*Global Ecology and Biogeography*

WILEY

Global Ecology and Biogeography | A Journal of Macroecology

PERSPECTIVE

# Leveraging Remote Sensing and Theory to Predict Tree Size Abundance Distributions Across Space

Adam J. Eichenwald[1] | John M. Grady[2,3] | Jonathan A. Knott[4] | J. Marcos Rodriguez[1†] | Ben Weinstein[5] | David A. Orwig[6] | Sydne Record[1]

[1]Department of Wildlife, Fisheries, and Conservation Biology, University of Maine, Orono, Maine, USA | [2]Maine Agricultural and Forest Experiment Station, Orono, Maine, USA | [3]Living Earth Collaborative, Washington University in St. Louis, St. Louis, Missouri, USA | [4]USDA Forest Service, St. Paul, Minnesota, USA | [5]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida, USA | [6]The Harvard Forest, Harvard University, Petersham, Massachusetts, USA

**Correspondence:** Adam J. Eichenwald (adam.eichenwald@maine.edu)

**Received:** 24 September 2024 | **Revised:** 9 June 2025 | **Accepted:** 20 June 2025

**Handling Editor:** Thiago Sanna Freire Silva

**Funding:** This work was supported by National Institute of Food and Agriculture, MEO-022425; National Aeronautics and Space Administration, 80NSSC23K0421.

**Keywords:** allometry | Bayesian | canopy | forest abundance | ForestGEO | NEON | satellite | scaling | size distributions | understory interpolation

## ABSTRACT

**Background:** Remote sensing (RS) technologies provide unprecedented opportunities to assess forest structure at broad spatial scales. Light detection and ranging (LiDAR) is a powerful tool that offers detailed vertical information, but its consistent high-resolution coverage can be limited across vast areas and quantifying understory vegetation remains challenging due to occlusion. Conversely, while high-resolution RGB imagery is more accessible and valuable for large-scale analyses, it comes with higher uncertainty and only captures canopy-level information. Integrating RGB data with ecological theory and existing LiDAR-derived products (e.g., canopy height models, CHM) allows us to enhance predictions and broaden the applicability of forest size structure mapping from the understory to the canopy, particularly through tree size–abundance relationships.

**Theory:** Tree size and abundance generally follow known distributions, where smaller trees are exponentially more common than larger trees. This pattern emerges from fundamental ecological constraints, including volume packing, metabolic scaling and light competition, which collectively govern forest self-organisation across spatial scales.

**Proposed Approach:** We present a workflow that integrates size-abundance scaling theory with RGB data and CHMs to improve predictions of forest structure. By estimating tree size distributions from RGB imagery, we extend the applicability of ecological scaling models to broader spatial scales and offer an approach that complements traditional methods.

**Assessment:** We apply this approach to forest monitoring networks, including NEON and ForestGEO, to assess its accuracy and generalisability across a range of forests (e.g., subtropical pine, boreal, low and dry, tall and dense) in North America. Our results show that RGB-based estimates can successfully recover size-abundance distributions from the understory to the canopy.

**Future Directions:** Further refinement of our approach could enhance predictions by incorporating species classifications from hyperspectral data and using more spatially or taxonomically specific allometric equations. These additions would enable more precise scaling of tree diameter from remote sensing data.

[†]Deceased

Post acceptance, Quentin D. Read was removed as an author from this article by mutual agreement of the authors. However, J. Marcos Rodriguez was deceased at the time of this decision, and as such, agreement to change the author list could not be obtained from them.

## 1 | Introduction

The advent of remote sensing (RS) technologies offers unprecedented opportunities to characterise and predict vegetation at large biogeographic scales (Pettorelli et al. 2016; Read et al. 2020) (Figure 1). This advancement is crucial, as trees play a key role in global change by shaping land-atmosphere exchanges, driving carbon sequestration, supporting biodiversity and regulating hydrological processes (Bonan 2008; Harris et al. 2021; Pan et al. 2024). Successful predictions, however, require integrating RS data with precise ground measurements, and harmonising temporal and spatial scales between RS and in situ data is challenging (Anderson 2018; Record et al. 2020; Knott et al. 2023). Furthermore, airborne or spaceborne RS methods rely on information that is observable from above (such as tree canopies), whereas many ecological processes such as recruitment and competition happen below the canopy. As a result, there are physical limitations on our capacity to use RS to investigate underlying mechanisms shaping plant communities across different spatial scales.

Fortunately, ecological theory provides a foundation for predicting ecological mechanisms based on measurable plant characteristics, and RS offers a means to assess these characteristics at large spatial extents. Tree size, for example, is a key RS-measurable trait that governs forest structure, abundance patterns and carbon storage (West et al. 2009; Hubau et al. 2019). Ecological scaling theory predicts that tree size and abundance approximately follow a negative power law (Yoda et al. 1963; Enquist et al. 1999; Picard and Gasparotto 2016) and recent work has successfully applied this theory using airborne LiDAR on forest canopies to recover size-abundance distributions (Fischer et al. 2020; Taubert et al. 2021). While airborne LiDAR data provide highly detailed structural information, their acquisition and processing can be expensive and computationally intensive, limiting their practical application for frequent or large-scale monitoring. Therefore, complementary methods that use data sources with greater temporal frequency and/or broader spatial coverage can improve predictions for regions and time periods where dedicated LiDAR acquisitions are not available. High-resolution RGB imagery, for example, is more readily available at broad spatial extents and offers unique advantages for individual tree delineation and frequent monitoring. Developing methods to estimate tree size-abundance distributions from RGB data could greatly enhance our ability to apply ecological scaling theory to predict vegetation structure across geographic scales and data availability scenarios.

Here, we propose integrating size-abundance scaling theory with individual tree crown delineations from high-resolution RGB imagery to improve predictions of forest size-abundance distributions. This approach first delineates individual tree crowns from high-resolution RGB imagery to create an initial tree size distribution of visible canopy trees. Subsequently, scaling theory is applied to infer the abundance of occluded understory trees. This method's robustness and broad applicability are evidenced by its successful testing across a wide range of forests (e.g., subtropical pine forests in Florida, boreal forests in Alaska, tall and dense canopies in the Pacific Northwest, low and dry forests in Puerto Rico). We present a workflow and case study demonstrating the potential of this approach and outline some of the technical opportunities and challenges therein. Finally, we discuss possible extensions of this approach.

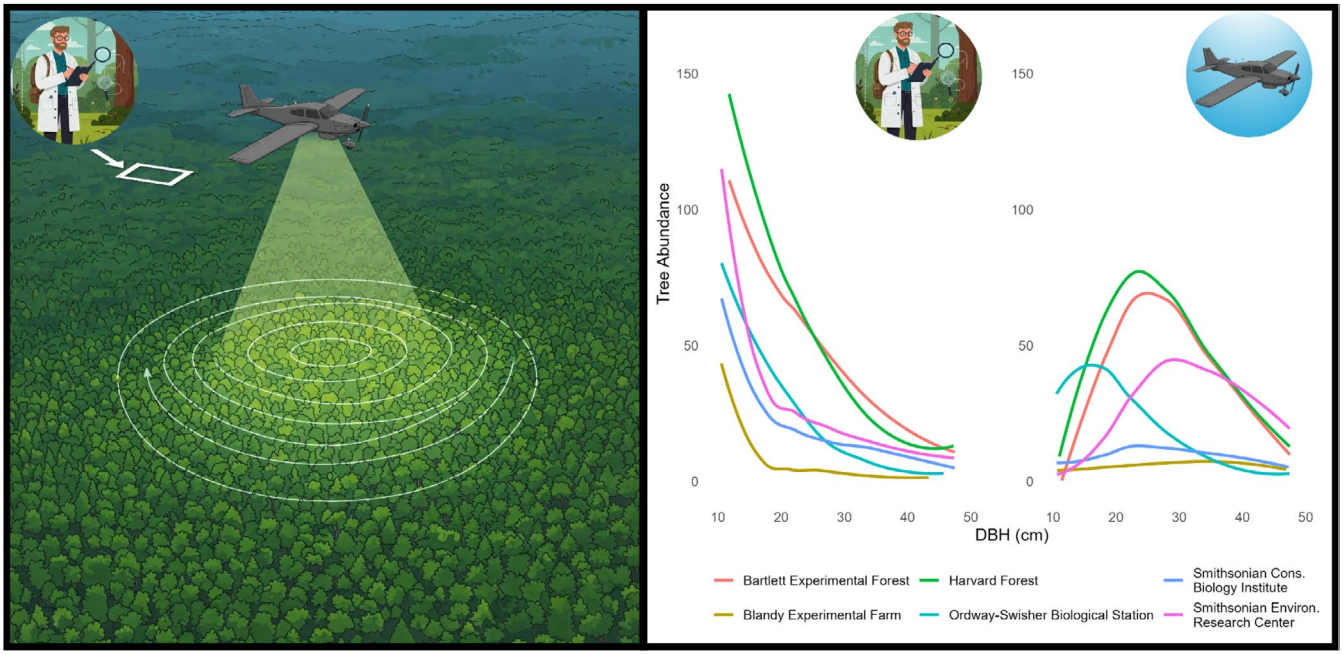


**FIGURE 1** | (Left) Field surveys provide high-quality forest inventory data but cover limited areas, leaving large gaps that could otherwise inform further analyses. Remote sensing (RS) can fill these gaps, measuring vast areas efficiently. However, (right) while field-surveyed data follow known distributions, remotely sensed data deviate below a site-specific threshold due to canopy occlusion, which obscures smaller understory trees. Graphs plot data from select National Ecological Observatory Network (NEON) sites. Clipart images were created with the assistance of Google's Imagen, an AI language model.

