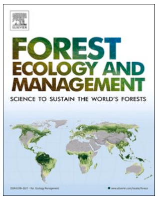
Forest Ecology and Management 537 (2023) 120943

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco



Scaling up uncertainties in allometric models: How to see the forest, not the trees

Jian Lin[a], Javier G.P. Gamarra[b], John E. Drake[c], Anibal Cuchietti[b,d], Ruth D. Yanai[c,*]

[a] Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, Hong Kong, China
[b] Forestry Division, Food and Agriculture Organization of the United Nations, Rome, Italy
[c] Department of Sustainable Resources Management, SUNY College of Environmental Science and Forestry, Syracuse, NY 13210, USA
[d] Dirección Nacional de Bosques (DNB), Ministerio de Ambiente y Desarrollo Sostenible (MAyDS), Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Allometric uncertainty
bootstrap
Monte Carlo
Bayesian
forest carbon budget

ABSTRACT

Quantifying uncertainty in forest assessments is challenging because of the number of sources of error and the many possible approaches to quantify and propagate them. The uncertainty in allometric equations has sometimes been represented by propagating uncertainty only in the prediction of individuals, but at large scales with large numbers of trees uncertainty in model fit is more important than uncertainty in individuals. We compared four different approaches to representing model uncertainty: a formula for the confidence interval, Monte Carlo sampling of the slope and intercept of the regression, bootstrap resampling of the allometric data, and a Bayesian approach. We applied these approaches to propagating model uncertainty at four different scales of tree inventory (10 to 10,000 trees) for four study sites with varying allometry and model fit statistics, ranging from a monocultural plantation to a multi-species shrubland with multi-stemmed trees. We found that the four approaches to quantifying uncertainty in model fit were in good agreement, except that bootstrapping resulted in higher uncertainty at the site with the fewest trees in the allometric data set (48), because outliers could be represented multiple times or not at all in each sample. The uncertainty in model fit did not vary with the number of trees in the inventory to which it was applied. In contrast, the uncertainty in predicting individuals was higher than model fit uncertainty when applied to small numbers of trees, but became negligible with 10,000 trees. The importance of this uncertainty source varied with the forest type, being largest for the shrubland, where the model fit was most poor. Low uncertainties were observed where model fit was high, as was the case in the monoculture plantation and in the subtropical jungle where hundreds of trees contributed to the allometric model. In all cases, propagating uncertainty only in the prediction of individuals would underestimate allometric uncertainty. It will always be most correct to include both uncertainty in predicting individuals and uncertainty in model fit, but when large numbers of individuals are involved, as in the case of national forest inventories, the contribution of uncertainty in predicting individuals can be ignored. When the number of trees is small, as may be the case in forest manipulation studies, both sources of allometric uncertainty are likely important and should be accounted for.

1. Introduction

As one of the nature-based climate solutions, forests have the capacity to store and accumulate carbon, offsetting greenhouse gas emissions (Fargione et al., 2018). To incorporate the forest sector into climate mitigation efforts, it is important to quantify their carbon stocks and fluxes of greenhouse gases (IPCC, 2019). Reporting forest carbon emission reductions is required by international commitments, such as the Paris Agreement, to achieve global mitigation efforts (UNFCCC, 2011). Quantifying forest carbon is also needed to evaluate progress towards land-based sustainable development goals such as sustainable forest management (Jandl et al., 2015).

Estimating forest carbon stocks and fluxes is challenging at any scale, from forest stands to entire countries, and there are many sources of uncertainty in the estimates. Quantifying uncertainty is essential to understanding the significance of differences, including those associated with efforts to reduce deforestation and degradation. Countries participating in Reducing Deforestation and Forest Degradation

---

* Corresponding author.
E-mail address: rdyanai@syr.edu (R.D. Yanai).

https://doi.org/10.1016/j.foreco.2023.120943
Received 21 December 2022; Received in revised form 14 March 2023; Accepted 15 March 2023

(REDD+) programs are required to include uncertainty in their reports, and the uncertainties can affect the payments made for the reported emission reductions (Yanai et al., 2020).

