

LETTER • OPEN ACCESS

Vicarious calibration of GEDI biomass with Landsat age data for understanding secondary forest carbon dynamics

To cite this article: Nidhi Jha *et al* 2024 *Environ. Res. Lett.* **19** 044062

View the [article online](#) for updates and enhancements.

You may also like

- [Inferring alpha, beta, and gamma plant diversity across biomes with GEDI spaceborne lidar](#)
C R Hakkenberg, J W Atkins, J F Brodie et al.
- [Assessing GEDI data fusions to map woodpecker distributions and biodiversity hotspots](#)
Lisa H Elliott, Jody C Vogeler, Joseph D Holbrook et al.
- [Assessing canopy height measurements from ICESat-2 and GEDI orbiting LiDAR across six different biomes with G-LiHT LiDAR](#)
Qiuyan Yu, Michael G Ryan, Wenjie Ji et al.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED

15 September 2023

REVISED

18 March 2024

ACCEPTED FOR PUBLICATION

21 March 2024

PUBLISHED

2 April 2024

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

Vicarious calibration of GEDI biomass with Landsat age data for
understanding secondary forest carbon dynamicsNidhi Jha^{1,*} , Sean P Healey² , Zhiqiang Yang², Göran Ståhl³ and Matthew G Betts¹ ¹ College of Forestry, Oregon State University, Corvallis, OR 97331, United States of America² US Forest Service Rocky Mountain Research Station, Ogden, UT 84401, United States of America³ Swedish University of Agricultural Sciences, Umeå, Sweden

* Author to whom any correspondence should be addressed.

E-mail: nidhi23aug@gmail.com**Keywords:** biomass recovery, GEDI vicarious calibration, secondary forests

Abstract

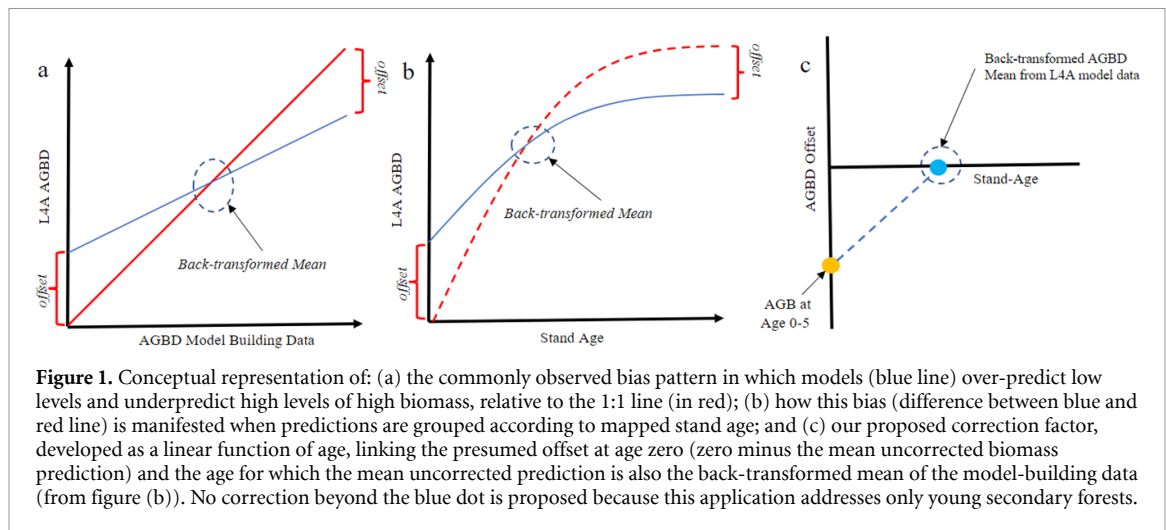
The recovery of biomass in secondary forests plays a vital role in global carbon sequestration processes and carbon emission mitigation. However, accurately quantifying the accumulation rate of aboveground biomass density in these forests is challenging owing to limited longitudinal field data. An alternative monitoring strategy is characterizing the mean biomass at a single point in time across stands with a range of known ages. This chronosequence approach can also be used with remotely sensed data by combining biomass measured with platforms such as NASA's Global Ecosystem Dynamics Investigation (GEDI) mission with forest age strata provided by historic Landsat imagery. However, focusing on the low-biomass conditions common in newly regenerating forests will accentuate commonly observed over-prediction of low biomass values. We propose a vicarious calibration approach that develops a correction for GEDI's biomass models in young forests, which may be mapped using Landsat time series, using an assumption that the aboveground biomass of newly cleared forests is zero. We tested this approach, which requires no additional local field data, in the U.S. Pacific Northwest, where extensive inventory data from the USDA Forest Service are available. Our results show that the calibration did not significantly improve the fit of predicted biomass as a function of age across 12 ecoregions (one-side *t*-test; $p = 0.20$), but it did significantly reduce bias for the youngest age groups with respect to reference data. Calibrated GEDI-based biomass estimates for < 20 year old forests were more accurate than 2006 IPCC defaults in most ecoregions (with respect to authoritative inventory estimates) and may represent a basis for refining carbon storage expectations for secondary forests globally.

1. Introduction

Secondary forests play a significant role in climate change mitigation due to their carbon sequestration potential (Pugh *et al* 2019, Koch and Kaplan 2022). Understanding aboveground biomass accumulation rates in such forests is crucial for subnational-to national-level reporting of forest carbon fluxes (Dobor *et al* 2018). In this paper, we propose a method for calibrating ecosystem-specific biomass accumulation curves—that is, sets of biomass predictions as a function of stand age—using lidar data from NASA's GEDI mission (Dubayah *et al* 2020). We test that approach in 12 ecoregions in the northwestern US for which comprehensive field data are available.

Practically, using global default carbon recovery rates to estimate carbon storage after forest disturbance instead of locally representative dynamics can have significant implications on estimates and their uncertainties in the context of the REDD+ (Reducing Emissions from Deforestation and Forest Degradation) framework (Jha *et al* 2020). Reduction of potential estimation errors in this context may affect forest management incentives through requirements related to 'uncertainty buffers' (ART Secretariat 2020, FCPF 2021).

Bias is a central consideration when using remotely sensed data to infer a population parameter such as mean forest biomass. The GEDI mission employs model-based predictors that are presumed



to be unbiased with respect to the lidar metrics GEDI delivers (Patterson *et al* 2019). However, Ståhl *et al* (2024) outline how such models may in practice lead to predictions which exhibit systematic errors (often over-prediction of low values and under-prediction of high values) when compared against independent ‘truth’ data. Expressed formally, although the predictions from GEDI models (and other similar models) are approximately model-unbiased, they are at the same time design-biased (*ibid.*). The design-bias is most severe for small and large true values. Depending upon the strength of the modeled relationship, this latter type of bias can have negative consequences on applications such as inferring mean biomass for a particular forest age group, as is the challenge motivating our study. Ståhl *et al* (2024) recommend application-specific calibration, as has been developed across multiple fields (Shukla 1972, Tellinghuisen 2000, Lindgren *et al* 2022).

The target of our research is a design-unbiased representation of mean biomass as it changes through different age groups, essentially linking inferences for a series of subpopulations identified by independent age maps. Compiling median measurements across age-based strata to understand ecosystem change, a tactic frequently called chronosequencing, is a common alternative to longitudinal study in the field of forest inventory (Rozendaal and Chazdon 2015, Poorter *et al* 2016, Lozada Dávila *et al* 2020). This prioritizes unbiasedness not in relation to remotely sensed predictor data or with respect to known biomass references (as described above), but instead conditional on an ancillary variable—stand age.