### 1.1 | Size-Abundance Scaling Theory

Scaling approaches to forest structure quantify the size dependence of abundance from saplings to the canopy, linking individual traits (size) to population attributes (abundance). In general, forests exhibit a pattern where small trees are more abundant than large trees, although the exact functional form of this relationship remains debated (Yoda et al. 1963; Enquist and Niklas 2001; Muller-Landau et al. 2006; West et al. 2009; Picard and Gasparotto 2016). Several interacting mechanisms have been proposed to explain this trend, including volume packing constraints (Yoda et al. 1963), stochastic packing theory (Taubert et al. 2015), growth–mortality trade-offs (Muller-Landau et al. 2006), energy availability (Damuth 1981; West et al. 2009) and light limitation in canopy gaps (Farrior et al. 2016). These factors likely work in combination to shape observed size–abundance relationships in forests.

Various statistical distributions have been used to model these relationships. Power-law models, such as the Pareto distribution, are commonly applied due to their simplicity and theoretical grounding in self-thinning and metabolic scaling theories (West et al. 2009). These models assume a constant scaling exponent across size classes, making them relatively easy to estimate and interpret in the context of theory. However, they often fail to capture deviations at the smallest and largest tree sizes (i.e., there are lower and upper bounds to the distribution). Alternative models, such as the Weibull distribution, can provide a more flexible approach, allowing for different rates of decline across size classes that may better capture the full range of observed tree sizes (Muller-Landau et al. 2006).

### 1.2 | A Workflow Integrating Theory and Remotely Sensed RGB Imagery

Estimating size-abundance scaling from remote sensing data generally involves three key steps: individual tree crown delineation, inference of tree dimensions and the fitting of appropriate size-abundance distributions. Our workflow is adaptable to any method providing tree segmentation (e.g., from LiDAR, hyperspectral or RGB imagery). In this study, delineation is performed using high-resolution RGB imagery. Our approach incorporates a Bayesian model designed to account for occlusion, explicitly adjusting for the fact that only a subset of the full tree size distribution is directly observed. The probability distribution used (e.g., Pareto) can be substituted with alternative distributions (e.g., Weibull) within the same framework, but we showcase the method using the Pareto distribution due to its relative simplicity. We fit the Pareto by truncating the smallest and largest trees to ensure a reliable fit to the intermediate size classes.

Successfully interpolating the size-abundance relationship for a given forest plot using a Pareto distribution requires us to estimate two parameters: $\alpha$ and $N_{tot}$. The Pareto distribution is defined as:

$$f(x) = \frac{\alpha x_{\min}^{\alpha}}{x^{\alpha+1}}, x \geq x_{\min}, \alpha > 0$$

Here, the shape parameter $\alpha$ describes the steepness or "tail-heaviness" of the distribution. The scale parameter $x_{min}$ is the minimum value for $x$ (tree size, represented as diameter at breast height (DBH)), and the distribution is defined only for $x \geq x_{min}$. The parameter $\alpha$ is often used to describe the shape of size-abundance relationships (Grady et al. 2024). $N_{tot}$ represents the total number of trees at the site above $x_{min}$, which scales the size-abundance relationship from a relative density function to an absolute abundance estimate.

We would generally recommend setting $x_{min}$ based on the constraints of the observed in situ data used to parameterise models and ecological knowledge of the system. For example, the United States Forest Inventory and Analysis (FIA) programme utilises a variable plot design for trees of DBH below 12.7 cm, so $x_{min}$ for predictive models parameterised with FIA data should not be lower than 12.7 cm. Taubert et al. (2021) found that size-abundance scaling for trees deviates from power law expectations for individuals below 10 cm DBH, so we caution against setting $x_{min}$ lower than 10 cm. With $x_{min}$ pre-set, $\alpha$ is the only parameter needed for an estimation of the probability distribution.

### 1.3 | Estimating $\alpha$

Our method of calculating $\alpha$ is split into 3 parts: (1) determining $x_{breakpoint}$, (2) calculating informed priors and (3) fitting a Bayesian model.

#### 1.3.1 | Determining $x_{breakpoint}$

Directly fitting a distribution to remotely sensed data incorrectly generates a shape reminiscent of a normal distribution (Figure 1). This is because DBH is allometrically linked to height; taller trees generally have larger diameters. Consequently, as DBH decreases, the likelihood of smaller trees being occluded by the canopies of taller trees increases (Figures 1 and 2c). There is a certain point $x_{breakpoint}$ above which remotely sensed trees should generally not be occluded due to their size and position in the canopy. Therefore, only remotely sensed individuals with sizes $\geq x_{breakpoint}$ have crowns that may be fully delineated with RS. We expect $x_{breakpoint}$ to vary with site and plot location due to changes in crown architecture of the target trees and their neighbours. We also expect upper limits on the applicability of the Pareto distribution (Farrior et al. 2016), which we define as $x_{upper}$. For simplicity, and based on previous research, we set $x_{upper} = 50$ cm (Taubert et al. 2021; Grady et al. 2024; Eichenwald et al. In Review).

Identifying the truncation point $x_{breakpoint}$ is circular: we need to spot deviations from expected densities to set the cutoff, but estimating those densities depends on fitting the distribution. We addressed this using kernel density estimation (KDE), a non-parametric method that approximates the data distribution from observed DBH values. We selected bandwidth via Sheather and Jones' (1991) method. Gaussian kernel density analysis can show artificial dips in density where there are gaps in our observed DBH data (i.e., zero trees of certain

