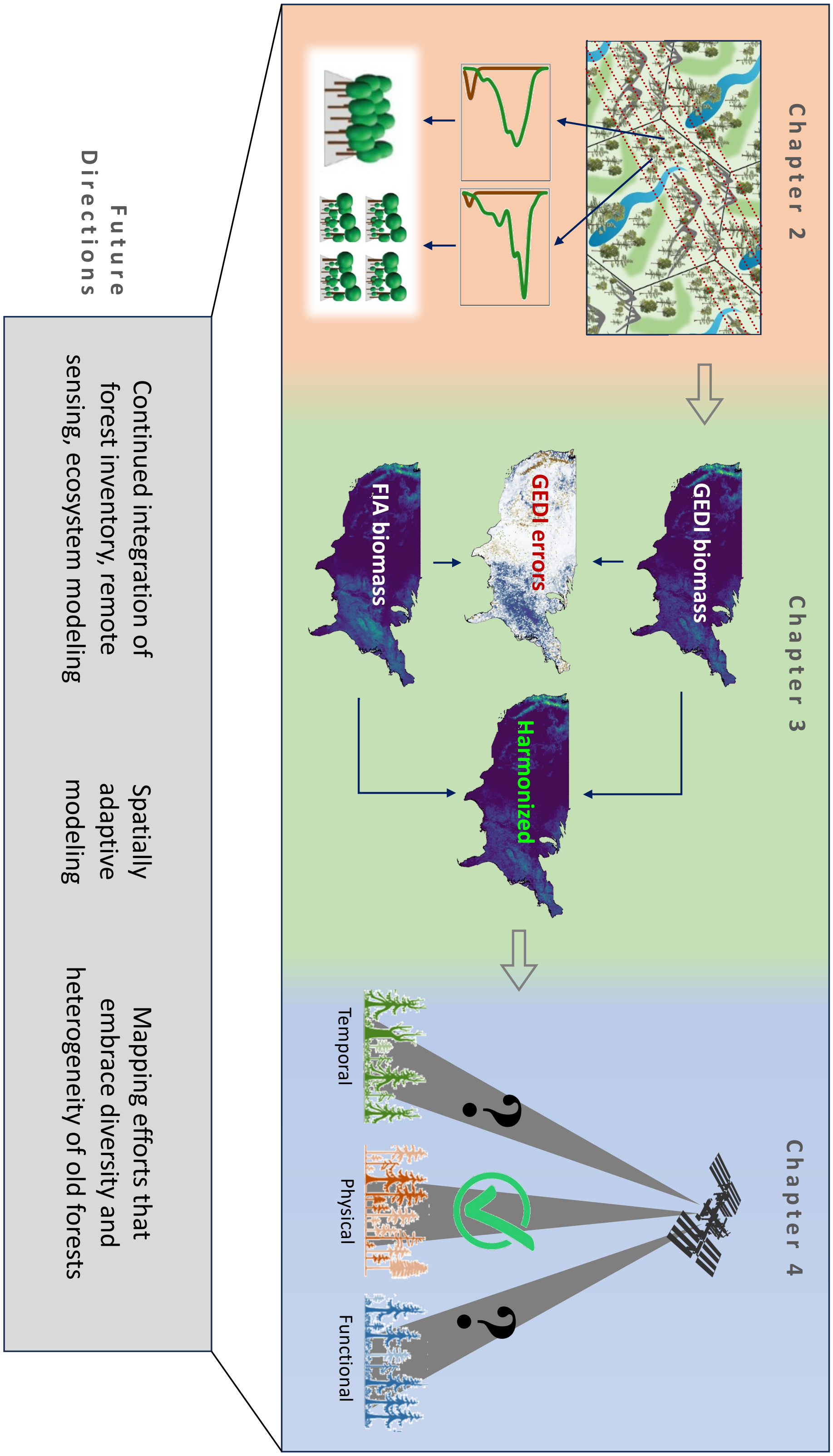

Figure 5.1: Summary of how the analyses and results from chapters 2-4 fit together and inform future research.

Chapter 2 considers the possibility that GEDI waveform shapes may not have a unique, one-to-one relationship with aboveground biomass density. I propose the concept of non-uniqueness, which I defined as the range of potential biomass values a given waveform shape is associated with. I found that some waveforms are associated with very small ranges of potential biomass, while others were associated with larger ranges. The discovery of different amounts of non-uniqueness for GEDI waveforms suggests a limit to the certainty of biomass predictions that is independent from traditional methods of quantifying regression uncertainty. I also discovered that the successional stage of a forest stand impacts the extent to which its waveform shape has a non-unique relationship with biomass. Stands further along in ecological succession (typically with higher biomass and more variation in tree heights and diameters) produced waveforms with a larger magnitude of non-uniqueness than early-successional stands (those with lower biomass and less variation in tree sizes). This suggests that the precision of GEDI's biomass predictions is dependent on the successional stage of the forests it observes.

Given the uncertainties in biomass for individual waveforms discovered in chapter 2, I then explored GEDI's ability to accurately characterize biomass distributions at large scales in chapter 3. I discovered that systematic differences between GEDI and FIA biomass maps were due to misspecification of GEDI's biomass models. This discovery suggests that the relationship between forest structure and biomass varies spatially, and that the difficulty in appropriately mapping landscape-scale gradients in biomass with GEDI's current prediction framework is due to a lack of spatially representative training data. I then explored methods to mitigate this limitation, and discovered that a relationship between GEDI forest structure metrics and biomass can be characterized in a manner that is scale-invariant. This discovery enabled calibration of GEDI biomass models using FIA data at an aggregate spatial scale, and resulted in unbiased models and predictions at the footprint scale that were harmonized with FIA plot level biomass. I used these models to produce unbiased biomass estimates at multiple spatial scales. This work demonstrates the value of fusion between remote sensing and forest inventory data; the resultant GEDI-based maps gain the quality of being unbiased through harmonization with FIA, while GEDI's sampling density enables high estimate precision for large scale estimates (hexagon), as well as maps at a much higher spatial resolution (1 km) than possible with FIA data alone.

The discoveries from chapters 2 and 3 relate to the scale-dependent variability in biomass with respect to forest structure. Here, a noisy relationship is the same as non-uniqueness, as similar measurements of forest structure could come from stands with different values of biomass. In chapter 2, I demonstrated that non-uniqueness increases with forest biomass and is unavoidable at the scale of GEDI footprints. I suggested that non-uniqueness may explain the heteroscadasticity common to linear regression biomass models (see Zolkos et al., 2013), and speculated that it may decrease with increasing spatial scale. Evidence from chapter 3 supports this speculation. Individual GEDI measurements of forest structure and FIA plot level biomass values were substantially more variable than their spatially aggregated counterparts that were six orders of magnitude larger than GEDI footprints (Appendix B). Spatial aggregation increased the consistency of the structure to biomass relationship by averaging out the non-uniqueness effect, and there was minimal heteroscadasticity during SISA model calibration (figs. B.5 and B.6). From these results I infer that the potential for non-uniqueness at large spatial scales is less than at the scale of a GEDI footprint.

Acknowledging the limitations of individual GEDI waveforms within late successional forests (chapter 2), I leveraged the improvements to landscape level biomass characterization (chapter 3) within an experiment to analyze old-growth forests at large spatial scales in chapter 4. Old-growth forests are a critically important component of forest development because they are the manifestation of many years of forest change, yet manifold definitions of old-growth have complicated efforts to identify and protect them. I demonstrated that definition criteria has a large impact on the amount and spatial distribution of old-growth estimated throughout the eastern US. I found evidence supporting my hypothesis of a mapping hierarchy with respect to old-growth definitions, using GEDI measurements of forest structure to model estimates of old-growth proportion at coarse spatial scales. Old-growth estimates based on a physical definition were predictable and had a clearer and more consistent relationship with GEDI forest structure variables than estimates from either the temporal or functional old-growth definitions. It was somewhat surprising that neither temporal or functional old-growth had structural signatures detectable by GEDI, indicating more structural variability in temporal and functional old-growth relative to physical old-growth.

Footprint-level biomass predictions from chapter 3 were effective predictors of old-growth in chapter 4 (fig. 4.8). This finding was consistent for physical old-growth throughout the entire eastern US. Biomass and FHD were meaningful predictors of temporal old-growth in the Northern Forests and Central Mixwood Plains regions. This result supports an empirically based theory of late successional biomass development for northern hardwood forests in the northeast from Keeton et al. (2011). The finding that increased biomass and stand complexity (FHD) are associated with temporal old-growth suggests forests that are not yet temporally old have not reached their maximum possible carbon storage capacity. This inference is in agreement with the conclusions of Rhemtulla et al. (2009), in that biomass still has potential for increase throughout secondary forests initiated by land use transitions. Furthermore, the finding that there is very little functional old growth throughout the Northern Forests and Central Mixwood Plains regions is more evidence of biomass accumulation over relatively long timescales, as the age thresholds for functional old-growth are larger than those for temporal old-growth in these regions.

The finding of non-uniqueness in late successional stands is relevant to old-growth theory. Pesklevits et al. (2011) argue that old-growth is ‘heterogeniously heterogenious’, a statement supported by the results from chapters 2 and 4. Old-growth is structurally heterogeneous at multiple spatial scales; gap dynamics and large variations in tree ages and sizes drive structural heterogeneity at small spatial scales (Runkle, 1981; McEwan et al., 2014), while large scale structural heterogeneity is indicated by the variable importance results in chapter 4. The apparent relationship between temporally old, late-successional stands and non-uniqueness adds yet another dimension of heterogeneity to these forests with respect to biomass and waveform lidar, in that highly similar waveform shapes from a structurally complex old growth forest could be associated with very different biomass values.

5.2 Research challenges and limitations

Here I discuss the limitations of my research, and what can and cannot be inferred from my discoveries. Note, these limitations are different from the limitations associated with using GEDI to analyze forest dynamics, which are the primary focus of this dissertation.

The discovery of non-uniqueness presents a challenge with respect to aboveground biomass density prediction and estimation from waveform lidar. My findings suggest non-uniqueness is inherent to waveform lidar and the relationship to biomass at high spatial resolutions. A practical limitation of this work is that characterizing the extent of non-uniqueness for different forest types or locations is unlikely without performing a waveform matching analysis with a simulated database of potential forest structure. It is possible that environmental covariates impact the likelihood or magnitude of non-uniqueness, however this was not explored in chapter 2. Aggregating GEDI data to coarser spatial resolutions may decrease the limitations associated with non-uniqueness, while maintaining a relatively high spatial resolution for forest mapping using spaceborne data.

I also demonstrated that, despite being a statistically valid assessment of prediction uncertainty based on the calibration data, OLS prediction intervals do not adequately characterize prediction uncertainty associated with non-uniqueness (fig. 2.5). However, there was no acknowledgement of non-uniqueness driven uncertainty in the prediction framework used in chapter 3. At present, there is no way to quantify or validate uncertainty from non-uniqueness at scale for any given location, or to incorporate such information within GEDI's uncertainty estimation framework. However, when using GEDI data at aggregated spatial scales (e.g. 1 km mapping), non-uniqueness is likely less impactful than it is for a individual waveform's biomass prediction.

The conceptual forest oldness framework that underpins chapter 4 simplifies how forests change over time. First, old-growth dimensions are characterized by a suite of forest attributes, and representation by a single attribute reduces the complexity of each dimension. Second, change in any given forest attribute is not necessarily monotonic. For example, a stand's mean age decreases when the oldest pioneer species dies, yet the tree's death may actually progress succession by initiating a growth release in advance regeneration that enables shade tolerant species to takes its place in the canopy. Here, an incremental decrease in mean stand age coincides with an incremental increase in successional development. Third and most important, binary classification of a continuous variable simplifies the process it represents. The classification approach taken in chapter 4 was necessary given the existing old-growth definition framework that was the focus of the analysis, but it may have decreased the capacity to infer structural properties associated with temporal and functional old-growth. Classification runs contrary to current theories of old-growth as non-normative and heterogeneous. Thus, the inferential capacity of predicting landscape estimates of old forest proportion may be limited by the mismatch between old-growth theory and the prediction framework.