Sources of uncertainty in forest carbon stocks include measurement error, sampling error, and allometric equation error (Cunia 1987; Hill et al., 2013). Measurement error is normally controlled through quality assurance procedures (Pollard et al., 2005; Ferretti et al., 2009; Barker et al., 2015; Birigazzi et al., 2019), which can be used to quantify this source (Campbell et al., 2019; Yanai et al., 2022). Sampling error depends on the variability in forest biomass across the landscape and on sample size and sampling design (Köhl et al., 2011). Converting forest inventory data to forest biomass requires allometric equations, which also have uncertainty, although this source is rarely propagated (Chave et al., 2004; Wayson et al., 2015; McRoberts and Westfall, 2016). Uncertainty associated with the choice of which allometric equations to use is even more rarely addressed (Melson et al., 2011; Picard et al., 2015).

One reason that allometric uncertainty is rarely considered in forest carbon accounting is that it is more difficult to quantify than measurement error or sampling error. One source of confusion is whether to propagate uncertainty in the prediction of individuals (Paré et al., 2013) or in the model fit (Yanai et al., 2010). Presumably, the uncertainty in individuals is most important in experimental studies where sample sizes are small, while estimates at the scale of landscapes or entire countries should consider model fit (Yanai et al., in review). To our knowledge, there has been no systematic evaluation of the importance of these two sources of allometric uncertainty—prediction of individuals vs. model fit—in any forest system.

The purpose of this paper is to explore the relative importance of the uncertainty in the allometric model fit compared with the uncertainty in predicting the biomass of individual trees. We expected to demonstrate that individual random error becomes negligible with large numbers of trees, while model uncertainty is not reduced by applying the calculation to larger samples. We also compared four different approaches to representing model uncertainty: a formula for the confidence interval, Monte Carlo sampling of the slope and intercept of the regression, bootstrapping the model fit, and a Bayesian approach. We applied these approaches to model uncertainty in four study sites with varying allometry and model fit statistics, ranging from a monocultural plantation to a multi-species shrubland with multi-stemmed trees.

2. Methods

2.1. Study sites

We used previously collected data from four sites (Fig. 1) that differed in stand characteristics (Table 1). Each site had tree inventory data and allometric equations based on a harvested sample of trees (Fig. 2).

2.1.1. Plantation (Hawaii)

The site is a tropical *Eucalyptus* plantation located at 450–510 m elevation on the northeast coast of the island of Hawaii (19°30′ N, 155°15′ W). Annual rainfall is ~ 4000 mm, and the mean annual temperature is 21 °C. The soils are classified as thixotropic isomesic Typic Hydrudands formed in volcanic ash (Kaiwiki series). *Eucalyptus saligna* were planted in a randomized complete block design with four blocks and seven levels of an N-fixing plant (*Falcataria mollucana*) at 2 m × 2 m spacing in January 1982 after plowing and herbiciding (Boyden et al., 2005).