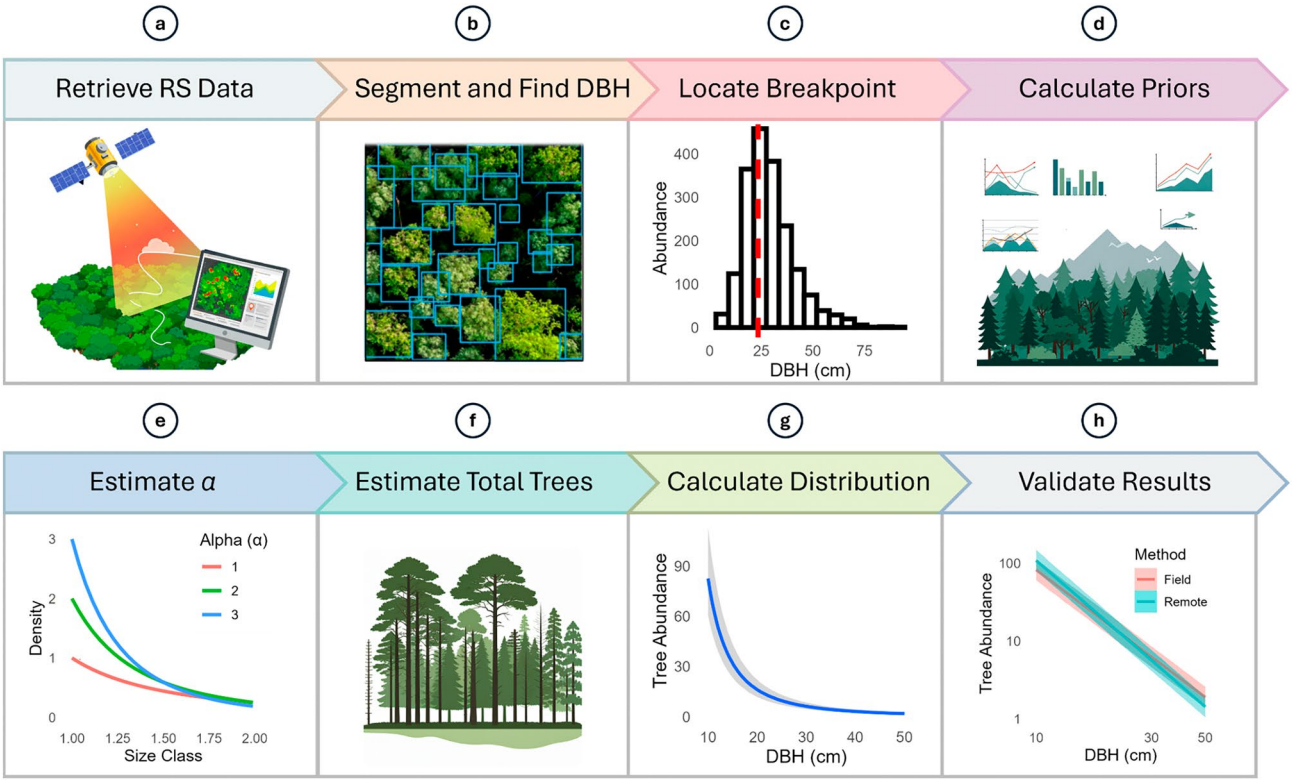


**FIGURE 2** | Visual representation of our method for estimating the size-abundance curve for a given forest based on remotely sensed (RS) data from the canopy. (a) RS data, including canopy height and red-green-blue (RGB) imagery, must be retrieved, either from an online database or gathered by drone. (b) A canopy segmentation algorithm is applied to create polygons of individual trees, and then the DBH of each tree is recovered via allometric equations utilising both canopy height and diameter as parameters. (c) The breakpoint $x_{\text{breakpoint}}$ for the RS data, representing the DBH above which the size-abundance curve from RS data follows an known size-abundance distribution, is estimated (trees below this threshold are assumed to be occluded, and only data with DBh $\geq x_{\text{breakpoint}}$ is used in subsequent steps). (d) Informed priors for the $\alpha$ parameter of a Pareto distribution are derived from environmental data for each plot (e.g., maximum tree height, elevation), using a random forest model trained on forest inventory data. (e) The posterior distribution for $\alpha$ is then estimated with a Bayesian model. (f) Uncertainty in the posterior distribution for $\alpha$ is propagated into the next step, where a separate Bayesian model is used to estimate a posterior distribution for the total number of trees ($N_{tot}$) in the plot. (g) A final size-abundance distribution with uncertainty is created by fitting the Pareto distribution with samples from the estimated $\alpha$ posterior distribution and multiplied by samples from the estimated $N_{\text{tot}}$ posterior distribution. Because the interpolated distribution is only estimated up to the threshold of 50 cm DBH, it can be extended by directly appending empirically observed trees from remote crown segmentation for DBH > 50 cm. This approach avoids double-counting, as the interpolation uses only trees ≤ 50 cm, while the segmentation data are reliable only for larger trees. (h) The final size-abundance distribution(s) can then be validated with in situ data (which will likely cover a smaller subset of a given area of interest). The end result allows for a large-scale raster dataset predicting forest size-abundance distributions. Clipart in steps (a), (d), and (f) were created with the assistance of Google's Imagen, an AI language model.

sizes). Since these dips reflect unobserved ranges rather than a true absence of trees, we filtered the density output, retaining only the values corresponding to observed DBH data. This ensured our density estimates were based only on actual measurements, preventing data gaps from being mistaken for real declines in tree density.

The KDE curve typically exhibits local maxima, one of which corresponds to $x_{\text{breakpoint}}$. This reflects the pattern that, above $x_{\text{breakpoint}}$, tree sizes follow a Pareto distribution with decreasing density, while occlusion causes reduced densities below $x_{\text{breakpoint}}$ (Figures 1 and 2c). However, additional peaks can occur depending on the data, requiring further steps to identify the correct $x_{\text{breakpoint}}$. To isolate $x_{\text{breakpoint}}$, we leveraged a key property of the Pareto distribution: in $\log_{10}$-$\log_{10}$ space, it forms a straight line with a slope of $-(\alpha+1)$ above $x_{\text{breakpoint}}$. We fit a piecewise regression to the $\log_{10}$-transformed data using the segmented() function in the *segmented* package in R (Muggeo 2020), allowing the breakpoint to be estimated from the data. We identified $x_{\text{breakpoint}}$ as the KDE peak nearest to this breakpoint.

### 1.3.2 | Calculating Informed Priors

Our priors for $\alpha$ were informed by established relationships between environmental conditions and forest structure. Duncanson et al. (2015) showed that tree size distributions vary predictably across environmental gradients in the United States. Building on this, we trained a random forest model using FIA data to predict $\alpha$ from maximum tree height (m), latitude, longitude and elevation. To better account for spatial autocorrelation, we evaluated model performance using spatial cross-validation (10 spatial folds, 75% training fraction) Ploton et al. 2020. The

model exhibited moderate explanatory power, with a median $R^2$ of 0.323 (Median Absolute Deviation (MAD) = 0.044) and a root mean square error (RMSE) ranging from 0.548 to 0.660. Due to the large propagated uncertainties, the prior is only weakly constrained.

#### 1.3.3 | Fitting a Bayesian Model

To recover $\alpha$ for a given site, we employed a Bayesian statistical framework to directly fit a truncated Pareto distribution. This is achieved by normalising the PDF over the observed range of sizes in units of DBH using the cumulative distribution function (CDF):

$$p_{\text{trunc}} = F\left(x_{\text{upper}} \mid x_{\min}, \alpha\right) - F\left(x_{\text{breakpoint}} \mid x_{\min}, \alpha\right)$$

where $F(x|x_{\min}, \alpha)$ is the CDF of the Pareto distribution and $p_{\text{trunc}}$ is the normalising constant that adjusts the probability distribution for truncation. The CDF gives the cumulative probability of observing a value less than or equal to a specific diameter, allowing for the calculation of the proportion of the total probability mass that lies within the observed range. This proportion, referred to as the normalising constant, is the difference in the CDF values at the upper and lower truncation points. By normalising the probability density function (PDF) with respect to this constant, we adjust the likelihood of observing each data point within the truncated range. This ensures that the total probability within the observed range sums to 1, accounting for the fact that the data are only available within a specific range of diameters. The likelihood for each observed data point is then:

$$\log\ \text{likelihood} = \text{adjust} * \log\left(F\left(x \mid x_{\min}, \alpha\right)\right) - \log\left(p_{\text{trunc}}\right)$$

For each observed data point, the likelihood is the probability density that the data point comes from the Pareto distribution with the given parameters (see below for explanation of adjustment factor). By summing the log-likelihoods across all observations, we obtain the total likelihood of the data given the model parameters. The model then maximises this total likelihood to find the parameter values that best explain the observed data.

This method assumes perfect segmentation for trees with DBH $\geq x_{\text{breakpoint}}$. However, segmentation accuracy declines in dense canopies (often measured by Leaf Area Index (LAI) Yang et al. 2019). High LAI increases two common errors: (1) under-segmentation, where multiple trees are grouped under a dominant crown, and (2) over-segmentation, where a single crown is split into multiple segments (Breidenbach and Astrup 2014). Our approach accounts for under-segmentation by assuming trees below $x_{\text{breakpoint}}$ are obscured, but over-segmentation becomes more likely above $x_{\text{breakpoint}}$ in dense forests. Higher LAI should therefore increase over-segmentation of larger trees. The position of $x_{\text{breakpoint}}$ relative to $x_{\min}$ introduces additional uncertainty. When $x_{\text{breakpoint}}$ is far from $x_{\min}$, more of the size-abundance distribution is inferred rather than observed, increasing reliance on model-based interpolation. Closer alignment between $x_{\text{breakpoint}}$ and $x_{\min}$ reduces this uncertainty. These effects interact: a high LAI with $x_{\text{breakpoint}}$ near $x_{\min}$ requires less adjustment, whereas a high LAI and distant $x_{\text{breakpoint}}$ amplify uncertainty.

To address these sources of uncertainty, we incorporated an adjustment factor that accounts for LAI and the location of $x_{\text{breakpoint}}$ relative to $x_{\min}$, defined as:

$$\text{adjust} = 1 - \sqrt{\text{LAI}_{\text{norm}} * x_{\text{breakpoint,norm}}}$$

where $\text{LAI}_{\text{norm}}$ and $x_{\text{breakpoint,norm}}$ are normalised metrics ranging from 0 to 1. $\text{LAI}_{\text{norm}}$ is calculated by dividing the observed LAI for the site as measured by the MODIS MOD15A2H product (Myneni et al. 2021) by 100, which is the global maximum for the MODIS data product. Similarly, $x_{\text{breakpoint,norm}}$ is calculated as $\frac{(x_{\text{breakpoint}} - x_{\min})}{x_{\max} - x_{\min}}$ with higher values indicating that $x_{\text{breakpoint}}$ is farther from $x_{\min}$. The adjustment factor penalises the likelihood, reducing the influence of observations from areas with high LAI and distant $x_{\text{breakpoint}}$ values that have greater uncertainty.