5.3 Conclusions and future implications

Variability is a theme to the discoveries in this dissertation. Non-uniqueness is variability in the relationship between biomass and forest structure measured by waveform lidar, and its likelihood changes with spatial scale. When spatial scale is held constant, relationships between GEDI forest structure variables and biomass were shown to change across space. Temporally and functionally old forests have diverse structural signatures within and between ecoregions, to the extent that mapping them with GEDI data is ineffective within a spatially coarse modeling framework.

While variability in forests is not a new concept, it highlights a dilemma in forest remote sensing. A major benefit of remote sensing is the ability to gather information on forests without physically visiting every pixel or observation location. However, without *a posteriori* knowledge of variability, it at best introduces limitations to what can be inferred from remotely sensed information, and at worst it could violate assumptions, bias models, and invalidate predictions. The following are my suggestions for future work aimed at improving GEDI's inferential capacity in light of this challenge.

Future work involving biomass prediction via remote sensing at large spatial scales should develop modeling frameworks that explicitly account for spatial variation in relationships. Local and regional implementation of spatially varying coefficient models or other spatial process models have shown promise and may decrease the potential for spatially varying model bias (Babcock et al., 2015; Datta et al., 2016; Taylor-Rodriguez et al., 2019). The field would especially benefit from development of these methods at continental or global scales through integration with NFI data. Explicit acknowledgement of space as a covariate may also help simplify aspects of biomass modeling frameworks. A spatially adaptive model (e.g. May et al., in review) may take the place of regional stratification and prediction in each strata using a different, spatially static model (e.g. Duncanson et al., 2022).

Old-growth mapping efforts are increasingly using remote sensing data, and emphasis should be placed on generating a diverse stack of remotely sensed predictors that incorporate other forest attributes in addition to vertical structure (Hirschmugl et al., 2023). Variability in old-growth structure, function, and composition should recognized within analytic frameworks, and mapping efforts should resist the temptation to classify old forests. Analyses that consider continuous variables and heterogeneity in forest attributes may be more successful than methods that reduce variation and complexity and introduce arbitrariness (Gray et al., 2023). Mapping analyses should characterize distributions of forest attributes from high-resolution remote sensing data at coarser landscape scales that are more relevant to old-growth processes. Such efforts will retain more information about the complexity and heterogeneity of forest attributes, which is necessary for inference into old forest conditions and processes. Furthermore, analyses of old-growth that incorporate ecosystem modeling constrained by inventory data and remote sensing hold great promise, as a calibrated simulation environment is well suited for investigating the many and diverse manifestations of forest change over very long time periods.

Much remains to be learned about how forest properties change over time. From a scientific perspective old forests provide opportunities to advance knowledge of forest processes and potential future conditions. From a human perspective old forests offer an extraordinary connection to our past and future. Future investigations of successional dynamics and the development of old forest characteristics must integrate forest inventories, remote sensing, and ecosystem modeling. The research presented here provides examples of how these areas of research may be integrated, and the results enable new insight into patterns of forest structure and how it changes over time.

Appendix A: Appendix to Chapter 2

A.1 FORMIND calibration

We calibrated a regional version of FORMIND to represent the tree geometry and stand structure for Northeastern US forests using data from the United States Forest Service Forest Inventory and Analysis (FIA) program, accessed with the rFIA software package (Stanke et al., 2020). Using FIA data we segmented the region's 27 most abundant tree species (based on total species basal area) into 9 different plant functional types (PFTs) (table A.1). Together these 27 species account for upwards of 85% the region's total basal area (Iverson et al., 2008). We calibrated the suite of allometric equations required by FORMIND for tree construction for each PFT using FIA measurements, and we calibrated tree biomass as a function of diameter according to the national-scale allometric equations from Jenkins et al. (2003). As the FF simulates forest structure, not growth, of trees, we only needed to calibrate physical tree geometry equations, which are straightforward. Calibrations of growth rates, photosynthesis, and other dynamic processes are more involved, but were not necessary for this analysis, and ignored.

A.2 Forest Factory

The Forest Factory (FF) simulates unique, 20 m x 20 m stem maps, and the structure of these stem maps is then generated by FORMIND in the same manner as the *in situ* stem maps. The purpose of the FF is to simulate the diversity in forest structure within a region, based on varying the PFT compositions and stem-size distributions across a large number of 20 m x 20 m simulated forest stands. During the construction of a forest stand, the FF iteratively adds trees to the plot by sampling from one of several predefined stem size distributions, and assigns each tree to a PFT. With the addition of each tree the FF checks to make sure there is space and resources (light, water, etc.) available to accommodate the new tree. In this way the FF ensures that stands are not unrealistically crowded with trees, and could theoretically exist in reality. For more on the Forest Factory see Bohn and Huth (2017).

The FF outputs 100 simulated stands at once, arranged in a 4 ha, 10 x 10 grid of 20 m x 20 m stands. While each stand is generated independently from the others, it is output directly next to other stands in the simulation space, and thus the crowns of trees near the edges of each stand extend over the plot boundaries into the adjacent stands. To avoid edge effects that would otherwise alter the waveform shape of each stand, we implemented the following process for every simulated stand in the FF database. We generated each FF stand's LiDAR point cloud with periodic boundaries, meaning that the tree crowns which extend over a plot boundary enter back into the plot on the opposite edge and do not enter into the adjacent stands, following the procedure of Knapp et al. (2021). This point cloud was then replicated into a 3 x 3 grid, so that when positioned directly to each other the periodic boundaries align and the nine 20 m x 20 m point clouds represent a homogeneous 60 m x 60 m point cloud with structural characteristics identical to the original 20 m x 20 m stand. The GEDI waveforms were then simulated in the center of this expanded point cloud, so that no tree crown influenced the waveform that was not accounted for in the stem map.

Table A.1: The 27 most abundant tree species within the Northeast US are grouped into nine PFTs based on common structural characteristics and shade tolerances. Maximum height and diameter values were derived by assessing the distributions of each variable for each PFT using all the most recent FIA data for the Northeast USA. N indicates the number of samples used for each PFT to calibrate the tree geometry equations needed in FORMIND (Fischer et al., 2016).

PFT	Name	Scientific	Common	Tolerance	Maximum height	Maximum diameter	N
1	fir-cedar	Abies balsamea Thuja occidentalis Picea mariana	balsam fir northern white-cedar black spruce	tolerant	35	0.87	71544
2	hemlock-spruce	Tsuga canadensis Picea rubens Picea glauca Picea abies	eastern hemlock red spruce white spruce Norway spruce	tolerant	43	1.11	67419
3	pine	Pinus strobus	eastern white pine	intermediate	54	1.60	21362
4	maple-beech	Acer saccharum Fagus grandifolia Tilia americana	sugar maple American beech American basswood	tolerant	45	1.34	56484
5	maple-ash	Acer rubrum Fraxinus americana Prunus serotina Fraxinus pennsylvanica	red maple white ash black cherry green ash	intermediate	45	1.13	66199
6	oak	Quercus rubra Quercus alba Quercus velutina Quercus prinus	northern red oak white oak black oak chestnut oak	intermediate	46	1.57	12243
7	birch	Betula alleghaniensis	yellow birch	intermediate	40	1.28	19362
8	birch-aspen	Betula papyrifera Populus tremuloides Betula lenta Populus grandidentata	paper birch quaking aspen sweet birch bigtooth aspen	intolerant	39	0.93	26622
9	understory	Acer pensylvanicum Ostrya virginiana Acer spicatum	striped maple eastern hophornbeam mountain maple	tolerant	29	0.50	5023

A.3 OLS method and results

We did not test all the possible combinations of model equations that could be used to predict AGBD from waveform RH metrics in 10m increments (RH10 - RH90 and RH98). Instead, we developed a set of 18 candidate models; each model contained one of three different upper-canopy metrics (RH98, RH90, RH80) and one of three different mid-canopy metrics (RH50, RH40, RH30), as the combination of upper and lower canopy metrics has been effective in modeling AGBD from GEDI waveforms (Duncanson et al., 2020, 2022). The first nine models were all the possible combinations between these groups, and the second nine models were the same as the previous and also included the interaction term between the upper- and mid-canopy variables. Performance of the top five candidate OLS models was similar when tested on the 50% of FF stands that were not used in model training, and final model selection (fig. A.1) was based on the lowest nRMSE value (table A.2). Once the final model was selected, we performed the back transformation on the fitted values and 90% prediction interval to obtain the predicted values and prediction interval in meaningful units (Mg ha^{-1}).

A.4 WFM method and results

Forty eight stem map waveforms did not have at least 100 FF waveforms with an $r >$ 0.75. The relative overlap algorithm is highly sensitive to slight differences in waveform shape, and we intentionally set this strict criteria (100 matches with $r >$ 0.75) to ensure we only analyzed stands with 100 similar waveforms. Waveform shape is also highly sensitive to slight differences in forest structure within a plot, and it was our expectation that all 428 plots may not have 100 matches. We tested various r thresholds and the number of matches required for each *in situ* stem map, and arrived at this combination as a balance between a strict enough matching requirement to ensure a high degree of similarity between the matched waveforms, and not being too restrictive and eliminating more *in situ* plots from the analysis. Tests using at least 50, 100, 200 and 500 did not yield very large differences in AGBD prediction or uncertainty, and 100 was chosen to include enough FF matches to ensure a robust AGBD distribution, while keeping the number of matches relatively small to ensure only the waveforms with the highest degree of similarity were used.

A.5 Stem maps and biomass

The 380 *in situ* stem maps with at least 100 best matches had AGBD values that ranged from 2.2 Mg ha^{-1} to 436.1 Mg ha^{-1}, with a mean and median of 171.2 Mg ha^{-1} and 169.7 Mg ha^{-1} respectively (fig. A.4A). The database of FF stands had AGBD values that ranged from 0.01 Mg ha^{-1} to 2713.6 Mg ha^{-1}, with a mean and median of 469.0 Mg ha^{-1} and 387.8 Mg ha^{-1} (fig. A.4C). As a whole, the 38,000 forest stands that were identified as best matches (100 matches across 380 sites) had biomass values that ranged from 0.3 Mg ha^{-1} to 1103.6 Mg ha^{-1},

Table A.2: Top five OLS models, sorted by nRMSE. In the formulas, y is the response variable, AGBD. The reported nRMSE and R^2 were calculated from the back-transformed predictions. Model 15 was selected as the final model.

Model	**nRMSE (%)**	**R^2**	**Formula**
15	32.08	0.829	$\sqrt{y}$ ~rh98 + rh40 + rh98*rh40
12	32.19	0.828	$\sqrt{y}$ ~rh98 + rh50 + rh98*rh50
14	32.47	0.825	$\sqrt{y}$ ~rh90 + rh40 + rh90*rh40
11	32.56	0.825	$\sqrt{y}$ ~rh90 + rh50 + rh90*rh50
2	32.59	0.824	$\sqrt{y}$ ~rh90 + rh50

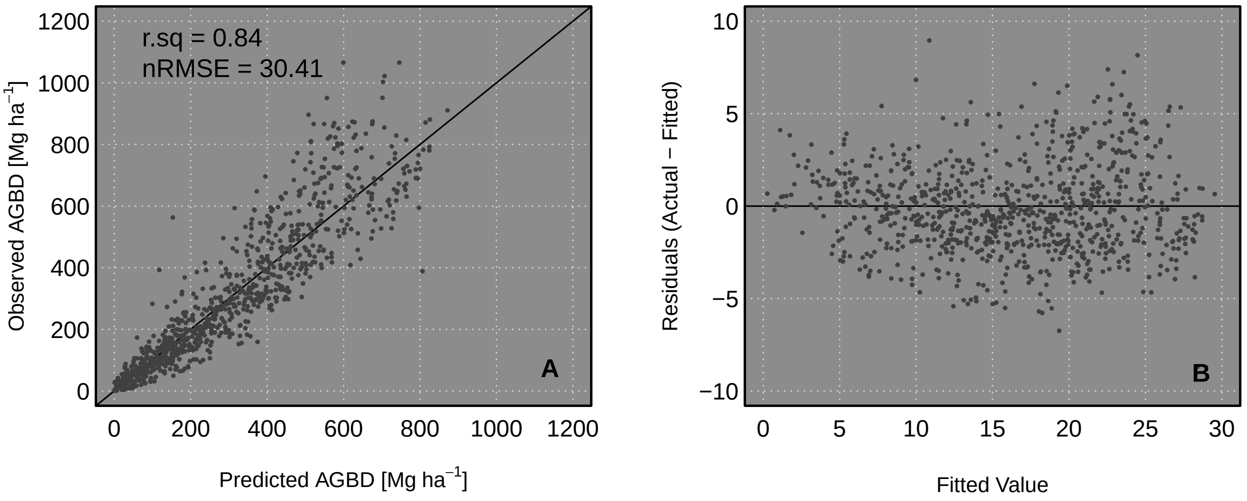

Figure A.1: The final OLS model applied to one of the 500 testing subsets (A), performs well and satisfies the assumption of constant variance. The residuals plot (B) is in square-root transformed units.

with a mean and median of 170.2 Mg ha^{-1} and 165.1 Mg ha^{-1} respectively (fig. A.4B).