Trees from these sites have been used to describe tree allometry

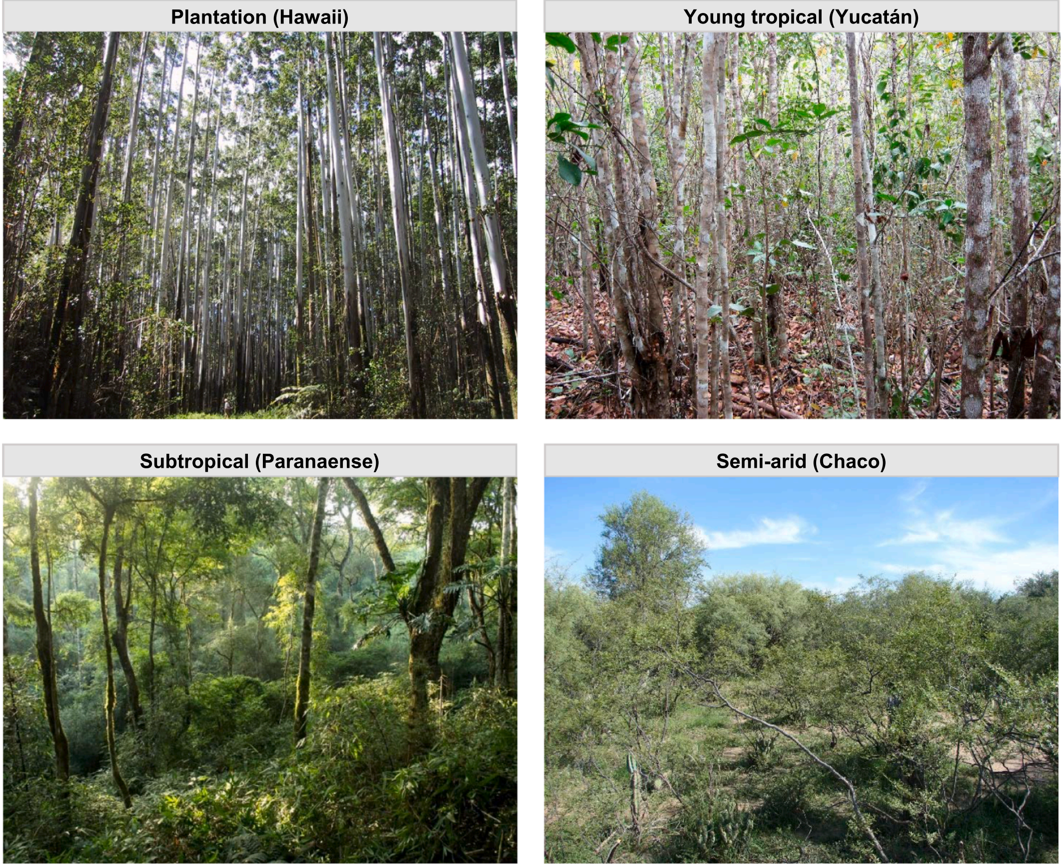


Fig. 1. Images of the study sites: an even-aged *Eucalyptus* plantation in Hawaii, secondary, semi-deciduous tropical forest on the Yucatán Peninsula, a subtropical jungle in the Selva Paranaense, and multi-stemmed individuals in the semi-arid Chaco.

Table 1
Characteristics of the four study sites. Diameter was taken at breast height (DBH, in cm), except at Chaco, where crown area (m^2) was measured.

Study site	Forest type	Mean annual temperature (C°)	Mean annual precipitation (mm)	Allometric data		Inventory data	
				Tree number	DBH or crown area (range)	Tree number	DBH or crown area (range)
Hawaii	Eucalyptus plantation	21	4000	93	3.5–81.7	1850	0.5–49.7
Yucatán	Young tropical moist forest	26	1200	48	3.2–160	21,871	2.5–105.1
Paranaense	Subtropical moist mixed forest	21	1600–2000	655	10.4–107.7	6732	10–190.9
Chaco	Subtropical semi-arid forest/open woodland	18	350–540	245	0.1–16.1	1070	0.9–195.9

(Debell et al., 1989; DeBell et al., 1997; Whitesell et al., 1988). We used a dataset of 93 *Eucalyptus* trees that cover the full range of tree diameters (3–82 cm dbh) (Binkley, personal communication).

For the tree inventory, we used the four replicate 30 m × 30 m plots that were *Eucalyptus* only (i.e., the 0% *Falcataria* treatment; Binkley et al., 2003; Boyden et al., 2005), resulting in 1850 trees in the inventory data set, at a density of trees with DBH > 0.5 cm DBH was 1528 stems ha^{-1}.

2.1.2. *Young tropical (Yucatán)*

This secondary, semi-deciduous tropical forest (Ochoa-Franco et al., 2019) is located in the municipality of Felipe Carrillo Puerto in Quintana Roo, Mexico, on the Yucatán Peninsula (88°14′W, 18°53′N). The climate is lowland tropical monsoonal, with annual rainfall of 1200 mm concentrated from May to October (Hernández-Stefanoni et al., 2014) and mean annual temperature of 26 °C. The soils are classified as gleysols, vertic cambisols and vertic luvisols. Common species include *Manikara zapota, Bursera simaruba, Metopium brownei, Lysiloma latisiliquum, and Piscidia piscipula,* among others.