### 1.4 | Estimating $N_{\text{tot}}$

To estimate the total abundance of trees $N_{\text{tot}}$ with DBH between $x_{\min}$ and $x_{\max}$, we developed a Bayesian hierarchical model that integrates observed tree counts across DBH bins with a truncated Pareto distribution. Tree DBH observations ($N_{\text{obs},k}$) were grouped into $k$ bins, each defined by a minimum ($x_{\min,k}$) and maximum ($x_{\max,k}$) DBH. Although using a single bin from $x_{\min}$ to $x_{\max}$ would theoretically simplify the model by directly comparing the total observed counts $N_{\text{obs}}$ to the expected value across the range, aggregating all data into one bin would cause the likelihood to depend heavily on the exact $N_{\text{tot}}$. Using a larger number of bins (we chose to use 8 bins, see Figure S1 for a sensitivity analysis for bin number) mitigates this issue by spreading the likelihood contribution across multiple segments of the size distribution. This approach allows the model to capture patterns in the distribution of tree sizes rather than overemphasising the total count alone. We use the *logbin* function from the *forestscaling* R package (White et al. 2008; Grady et al. 2024) to bin tree diameters. The expected count of trees in each bin ($\lambda_k$) was calculated as:

$$\lambda_k = N_{\text{tot}} * \frac{x_{\min,k}^{1-\alpha} - x_{\max,k}^{1-\alpha}}{x_{\min}^{1-\alpha} - x_{\max}^{1-\alpha}}$$

where the numerator accounts for the proportion of the distribution that falls within bin $k$. The denominator normalises $\lambda_k$ over the full range of the truncated Pareto. To account for uncertainty in the parameter $\alpha$, we marginalised over a set of pre-sampled values ($n_{\text{alpha}}$) drawn from the posterior distribution calculated in the previous section. For each sampled $\alpha$, the likelihood of the observed count ($N_{\text{obs}}$) was calculated using a normal distribution:

$$N_{\text{obs},k} = \text{Normal}\left(\lambda_k, \sigma\right)$$

where $\sigma$ represents the standard deviation of the observation error, shared across all bins. The marginal likelihood was then computed by summing over the log-likelihoods for all sampled

values of $\alpha$. We utilise a weakly informative prior (Lemoine 2019) assuming a normal distribution with a mean based on typical tree densities observed in forests across the conterminous United States, and a standard deviation that reflects the variability in these estimates ($1250 \pm 625$). This prior is informed by the TreeMap dataset, which provides tree density estimates (trees per acre, TPA) for the U.S. using a random forest machine-learning approach (Riley et al. 2021). The mean is selected to reflect central tendencies in the TreeMap data, whereas the standard deviation accounts for the spread in tree density estimates across different forests. Alternatively, this prior can also be informed by a smaller, in situ observed validation dataset from a subset of a given site. Finally, the model was coded with the Stan programming language in R and fit using Hamiltonian Monte Carlo (HMC) sampling (4 chains, warmup = 2000, iterations = 5000).

As with the method to estimate $\alpha$ in the previous section, this method is still prone to overestimating $N_{tot}$ under certain forest conditions if used without adjustment. We again use an adjustment factor based on normalised LAI and the breakpoint:

$$N_{tot} = N_{tot,adjust} * \text{adjust}$$

### 1.5 | Assessment

To test our approach, we recovered size-abundance relationships from RS data and compared the results to observed in situ data from forest plots collected by the National Ecological Observatory Network (NEON) and the Smithsonian ForestGEO network.

#### 1.5.1 | NEON

NEON is a long-term ecological monitoring programme collecting open-access data across U.S. ecosystems (Kampe 2010). NEON gathers environmental data from sites spanning multiple biomes. NEON field science staff collect information on vegetation structure by emulating the Carolina Vegetation Sampling protocol, generally collecting between 20 and 30 plots per site (Barnett et al. 2019). Plots are generally either 400 or 1600 $m^2$. Trees with a DBH $\geq 10$ cm are mapped, while those $< 10$ cm are sampled in smaller nested subplots without spatial mapping. We exclude trees with DBH $< 10$ cm and set $x_{min}$ to 10 cm when estimating parameters for NEON sites. Of the NEON sites measured, only 23 had enough measurements of trees with DBH $\geq 10$ (approximately $\geq 25$ trees) to apply a size-abundance scaling recovery method.

We used pre-existing canopy segmentation of RS data for NEON sites derived from imagery collected by the NEON Airborne Observation Platform (Weinstein et al. 2021; National Ecological Observatory Network (NEON) 2025a; National Ecological Observatory Network (NEON) 2025b). Since the RGB image extents exceeded the boundaries of the NEON in situ vegetation structure plots, we created a polygon shapefile using plot corners and UTM data for all plots. We then selected the segmented trees that overlapped with NEON plots using the Select Layer by Location tool in ArcGIS Pro and calculated their perimeter and area, ensuring that the trees were not cropped and their polygons remained intact. Crown diameter was estimated as $\text{Diameter} = 0.5 * \sqrt{\text{Perimeter}^2 - (8 * \text{Area})}$, and tree height was estimated from the NEON LiDAR-derived Canopy Height Model (CHM) (NEON (National Ecological Observatory Network) 2024) within each tree segmented polygon. While high-resolution CHMs can be derived from RGB data Wagner et al. 2024, our primary focus here is to evaluate our RGB method for estimating size-abundance relationships. Therefore, we use LiDAR-derived height as a reference, and we discuss the potential for RGB-based CHMs in future applications for a fully RGB-based workflow (see Section 2). Finally, we calculated DBH with the *itcSegment* package (Dalponte 2016) in R, which utilises allometric equations from Jucker et al. (2017).

We calculated abundance for each combination of size class, method (field survey or RS), and site by sampling $\alpha$ and $N_{tot}$ from their posterior distributions, propagating uncertainty into the final abundance estimates. For the field survey data, we first preprocessed the raw NEON data (removing multibole entries and excluding trees with DBH $< 10$ cm) to ensure consistent and comparable measurements. We then used the Stan code and fitting procedure from Grady et al. (2024), which fits a Pareto distribution with a lognormal prior on $\alpha$, to model the distribution for each resulting plot. We used a Bayesian linear regression model via *brms* (Bürkner 2017) to examine how data collection methods (remote sensing vs. field survey) influence the relationship between tree abundance and size class, while accounting for site-specific variability. Tree abundance and DBH were $\log_{10}$-transformed to fit the assumptions of a linear model. However, $\log_{10}$ transformation is not a linear rescaling—it compresses differences at higher values while expanding differences at lower values, which can introduce bias when back-transforming predictions. By applying the inverse transformation within the posterior prediction process, rather than after summarising results, we ensured that both model uncertainty and the effects of this transformation asymmetry were carried through to the final estimates. This approach preserved not only the modelled variability but also any skew introduced by the log transformation, preventing systematic distortions in the back-transformed values. The model includes an interaction between size class and method to assess how methods affect the abundance-size relationship. Random intercepts and slopes for size class were included for each site to account for site-specific differences.

The models used to estimate both total tree abundance and $\alpha$ from RS data were successful at recovering each parameter (Figure 3). To assess accuracy, a separate regression compared estimated and observed parameter values across plots. For total tree abundance, the Bayesian regression model explained 71.7% of the variance (Bayesian $R^2 = 0.717$), suggesting a strong correspondence between the modelled and observed values. The intercept was estimated at −35.11 (95% Credible Interval: −320.70, 273.56), indicating no substantial systematic bias. The slope was estimated at 1.06 (95% Credible Interval: 0.75, 1.35), indicating a strong 1:1 agreement between predicted and observed total tree abundance. The model's Root Mean Squared Error (RMSE) was 1000.39, providing an estimate of the typical error in total tree abundance predictions. The Mean Error (ME) or bias was −858.36, suggesting a slight overall tendency for the model to underestimate. For the estimation of $\alpha$, the Bayesian regression model