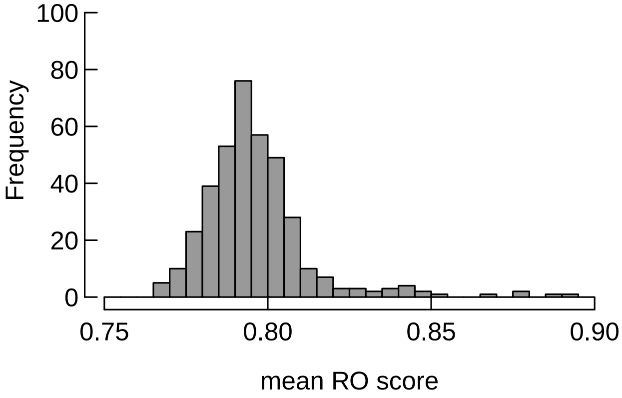

Figure A.2: FF produces many stands with waveforms that match the stem mapped field plots well. This distribution shows the mean relative overlap score across the 380 sites with at least 100 matches with a relative overlap above 0.75

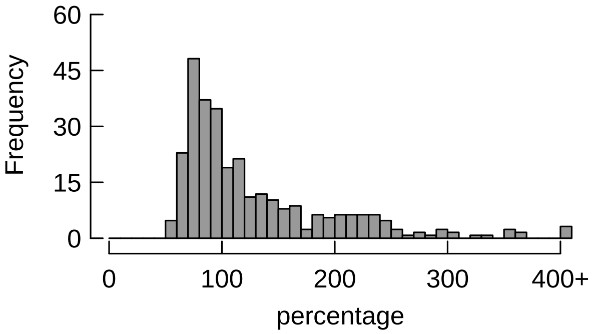

Figure A.3: WFM AGBD prediction uncertainty as a percentage of the predicted AGBD value is shown for the 380 *in situ* plots with at least 100 best matches; the uncertainty tends to be similar in magnitude to the predicted value, although some sites have uncertainties several times the predicted AGBD.

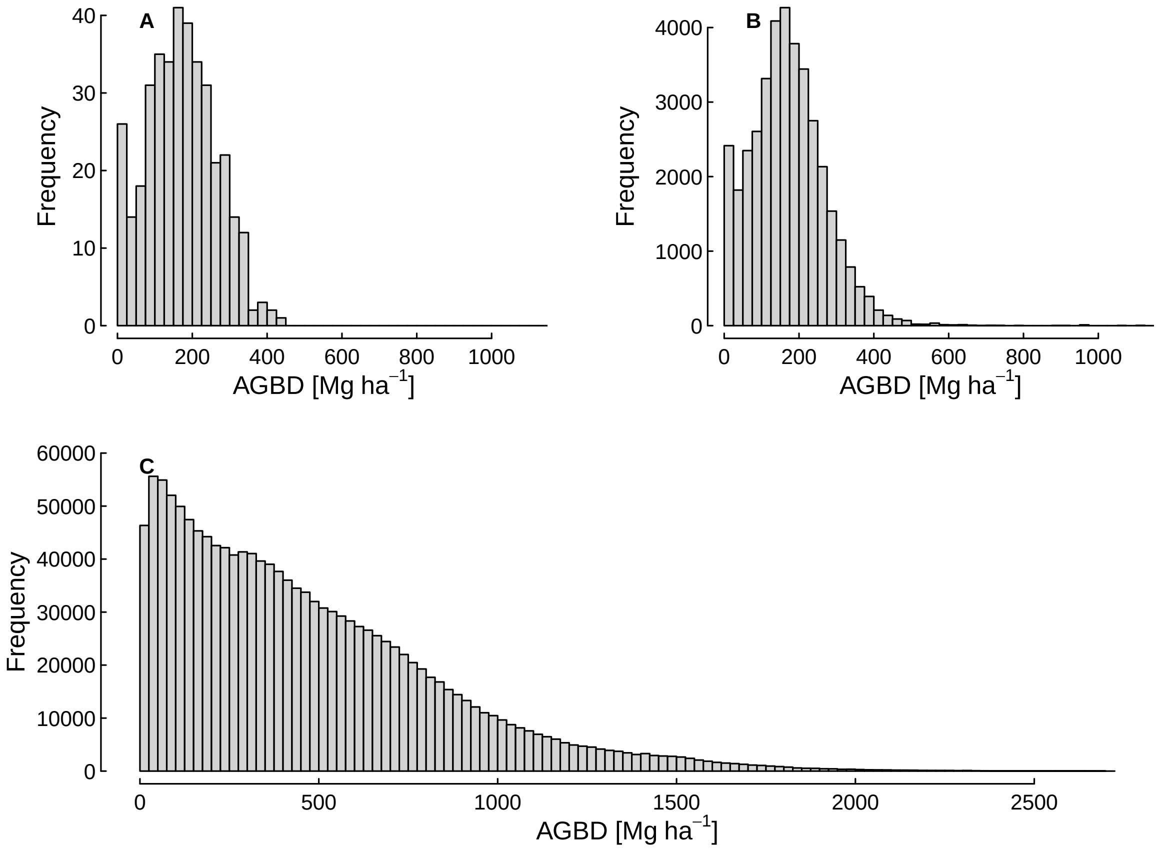

Figure A.4: The distribution of AGBD across the 380 *in situ* stem maps with at least 100 best matches (A) is highly similar to the distribution of AGBD across all the FF stands identified as best matches (B), despite a very different distribution of AGBD across the entire FF database used in the WFM process (C).

Appendix B: Appendix to Chapter 3

B.1 GEDI's standard quality filtering

Below is a summary of GEDI's existing data quality filtering practices taken from Dubayah et al. (2022b). These criteria are used to identify high-quality observations used in creating GEDI's hybrid AGBD estimates. For complete details please refer to the GEDI Level 4B Algorithm Theoretical Basis Document (Healey et al., 2022).

1. Shots flagged as quality by the GEDI L2A Footprint Height and Elevation metric product which identifies surface waveforms with high fidelity (Hofton et al., 2019).
2. Only shots with a beam sensitivity ¿0.95 were included. Beam sensitivity was calculated using a 3-sigma signal threshold and thresholds were selected to provide a sufficiently high signal-to-noise ratio to penetrate the highest canopy cover expected in these regions (Tang et al., 2019).
3. Shots with high degradation of geolocation performance were excluded from the sample since these may fall outside the geographic extent of a 1 km cell.
4. Orbit granules affected by low cloud or fog, which were identified using an iterative local outlier detection algorithm.

B.2 Additional Filtering Criteria

This paragraph provides the complete details of our topography filtering algorithm. First, we used 30 meter spatial resolution data sets of Landsat Tree Canopy Cover and SRTM Slope (Observation and Center, 2017) to identify candidate GEDI observations for removal, which we determined to be heavily impacted by topography relative to the presumed amount of vegetation within the footprint. For each range of canopy cover in table B.1, we calculated the 99th percentile of canopy height from all GEDI shots on slopes $<10°$, and compared this value to the shots from each of the combinations of slope and canopy cover in table B.1. If a sloped GEDI shot's canopy height was larger than the 99th percentile of canopy heights from the "flat" ground (slope $<10°$) shots in the same canopy cover group, we deemed this shot potentially too heavily affected by topography for the L4A to be valid. In regards to biomass estimation, we deal with these topography affected shots in two different ways: 1) for shots with a canopy cover of 0% we set the AGBD prediction to 0 Mg ha^{-1}, because we have high confidence there is no vegetation in the waveform (n=14,253,766; 2.6% of all GEDI observations in the US that passed GEDI's original quality filters), and 2) for shots with canopy cover greater than 0%, we remove this observation from the GEDI sample used for biomass estimation (n=437,624; 0.08% of all GEDI observations in the US that passed GEDI's initial quality filters). This decision was made to balance potential AGBD underestimation that would result from setting the footprint level predictions to 0 Mg ha^{-1}for shots we are confident have some biomass but are also heavily impacted by topography with the potential overestimation that would result from keeping these same shots in the sample when we believe the L4A prediction to be biased high due to topography. In other words, the best way to deal with these mixed vegetation-topography shots was to remove them from the sample.

B.3 SISA model development

B.3.1 Prediction Strata

The first step in SISA model calibration was to delineate SISA prediction strata that have consistent relationships between AGBD and GEDI derived forest structure metrics, to satisfy the first assumption necessary for the SISA models. We accomplished this by maximizing within strata similarity in forest type and climate at hexagon level, to match the resolution at which we calibrated the SISA models. The 2019 National Land Cover Dataset (NLCD) at 30 meter resolution has needleleaf (NT), deciduous broadleaf (DBT) and mixed (MIX) forest type classifications, and we calculated the proportions of each class within each hexagon (Wickham et al., 2021). We then assigned each hexagon to one of those three forest types as follows. If the proportion of MIX was the largest within the hexagon, or, if the larger proportion of non-mixed type (either NT or DBT) was less than the sum of the MIX proportion and the smaller non-mixed proportion, we classified it as MIX. If MIX was not the largest proportion and the larger proportion of non-mixed type was more than the sum of the MIX proportion and the other non-mixed proportion, we classified it as the forest type with the largest proportion. The result was a classification map of NLCD forest types at the hexagon level. We then intersected this map with the US Environmental Protection Agency (EPA) ecoregion classification level II, to designate ecoregional classifications with each of the MIX, NT, and DBT classes (Omernik and Griffith, 2014). This final stratification yielded ten different forest strata.

B.3.2 Calibration

To meet the assumption of a linear relationship at the unit (GEDI footprint) level, we used the GEDI calibration sites to assess all possible relationships between GEDI forest structure metrics and AGBD. We spatially intersected the calibration sites with the 30 meter NLCD map to classify each inventory plot as either DBT, NT, or MIX in a manner consistent with our forest strata. Since there were not enough plots in the database to test for linear relationships within each strata, we grouped the calibration sites by NLCD forest type (NT, DBT, or MIX). For each forest type we plotted the calibration sites' AGBD against the GEDI waveform relative height (RH) metrics. We used the RH metrics in 10 percent increments from RH10 to RH90, as well as RH98, which is a more stable measurement of maximum canopy height than RH100, and all the possible two-way interaction terms, and we raised each variable (individual RH metrics and the interaction terms) to the one-half, first, second, and third powers. We then fit a linear model between each variable and AGBD, and visually determined whether it was linear based on the scatter plots, because in many cases it was obvious to judge whether or not the relationship was linear. When it was not obvious, we then fit a first order non-linear polynomial model to the relationship, and used an analysis of variance to determine if the non-linear model explained significantly more variance than the linear model (α = 0.05). If not, we determined the relationship to be linear. If multiple transformations for a single variable resulted in a linear relationship, we selected the least transformed variable. The output of this process was a set of GEDI waveform variables and the associated transformations that were determined to have a linear relationship with AGBD at the footprint level, for each of the DBT, NT, and MIX forest types (see fig. B.2,fig. B.3,fig. B.4 for the linear diagnostics for the predictor variables used in the final SISA models in table B.2).

Next, we calculated the hexagon-level aggregations of each variable from the above list for each corresponding forest strata, taking special care to apply the interactions and transformations at the unit level of GEDI footprints prior to aggregation. This is important, because applying transformations or interaction effects after aggregation would violate the linearity constraint necessarily to ensure the resultant aggregate models were applicable at the unit level. We aggregated only the GEDI shots that were deemed suitable for biomass estimation, to ensure we used exactly the same GEDI observations for variable aggregation as for biomass estimation. The end result was a data set containing the average value within each hexagon for all GEDI metrics had linear relationships with AGBD at the plot level.