This study was conducted with data from the Mexican network of Intensive Carbon Monitoring Sites (MEX-SMIC) which includes forested sites located in strategic landscapes for REDD+ activities. The area includes three land uses: slash and burn agriculture, selective logging, and fire-affected forest.

To develop the allometric equation, 48 trees were selected in proportion to their abundance in diameter classes 20 cm wide (Table 1). These trees were felled during selective logging in areas under management plans and with landowner permission in areas not subject to logging plans.

The tree inventory involved 32 cluster plots, each 0.22 ha in size, distributed over an area of 9 km^2, following the design of Mexico's national forest inventory (CONAFOR, 2010). The location of the plots was systematic with some adjustment to avoid the boundaries between the land-use types. In each cluster plot, trees 2.5–7.5 cm DBH were measured on 0.032 ha, trees 7.5–20 cm were measured on 0.16 ha, and trees > 20 cm were measured on 0.22 ha. The combined stem density of trees with DBH > 2.5 cm was 6558 ha^{-1}.

2.1.3. *Subtropical (Paranaense)*

This type of jungle occurs in the northeast region of Argentina (25°30́-29°7́S and 53°40́-57°1́W) and continues through eastern Paraguay and Brazil. The climate is warm and humid, with rainfall throughout the year, with a total that varies from 1564 mm to 2012 mm per year. The average annual temperature varies between 20 and 21 °C;