Specifically, it would be insufficient to calibrate individual GEDI biomass predictions (Kellner *et al* 2023), essentially making the blue line in figure 1(a) look more like the red line, because the relevant information is derived only when predictions are stratified by ancillary age data (figure 1(b)). Instead, we propose a calibration approach focusing on eliminating bias as a function of age, using: (1)

the difference between uncorrected GEDI predictions and zero biomass at locations where age is known to be zero; and (2) an assumption that there is no bias at the age associated with the biomass density reflecting the mean of the training data (figure 1(c)). The resulting age-dependent linear correction (figure 1(c)) may be applied to individual GEDI biomass predictions or directly to age group means if observations are given equal weight in calculation of the mean.

Development of calibration offsets from retrievals with known ground properties (in this case, quantifying overprediction of areas presumed to have age zero) is an example of vicarious calibration, often used to adjust radiometric, spectral, and geometric performance of Earth-observing sensors (Kabir *et al* 2020, Pearlshtien *et al* 2023). Significantly, this approach does not require field data beyond the observations used to fit GEDI’s biomass model (Duncanson *et al* 2022), an advantage because no such observations are available for many parts of the world (Chave *et al* 2019, Schepaschenko *et al* 2019).

The vicarious calibration we propose relies upon an assumption that zero aboveground biomass remains following stand-clearing disturbances, which are readily detectable using Landsat time series (Healey *et al* 2008). Landsat time series allow monitoring of disturbance history across 50 years (Wulder *et al* 2022), capturing dynamics ranging from complete forest clearance to subtle regrowth (Cohen *et al* 2010). The zero-age/zero-biomass locations we use in vicarious calibration may be obtained throughout GEDI’s observed latitudes. We evaluated effects of our vicarious calibration on chronosequence accuracy using data derived from the US Forest Service’s FIA (Forest Inventory and Analysis) Program (Bechtold and Patterson 2005). We likewise compared implied carbon accumulation rates with default 2019 IPCC values. This research aims to enhance the use of empirical data in generating more locally representative default rates of carbon gain while recognizing

practical limitations on available calibration data. The performance of vicarious calibration in this context across thoroughly inventoried ecosystems may inform its use elsewhere.

2. Methods

2.1. Site selection and biomass data

The study area covers approximately 23 million hectares in 12 ecoregions across portions of the US states of Washington, Oregon, and California (figure 2). The 12 ecoregions are based on the physiographic and geographic provinces introduced by (Franklin and Dyrness 1973) and later modified and adopted by the Forest Ecosystem Management Team (1993). Ecoregions included California Cascades (CaCAS), California Coast Range (CaCOA), California Klamath (CaKLA), Oregon Coast Range (OrCOA), Oregon Eastern Cascades (OrECO), Oregon Klamath (OrKLA), Oregon Western Cascades (OrWCO), Oregon Willamette Valley (OrWIL), Washington Eastern Cascades (WaECW), Washington Olympic Peninsula (WaOLY), Washington Western Cascades (WaWCW), Washington Western Lowlands (WaWLO).

We used publicly available FIA plot-level biomass data with stand age to validate our results (<https://apps.fs.usda.gov/fia/datamart/datamart.html>). Attribution of each plot to an ecoregion used publicly distributed ‘fuzzed’ (to within 1 km) coordinates, a level of imprecision we considered trivial for this analysis. We grouped plots according to the age groups available from the Landsat disturbance history maps (below). The total number of field plots used, along with the number of GEDI samples in each age-group, is given in table 1. We calculated the mean and standard error of the mean aboveground live tree biomass in megagrams per hectare (Mg ha^{-1}) for each age group. The estimates incorporated plot weights derived from inventory expansion factors that account for the sample design (Burrill *et al* 2021). Plots with multiple conditions were included, weighted by the proportion of the plot represented by each condition.

2.1.1. Landsat disturbance data

We combined two historical disturbance maps derived from Landsat MSS, TM, ETM+, and OLI for the study area. From 1972 to 1984, we used stand replacing disturbance by (Healey *et al* 2008). For 1984–2021, we used disturbance, land use and land cover products from the Landscape Change Monitoring System Disturbance Maps (LCMS) (USDA Forest Service 2022).

To accurately identify stand initiation dates, we applied a series of filters to the maps, restricting analysis to areas with a high certainty of disturbance at a known time. The 1972–84 map exclusively targeted stand-initiating disturbance, while disturbed

pixels from 1984 to 2021 were included in the analysis if they were forested (according to corresponding land use maps) in the pre-disturbance year, non-tree covered immediately after disturbance, and ultimately remained in a forested use condition for six to eight years post-disturbance.

We explicitly avoided partial disturbances and multiple cohorts stands. This was done via the application of three additional filters, all of which operated only on the most recent disturbance mapped for a given pixel. First, we excluded all pixels labeled in the LCMS dataset as disturbed by any means other than harvest. Second, we eliminated all pixels on the boundary of harvest patches to avoid mixed pixels and to minimize the effect of spatial uncertainty in both the GEDI and Landsat data. Finally, we removed isolated pixel clusters smaller than 1 ha. We also discarded pixels that underwent more than one shift during the entire study period. Before 1984, Landsat MSS-derived disturbances were at a $60 \text{ m} \times 60 \text{ m}$ spatial resolution. To ensure consistent time series data for merging, we aligned all post-1984 LCMS disturbances using the reference image from 1972 to 1977 derived disturbances and aggregated each image to $60 \times 60 \text{ m}$. Finally, we assigned stand age as the time since the mapped stand-clearing disturbance. This definition of stand age roughly corresponds to the measures used by FIA to assign an age to a surveyed forest condition. Utilizing precise coordinates from certain FIA plots that align with Landsat-derived age patches, we observed that, on average, the Landsat-derived stand age deviates by about 3–5 years compared to the FIA mean age by ecoregions. To enhance comparability, we organized map age classes in 5 year intervals, except 1977–1984 interval, which was chosen by the producers of the original map (Healey *et al* 2008).

2.1.2. GEDI L4A data processing and evaluation

The GEDI instrument is on board the International Space Station and equipped with two lasers in full-power beams and one laser split in two coverage beams in the near-infrared region (Dubayah *et al* 2020). For this study, we have used the footprint data set of Level 4AAGBD, Version 2.1 (L4A) (Dubayah *et al* 2022) containing a set of quality flags that allow us to remove the footprints with low sensitivity and potential atmospheric or ground-finding problems (Dubayah *et al* 2020). The aboveground biomass density (AGBD) data for GEDI L4A was obtained by utilizing a calibration dataset that compiled height metrics derived from GEDI Level 2A (L2A) simulated waveforms and field estimates across diverse regions and plant functional types (PFTs) (Duncanson *et al* 2022). The L4A data covering our study area from 2019 to 2021 were downloaded using the NASA-Earth Data ORNL DAAC platform (<https://search.earthdata.nasa.gov>, accessed 12 May 2022). We used the ‘GEDI4R’ package (Vangi *et al* 2022) to process the