This data set was then used as the predictor variable(s) $\hat{x_j}$ for SISA model calibration (eq. 9), which was repeated independently for each forest strata. We used the design-based estimates of AGBD and associated sampling errors from Menlove and Healey (2020) within each stratum as the response variable $\hat{y_j}$ and sampling error δ_j, excluding hexagons with fewer than 2 forested plots, as the reported sampling errors for these plots was 0. We fit a univariate SISA model for every predictor variable, as well as bivariate SISA models using every possible combination of two predictor variables. After model fitting, we selected six candidate models for further evaluation, to determine which was best suited for biomass prediction at the footprint level; we chose the three models with the lowest residual standard error from the model fitting, as well as the three models with the lowest residual standard error from the linear models fit at the footprint level used to assess which unit level relationships were linear. These six candidate models were then individually applied to the on-orbit GEDI observations within the associated forest strata, resulting in a set of six different footprint-level AGBD predictions for every GEDI observation. We compared the distribution of footprint level AGBD predictions from each model with the distribution of FIA plot-level AGBD values from that region. Each region's final SISA model was the one that produced the most similar unit-level distribution of biomass when compared to that regions' distribution of FIA plot-level biomass, based on quantile-quantile plots, side by side comparisons, and a Kolmogorov-Smirnov test (table B.2, fig. B.5, fig. B.6). The predictions and parameter estimate covariance matrices from the final SISA models were then used to generate hexagon level AGBD estimates using GEDI's hybrid mean and variances estimators. We refer to these estimates as the GEDI-FIA fusion estimates, or simply the 'fusion' estimates.

In rare instances (4.5% of all shots) the final SISA models resulted in slightly negative predictions of AGBD at the unit level, which we set to 0 Mg ha^{-1}. On average the negative predicted value was -3.1 Mg ha^{-1}(median of -2.6 Mg ha^{-1}), and the average L4A prediction for these same GEDI observations was 10.2 Mg ha^{-1}(median of 8.7 Mg ha^{-1}), indicating these occurrences are limited to bare ground or sparsely vegetated areas.

B.4 Supplementary Tables and Figures

Table B.1: Slope filtering algorithm settings

Slope threshold	% Canopy Cover range
≥20	0-15
≥30	15-30
≥35	30-40
≥40	40-50
≥45	50-60
≥50	60-70

Table B.2: SISA models by strata. All variables used were interaction terms, indicated by ‘name’. The ‘power’ values indicate the transformation, if any, in the form of what power the variable was raised to. The ‘adjustment’ values show the constant used to divide the squared interaction terms by, if any, to reduce the variables’ magnitudes after transformations were applied

region	R^2	RSE	intercept	variable 1				variable 2			
				name	power	β	adjustment	name	power	β	adjustment
DBT_C	0.867	4.656	-858.914	rh40_rh90	½	8.545	-	-	-	-	-
DBT_N	0.929	10.426	-1101.059	rh40_rh90	½	10.921	-	-	-	-	-
DBT_S	0.84	14.98	-1101.846	rh40_rh70	½	10.92	-	-	-	-	-
MIX_N	0.945	6.538	-1109.06	rh30_rh90	½	11.111	-	-	-	-	-
MIX_S	0.889	4.527	-898.609	rh40_rh80	½	9.004	-	-	-	-	-
NT_C	0.901	3.478	-284.897	rh40_rh90	2	16.434	1E6	rh60_rh80	2	-13.598	1E6
NT_M	0.942	17.74	-130.531	rh50_rh90	2	-5.212	1E6	rh70_rh98	2	6.092	1E6
NT_N	0.908	3.24	-357.047	rh40_rh98	2	18.968	1E6	rh50_rh98	2	-15.457	1E6
NT_S	0.894	5.638	-129.077	rh50_rh98	2	2.72	1E6	rh60_rh90	2	-1.417	1E6
NT_W	0.927	24.895	-288.935	rh70_rh90	2	2.324	1E6	-	-	-	-

Table B.3: Percentage of calibration sites used to train the GEDI DBT and NT L4A models within each SISA model strata. Columns sum to 100%

	DBT	NT
DBT_N	0.8	0.0
DBT_S	0.3	0.1
MIX_N	85.3	24.7
MIX_S	7.9	3.8
NT_C	4.7	0.0
NT_M	0.2	3.5
NT_N	0.6	67.9
NT_S	0.1	0.1

B.4.1 Figures

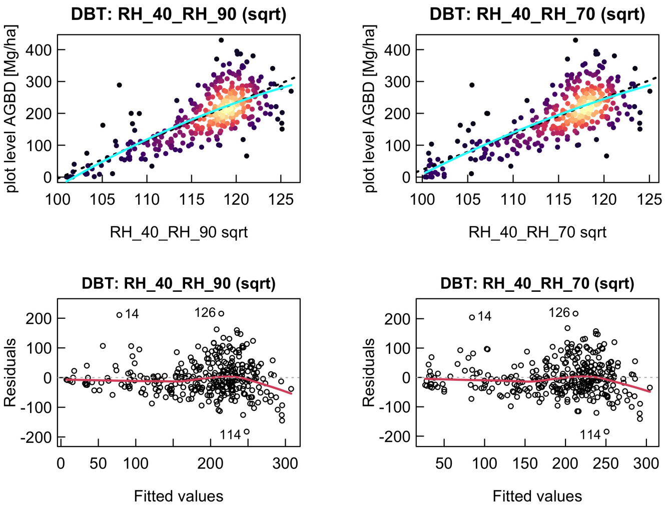

Figure B.2: Unit-level relationships between final SISA model predictor variables used in the DBT_C, DBT_N, and DBT_S SISA models, from the GEDI calibration plots (table B.2). The top row shows the GEDI predictor plotted against AGBD, with the regression line in dotted black, and a first order polynomial model fit in cyan. The bottom row shows the fitted values against the residuals from the linear model fit in the top row.

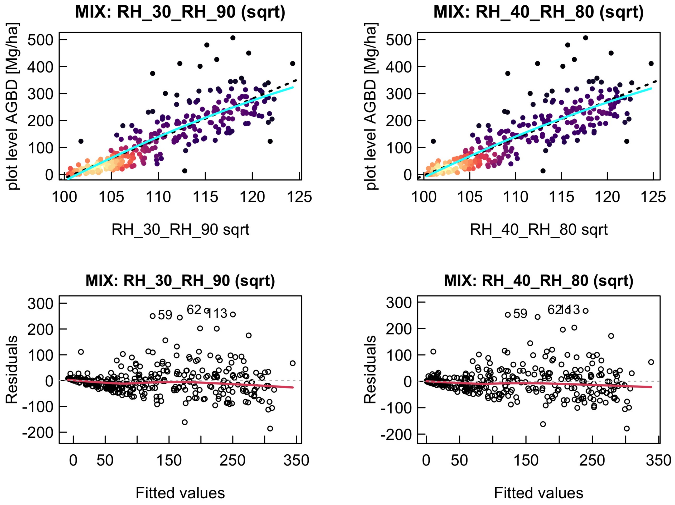

Figure B.3: Unit-level relationships between final SISA model predictor variables used in the MIX_N and MIX_S SISA models, from the GEDI calibration plots (table B.2). The top row shows the GEDI predictor plotted against AGBD, with the regression line in dotted black, and a first order polynomial model fit in cyan. The bottom row shows the fitted values against the residuals from the linear model fit in the top row.

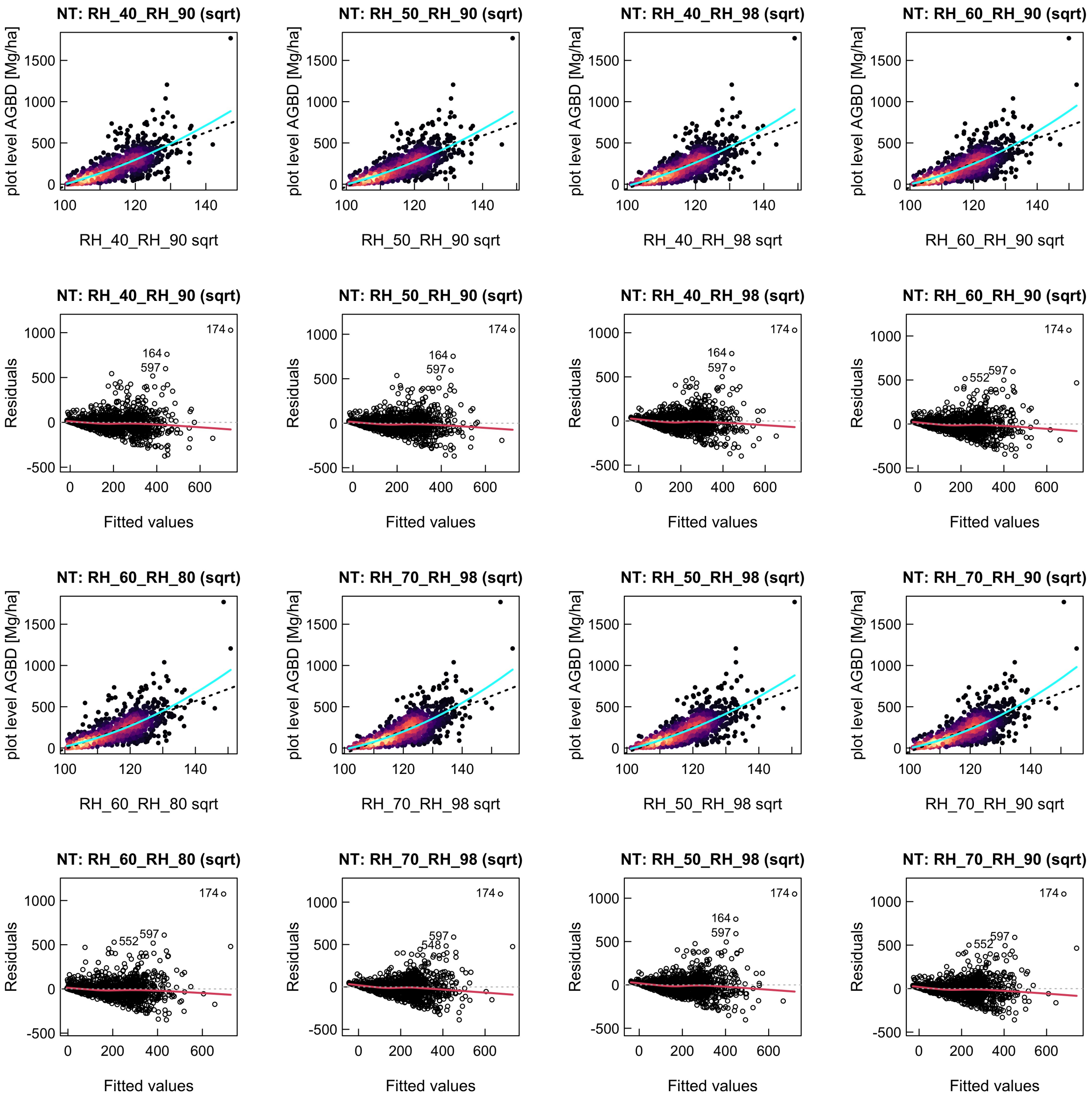

Figure B.4: Unit-level relationships between final SISA model predictor variables used in the NT_C, NT_M, NT_N, NT_S and NT_W SISA models, from the GEDI calibration plots (table B.2). The colored plots show the GEDI predictor plotted against AGBD, with the regression line in dotted black, and a first order polynomial model fit in cyan. The black open circle plots show the fitted values against the residuals from the linear model fit in the top row.