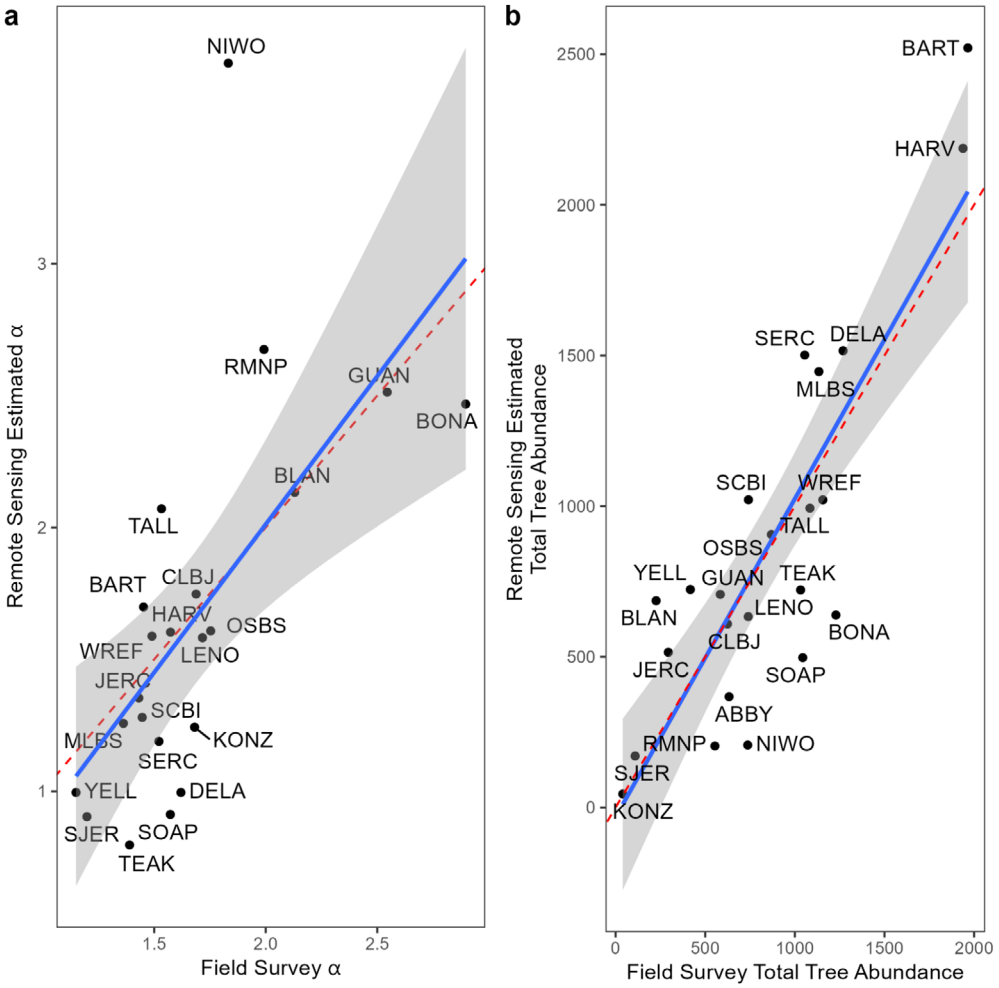


**FIGURE 3** | Predictions for $\alpha$ (a) and $N_{tot}$ (b) for each NEON site, plotted with the in situ derived values for each parameter. The 1:1 line (slope = 1 and intercept = 0) is shown as a dashed red line. The linear model average prediction line from a model fit with RS data is shown as a solid blue line. Shaded areas represent 95% confidence intervals for the RS-derived model.

explained 41.3% of the variance (Bayesian $R^2 = 0.413$). The intercept was estimated at −0.23 (95% Credible Interval (CI): −1.30, 0.90), suggesting no substantial systematic bias. The slope was estimated at 1.12 (95% CI: 0.47, 1.75), indicating that the relationship between predicted and observed $\alpha$ values is largely consistent with a 1:1 agreement. The RMSE was 0.531, providing an estimate of the typical error in $\alpha$ predictions. Furthermore, the ME, or bias, was 0.007, which indicates a negligible overall bias in the estimates. Removing two outlier sites from the Colorado Rocky Mountains (NIWO and RMNP, Figure 4, see discussion for explanation), where our allometric equations poorly recovered DBH, substantially improved model performance for $\alpha$ estimation. The revised Bayesian model explained 64.4% of the variance (Bayesian $R^2 = 0.644$). The intercept was estimated at −0.08 (95% CI: −0.72, 0.54), indicating a negligible systematic bias. The slope was estimated at 0.95 (95% CI: 0.59, 1.32), suggesting a strong 1:1 agreement between predicted and observed $\alpha$ values. The RMSE was 0.293, providing an estimate of the typical error in $\alpha$ predictions. Furthermore, the ME or bias was 0.001, which further confirms a negligible overall bias in the estimates. These predictions of $\alpha$ based on incomplete remote sensing data are significantly better than estimates derived solely from a random forest model using environmental data (these estimates serve as our priors) (Figure S2).

Our final model predicting size distributions from RS estimated parameters and comparing to field measurements included the effect of site as a random effect (Table S1). Tree abundance estimates from RS matched those from field surveys, with overlapping uncertainty intervals at key size classes (Figure 5). At $x_{min}$, field surveys estimated an average of 120.8 trees (95% CI: 49.0–301.3), while RS produced a comparable estimate of 93.4 trees (95% CI: 37.7–231.9), a difference of 27.4 trees on average. Due to the model's log-transformed scale, the relative difference is more appropriately expressed as a ratio: field estimates were on average 1.29 times those from Remote Sensing (95% HPD CI: 1.22–1.37). The larger CIs for each method are the result of the random effect of site on the model (one NEON site could have many fewer trees at $x_{min}$ than another, see Figure 1). Intra-site predictions have much narrower CIs, as shown in the ForestGEO assessment in the next section. A similar pattern was observed at the average $x_{breakpoint}$ size (18.69 cm DBH), where field-based estimates averaged 21.2 trees (95% CI: 8.7–53.1), compared to 18.0 trees (95% CI: 7.2–45.2) from RS. The average difference was 3.3 trees, and field estimates were on average 1.18 times

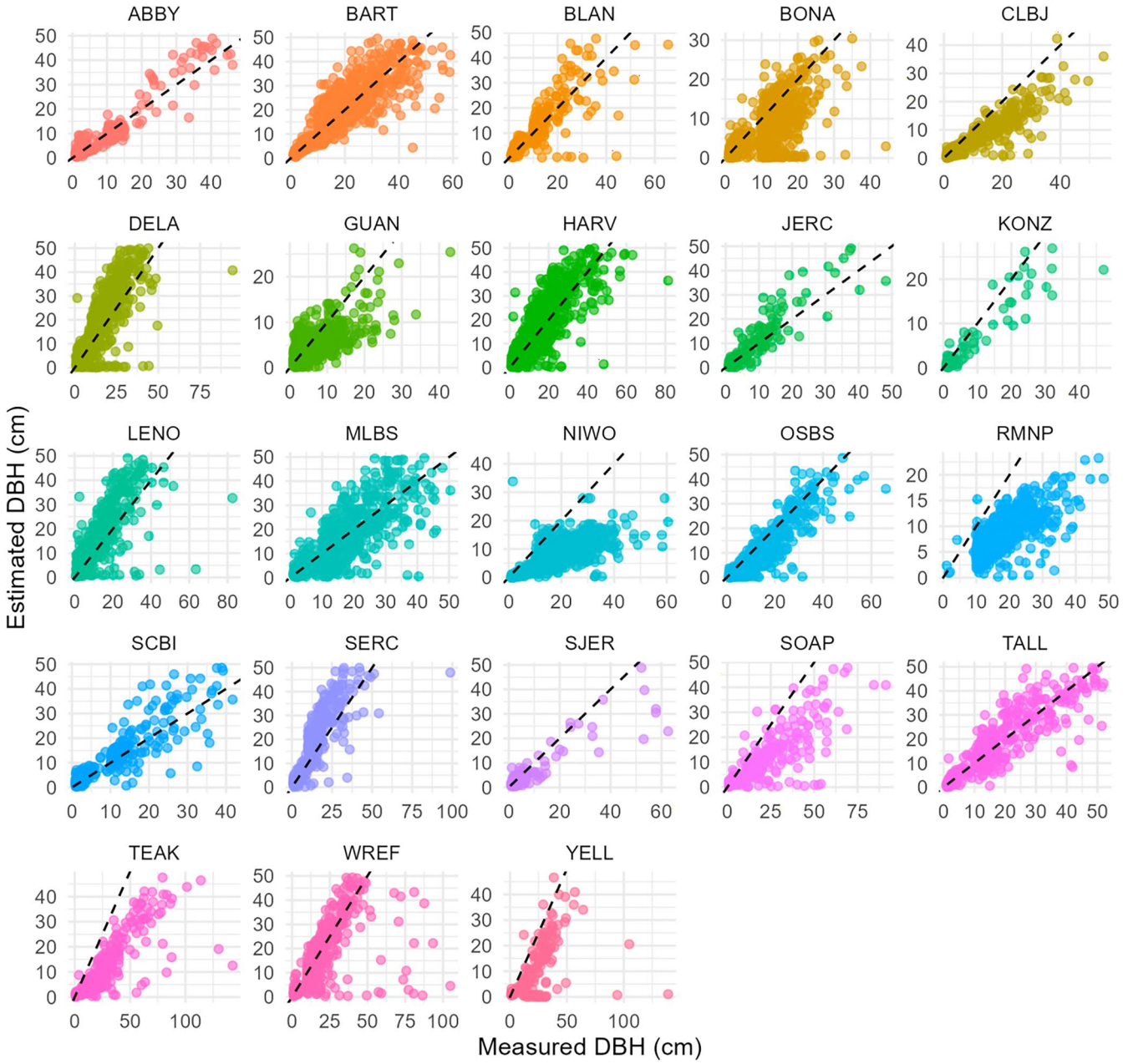


**FIGURE 4** | Observed DBH measured by NEON field staff versus allometrically estimated DBH using the equation from Jucker et al. (2017), by site. Both *x* and *y* represent in situ data, as in this figure the allometric estimate of DBH uses direct measurements of crown diameter and height. A perfect 1:1 ratio is plotted with a dashed line. Sites with worse predictions of DBH (where the equation does not appear to recover DBH from height and crown diameter well) generally correspond to the sites with the largest residuals in estimations of $\alpha$ from RS data (e.g., NIWO, RMNP, SOAP, TEAK).

those from Remote Sensing (95% HPD CI: 1.15–1.22). The overlapping confidence intervals suggest that RS effectively captures tree abundance patterns from field-based measurements. The model converged acceptably (Rhat = 1.00 for all parameters, all ESS > 1000), and residual variability was low ($\sigma = 0.17$).