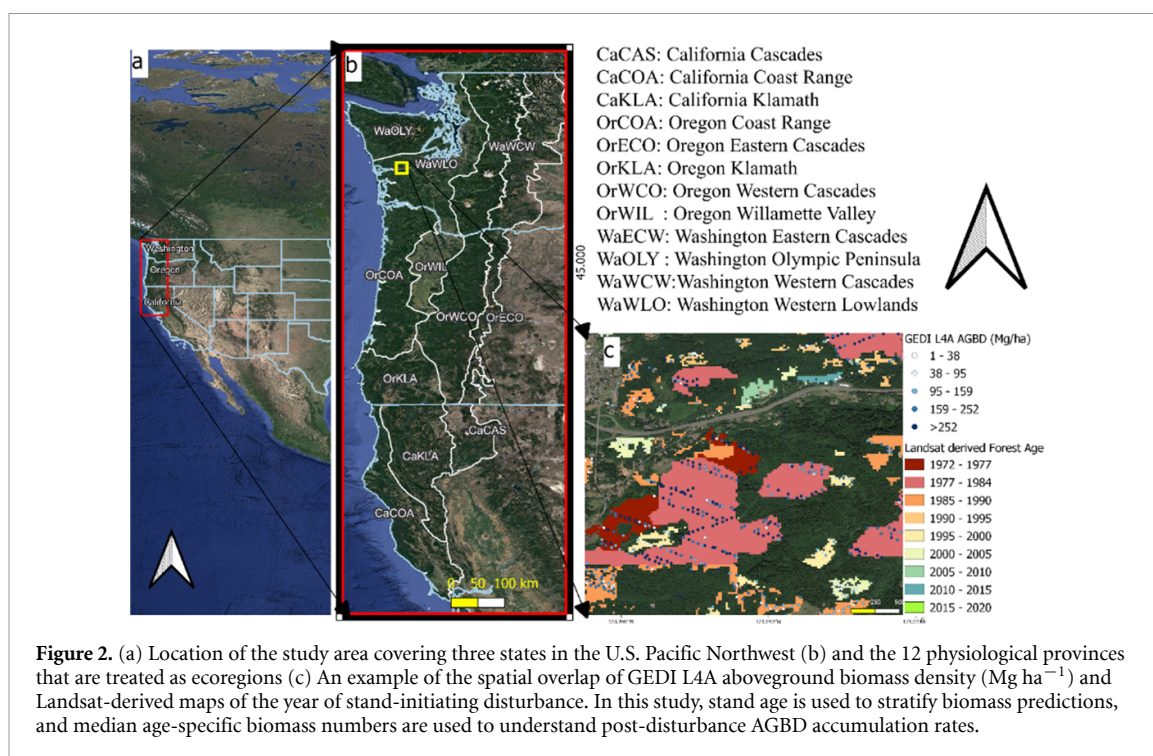


Table 1. Summary of number of FIA field plots available for each age group with GEDI Samples provided in [].

Prov/ Age-group	0–5	10–15	15–20	15–20	20–25	25–30	30–35	36–43	43–48
CaCAS	14[415]	7[2433]	5[3982]	5[2106]	10[1515]	9[5435]	6[2591]	7[349]	5[235]
CaCOA	14[97]	9[677]	11[759]	14[878]	11[1347]	20[917]	21[5337]	49[4838]	31[864]
CaKLA	61[2423]	20[3903]	12[6336]	13[4677]	16[2069]	24[2116]	15[7545]	38[943]	30[4782]
OrCOA	124[3245]	56[17 242]	62[24 562]	69[26 473]	96[22 282]	108[25 421]	88[32 484]	139[97 822]	72[33 422]
OrECO	46[496]	7[1596]	13[2400]	21[6411]	28[2245]	43[2228]	27[5120]	28[15 930]	21[5501]
OrKLA	53[879]	32[4271]	50[5429]	37[12 158]	44[3487]	35[4153]	40[9627]	55[33 626]	41[8886]
OrWCO	98[1977]	34[10 419]	56[13 531]	71[13 253]	125[9117]	122[9010]	124[26 966]	180[96 573]	89[37 606]
OrWIL	11[169]	5[1097]	7[1705]	10[2101]	10[1914]	6[1650]	2[1436]	5[1570]	6[400]
WaECW	133[1117]	22[4751]	35[8938]	30[11 127]	40[7526]	38[5052]	34[6748]	66[20 239]	21[9270]
WaOLY	32[856]	18[5911]	19[8088]	28[9760]	34[7990]	45[8985]	57[12 899]	49[50 924]	32[22 713]
WaWCW	70[1849]	25[9652]	31[10 567]	42[12 050]	75[11 618]	101[13 196]	123[25 544]	125[89 094]	73[61 462]
WaWLO	95[2457]	39[15 944]	66[24 599]	93[29 938]	97[27 171]	78[30 609]	68[38 031]	58[133 236]	22[43 350]

downloaded HDF5 GEDI data under the 4.0.3 version of the R statistical software (R Core Team 2022).

We applied quality control measures to filter the GEDI data. We excluded footprints with a l4 quality flag = 0, indicating issues like low sensitivity, leaf-off conditions for deciduous trees, or shots over water or urban areas. We only retained shots within secondary forest pixels identified in the Landsat disturbance map, discarding boundary samples. After filtering, 1490 720 shots were used in the analysis. (table 1).

To initially assess the suitability of GEDI's L4A biomass model in our chosen ecoregions, we linked FIA plot data from 2015 to 2019 with GEDI samples within 200 meters of each plot center. We used confidential FIA plot coordinates for this analysis in a secure environment, acknowledging imperfect spatial alignment with the GEDI shots.

Approximately 2% of the region's land area falls in one of the analysis' age strata following the filters described above. Approximately 51% of mapped stand-clearing harvests happened before 1984. Approximately 64% of disturbance patches had no GEDI sample shots. Out of 1.5 million selected GEDI shots, 36% were retrieved in 2020, 30% in 2021, 23% in 2019 and 11% in 2022. The mean of all the selected GEDI AGBD predictions at the 25 m footprint scale was 160 Mg ha^{-1} ($\pm 139 \text{ Mg ha}^{-1}$), with a range of 0–4394 Mg ha^{-1} .

2.2. GEDI biomass recovery and vicarious calibration

We produced uncalibrated chronosequences by calculating both the median and interquartile range for GEDI L4A AGBD predictions (2019–2021) in each mapped age group within each ecological province.

Medians were used in this analysis instead of means to minimize outliers resulting from Landsat map error and/or GEDI geolocation error that resulted in GEDI shots from mature forests being assigned to classes representing young forests.

Our calibration is designed to correct over-prediction of low levels of biomass in the GEDI L4A product, specifically in the space where GEDI L4A predictions are grouped by mapped age (refer figure 1(b)). Here we elaborate upon our vicarious calibration method and its underlying assumptions. In figure 1(b) we expect that the bias originates from the original GEDI L4A prediction space as depicted in figure 1(a). We made two assumptions that allowed vicarious calibration of GEDI's biomass predictions in the context of young forest carbon accumulation. First, we assumed that new clearcuts (age-group 0–5) have no AGBD, and the appropriate offset ($AGBD_{\text{Offset}}$) for this age group (yellow dot in figure 1(c)) is determined by subtracting the median GEDI L4A AGBD prediction in that age-group from zero. Although not all clearcuts lead to total biomass removal (e.g. Oregon Revised Statute (ORS) (2023), § 527.676 mandates the retention of specific elements), we presume such restricted biomass retention in clearcut scenarios is deemed relatively insignificant. Second, we assume that there is no consistent bias in L4A predictions associated with the mean of the training dataset. We further identified the age (AGE_{mean} in equation (1) associated with mean of the L4A reference dataset (135 Mg ha^{-1} , after backtransformation; see figure 1(a)), and declared the needed offset at that age to be zero (blue dot, figure 1(c)). An age-specific offset was then added to all L4A biomass predictions, using an offset function linearly interpolated between the yellow and blue points (equation (2)). No offset was added to points above the age associated with the mid-point of the modeling dataset (equation (3)) because of a lack of information on high-biomass shots and a recognition that most secondary forests contain relatively few high-biomass conditions. In some regions, it might be necessary to perform this calibration separately for every PFT (different functional types have different L4A models: Duncanson *et al* 2022), but we treated the entire study area as 'evergreen needleleaf trees', the dominant model used by the GEDI mission in the study area (60%–99%, depending upon ecoregion),

$$Gedi_{\text{Calibrated}} = Gedi_{\text{predicted}} + AGBD_{\text{Offset}} \quad (1)$$

$$AGBD_{\text{Offset}} = a + b \times \text{age}, \text{ when age} < AGE_{\text{mean}} \quad (2)$$

$$AGBD_{\text{Offset}} = 0, \text{ when for age} \geq AGE_{\text{mean}} \quad (3)$$

where a and b are calibration parameters fit by ecoregion and a is a negative coefficient.