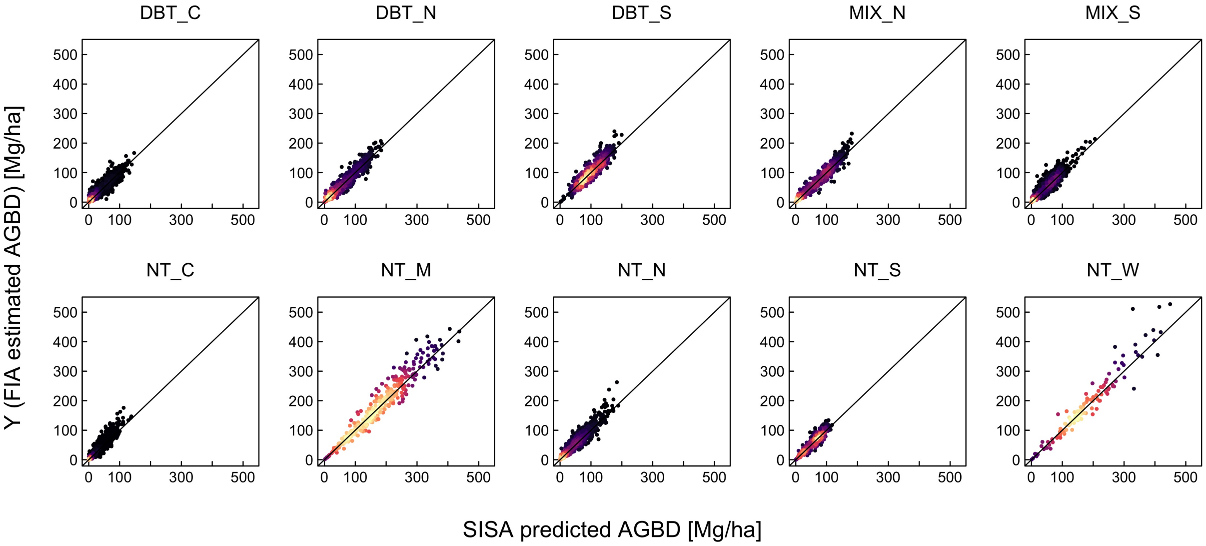

Figure B.5: Predicted and observed AGBD values from SISA model calibration at the hexagon scale. The observed values are the post stratified FIA estimates of AGBD, and the predicted values are the hexagon level predictions resultant from each region's final SISA model.

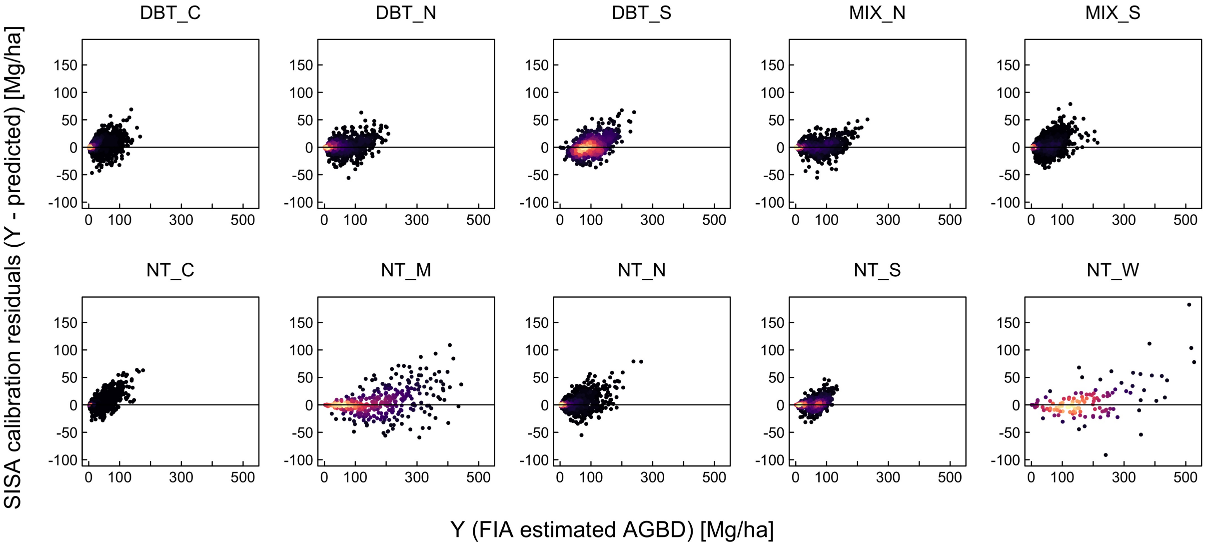

Figure B.6: SISA model calibration residuals (Y - predicted) plotted against the response variable (FIA hexagon estimates).

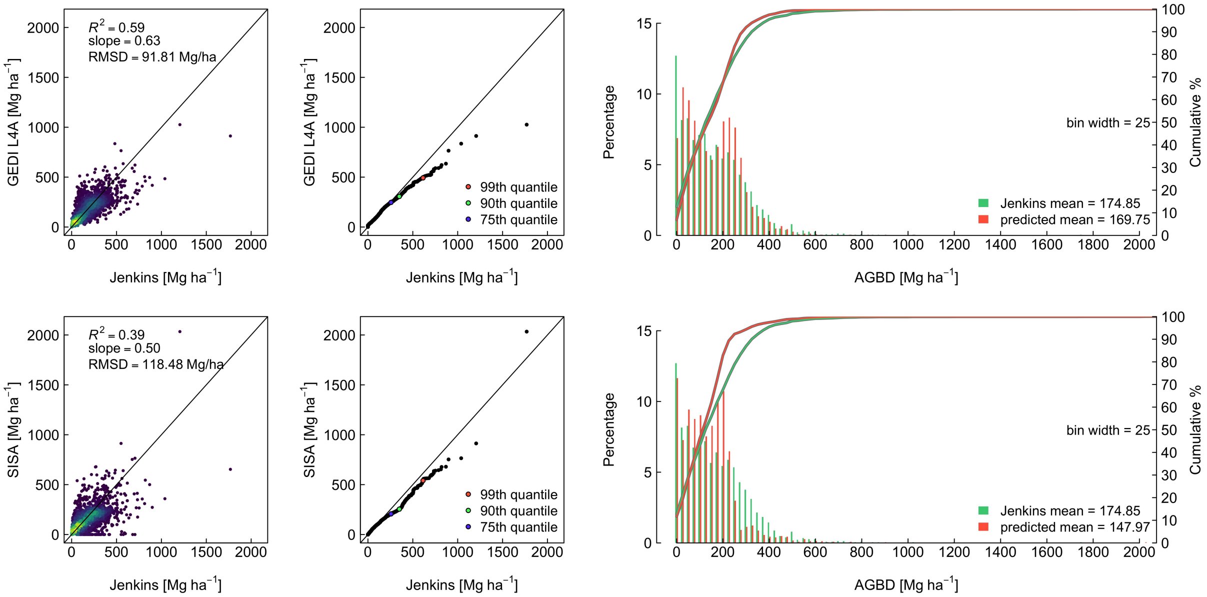

Figure B.7: L4A models generally outperform the SISA models when applied to the database of GEDI calibration sites used to train the L4A models.

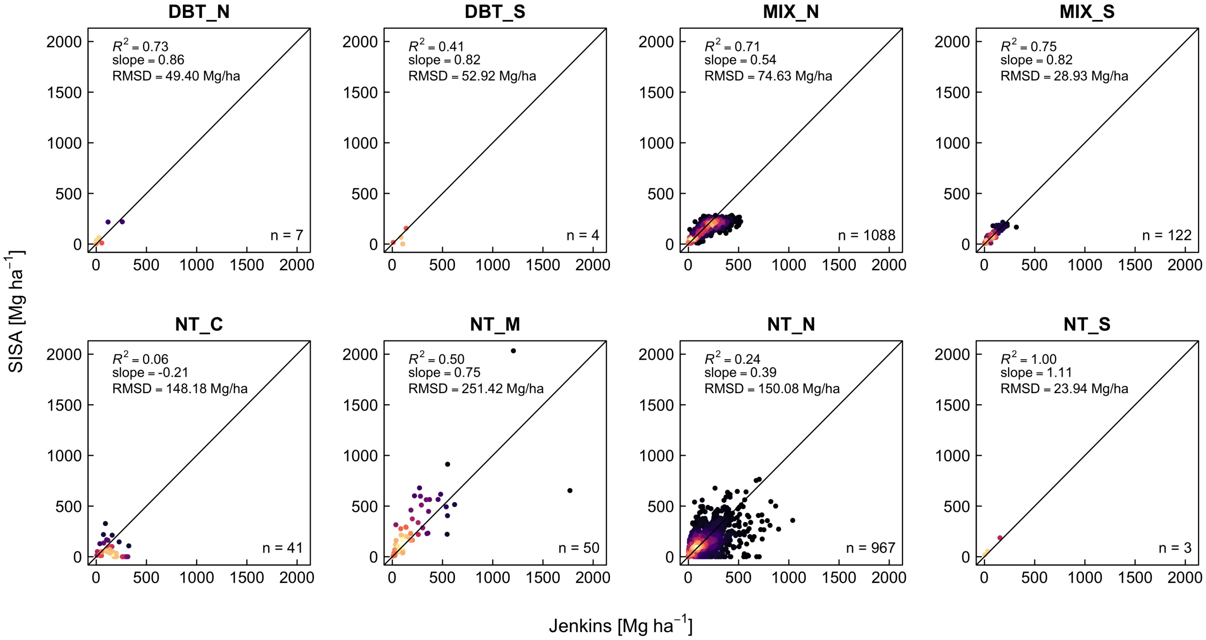

Figure B.8: Unit-level SISA predictions plotted against inventory-based AGBD (bottomleft plot in fig. B.7), stratified by SISA prediction strata.

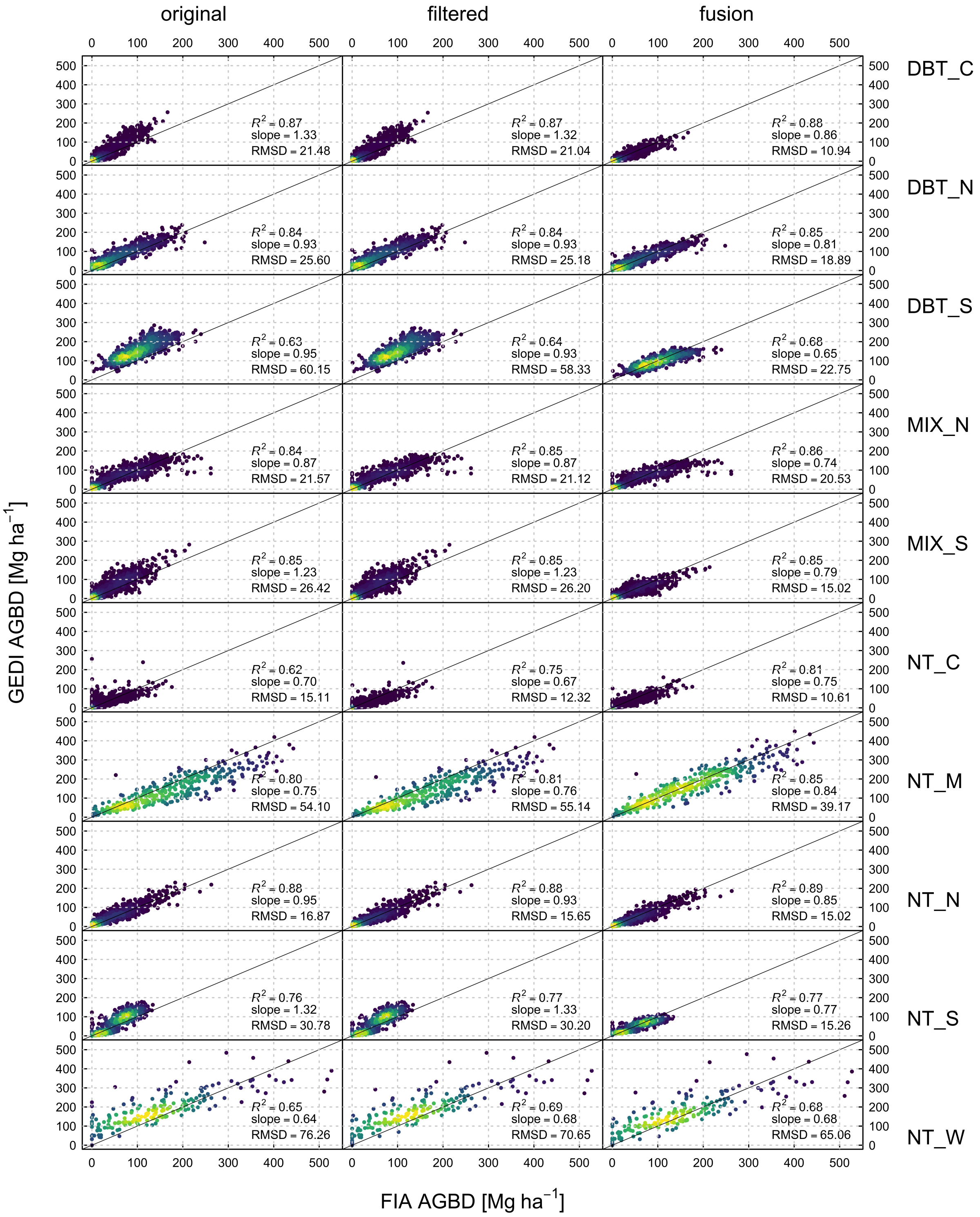

Figure B.9: Hexagon estimate comparisons between the FIA estimates (x-axis) and GEDI (y-axis) original, filtered, and fusion estimates, stratified by SISA prediction strata.