### 1.5.2 | ForestGEO

The Center for Tropical Forest Science (CTFS)—Forest Global Earth Observatory (ForestGEO) is a global network of forest research sites with a standardised protocol for measuring trees (Anderson-Teixeira et al. 2015). Some ForestGEO sites also serve as NEON sites but monitor vegetation in a larger area. Therefore, some of the pre-existing canopy segmentation data (Weinstein et al. 2021) also cover the larger ForestGEO plots. We applied the same method as in the NEON sites to two ForestGEO sites in the northeastern US, Harvard Forest (HARV, Petersham, Massachusetts) and the Smithsonian Environmental Research Center (SERC, Edgewater, Maryland), splitting each into 1-ha blocks. Some 1-ha blocks in HARV overlap with gaps in the RS data where there is a swamp, and these blocks were removed from our analysis. Each block was treated as a random effect (Table S1).

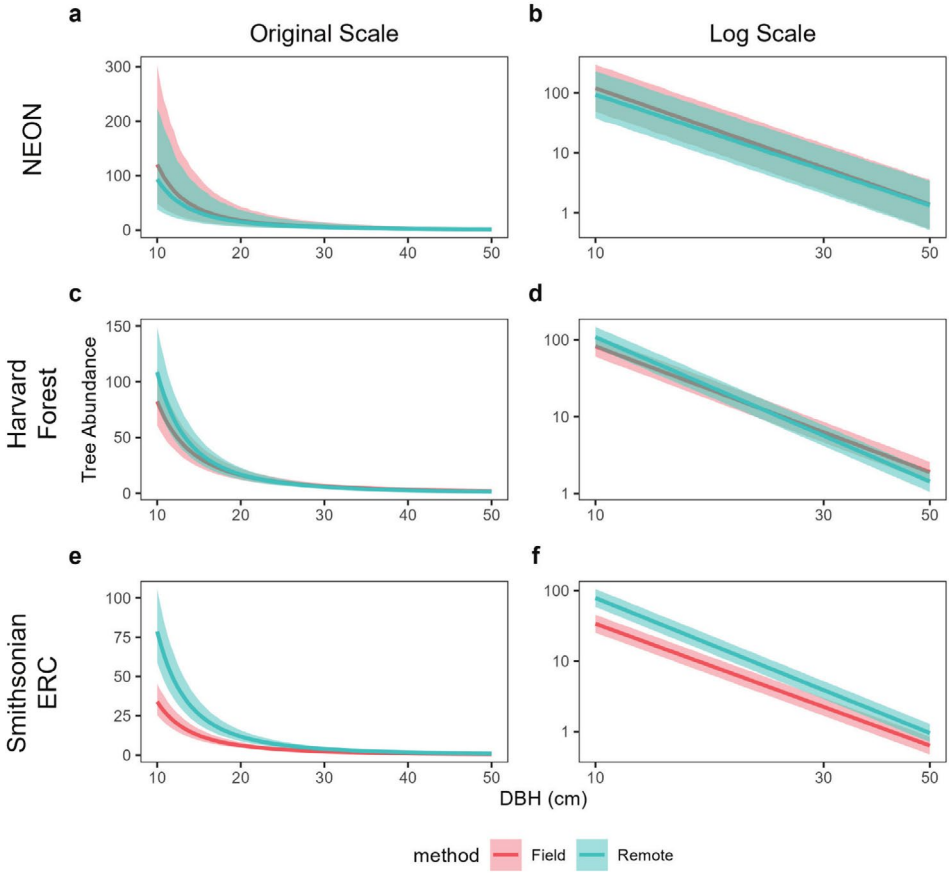


**FIGURE 5** | Predicted tree abundance across diameter size classes (DBH) from Bayesian hierarchical models fit to data from NEON plots (top), Harvard Forest plots (middle), and Smithsonian Environmental Research Center (SERC) plots (bottom). The left column (a, c, e) displays predictions on the original scale, while the right column (b, d, f) presents the same predictions on a log–log scale to aid visualisation of size-abundance patterns. Shaded bands represent 95% credible intervals. Predictions integrate field-collected and remotely sensed tree data, with a random effect of site for NEON plots and hectare ID for SERC and Harvard Forest plots to account for location-specific variation.

At $x_{min}$, field surveys at HARV estimated an average of 82.9 trees (95% CI: 60.6–113.0), while RS returned a slightly higher estimate of 108.7 trees (95% CI: 79.0–148.6). Similarly, at SERC, field surveys estimated 34.0 trees (95% CI: 25.2–45.5), while RS produced a higher estimate of 78.9 trees (95% CI: 58.9–105.9). At the average $x_{breakpoint}$ for each site—28.54 cm at HARV and 35.81 cm at SERC—field surveys estimated 7.1 trees (95% CI: 5.2–9.6) and 1.5 trees (95% CI: 1.1–1.9), respectively. RS returned similar estimates of 6.5 trees (95% CI: 4.7–8.9) at HARV and 2.4 trees (95% CI: 1.8–3.2) at SERC. Our RS-based estimates for SERC were poorer than HARV in matching in situ observations, likely due to the higher species richness of the forest at that location (see Section 2).

## 2 | Considerations

Although our estimation method overall captures tree abundance patterns, certain factors introduce limitations that warrant consideration. For example, we did not propagate potential error from the original allometric equations. Regionally or species-specific allometric equations for estimating DBH would likely yield more accurate results. Indeed, the NEON sites where our $\alpha$ predictions were notably poor—Niwot Ridge (NIWO) and Rocky Mountain National Park (RMNP)—are both in the Rocky Mountains of Colorado. Our result may stem from relatively poor tree crown segmentation in this region. Weinstein et al. (2020) evaluated segmentation recall—the fraction of reference trees correctly detected by the segmentation process—and found that NIWO had the second lowest recall among all tested sites. We also note that analysis of FIA data for this elevation range reveals a different height-DBH relationship than predicted by our global equation for estimating DBH from crown diameter and height (Jucker et al. 2017), which is not well-suited to this region (Figure 4). This conclusion is further supported by our ForestGEO predictions, where estimates at SERC were less accurate than at Harvard Forest. Tree species richness at SERC (74 species with at least one tree $\geq$10 cm DBH) is roughly 2.5 times that of Harvard Forest (30 species with at least one tree $\geq$10 cm DBH). While even a monoculture community could significantly deviate from allometrically expected DBH if its dominant species is poorly represented by a general equation, this higher diversity at SERC should increase the probability of encountering multiple species that deviate from a species-agnostic equation like Jucker et al. (2017), thus contributing to poorer overall allometric fits.

Beyond species composition and allometric deviations, differences in dominant forest types likely play a role as well. Harvard

Forest exhibits a combination of deciduous, evergreen and mixed forests, whereas SERC is predominantly deciduous forest. Deciduous forests often present greater challenges for individual tree segmentation, particularly in leaf-on conditions, due to large variation in tree shapes and sizes, denser canopies and a higher number of species, making individual trees harder to detect (Hamraz et al. 2016). We attempted to account for this with our adjustment factor; however, this may result in underestimations of canopy trees in highly delineated evergreen canopies. Indeed, at sites like SOAP and TEAK, which are completely evergreen forests, our $\alpha$ predictions deviated from field data in the opposite direction from the deviations shown by NIWO and RMNP. While designed to improve accuracy in typically denser or more complex deciduous canopies, this correction could lead to an artificial reduction in detected visible trees for evergreen forests. This suggests that future refinements of our method could benefit significantly from correction factors that are not only contingent on canopy density, but also on forest type (e.g., deciduous vs. evergreen) or even dominant species, allowing for more precise allometric and abundance estimates across diverse forest types. These segmentation difficulties, combined with species diversity and potential allometric mismatches, reinforce the need for locally calibrated or species-specific equations for robust estimates.

Therefore, enhancing tree and species detection from RS data for the application of more specific allometric equations and refinement of our truncated method will be critical. Indeed, leveraging more localised allometries to convert crown dimensions into DBH, for example, from databases like Tallo (Jucker et al. 2022) or the Legacy Tree Database (Radtke et al. 2015) represents a significant opportunity to further strengthen our method. This enhancement could also make robust biomass estimation from the abundances derived from this approach a more viable and valuable application. Recent studies also demonstrate the potential of combining RGB imagery for crown segmentation with high-resolution hyperspectral imagery for species classification (Marconi et al. 2022; Weinstein et al. 2024). When species identification is uncertain, foliar trait information from hyperspectral data could instead be used to classify crowns into life history guilds (Serbin and Townsend 2002), or other geospatial products could provide coarser resolution predictions of forest composition based on FIA data (e.g., TreeMap; Riley et al. 2021). Future applications of hyperspectral data for crown classification are especially promising given NASA's planned investment in the Surface Biology and Geology mission, provided the spatial resolution allows for precise tree delineation (Stavros et al. 2023).