We conducted a comparison between the chronosequence derived from uncalibrated GEDI L4A and calibrated GEDI L4A against the chronosequence

derived from FIA's estimates of age-specific means (with standard errors) for each province. The comparison involved assessing the overall fit using a one-sided paired t -test to evaluate if the proposed calibration process improved the correspondence of GEDI and FIA (i.e. reference) chronosequences. Specifically, the RMSE (equation (4)) of calibrated and uncalibrated chronosequences were evaluated in pairs by ecoregion,

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (FIA - Gedi)^2}{N}}. \quad (4)$$

Furthermore, we performed age-group specific t -tests to evaluate the difference in absolute bias across our 12 ecoregions between the uncalibrated and calibrated L4A estimates. Finally, we performed one-sided t -tests on the mean differences between FIA and calibrated GEDI values, and between FIA and default 2019 Refinement to the 2006 IPCC rates for aggregated <20 and 20–48 year age groups, using IPCC default values for Temperate Secondary Forest (table 4.7, volume 4: Agriculture, Forestry and Other Land Use, Buendia *et al* 2019, Shukla *et al* 2019). The IPCC value used in comparison with our 20–48 year class is from the class: secondary forest, 20+ years. Note that the FIA data we use as a reference at the ecoregion level (in addition to Canadian data) for IPCC regional-scale reference levels.

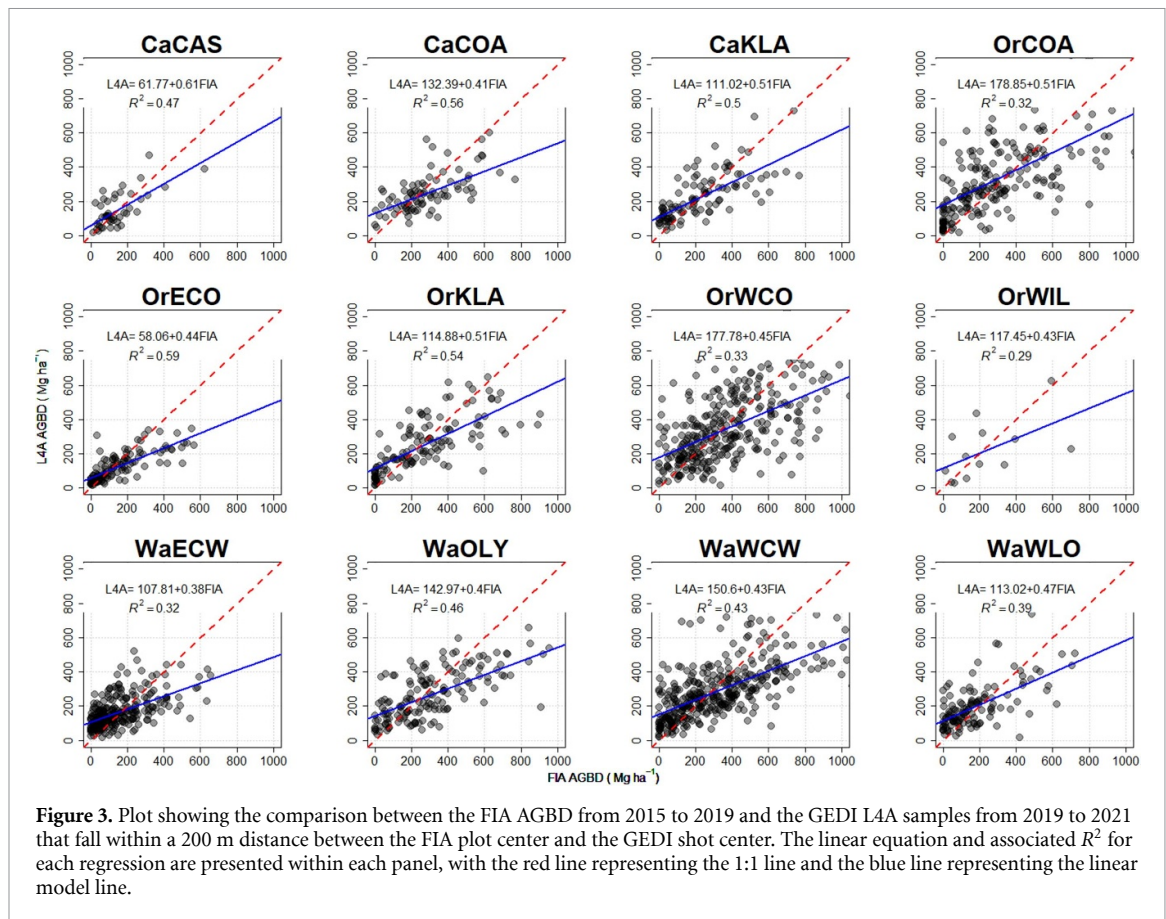
3. Results

3.1. Preliminary evaluation of GEDI AGBD

The preliminary comparison between FIA and uncalibrated GEDI L4A AGBD values in each ecoregion, presented here for context, demonstrated varying levels of agreement (figure 3). It must be noted that the spatial overlay of FIA plots and GEDI shots was only approximate (<200 m between plot and shot center points), which likely contributed to R^2 below 0.5 in most ecoregions. As expected, high biomass values tended to be under-predicted in the uncalibrated set of GEDI predictions, while low-values tended to be over-predicted ranging from 58.06 to $177.78 \text{ Mg ha}^{-1}$ demonstrating the issue of occurrence of non-uniform bias as shown in the conceptual representation in figure 1(a).

3.2. GEDI biomass recovery and vicarious calibration

The ecoregions represented a range of productivity, with some reaching only 50 Mg ha^{-1} (by median) in 48 years or $1.04 \text{ Mg ha}^{-1} \text{ yr}^{-1}$ (e.g. California Cascades), and others reaching 250 Mg ha^{-1} or $5.2 \text{ Mg ha}^{-1} \text{ yr}^{-1}$ (Washington's Olympic Peninsula and Oregon's Coast Range). In several ecoregions, carbon accumulation slowed down or reversed from



43 to 48 years of age, according to both FIA and GEDI (figure 4).

We evaluated the overall agreement between uncalibrated and calibrated GEDI-based chronosequences with reference FIA-based chronosequences using the root mean square error (RMSE) metric (table 2). A one-sided paired t -test comparing the mean RMSE (evaluated at each age group against the plot-based FIA estimate) between the two groups ($t = 0.85$, p -value = 0.20) revealed no significant enhancement in the overall fit following calibration. Uncalibrated GEDI-based chronosequence displayed a pattern similar to the conceptual representation in figure 1(b).