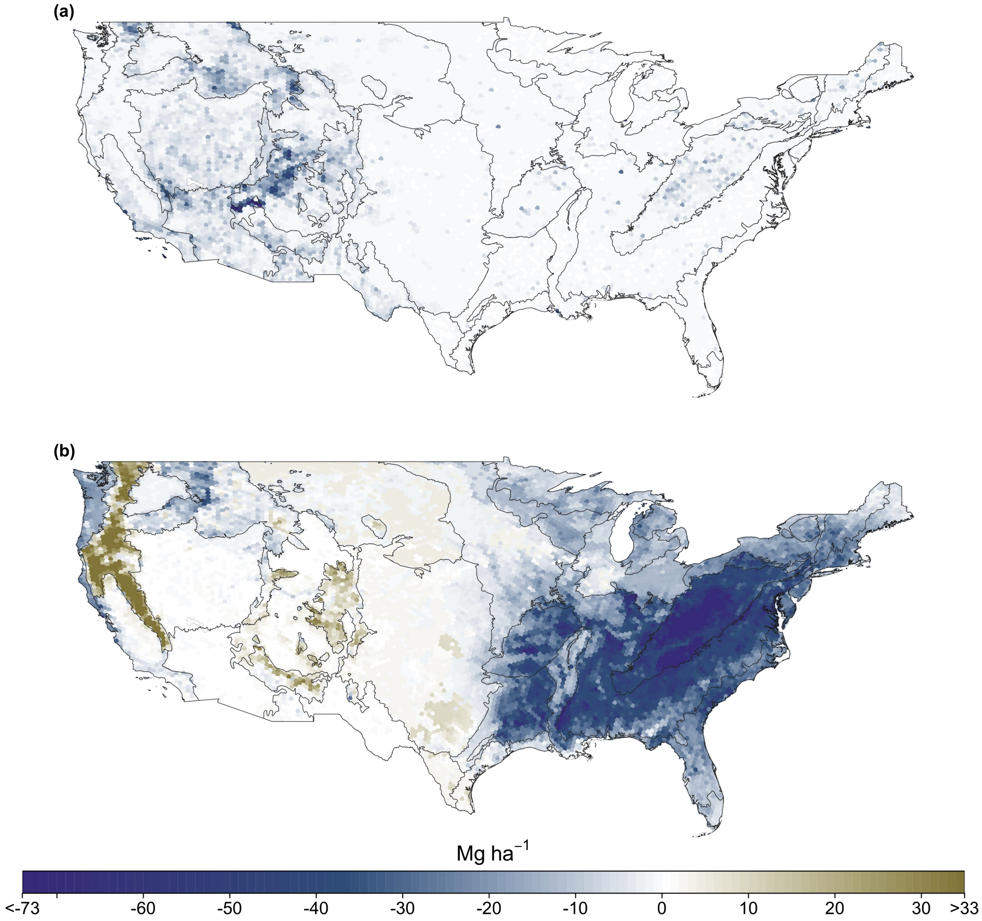

Figure B.10: Changes to GEDI hexagon estimates between the original and filtered estimates (a), and between the filtered and fusion estimates (b). Topography and cloud filtering results in moderate change to GEDI's hybrid AGBD estimates (a), primarily in clustered areas in the west. The SISA models resulted in widespread and substantial changes to GEDI's hybrid estimates (b), most notably decreasing the estimates throughout eastern deciduous and mixed forests and increasing the estimates in conifer dominated forests within the Sierra Nevada, Cascade, and Rock Mountain ranges.

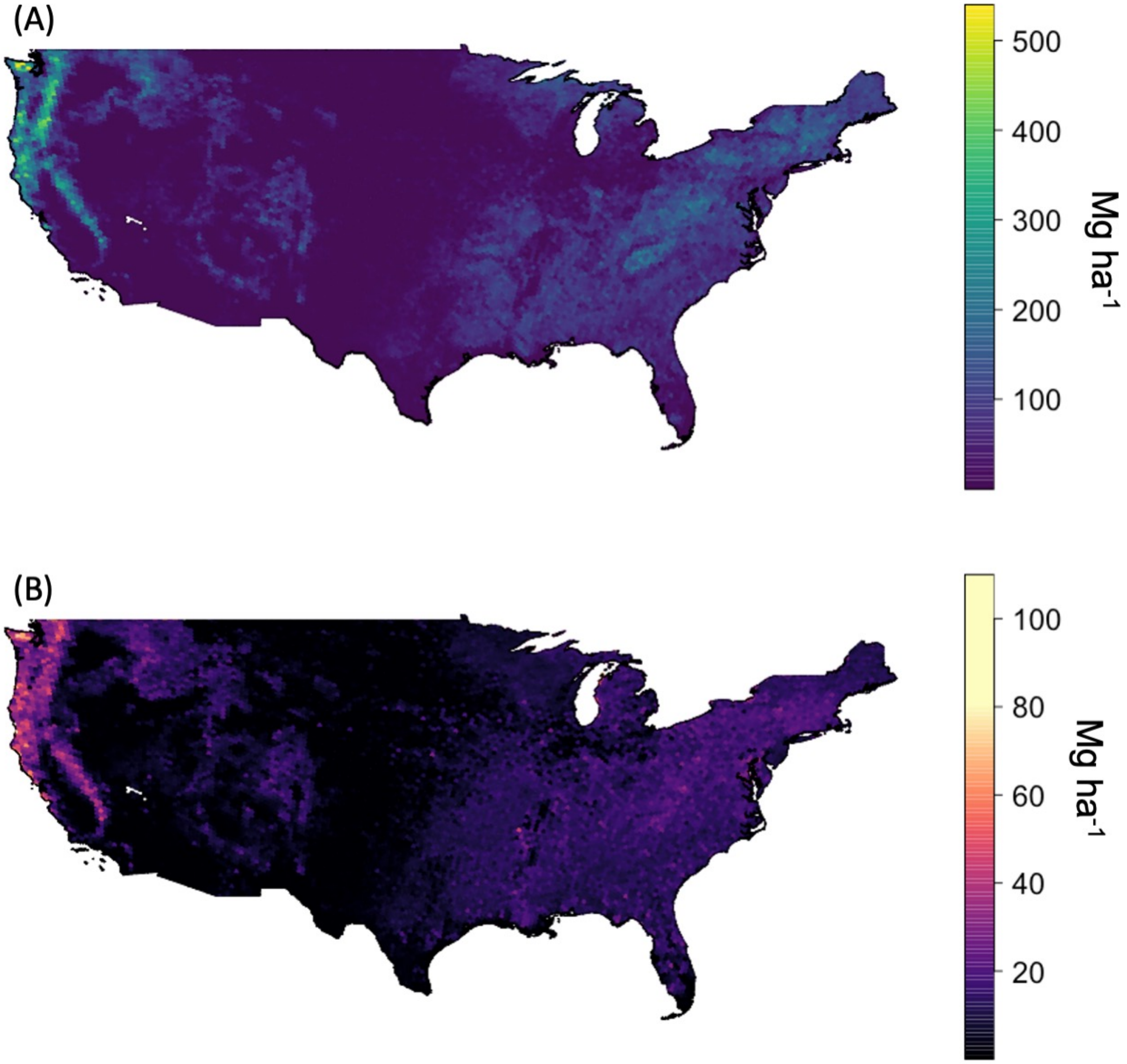

Figure B.1: FIA post stratified hexagon estimates of mean AGBD (A) and the associated standard error (B).

Bibliography

Ahmed, R., Siqueira, P., Hensley, S., and Bergen, K. (2013). Uncertainty of forest biomass estimates in north temperate forests due to allometry: implications for remote sensing. *Remote Sensing*, 5(6):3007–3036.

Avitabile, V., Herold, M., Heuvelink, G. B., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Ashton, P. S., Banin, L., Bayol, N., et al. (2016). An integrated pan-tropical biomass map using multiple reference datasets. *Global change biology*, 22(4):1406–1420.

Babcock, C., Finley, A. O., Bradford, J. B., Kolka, R., Birdsey, R., and Ryan, M. G. (2015). Lidar based prediction of forest biomass using hierarchical models with spatially varying coefficients. *Remote Sensing of Environment*, 169:113–127.

Babcock, C., Finley, A. O., Cook, B. D., Weiskittel, A., and Woodall, C. W. (2016). Modeling forest biomass and growth: Coupling long-term inventory and lidar data. *Remote Sensing of Environment*, 182:1–12.

Babcock, C., Finley, A. O., Andersen, H.-E., Pattison, R., Cook, B. D., Morton, D. C., Alonzo, M., Nelson, R., Gregoire, T., Ene, L., et al. (2018). Geostatistical estimation of forest biomass in interior Alaska combining Landsat-derived tree cover, sampled airborne lidar and field observations. *Remote Sensing of Environment*, 212:212–230.

Barndt, S., Gray, A., Hayward, G., Hiemstra, C., Kamoske, A., Kleinsmith, S., Krueger, J., Palmer, M., Pelz, K., Salverson, W., Schuler, T., Schumacher, C., Tilton, K., and Woodall, C. Mature and Old-Growth Forests: Definition, Identification, and Initial Inventory on Lands Managed by the Forest Service and Bureau of Land Management, (2023).

Barnett, K., Aplet, G. H., and Belote, R. T. (2023). Classifying, inventorying, and mapping mature and old-growth forests in the United States. *Frontiers in Forests and Global Change*, 5:1070372.

Barton, A. M. and Keeton, W. S. *Ecology and recovery of eastern old-growth forests*. Island Press, (2018).

Bauer, L., Knapp, N., and Fischer, R. (2021). Mapping amazon forest productivity by fusing gedi lidar waveforms with an individual-based forest model. *Remote Sensing*, 13(22):4540.

Bechtold, W. A. and Patterson, P. L. *The enhanced forest inventory and analysis program–national sampling design and estimation procedures*. Number 80. USDA Forest Service, Southern Research Station, (2005).

Bohn, F. J. and Huth, A. (2017). The importance of forest structure to biodiversity–productivity relationships. *Royal Society open science*, 4(1):160521.

Bohn, F. J., Frank, K., and Huth, A. (2014). Of climate and its resulting tree growth: Simulating the productivity of temperate forests. *Ecological modelling*, 278:9–17.

Bruening, J. M., May, P. B., Armston, J. D., and Dubayah, R. O. (2023). Precise and Unbiased Biomass Estimation From GEDI Data and the US Forest Inventory. *Frontiers in Forests and Global Change*, 6:1149153.

Canadell, J. G., Le Quéré, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., Conway, T. J., Gillett, N. P., Houghton, R., and Marland, G. (2007). Contributions to accelerating atmospheric co2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the national academy of sciences*, 104(47):18866–18870.

Caspersen, J. P., Pacala, S. W., Jenkins, J. C., Hurtt, G. C., Moorcroft, P. R., and Birdsey, R. A. (2000). Contributions of land-use history to carbon accumulation in US forests. *Science*, 290 (5494):1148–1151.

Chen, X., Disney, M., Lewis, P., Armston, J., Han, J., and Li, J. (2014). Sensitivity of direct canopy gap fraction retrieval from airborne waveform lidar to topography and survey characteristics. *Remote Sensing of Environment*, 143:15–25.

Cogbill, C. V., Burk, J., and Motzkin, G. (2002). The forests of presettlement New England, USA: spatial and compositional patterns based on town proprietor surveys. *Journal of Biogeography*, 29(10-11):1279–1304.

Crockett, E. T., Atkins, J. W., Guo, Q., Sun, G., Potter, K. M., Ollinger, S., Silva, C. A., Tang, H., Woodall, C. W., Holgerson, J., et al. (2023). Structural and species diversity explain aboveground carbon storage in forests across the United States: Evidence from GEDI and forest inventory data. *Remote Sensing of Environment*, 295:113703.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.