While refining allometric equations can improve accuracy in certain regions, there are still other practical challenges: obtaining high-resolution satellite data and acquiring suitable validation data. Tree segmentation becomes increasingly unreliable at resolutions coarser than even 30 cm (Weinstein et al. 2020), yet ultra-high-resolution satellite imagery (e.g., 1 cm) remains difficult to obtain and work with at large scales. RGB imagery is becoming more available at 30 cm resolution via NASA's Commercial Satellite Data Acquisition programme, and advances in drone technology provide a promising alternative for capturing high-resolution imagery at localised scales (Tang and Shao 2015). Field validation requires comprehensive tree inventories, including DBHs down to the $x_{\min}$ threshold. Forest inventories often present challenges for such validation due to the frequent use of variable subplots, where the effective sampling area changes based on tree DBH, or spatial and/or temporal mismatches between inventory data collection and remote sensing (Knott et al. 2023). Given these limitations, independently conducting targeted tree inventories may be a more effective approach. Emerging technologies, such as LiDAR-equipped smartphones, now enable rapid, large-scale inventories, with studies reporting up to a 93.2% reduction in person-hours for 1-ha plots (Tatsumi et al. 2023).

Integrating high-resolution LiDAR data with widely available RGB imagery offers a powerful and increasingly viable approach for deriving detailed forest structure estimates. This combined approach is particularly valuable in regions where dedicated LiDAR surveys are not available, or when the aim is to obtain estimates at shorter, more frequent intervals than typical LiDAR acquisition campaigns allow. While our study's allometric equations for DBH estimation currently rely on a CHM parameterised using airborne high-resolution LiDAR data, our broader focus emphasises methods that reduce reliance on new LiDAR acquisitions for ongoing monitoring. This approach becomes feasible through recent advances in remote sensing that offer promising alternatives to direct LiDAR acquisition. Novel methods now allow for the estimation of high-resolution CHMs directly from RGB imagery. For instance, Wagner et al. (2024) developed a model trained on NEON LiDAR data that successfully generated a high-resolution model of canopy height from RGB National Agricultural Imagery Program (NAIP) imagery for the entire state of California. Similarly, Schroeder et al. (2022) show that NAIP digital aerial photogrammetric point clouds that have been developed from stereo NAIP imagery show high correlations with LiDAR and FIA height data. Such approaches demonstrate that, rather than requiring new airborne LiDAR acquisitions, pre-existing high-resolution airborne LiDAR datasets can be leveraged to train models that infer canopy height from RGB data. This reduces dependence on costly airborne LiDAR campaigns while expanding the applicability of our approach to regions where only RGB imagery is available.

Additionally, the tree segmentation procedure we relied on creates bounding boxes to delineate individual trees, which, while efficient, may lead to less accurate estimates of crown diameter compared to more sophisticated methods. Bounding boxes approximate the outermost extent of the tree's canopy but may not capture the true shape of the tree, potentially introducing errors in crown size estimation. These inaccuracies can propagate through to allometrically derived DBH estimates, affecting the overall accuracy of tree size distributions and abundance predictions. At the time of writing, the bounding box method used in this paper had higher precision and recall than several commonly used LiDAR-based methods and performed competitively in a benchmark crown delineation comparison (Weinstein et al. 2021). As object detection algorithms for canopy segmentation continue to improve (e.g., Ball et al. 2023), incorporating methods that generate convex hulls or more precise crown delineations could enhance segmentation accuracy (Feng et al. 2025), provided that the methods maintain both precision and computational efficiency. This shift could lead to more reliable tree size estimates and improve the performance of our modelling approach in future applications.

Finally, while our workflow primarily utilises the Pareto distribution to model size-abundance scaling, there are situations where alternative distributions, such as the Weibull, may be more appropriate. The Pareto distribution assumes a strict power-law decline in abundance across size classes, but in many forests, the observed distribution deviates from this expectation, particularly at the smallest and largest tree sizes. Because our analysis inherently excludes the smallest trees due to occlusion, applying a three-parameter Weibull distribution with a minimum truncation may provide a biologically reasonable alternative when capturing the larger trees. This flexibility is already built into our workflow, allowing for the substitution of alternative distributions like the Weibull where appropriate without requiring fundamental changes to the overall methodology.

## 3 | Applications and Conclusion

Our workflow provides a scalable method for characterising size-abundance distributions across broad spatial extents, which shape forest stand structure and, in turn, influence forest structural diversity (FSD) attributes such as canopy complexity, vertical stratification and habitat heterogeneity. FSD attributes have been linked to applications in sustainable forest management, wildlife conservation and ecosystem monitoring, with ties to species distributions, forest productivity and community composition (Atkins et al. 2023). By generating standardised, spatially explicit FSD metrics, our approach could contribute to efforts aimed at integrating structural attributes into biodiversity assessments. Additionally, FSD layers derived from our method could provide ecologically relevant predictor variables in species distribution models (SDMs), which are widely used in conservation to assess habitat suitability, predict species responses to environmental change and inform management decisions (Araújo et al. 2019). By linking remote sensing and ecological theory, our approach strengthens our capacity to model and predict ecological patterns and processes under changing environmental conditions at large scales.


### Acknowledgements

This research was supported by NASA's Biodiversity and Ecological Conservation programme (grant 80NSSC23K0421) and by Hatch project award no. [MEO-022425], from the U.S. Department of Agriculture's National Institute of Food and Agriculture. Dr. Quentin Read provided valuable statistical analysis and constructive feedback on the manuscript. We also acknowledge Drs. Sean McMahon and Geoff Parker for their efforts in collecting data from the SERC plot. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy.


### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The data that support the findings of this study are available in the ForestGEO data portal at http://ctfs.si.edu/datarequest/, the FIA database at https://apps.fs.usda.gov/fia/datamart/datamart.html and the NEON database at https://doi.org/10.48443/653k-mf13. We recommend using the rFIA R package to download FIA data. The R code used to analyse these data is available at https://github.com/ForestScaling/ScalingFromSky (doi: 10.5281/zenodo.15593238).

### References


Anderson, C. B. 2018. "Biodiversity Monitoring, Earth Observations and the Ecology of Scale." *Ecology Letters* 21: 1572–1585.

Anderson-Teixeira, K. J., S. J. Davies, A. C. Bennett, et al. 2015. "CTFS-ForestGEO: A Worldwide Network Monitoring Forests in an Era of Global Change." *Global Change Biology* 21: 528–549.

Araújo, M. B., R. P. Anderson, A. Márcia Barbosa, et al. 2019. "Standards for Distribution Models in Biodiversity Assessments." *Science Advances* 5: eaat4858.

Atkins, J. W., P. Bhatt, L. Carrasco, et al. 2023. "Integrating Forest Structural Diversity Measurement Into Ecological Research." *Ecosphere* 14: e4633.

Ball, J. G. C., S. H. M. Hickman, T. D. Jackson, et al. 2023. "Accurate Delineation of Individual Tree Crowns in Tropical Forests From Aerial RGB Imagery Using Mask R-CNN." *Remote Sensing in Ecology and Conservation* 9: 641–655.

Barnett, D. T., P. B. Adler, B. R. Chemel, et al. 2019. "The Plant Diversity Sampling Design for the National Ecological Observatory Network." *Ecosphere* 10: e02603.

Bonan, G. B. 2008. "Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests." *Science* 320: 1444–1449.

Breidenbach, J., and R. Astrup. 2014. "The Semi-Individual Tree Crown Approach." In *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*, edited by M. Maltamo, E. Næsset, and J. Vauhkonen, 113–133. Springer.

Bürkner, P.-C. 2017. "Brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80: 1–28.

Dalponte, M. 2016. "itcSegment: Individual Tree Crowns Segmentation. R Package Version 1.0."

Damuth, J. 1981. "Population Density and Body Size in Mammals." *Nature* 290: 699–700.

Duncanson, L. I., R. O. Dubayah, and B. J. Enquist. 2015. "Assessing the General Patterns of Forest Structure: Quantifying Tree and Forest Allometric Scaling Relationships in the United States." *Global Ecology and Biogeography* 24: 1465–1475.

Eichenwald, A. J., J. M. Grady, J. A. Knott, Q. D. Read, and S. Record. In Review. "The Impact of Disturbance History on Size-Abundance Scaling Patterns in Forests."

Enquist, B. J., and K. J. Niklas. 2001. "Invariant Scaling Relations Across Tree-Dominated Communities." *Nature* 410: 655–660.

Enquist, B. J., G. B. West, E. L. Charnov, and J. H. Brown. 1999. "Allometric Scaling of Production and Life-History Variation in Vascular Plants." *Nature* 401: 907–911.