However, age group-specific one-sided t -tests on the absolute value of GEDI differences from FIA suggested that calibration significantly reduced model bias in the 0–5 and 15–20 year age groups (figure 5). While vicarious calibration did not significantly improve the agreement between GEDI- and FIA-based biomass estimates across all age groups, this result shows significant reduction in prediction bias in the youngest age classes.

3.3. Comparison with 2019 refined IPCC default AGBD accumulation rates

Using FIA chronosequences as ‘truth’, we conducted a comparison of AGBD accumulation rates, derived from both uncalibrated and calibrated GEDI L4A estimates as well as default values provided by the

2019 Refinement to the 2006 IPCC (Shukla *et al* 2019) (figure 6). Following IPCC precedent, estimates were aggregated to 0–19 years old and 20–48 years old (IPCC groups all secondary forests over 20 year old). All of the ecoregions tested here fall into the same IPCC domain and are therefore compared against the same default value. For the AGBD gained in the first 20 years, that IPCC value was greater than the aggregate FIA value in each of the 12 ecoregions (figure 6). Calibration significantly reduced the difference between the FIA and GEDI estimates ($t = 6.23$, $p < 0.0001$). However, calibration had no significant benefit in the older age group compared to uncalibrated GEDI estimates ($t = 1.7579$, p -value = 0.10). For the 20–48 year age group, the default 2019 IPCC recovery rate of 129 Mg ha^{-1} was lower than the FIA estimate for the most productive ecoregions.

Paired t -tests indicated that the differences between 2019 IPCC and reference FIA values were consistently significantly higher ($p < 0.05$) for both age groups than the differences between FIA and calibrated GEDI values.

4. Discussion

In much of the world, forest inventories do not support the kind of observation-based chronosequences provided here by FIA. In this study, we proposed and

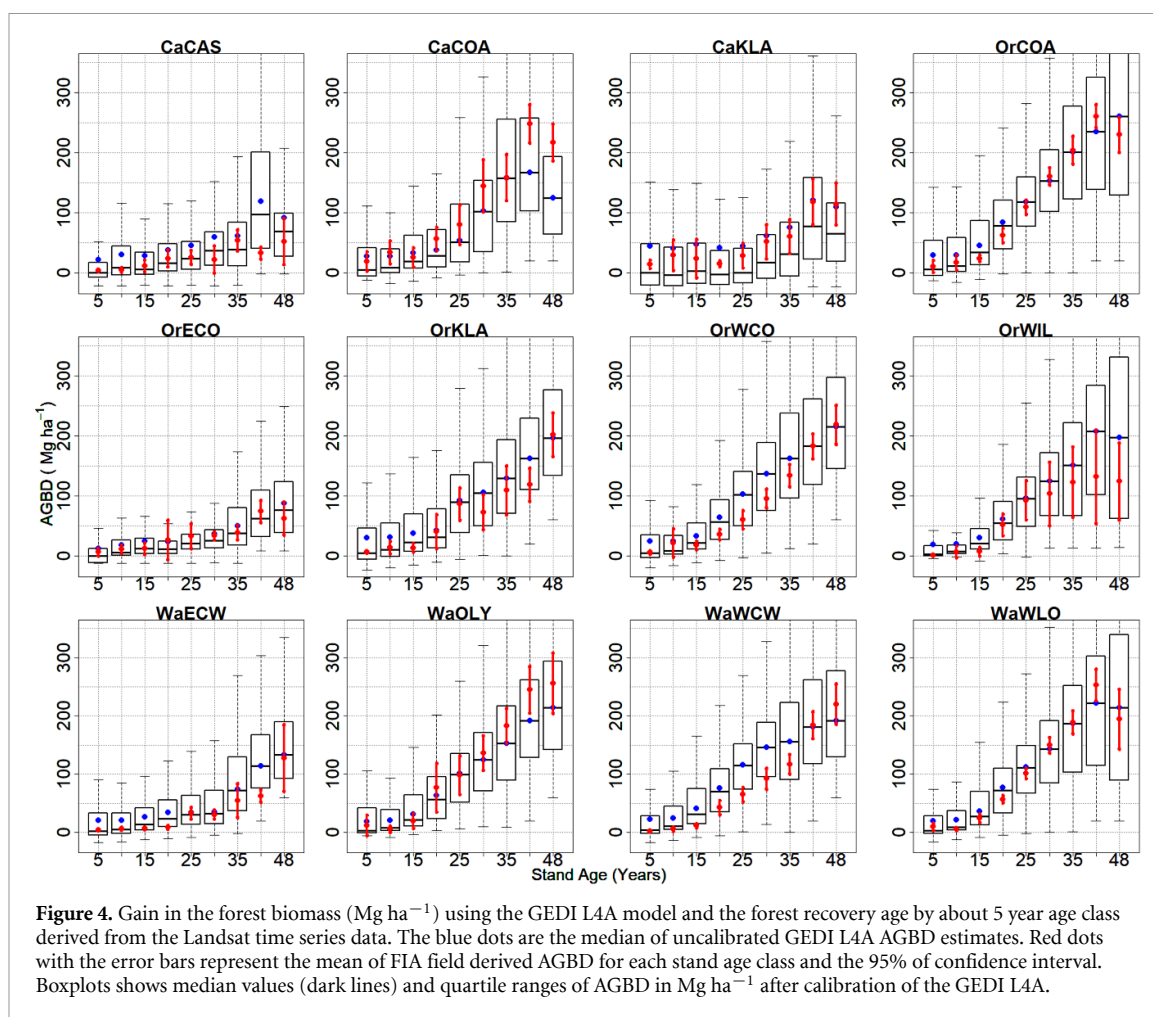


Table 2. Root means square error of AGBD estimates derived from both uncalibrated and calibrated GEDI-based methods, across age groups as assessed against FIA estimates.

S. No	Province	RMSE GEDI L4A uncalibrated (Mg ha^{-1})	RMSE GEDI L4A calibrated (Mg ha^{-1})
1	CaCAS	36.9	23.6
2	CaCOA	44.7	46.4
3	CaKLA	17.7	31.7
4	OrCOA	18.7	15.4
5	OrECO	10.7	9.7
6	OrKLA	23.2	19.7
7	OrWCO	25.4	23.4
8	OrWIL	38.6	37.2
9	WaECW	22.6	19
10	WaOLY	26.2	26
11	WaWCW	34.2	31.4
12	WaWLO	16.5	14.1

evaluated a vicarious calibration method specifically designed to better use GEDI-derived AGBD predictions to understand carbon dynamics in young, secondary forests. While calibration does not always lead to significantly improved agreement with FIA across

50 years of recovery, we found that vicarious calibration is effective in eliminating bias in early age groups (<20 years).

The context of calibrating a chronosequence, and not simply biomass stock, introduces the unique requirement that we minimize bias conditional on mapped stand age. With adequate field data, we might be able to adapt for this purpose conventional calibration methods such as histogram matching (Gilichinsky *et al* 2012), restricted imputation (Barth *et al* 2012) or classical calibration (Shukla 1972, Tellinghuisen 2000, Lindgren *et al* 2022). Our vicarious approach, which does not require locally collected field data, achieved significant improvements in mean AGBD estimates for two of the three youngest age groups tested. Below, we consider practical mapping concerns, limitations to our approach, and potential applications.