Davis, M. B. (1993). Old growth in the east. *The Cenozoic Society, Richmond, VT*.

Davis, M. D. *Eastern old-growth forests: prospects for rediscovery and recovery*. Island Press, (1996).

Davis, R. J., Bell, D. M., Gregory, M. J., Yang, Z., Gray, A. N., Healey, S. P., and Stratton, A. E. Northwest Forest Plan—the first 25 years (1994–2018): status and trends of late-successional and old-growth forests. Technical report, (2022). URL `https://doi.org/10.2737%2Fpnw-gtr-1004`.

DellaSala, D. A., Mackey, B., Norman, P., Campbell, C., Comer, P. J., Kormos, C. F., Keith, H., and Rogers, B. (2022). Mature and old-growth forests contribute to large-scale conservation targets in the conterminous United States. *Frontiers in Forests and Global Change*, 5:979528.

Drake, J. B., Dubayah, R. O., Knox, R. G., Clark, D. B., and Blair, J. B. (2002). Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, 81(2-3):378–392.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., et al. (2020a). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of remote sensing*, 1: 100002.

Dubayah, R., Armston, J., Kellner, J., Duncanson, L., Healey, S., Patterson, P., Hancock, S., Tang, H., Bruening, J., Hofton, M., et al. GEDI L4A Footprint Level Aboveground Biomass Density, Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA, (2021a).

Dubayah, R., Armston, J., Kellner, J., Duncanson, L., Hofton, M., Blair, J., and Luthcke, S. Gedi l4a footprint level aboveground biomass density, version 1, (2021b).

Dubayah, R., Hofton, M., Blair, J., Armston, J., Tang, H., and Luthcke, S. (2021c). GEDI L2A elevation and height metrics data global footprint level V002. *NASA EOSDIS Land Processes DAAC*. doi: https://doi.org/10.5067/GEDI/GEDI02_A.002.

Dubayah, R., Tang, H., Armston, J., Luthcke, S., Hofton, M., and Blair, J. (2021d). GEDI L2B Canopy Cover and Vertical Profile Metrics Data Global Footprint Level V002. *NASA EOSDIS Land Processes DAAC*. doi: https://doi.org/10.5067/GEDI/GEDI02_B.001.

Dubayah, R., Armston, J., Healey, S., Yang, Z., Patterson, P., Saarela, S., Stahl, G., Duncanson, L., and Kellner, J. GEDI L4B Gridded Aboveground Biomass Density, Version 2, (2022a).

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. (2022b). GEDI launches a new era of biomass inference from space. *Environmental Research Letters*, 17(9):095001.

Dubayah, R. O. and Drake, J. B. (2000). Lidar remote sensing for forestry. *Journal of forestry*, 98(6):44–46.

Dubayah, S., Hofton, M., Blair, J., Armston, J., Tang, H., and Luthcke, S. (2020b). GEDI L2A elevation and height metrics data global footprint level V001. *NASA EOSDIS Land Processes DAAC*.

Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., Silva, C. A., Armston, J., Luthcke, S. B., Hofton, M., et al. (2020). Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sensing of Environment*, 242:111779.

Duncanson, L., Armston, J., Disney, M., Avitabile, V., Barbier, N., Calders, K., Carter, S., Chave, J., Herold, M., MacBean, N., et al. (2021). Aboveground woody biomass product validation good practices protocol. doi: doi:10.5067/doc/ceoswgcv/lpv/agb.001.

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. (2022). Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sensing of Environment*, 270:112845.

Falkowski, M. J., Evans, J. S., Martinuzzi, S., Gessler, P. E., and Hudak, A. T. (2009). Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. *Remote Sensing of Environment*, 113(5):946–956.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74 (366a):269–277.

Fischer, R., Bohn, F., de Paula, M. D., Dislich, C., Groeneveld, J., Gutiérrez, A. G., Kazmierczak, M., Knapp, N., Lehmann, S., Paulick, S., et al. (2016). Lessons learned from applying a forest gap model to understand ecosystem and carbon dynamics of complex tropical forests. *Ecological Modelling*, 326:124–133.

Foster, D. R. et al. *Wildlands and woodlands, farmlands and communities: broadening the vision for New England.* Harvard University Press, (2017).

Frazer, G., Magnussen, S., Wulder, M., and Niemann, K. (2011). Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of lidar-derived estimates of forest stand biomass. *Remote Sensing of Environment*, 115(2):636–649.

Frelich, L. E. and Reich, P. B. (2003). Perspectives on development of definitions and values related to old-growth forests. *Environmental Reviews*, 11(S1):S9–S22.

Friedl, M., Gray, J., and Sulla-Menashe, D. MCD12Q2 MODIS/Terra+ Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC, (2019).

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., et al. (2022). Global carbon budget 2022. *Earth System Science Data Discussions*, 2022:1–159.

Fuller, J. L., Foster, D. R., McLachlan, J. S., and Drake, N. (1998). Impact of human activity on regional forest composition and dynamics in central New England. *Ecosystems*, 1:76–95.

Fuller, W. A. *Measurement error models.* John Wiley & Sons, New York, NY, (2009).

Gaines, G., Arndt, P., Croy, S., Devall, M., Greenberg, C., Hooks, S., Martin, B., Neal, S., Pierson, G., and Wilson, D. (1997). Guidance for conserving and restoring old-growth forest communities on national forests in the southern region. *Forestry Report R8-FR*, 62.

Gillman, L. N., Wright, S. D., Cusens, J., McBride, P. D., Malhi, Y., and Whittaker, R. J. (2015). Latitude, productivity and species richness. *Global Ecology and Biogeography*, 24(1):107–117.

Gray, A. N., Brandeis, T. J., Shaw, J. D., McWilliams, W. H., Miles, P. D., et al. (2012). Forest inventory and analysis database of the United States of America (FIA). *Biodiversity and Ecology*, 4:225–231.

Gray, A. N., Pelz, K., Hayward, G. D., Schuler, T., Salverson, W., Palmer, M., Schumacher, C., and Woodall, C. W. (2023). Perspectives: The wicked problem of defining and inventorying mature and old-growth forests. *Forest Ecology and Management*, 546:121350.

Haavik, L. Northeastern US Forest Pests, (2019). URL `http://www.fs.fed.us/ccrc/topics/northeastern-us-forest-pests`.

Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L. I., Kellner, J. R., and Dubayah, R. (2019). The GEDI simulator: A large-footprint waveform lidar simulator for calibration and validation of spaceborne missions. *Earth and Space Science*, 6(2):294–310.

Hansen, A. J., Phillips, L. B., Dubayah, R., Goetz, S., and Hofton, M. (2014). Regional-scale application of lidar: Variation in forest canopy structure across the southeastern US. *Forest Ecology and Management*, 329:214–226.

Healey, S. P., Patterson, P. L., Saatchi, S., Lefsky, M. A., Lister, A. J., and Freeman, E. A. (2012). A sample design for globally consistent biomass estimation using lidar data from the Geoscience Laser Altimeter System (GLAS). *Carbon balance and management*, 7(1):1–9.

Healey, S. P., Patterson, P. L., and Armston, J. (2022). Algorithm Theoretical Basis Document (ATBD) for GEDI Level-4B (L4B) Gridded Aboveground Biomass Density. URL `https://daac.ornl.gov/daacdata/gedi/GEDI_L4B_Gridded_Biomass/comp/GEDI_L4B_ATBD_v1.0.pdf`.

Helmer, E. H., Brown, S., and Cohen, W. (2000). Mapping montane tropical forest successional stage and land use with multi-date Landsat imagery. *International journal of remote sensing*, 21(11):2163–2183.

Hilbert, J. and Wiensczyk, A. (2007). Old-growth definitions and management: A literature review. *Journal of Ecosystems and Management*.

Hirschmugl, M., Sobe, C., Di Filippo, A., Berger, V., Kirchmeir, H., and Vandekerkhove, K. (2023). Review on the possibilities of mapping old-growth temperate forests by remote sensing in Europe. *Environmental Modeling & Assessment*, pages 1–25.

Hofton, M., Blair, J., Story, S., and Yi, D. (2019). Algorithm Theoretical Basis Document (ATBD) for GEDI transmit and receive waveform processing for L1 and L2 products. *University of Maryland: College Park, MD, USA*, page 44.

Hoppus, M. and Lister, A. The status of accurately locating forest inventory and analysis plots using the global positioning system. In *In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; McWilliams, William H., eds. Proceedings of the seventh annual forest inventory and analysis symposium; October 3-6, 2005; Portland, ME. Gen. Tech. Rep. WO-77. Washington, DC: US Department of Agriculture, Forest Service: 179-184.*, volume 77, (2007).

Houghton, R. A., Baccini, A., and Walker, W. S. Where is the residual terrestrial carbon sink?, (2018).

Hurtt, G., Pacala, S. W., Moorcroft, P. R., Caspersen, J., Shevliakova, E., Houghton, R., and Moore, B. (2002). Projecting the future of the US carbon sink. *Proceedings of the National Academy of Sciences*, 99(3):1389–1394.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R., Feddema, J., Fischer, G., Fisk, J., Hibbard, K., Houghton, R., Janetos, A., et al. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic change*, 109:117–161.

Inman, H. F. and Bradley Jr, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics-Theory and Methods*, 18(10):3851–3874.

Iverson, L. R., Prasad, A. M., Matthews, S. N., and Peters, M. (2008). Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest ecology and management*, 254(3):390–406.

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., and Birdsey, R. A. (2003). National-scale biomass estimators for United States tree species. *Forest science*, 49(1):12–35.

Kane, V. R., Bakker, J. D., McGaughey, R. J., Lutz, J. A., Gersonde, R. F., and Franklin, J. F. (2010). Examining conifer canopy structural complexity across forest ages and elevations with lidar data. *Canadian Journal of Forest Research*, 40(4):774–787.

Keeton, W. S., Whitman, A. A., McGee, G. C., and Goodale, C. L. (2011). Late-successional biomass development in northern hardwood-conifer forests of the northeastern United States. *Forest Science*, 57(6):489–505.

Kellner, J., Armston, J., and Duncanson, L. (2023). Algorithm Theoretical Basis Document for GEDI footprint aboveground biomass density. *Earth and Space Science*, 10(4).

Kellogg, K., Hoffman, P., Standley, S., Shaffer, S., Rosen, P., Edelstein, W., Dunn, C., Baker, C., Barela, P., Shen, Y., Guerrero, A. M., Xaypraseuth, P., Sagi, V. R., Sreekantha, C. V., Harinath, N., Kumar, R., Bhan, R., and Sarma, C. V. H. S. NASA-ISRO Synthetic Aperture Radar (NISAR) Mission. In *2020 IEEE Aerospace Conference*, pages 1–21, (2020). doi: 10.1109/AERO47225.2020.9172638.

Knapp, N., Fischer, R., and Huth, A. (2018). Linking lidar and forest modeling to assess biomass estimation across scales and disturbance states. *Remote Sensing of Environment*, 205:199–209.

Knapp, N., Huth, A., and Fischer, R. (2021). Tree crowns cause border effects in area-based biomass estimations from remote sensing. *Remote Sensing*, 13(8):1592.

Kondo, M., Ichii, K., Patra, P. K., Poulter, B., Calle, L., Koven, C., Pugh, T. A., Kato, E., Harper, A., Zaehle, S., et al. (2018). Plant regrowth as a driver of recent enhancement of terrestrial co2 uptake. *Geophysical Research Letters*, 45(10):4820–4830.

Körner, C. (2017). A matter of tree longevity. *Science*, 355(6321):130–131.

Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M., and Zink, M. (2007). TanDEM-X: A satellite formation for high-resolution SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3317–3341.

Lorimer, C. G. and Halpin, C. R. (2014). Classification and dynamics of developmental stages in late-successional temperate forests. *Forest Ecology and Management*, 334:344–357.

Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Vaglio Laurin, G., and Saah, D. (2012). Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates. *International Journal of Forestry Research*, 2012.

Ma, L. *Advanced Modeling Using Land-Use History and Remote Sensing to Improve Projections of Terrestrial Carbon Dynamics*. PhD thesis, University of Maryland, College Park, (2021).

Ma, L., Hurtt, G., Ott, L., Sahajpal, R., Fisk, J., Lamb, R., Tang, H., Flanagan, S., Chini, L., Chatterjee, A., et al. (2022). Global evaluation of the ecosystem demography model (ed v3. 0). *Geoscientific Model Development*, 15(5):1971–1994.