Farrior, C. E., S. A. Bohlman, S. Hubbell, and S. W. Pacala. 2016. "Dominance of the Suppressed: Power-Law Size Structure in Tropical Forests." *Science* 351: 155–157.

Feng, Z., Z. Wang, S. I. Bueno, et al. 2025. "Static Segmentation by Tracking: A Frustratingly Label-Efficient Approach to Fine-Grained Segmentation."

Fischer, F. J., N. Labrière, G. Vincent, et al. 2020. "A Simulation Method to Infer Tree Allometry and Forest Structure From Airborne Laser Scanning and Forest Inventories." *Remote Sensing of Environment* 251: 112056.

Grady, J. M., Q. D. Read, S. Record, et al. 2024. "Life History Scaling in a Tropical Forest." *Journal of Ecology* 112: 487–500.

Hamraz, H., M. A. Contreras, and J. Zhang. 2016. "A Robust Approach for Tree Segmentation in Deciduous Forests Using Small-Footprint

Airborne LiDAR Data." *International Journal of Applied Earth Observation and Geoinformation* 52: 532–541.

Harris, N. L., D. A. Gibbs, A. Baccini, et al. 2021. "Global Maps of Twenty-First Century Forest Carbon Fluxes." *Nature Climate Change* 11: 234–240.

Hubau, W., T. De Mil, J. Van Den Bulcke, et al. 2019. "The Persistence of Carbon in the African Forest Understory." *Nature Plants* 5: 133–140.

Jucker, T., J. Caspersen, J. Chave, et al. 2017. "Allometric Equations for Integrating Remote Sensing Imagery Into Forest Monitoring Programmes." *Global Change Biology* 23: 177–190.

Jucker, T., F. J. Fischer, J. Chave, et al. 2022. "Tallo: A Global Tree Allometry and Crown Architecture Database." *Global Change Biology* 28: 5254–5268.

Kampe, T. U. 2010. "NEON: The First Continental-Scale Ecological Observatory With Airborne Remote Sensing of Vegetation Canopy Biochemistry and Structure." *Journal of Applied Remote Sensing* 4: 043510.

Knott, J. A., G. C. Liknes, C. L. Giebink, et al. 2023. "Effects of Outliers on Remote Sensing-Assisted Forest Biomass Estimation: A Case Study From the United States National Forest Inventory." *Methods in Ecology and Evolution* 14: 1587–1602.

Lemoine, N. P. 2019. "Moving Beyond Noninformative Priors: Why and How to Choose Weakly Informative Priors in Bayesian Analyses." *Oikos* 128: 912–928.

Marconi, S., B. G. Weinstein, S. Zou, et al. 2022. "Continental-Scale Hyperspectral Tree Species Classification in the United States National Ecological Observatory Network." *Remote Sensing of Environment* 282: 113264.

Muggeo, V. M. 2020. "Selecting Number of Breakpoints in Segmented Regression: Implementation in the R Package Segmented. Technical Report."

Muller-Landau, H. C., R. S. Condit, K. E. Harms, et al. 2006. "Comparing Tropical Forest Tree Size Distributions With the Predictions of Metabolic Ecology and Equilibrium Models." *Ecology Letters* 9: 589–602.

Myneni, R., Y. Knyazikhin, and T. Park. 2021. "MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V061."

National Ecological Observatory Network (NEON). 2025a. "High-Resolution Orthorectified Camera Imagery (DP1.30010.001)."

National Ecological Observatory Network (NEON). 2025b. "Vegetation Structure (DP1.10098.001)."

National Ecological Observatory Network (NEON). 2024. "Ecosystem Structure (DP3.30015.001), RELEASE-2024."

Pan, Y., R. A. Birdsey, O. L. Phillips, et al. 2024. "The Enduring World Forest Carbon Sink." *Nature* 631: 563–569.

Pettorelli, N., M. Wegmann, A. Skidmore, et al. 2016. "Framing the Concept of Satellite Remote Sensing Essential Biodiversity Variables: Challenges and Future Directions." *Remote Sensing in Ecology and Conservation* 2: 122–131.

Picard, N., and D. Gasparotto. 2016. "Liocourt's Law for Tree Diameter Distribution in Forest Stands." *Annals of Forest Science* 73: 751–755.

Ploton, P., F. Mortier, M. Réjou-Méchain, et al. 2020. "Spatial Validation Reveals Poor Predictive Performance of Large-Scale Ecological Mapping Models." *Nature Communications* 11: 4540.

Radtke, P. J., D. M. Walker, A. R. Weiskittel, J. Frank, J. W. Coulston, and J. A. Westfall. 2015. "Legacy Tree Data: A National Database of Detailed Tree Measurements for Volume, Weight, and Physical Properties. Pushing Boundaries: New Directions in Inventory Techniques and Applications: Forest Inventory and Analysis (FIA) Symposium, Pp. 25–30."

Read, Q. D., P. L. Zarnetske, S. Record, et al. 2020. "Beyond Counts and Averages: Relating Geodiversity to Dimensions of Biodiversity." *Global Ecology and Biogeography* 29: 696–710.

Record, S., K. M. Dahlin, P. L. Zarnetske, et al. 2020. "Remote Sensing of Geodiversity as a Link to Biodiversity." In *Remote Sensing of Plant Biodiversity*, 225–253. Springer International Publishing.

Riley, K. L., I. C. Grenfell, M. A. Finney, and J. M. Wiener. 2021. "TreeMap, a Tree-Level Model of Conterminous US Forests Circa 2014 Produced by Imputation of FIA Plot Data." *Scientific Data* 8: 11.

Schroeder, T. A., S. Obata, M. Papeş, and B. Branoff. 2022. "Evaluating Statewide NAIP Photogrammetric Point Clouds for Operational Improvement of National Forest Inventory Estimates in Mixed Hardwood Forests of the Southeastern U.S." *Remote Sensing* 14: 4386.

Serbin, S. P., and P. A. Townsend. 2002. "Scaling Functional Traits From Leaves to Canopies." In *Remote Sensing of Plant Biodiversity*, 43–82. Springer Open.

Sheather, S. J., and M. C. Jones. 1991. "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 53: 683–690.

Stavros, E. N., J. Chrone, K. Cawse-Nicholson, et al. 2023. "Designing an Observing System to Study the Surface Biology and Geology (SBG) of the Earth in the 2020s." *Journal of Geophysical Research: Biogeosciences* 128: e2021JG006471.

Tang, L., and G. Shao. 2015. "Drone Remote Sensing for Forestry Research and Practices." *Journal of Forestry Research* 26: 791–797.

Tatsumi, S., K. Yamaguchi, and N. Furuya. 2023. "ForestScanner: A Mobile Application for Measuring and Mapping Trees With LiDAR-Equipped iPhone and iPad." *Methods in Ecology and Evolution* 14: 1603–1609.

Taubert, F., R. Fischer, N. Knapp, and A. Huth. 2021. "Deriving Tree Size Distributions of Tropical Forests From LiDAR." *Remote Sensing* 13, no. 1: 131.

Taubert, F., M. W. Jahn, H.-J. Dobner, T. Wiegand, and A. Huth. 2015. "The Structure of Tropical Forests and Sphere Packings." *Proceedings of the National Academy of Sciences* 112: 15125–15129.

Wagner, F. H., S. Roberts, A. L. Ritz, et al. 2024. "Sub-Meter Tree Height Mapping of California Using Aerial Images and LiDAR-Informed U-Net Model." *Remote Sensing of Environment* 305: 114099.

Weinstein, B. G., S. Marconi, M. Aubry-Kientz, G. Vincent, H. Senyondo, and E. P. White. 2020. "DeepForest: A Python Package for RGB Deep Learning Tree Crown Delineation." *Methods in Ecology and Evolution* 11: 1743–1751.

Weinstein, B. G., S. Marconi, S. A. Bohlman, et al. 2021. "A Remote Sensing Derived Data Set of 100 Million Individual Tree Crowns for the National Ecological Observatory Network." *eLife* 10: e62922.

Weinstein, B. G., S. Marconi, A. Zare, et al. 2024. "Individual Tree Species Maps for the National Ecological Observatory Network." *PLoS Biology* 22: e3002700.

West, G. B., B. J. Enquist, and J. H. Brown. 2009. "A General Quantitative Theory of Forest Structure and Dynamics." *Proceedings of the National Academy of Sciences* 106: 7040–7045.

White, E. P., B. J. Enquist, and J. L. Green. 2008. "On Estimating the Exponent of Power-Law Frequency Distributions." *Ecology* 89: 905–912.

Yang, Q., Y. Su, S. Jin, et al. 2019. "The Influence of Vegetation Characteristics on Individual Tree Segmentation Methods With Airborne LiDAR Data." *Remote Sensing* 11: 2880.

Yoda, K., T. Kira, H. Ogawa, and K. Hozumi. 1963. "Self-Thinning in Overcrowded Pure Stands Under Cultivated and Natural Conditions." *Journal of Biology* 14: 107–129.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.