4.1. Chronosequence site selection

We created an age map by merging the stand replacement disturbance time-series from Landsat disturbance maps of 1972–1977, 1977–1984, and annual LCMS maps from 1984 to 2021. The utilization of

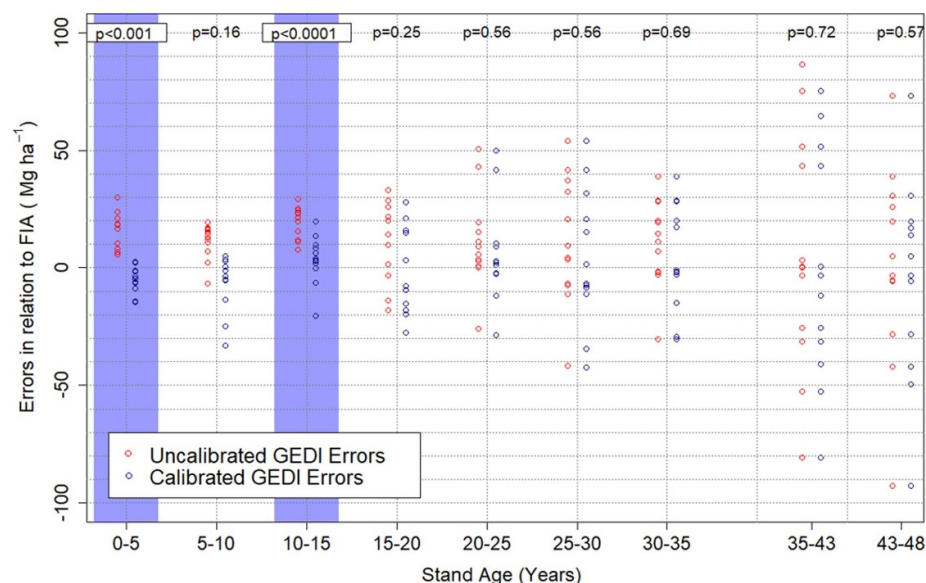


Figure 5. Errors in relation to FIA for uncalibrated and calibrated L4A estimates. Each circle point represents the difference for a specific ecoregion between the estimated FIA mean and the median of GEDI predictions with red points for uncalibrated and blue points for calibrated GEDI circle. The p -values displayed for each age group refer to a one-sided t -test comparing calibrated and uncalibrated absolute differences from FIA of median AGBD estimates across all 12 ecoregions. The vertical blue shading indicates the age-groups in which calibration significantly improved agreement with FIA.

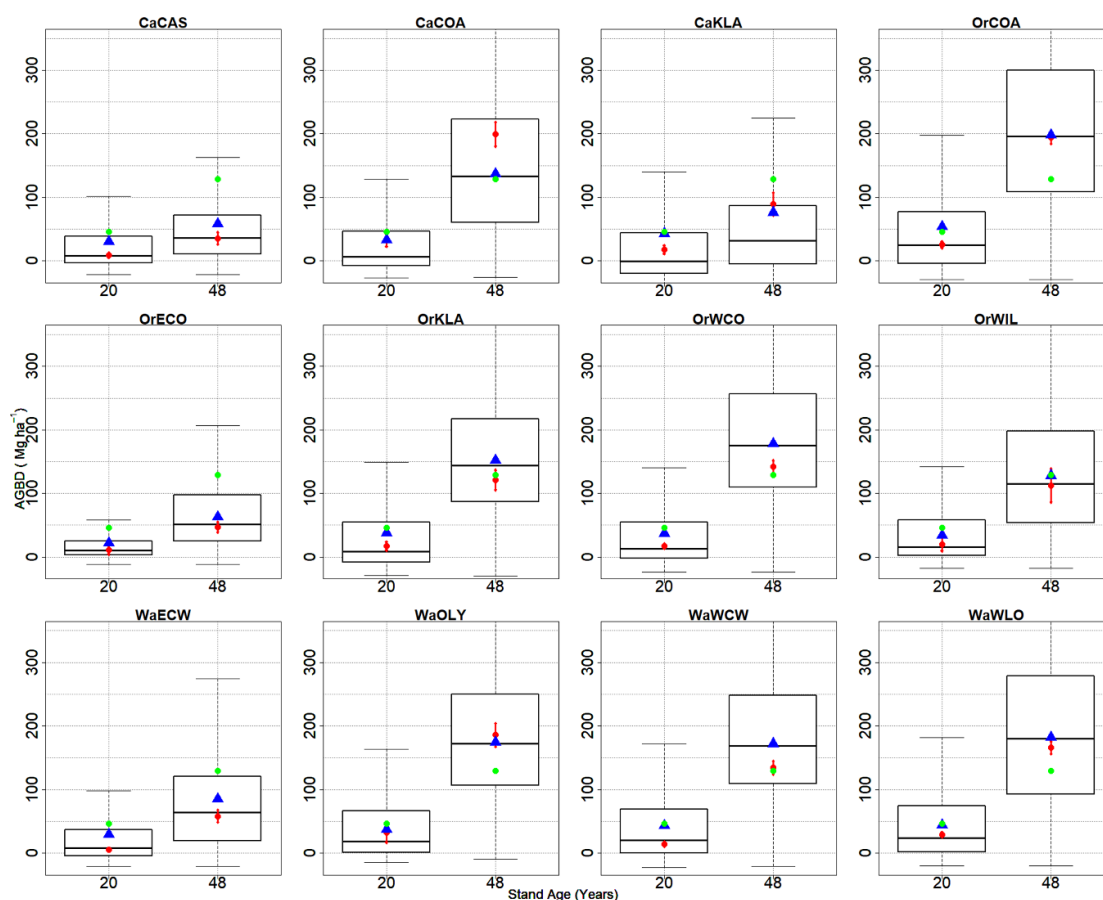


Figure 6. The biomass accumulation for the under 20 and 20–48 year age groups are depicted. The red dots, along with error bars, represent the mean values of FIA field derived AGBD with standard error and a 95% confidence interval. The blue triangle and green dots represent the median AGBD accumulation values of uncalibrated L4A estimates and the default AGBD accumulation value from 2019 IPCC, respectively. The boxplot illustrates the median values (indicated by dark lines) and quartile ranges of AGBD in Mg ha^{-1} after calibration of the GEDI L4A.

the historical MSS-derived product presents numerous challenges, including its lower resolution, inconsistencies in spatial, geometrical, and radiometric sensors, as well as a limited number of bands compared to more recent Landsat products (Banskota *et al* 2014). Despite these challenges, MSS contributed valuable insights into carbon gain during the earliest stages of stand development. A strength of the method we propose is that it only depends upon being able to identify stand-clearing disturbances, a feasible goal even with earlier MSS imagery (Renó *et al* 2011).

Our results are subject to limitations that apply to all chronosequences, as well as some factors that are specific to remotely sensed carbon recovery curves. Chronosequences depend upon simplified stand histories, and we explicitly omitted stands affected by non-stand replacing disturbances, which may in fact be common in countries attempting to account for carbon dynamics in secondary forests. Also, this kind of chronosequence approach may be ineffective in ecosystems with high tree mortality heterogeneity, leading to a multi-modal pattern of biomass accumulation (Harmon and Pabst 2015). Global change and large-scale species conversions may also affect how representative regional chronosequences might be of future forest carbon dynamics. Specific to our reliance upon GEDI, it is important to highlight that the correspondence of calibrated GEDI biomass estimates with reference data in the U.S. may not be representative of accuracy in regions where the GEDI L4A model-building data do not represent the forest population well. Our study area benefits from relatively well-fit L4A model across various ecoregions (figure 3). In areas where the GEDI L4 models may not represent local conditions, Bullock *et al* (2023) propose using available plot data opportunistically with on-orbit GEDI data to fit customized local models. Such models may require further calibration, as proposed here, when used for constructing chronosequences.