Maréchaux, I., Langerwisch, F., Huth, A., Bugmann, H., Morin, X., Reyer, C. P., Seidl, R., Collalti, A., Dantas de Paula, M., Fischer, R., et al. (2021). Tackling unresolved questions in forest ecology: The past and future role of simulation models. *Ecology and evolution*, 11(9): 3746–3770.

Margolis, H. A., Nelson, R. F., Montesano, P. M., Beaudoin, A., Sun, G., Andersen, H.-E., and Wulder, M. A. (2015). Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America. *Canadian Journal of Forest Research*, 45(7):838–855.

Martin, M., Cerrejón, C., and Valeria, O. (2021). Complementary airborne lidar and satellite indices are reliable predictors of disturbance-induced structural diversity in mixed old-growth forest landscapes. *Remote Sensing of Environment*, 267:112746.

May, P., McConville, K. S., Moisen, G. G., Bruening, J., and Dubayah, R. (2023). A spatially varying model for small area estimates of biomass density across the contiguous United States. *Remote Sensing of Environment*, 286:113420.

May, P., Dubayah, R., Bruening, J., and Gaines, G. (in review). An FIA-GEDI training data set for NFI-consistent lidar models. *International Journal of Applied Earth Observation and Geoinformation*.

McDowell, N. G., Allen, C. D., Anderson-Teixeira, K., Aukema, B. H., Bond-Lamberty, B., Chini, L., Clark, J. S., Dietze, M., Grossiord, C., Hanbury-Brown, A., et al. (2020). Pervasive shifts in forest dynamics in a changing world. *Science*, 368(6494).

McEwan, R. W., Pederson, N., Cooper, A., Taylor, J., Watts, R., and Hruska, A. (2014). Fire and gap dynamics over 300 years in an old-growth temperate forest. *Applied Vegetation Science*, 17(2):312–322.

McRoberts, R. E., Næsset, E., Saatchi, S., Liknes, G. C., Walters, B. F., and Chen, Q. (2019). Local validation of global biomass maps. *International Journal of Applied Earth Observation and Geoinformation*, 83:101931.

Menlove, J. and Healey, S. P. (2020). A comprehensive forest biomass dataset for the USA allows customized validation of remotely sensed biomass estimates. *Remote Sensing*, 12(24):4141.

Molina, I. and Marhuenda, Y. (2015). sae: An R Package for Small Area Estimation. *R J.*, 7(1): 81.

Moore, K. D. and Nelson, M. P. (2023). The perilous and important art of definition: the case of the old-growth forest. *Front Ecol Environ*, 21(6):264–265.

Nelson, R., Margolis, H., Montesano, P., Sun, G., Cook, B., Corp, L., Andersen, H.-E., deJong, B., Pellat, F. P., Fickel, T., et al. (2017). Lidar-based estimates of aboveground biomass in the continental US and Mexico using ground, airborne, and satellite observations. *Remote Sensing of Environment*, 188:127–140.

Ni-Meister, W., Lee, S., Strahler, A. H., Woodcock, C. E., Schaaf, C., Yao, T., Ranson, K. J., Sun, G., and Blair, J. B. (2010). Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research: Biogeosciences*, 115(G2).

Observation, E. R. and Center, S. E. Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global, (2017).

Oliver, C. D., Larson, B. C., and Oliver, C. (1996). Forest stand dynamics.

Omernik, J. M. and Griffith, G. E. (2014). Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environmental management*, 54(6):1249–1266.

Palik, B. J., D'Amato, A. W., Franklin, J. F., and Johnson, K. N. *Ecological silviculture: foundations and applications*. Waveland Press, (2020).

Pan, Y., Chen, J., Birdsey, R., McCullough, K., He, L., and Deng, F. (2011). Age structure and disturbance legacy of north american forests. *Biogeosciences*, 8(3):715–732.

Park, T., Kennedy, R. E., Choi, S., Wu, J., Lefsky, M. A., Bi, J., Mantooth, J. A., Myneni, R. B., and Knyazikhin, Y. (2014). Application of physically-based slope correction for maximum forest canopy height estimation using waveform lidar across different footprint sizes and locations: Tests on LVIS and GLAS. *Remote Sensing*, 6(7):6566–6586.

Patterson, P. L., Healey, S. P., Ståhl, G., Saarela, S., Holm, S., Andersen, H.-E., Dubayah, R. O., Duncanson, L., Hancock, S., Armston, J., et al. (2019). Statistical properties of hybrid estimators proposed for GEDI - NASA's Global Ecosystem Dynamics Investigation. *Environmental Research Letters*, 14(6):065007.

Pederson, N. (2010). External characteristics of old trees in the eastern deciduous forest. *Natural Areas Journal*, 30(4):396–407.

Pederson, N., Dyer, J. M., McEwan, R. W., Hessl, A. E., Mock, C. J., Orwig, D. A., Rieder, H. E., and Cook, B. I. (2014). The legacy of episodic climatic events in shaping temperate, broadleaf forests. *Ecological Monographs*, 84(4):599–620.

Pelz, K., Hayward, G., Gray, A., Berryman, E., Woodall, C., Nathanson, A., and Morgan, N. (2023). Quantifying old-growth forest of United States Forest Service public lands. *Forest Ecology and Management*, 549:121437.

Pesklevits, A., Duinker, P. N., and Bush, P. G. (2011). Old-growth forests: Anatomy of a wicked problem. *Forests*, 2(1):343–356.

Pinto, N., Simard, M., and Dubayah, R. (2012). Using InSAR coherence to map stand age in a boreal forest. *Remote Sensing*, 5(1):42–56.

Pratesi, M. and Salvati, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical methods and applications*, 17:113–141.

Pugh, S. A., Turner, J. A., Burrill, E. A., and David, W. (2018). The Forest Inventory and Analysis Database: Population Estimation User Guide. *USDA Forest Service, Washington DC, USA*.

Pugh, T. A. M., Lindeskog, M., Smith, B., Poulter, B., Arneth, A., Haverd, V., and Calle, L. (2019). Role of forest regrowth in global carbon sink dynamics. *P.N.A.S.*, page 201810512.

Qi, W., Saarela, S., Armston, J., Ståhl, G., and Dubayah, R. (2019). Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. *Remote Sensing of Environment*, 232:111283.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, (2022). URL `https://www.R-project.org/`.

Rao, J. N. and Molina, I. *Small area estimation*. John Wiley & Sons, New York, NY, (2015).

Rejou-Mechain, M., Muller-Landau, H. C., Detto, M., Thomas, S. C., Le Toan, T., Saatchi, S. S., Barreto-Silva, J. S., Bourg, N. A., Bunyavejchewin, S., Butt, N., et al. (2014). Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. *Biogeosciences*, 11(23):6827–6840.

Rhemtulla, J. M., Mladenoff, D. J., and Clayton, M. K. (2009). Historical forest baselines reveal potential for continued carbon sequestration. *Proceedings of the National Academy of Sciences*, 106(15):6082–6087.

Rödig, E., Knapp, N., Fischer, R., Bohn, F. J., Dubayah, R., Tang, H., and Huth, A. (2019). From small-scale forest structure to Amazon-wide carbon estimates. *Nature communications*, 10(1): 1–7.

Runkle, J. R. (1981). Gap regeneration in some old-growth forests of the eastern United States. *Ecology*, 62(4):1041–1051.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the national academy of sciences*, 108(24):9899–9904.

Santoro, M. and Cartus, O. (2021). Global datasets of forest above-ground biomass for the years 2010, 2017 and 2018, v2. *Centre for Environmental Data Analysis*, 10.

Scarth, P., Armston, J., Lucas, R., and Bunting, P. (2019). A structural classification of Australian vegetation using ICESat/GLAS, ALOS PALSAR, and Landsat sensor data. *Remote Sensing*, 11(2):147.

Shao, G., Shao, G., Gallion, J., Saunders, M. R., Frankenberger, J. R., and Fei, S. (2018). Improving lidar-based aboveground biomass estimation of temperate hardwood forests with varying site productivity. *Remote Sensing of Environment*, 204:872–882.

Shugart, H., Foster, A., Wang, B., Druckenbrod, D., Ma, J., Lerdau, M., Saatchi, S., Yang, X., and Yan, X. (2020). Gap models across micro-to mega-scales of time and space: examples of tansley?s ecosystem concept. *Forest Ecosystems*, 7:1–18.

Shugart, H. H., Asner, G. P., Fischer, R., et al. (2015). Computer and remote-sensing infrastructure to enhance large-scale testing of individual-based forest models. *Frontiers in Ecology and the Environment*, 13(9):503–511.

Spies, T. (1997). Forest stand structure, composition, and function. *Creating a forestry for the 21st century. Island Press, Washington, DC, USA*, pages 11–30.

Spies, T. A. (2004). Ecological concepts and diversity of old-growth forests. *Journal of forestry*, 102(3):14–20.

Spracklen, B. and Spracklen, D. V. (2021). Determination of structural characteristics of old-growth forest in Ukraine using spaceborne LiDAR. *Remote Sensing*, 13(7):1233.

Spracklen, B. D. and Spracklen, D. V. (2019). Identifying European old-growth forests using remote sensing: A study in the Ukrainian Carpathians. *Forests*, 10(2):127.

Stanke, H., Finley, A. O., Weed, A. S., Walters, B. F., and Domke, G. M. (2020). rFIA: An R package for estimation of forest attributes with the US Forest Inventory and Analysis database. *Environmental Modelling and Software*, 127:104664.

Tang, H. and Armston, J. (2019). Algorithm Theoretical Basis Document (ATBD) for GEDI L2B Footprint Canopy Cover and Vertical Profile Metrics. *University of Maryland: College Park, MD, USA*, page 39.

Tang, H., Armston, J., Hancock, S., Marselis, S., Goetz, S., and Dubayah, R. (2019). Characterizing global forest canopy cover distribution using spaceborne lidar. *Remote Sensing of Environment*, 231:111262.

Taylor-Rodriguez, D., Finley, A. O., Datta, A., Babcock, C., Andersen, H.-E., Cook, B. D., Morton, D. C., and Banerjee, S. (2019). Spatial factor models for high-dimensional and large spatial data: An application in forest variable mapping. *Statistica Sinica*, 29:1155.

Therneau, T. and Atkinson, B. *rpart: Recursive Partitioning and Regression Trees*, (2022). URL `https://CRAN.R-project.org/package=rpart`. R package version 4.1.19.

Therneau, T. M., Atkinson, B., and Ripley, M. B. (2010). The rpart package. *R Foundation for Statistical Computing: Oxford, UK.*

Thompson, J. R., Carpenter, D. N., Cogbill, C. V., and Foster, D. R. (2013). Four centuries of change in northeastern United States forests. *PloS one*, 8(9):e72540.

Tyrrell, L. E. *Information about old growth for selected forest type groups in the eastern United States*, volume 197. US Department of Agriculture, Forest Service, North Central Forest ?, (1998).

Uber, H. Hexagonal hierarchical geospatial indexing system, (2018).

Vorster, A. G., Evangelista, P. H., Stovall, A. E., and Ex, S. (2020). Variability and uncertainty in forest biomass estimates from the tree to landscape scale: the role of allometric equations. *Carbon balance and management*, 15:1–20.

White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., and Tompalski, P. (2016). Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing*, 42(5):619–641.

Wickham, J., Stehman, S. V., Sorenson, D. G., Gass, L., and Dewitz, J. A. (2021). Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. *Remote Sensing of Environment*, 257:112357.

Wirth, C., Messier, C., Bergeron, Y., Frank, D., and Fankhänel, A. *Old-Growth Forest Definitions: a Pragmatic View*. Springer Berlin Heidelberg, Berlin, Heidelberg, (2009). ISBN 978-3-540-92706-8.

Yang, W., Ni-Meister, W., and Lee, S. (2011). Assessment of the impacts of surface topography, off-nadir pointing and vegetation structure on vegetation lidar waveforms using an extended geometric optical and radiative transfer model. *Remote Sensing of Environment*, 115(11): 2810–2822.

Zhao, F., Guo, Q., and Kelly, M. (2012). Allometric equation choice impacts lidar-based forest biomass estimates: A case study from the Sierra National Forest, CA. *Agricultural and forest meteorology*, 165:64–72.

Zolkos, S., Goetz, S., and Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote sensing of environment*, 128:289–298.