4.2. AGBD recovery trajectory

The CaCAS, CaCOA, CaKLA, OrWIL, and WaWLO ecoregions showed a drop in AGBD in the 43–50 age group, suggesting possible density-dependent mortality or a temporary age-related decline (Xu *et al* 2012, Johnson *et al* 2016, Hancock *et al* 2022). Density-dependent mortality occurs when competition near canopy closure leads to a reduction in forest biomass, primarily due to decrease in non-dominant trees (Binkley 2004). FIA-derived chronosequence corroborated this temporary decline in biomass.

In terms of uncertainty assessment, it is worth noting that we do not provide formal inference for the uncertainty of GEDI estimates, instead relying upon quantile-based box plots as informal measures of uncertainty (figures 4 and 6). Unlike the hybrid estimators proposed by Patterson *et al* (2019), we use the median as a measure of central tendency. This

decision was intended to reduce the leverage of predictions of high biomass in young stands; such predictions may be traced to both spatial errors in GEDI coordinates and imperfections in the Landsat-based age strata, which sometime include seed-tree harvests that leave large trees in otherwise cleared areas. Formal modes of model-based inference based upon the median would need to be developed for applications such as ours.

4.3. Applications

Mechanisms like REDD+ have opened the door for avoided emissions and removals to be recognized for their contributions to climate change mitigation under multiple schemes (e.g. performance-based payments, voluntary carbon markets). However, producing data needed to support initiatives such as Verified Carbon Standard (VCS-VERRA), Forest Carbon Partnership Facility (FCPF), and BioCarbon Fund Initiative for Sustainable Forest Landscapes requires a level of effort beyond what can be deemed affordable or viable to countries and program developers. The ideal network of permanent plots, strategically designed to track carbon sequestration in recovering forests age would not only be extremely expensive but also take time we cannot afford as the climate crisis unfolds. The methods presented here represent a viable opportunity to elaborate and deliver data that enables their use towards carbon accounting and the issuance of verifiable emissions removals estimates. Estimates will not only benefit ex-post but also ex-ante estimates (e.g. West *et al* 2023) that will help with intervention design and management.

Disturbances can result in significant net carbon losses to the atmosphere that may take centuries to offset (Houghton *et al* 2009, Krankina *et al* 2014). Our reference data showed that the default AGBD accumulation rate for the <20 age-group, as refined in the 2019 IPCC guidelines (Shukla *et al* 2019) were consistent overestimates (figure 6). Vicarious correction of GEDI estimates was particularly effective in this youngest age range (figures 5 and 6), yielding consistently more accurate AGBD estimates than IPCC defaults and uncalibrated GEDI estimates. Furthermore, the large number of GEDI measurements allowed sample-based AGBD estimates at finer 5 year time increments, another potential refinement of currently available defaults.

On the other hand, results cannot be disaggregated by factors often deemed important in secondary forest carbon analysis, such as: harvested vs. burned stand history (McDowell *et al* 2020); natural vs. artificial regeneration (Paquette *et al* 2009) and specific accounting of carbon in afforestation projects (Fradette *et al* 2021). Our use of GEDI's sample is appropriate for empirically determining the central tendency of carbon gain in a given area, but exploratory comparisons are limited to variables (such as age

group here) that can be mapped for purposes of establishing subsamples.

5. Conclusions

Our study demonstrates the effectiveness of combining readily available remote sensing resources, comprehensive Landsat disturbance maps and GEDI footprint-level biomass data, to create chronosequences representing carbon recovery rates in secondary forests. While GEDI's L4A biomass predictions appeared relatively unbiased when matched to inventory plots measured within 200 m of shot center, substantial design bias occurred when predicted biomass was compared to field biomass as a function of mapped stand age. Our simple vicarious calibration process significantly improved GEDI agreement with inventory results in the youngest age groups, critical to understanding secondary forest regrowth and the carbon benefits that secondary forests convey. Replicating this methodology elsewhere is likely to enhance the uniformity and representativeness of greenhouse gas reporting and management.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This research was funded by NASA Grants 80HQTR21T0013 (GEDI Science Team), 80HQTR21T0020 (Land Cover and Land Use Change Program), and NNL22OB15A (GEDI mission). Support was also provided by the USDA Forest Service Rocky Mountain Research Station. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy. We are especially grateful to Dr Naikoa Aguilar Amuchastegui of the World Bank for constructive insights related to applications. We further thank the GEDI Mission Team, particularly Drs. John Armston, Laura Duncanson, and James Kellner.

ORCID iDs

Nidhi Jha  <https://orcid.org/0000-0002-2569-1595>
 Sean P Healey  <https://orcid.org/0000-0003-3498-4266>
 Göran Ståhl  <https://orcid.org/0000-0001-9030-8057>
 Matthew G Betts  <https://orcid.org/0000-0002-7100-2551>

References

- ART Secretariat 2020 The REDD+ environmental excellence standard (TREES) *Architecture for REDD+ Transactions Program*
- Banskota A, Kayastha N, Falkowski M J, Wulder M A, Froese R E and White J C 2014 Forest monitoring using Landsat time series data: a review *Can. J. Remote Sens.* **40** 362–84
- Barth A, Lind T and Ståhl G 2012 Restricted imputation for improving spatial consistency in landscape level data for forest scenario analysis *For. Ecol. Manage.* **272** 61–68
- Bechtold W A and Patterson P L (ed) 2005 The enhanced forest inventory and analysis program—national sampling design and estimation procedures *General Technical Reports SRS-80* (U.S. Department of Agriculture, Forest Service, Southern Research Station) p 080
- Binkley D 2004 A hypothesis about the interaction of tree dominance and stand production through stand development *For. Ecol. Manage.* **190** 265–71
- Buendia L, Miwa K, Ngara T and Tanabe K 2019 Agriculture, forestry and other land use *Intergovernmental Panel on Climate Change (IPCC) Guidelines for National Greenhouse Gas Inventories* vol 4 (IPCC National Greenhouse Gas Inventories Programme Technical Support Unit, Institute for Global Environmental Strategies) p Ap
- Bullock E L et al 2023 Estimating aboveground biomass density using hybrid statistical inference with GEDI lidar data and Paraguay's national forest inventory *Environ. Res. Lett.* **18** 085001
- Burrill E A et al 2021 The forest inventory and analysis database: database description and user guide version 9.0 for phase 2 (U.S. Department of Agriculture, Forest Service) (available at: www.fia.fs.fed.us/library/database-documentation/)
- Chave J et al 2019 Ground data are essential for biomass remote sensing missions *Surv. Geophys.* **40** 863–80
- Cohen W, Yang Z and Kennedy R 2010 Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—tools for calibration and validation *Remote Sens. Environ.* **114** 2911–24
- Dobor L, Hlásny T, Rammer W, Barka I, Trombik J, Pavlenda P, Šebeň V, Štěpánek P and Seidl R 2018 Post-disturbance recovery of forest carbon in a temperate forest landscape under climate change *Agric. For. Meteorol.* **263** 308–22
- Dubayah R et al 2020 The global ecosystem dynamics investigation: high-resolution laser ranging of the Earth's forests and topography *Sci. Remote Sens.* **1** 100002
- Dubayah R et al 2022 *GEDI L4A Footprint Level Aboveground Biomass Density, Version 2.1*. ORNL DAAC, Oak Ridge, Tennessee, USA (ORNL DAAC) (<https://doi.org/10.3334/ORNLDAC/2056>)
- Duncanson L et al 2022 Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission *Remote Sens. Environ.* **270** 112845
- FCPF 2021 Forest Carbon Partnership Facility: Buffer Guidelines Version 3.1
- Forest Ecosystem Management Assessment Team (US) 1993 *Forest Ecosystem Management: An Ecological, Economic, and Social Assessment* (The Service)
- Fradette O, Marty C, Faubert P, Dessureault P-L, Paré M, Bouchard S and Villeneuve C 2021 Additional carbon sequestration potential of abandoned agricultural land afforestation in the boreal zone: a modelling approach *For. Ecol. Manage.* **499** 119565
- Franklin J F and Dyrness C T 1973 Natural vegetation of Oregon and Washington (US Forest Service Research and Development) (available at: www.fs.usda.gov/research/treearch/26203)
- Gilichinsky M, Heiskanen J, Barth A, Wallerman J, Egberth M and Nilsson M 2012 Histogram matching for the calibration of kNN stem volume estimates *Int. J. Remote Sens.* **33** 7117–31

- Hancock M, Sitch S, Fischer F J, Chave J, O'Sullivan M, Fawcett D and Mercado L M 2022 Modelling the impact of wood density dependent tree mortality on the spatial distribution of Amazonian vegetation carbon *Biogeosci. Discuss.* **2022** 1–37
- Harmon M E and Pabst R J 2015 Testing predictions of forest succession using long-term measurements: 100 yrs of observations in the Oregon Cascades *J. Veg. Sci.* **26** 722–32
- Healey S, Cohen W, Spies T A, Moeur M, Pflugmacher D, Whitley M G and Lefsky M 2008 The relative impact of harvest and fire upon landscape-level dynamics of older forests: lessons from the northwest forest plan *Ecosystems* **1** 1106–19
- Houghton R A, Hall F and Goetz S J 2009 Importance of biomass in the global carbon cycle *J. Geophys. Res. Biogeosci.* **114** G00E03
- Jha N et al 2020 Forest aboveground biomass stock and resilience in a tropical landscape of Thailand *Biogeosciences* **17** 121–34
- Johnson M O et al 2016 Variation in stem mortality rates determines patterns of above-ground biomass in Amazonian forests: implications for dynamic global vegetation models *Glob. Change Biol.* **22** 3996–4013
- Kabir S, Leigh L and Helder D 2020 Vicarious methodologies to assess and improve the quality of the optical remote sensing images: a critical review *Remote Sens.* **12** 4029
- Kellner J R, Armston J and Duncanson L 2023 Algorithm theoretical basis document for GEDI footprint aboveground biomass density *Earth Space Sci.* **10** e2022EA002516
- Koch A and Kaplan J O 2022 Tropical forest restoration under future climate change *Nat. Clim. Change* **12** 279–83
- Krankina O, Dellasala D, Leonard J and Yatskov M 2014 High-biomass forests of the Pacific Northwest: who manages them and how much is protected? *Environ. Manage.* **54** 112–21
- Lindgren N, Nyström K, Saarela S, Olsson H and Ståhl G 2022 Importance of calibration for improving the efficiency of data assimilation for predicting forest characteristics *Remote Sens.* **14** 4627
- Lozada Dávila J R, Soriano P, Costa M, García-Quintero A M, Sánchez D, Villarreal A and Arends E 2020 Long-term carbon stock recovery in a neotropical-logged forest *Plant Biosyst.* **154** 241–7
- McDowell N G et al 2020 Pervasive shifts in forest dynamics in a changing world *Science* **368** eaaz9463
- Oregon Revised Statute (ORS) 2023 Pest control Title 44-Forestry and Forest Products vol 14 ch 527, § 527.676 (available at: <https://casetext.com/statute/oregon-revised-statutes/title-44-forestry-and-forest-products/chapter-527-pest-control-forest-practices/oregon-forest-practices-act/generally/section-527676-leaving-snags-and-downed-logs-in-harvest-type-2-or-3-units-green-trees-to-be-left-near-certain-streams>)
- Paquette A, Hawryshyn J, Senikas A V and Potvin C 2009 Enrichment planting in secondary forests: a promising clean development mechanism to increase terrestrial carbon sinks *Ecol. Soc.* **14**
- Patterson P L et al 2019 Statistical properties of hybrid estimators proposed for GEDI—NASA's global ecosystem dynamics investigation *Environ. Res. Lett.* **14** 065007
- Pearlshtien D H, Pignatti S and Ben-Dor E 2023 Vicarious CAL/VAL approach for orbital hyperspectral sensors using multiple sites *Remote Sens.* **15** 771
- Poorter L et al 2016 Biomass resilience of neotropical secondary forests *Nature* **530** 211–4
- 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 *Proc. Natl Acad. Sci.* **116** 4382–7
- R Core Team 2022 R: a language and environment for statistical computing (available at: <https://cir.nii.ac.jp/crid/1370294721063650048>)
- Renó V F, Novo E M L M, Suemitsu C, Rennó C D and Silva T S F 2011 Assessment of deforestation in the Lower Amazon floodplain using historical Landsat MSS/TM imagery *Remote Sens. Environ.* **115** 3446–56
- Rozendaal D M A and Chazdon R L 2015 Demographic drivers of tree biomass change during secondary succession in northeastern Costa Rica *Ecol. Appl.* **25** 506–16
- Schepaschenko D et al 2019 The forest observation system, building a global reference dataset for remote sensing of forest biomass *Sci. Data* **6** 198
- Shukla G K 1972 On the problem of calibration *Technometrics* **14** 547–53
- Shukla P R, Skea J, Calvo Buendia E, Masson-Delmotte V, Pörtner H O, Roberts D C and Malley J 2019 Climate change and land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (IPCC)
- Ståhl G et al 2024 Why ecosystem characteristics predicted from remotely sensed data are unbiased and biased at the same time—and how this affects applications *For. Ecosyst.* **11** 100164
- Tellinghuisen J 2000 Inverse vs. classical calibration for small data sets *Fresen. J. Anal. Chem.* **368** 585–8
- USDA Forest Service 2022 USFS Landscape Change Monitoring System V2021.7 (Conterminous United States and Southeastern Alaska)
- Vangi E, D'Amico G, Francini S and Chirici G 2022 GEDI4R: an R package for NASA's GEDI level 4 A data downloading, processing and visualization *Earth Sci. Inf.* **16** 1109–17
- West T A P, Wunder S, Sills E O, Börner J, Rifai S W, Neidermeier A N and Kontoleon A 2023 Action needed to make carbon offsets from tropical forest conservation work for climate change mitigation (arXiv:2301.03354)
- Wulder M A et al 2022 Fifty years of Landsat science and impacts *Remote Sens. Environ.* **280** 113195
- Xu C-Y, Turnbull M H, Tissue D T, Lewis J D, Carson R, Schuster W S F, Whitehead D, Walcroft A S, Li J and Griffin K L 2012 Age-related decline of stand biomass accumulation is primarily due to mortality and not to reduction in NPP associated with individual tree physiology, tree growth or stand structure in a Quercus-dominated forest *J. Ecol.* **100** 428–40