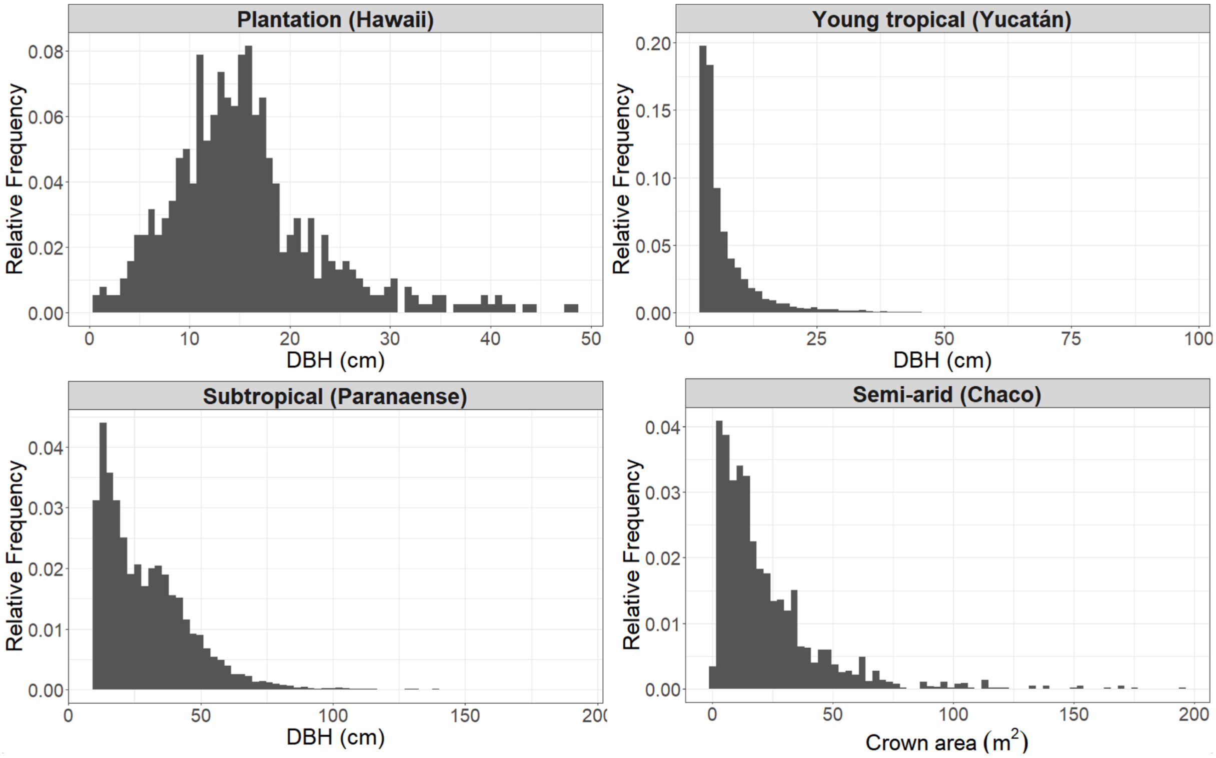


Fig. 2. Size distribution of trees in the forest inventory at each study site (crown area for Chaco, diameter for the other sites). The density is the fraction of trees in each size interval; the sizes were divided into 70 bins of equal width for each site.

winters are mild and summers are not excessively hot due to frequent rainfall (Cabrera 1971). The topography of the region is irregular, with low mountain peaks. The soils are lateritic, red and acid, and fine textured (Montagnini et al., 2006). The predominant vegetation type is subtropical jungle. Common species include *Nectandra* sp., *Balfourodendron riedelianum*, *Aspidosperma polyneuron* and *Araucaria angustifolia* (Cabrera 1994). The area includes different states of forest degradation and has been subjected to extensive logging with 20-year rotation periods.

Trees for volume models were chosen across the study area keeping climatic and soil conditions as constant as possible. Approximately 26 individuals for each of 25 species were selected to represent the wide range of plant sizes and the most abundant tree species of the Paranaense jungle (Secretaría de Desarrollo Sustentable y Política Ambiental, 2001). The interquartile range of tree heights was 17 m to 23 m. The range of tree diameters is given in Table 1.

For the inventory dataset, we used 6732 trees from 108 plots in this forest region from the First Argentinian National Inventory dataset (Secretaría de Ambiente y Desarrollo Sustentable, 2007). Trees ≥ 30 cm DBH were sampled on subplots of 0.5 ha and trees between 10 and 30 cm DBH were sampled on 0.2 ha subplots. The density of trees > 10 cm DBH was 191 ha^{-1}. The interquartile range of tree heights was 11 m to 24 m; the range of tree diameters is given in Table 1.

2.1.4. Semi-arid (Chaco)

This type of forest is located in central-western Argentina (31°16′–31°37′S, 65°25′–65°32′W). Soils are mainly sandy loam Aridisols of alluvial origin. Annual rainfall ranges between 250 and 540 mm, and the mean annual temperature is 18 °C. In this semiarid environment, the typical vegetation is xerophytic and often dominated by multi-stemmed shrubs. Common species include *Aspidosperma quebracho blanco*, *Prosopis flexuosa*, *Mimozyganthus carinatus*, *Senegalia gilliesii*, and *Larrea divaricata* (Cabido et al., 1992; Zuloaga et al., 2008). The study area includes fragments of conserved forest within a mosaic of different land uses, mainly logging and grazing (Zak et al., 2004; Baumann et al., 2018).

Individuals for biomass models were chosen across the study area keeping climatic and soil conditions as constant as possible. An average of 30 individuals per species were selected to include the range of plant sizes observed in the field (Conti et al., 2013). Several variables often used in allometric estimations of shrub biomass were measured prior to destructive sampling. In the present study, crown area was used in the biomass model, calculated from the maximum crown diameter (CD_1) and its perpendicular diameter (CD_2) as $\frac{\pi CD_1 \cdot CD_2}{4}$.

For the inventory data set, we used 1070 trees from 93 plots in this forest region from the Argentinian National Inventory dataset (Secretaría de Gobierno de Ambiente y Desarrollo Sustentable de la Nación, 2019). These were multi-stemmed trees with the crown diameter measurements needed for biomass estimation using the equation of Conti et al. (2013). Inventory plots were located on a grid. At each plot, trees ≥ 10 cm DBH were tallied in a 0.1 ha area; trees 5 to 10 cm DBH were tallied in a smaller concentric plot of 0.025 ha. Density was 322 individuals ha^{-1}, most of them multi-stemmed. Single-stemmed trees are not represented in our calculation of forest biomass, because they are modeled with a different allometric equation; these trees represent about 16% of the total biomass.

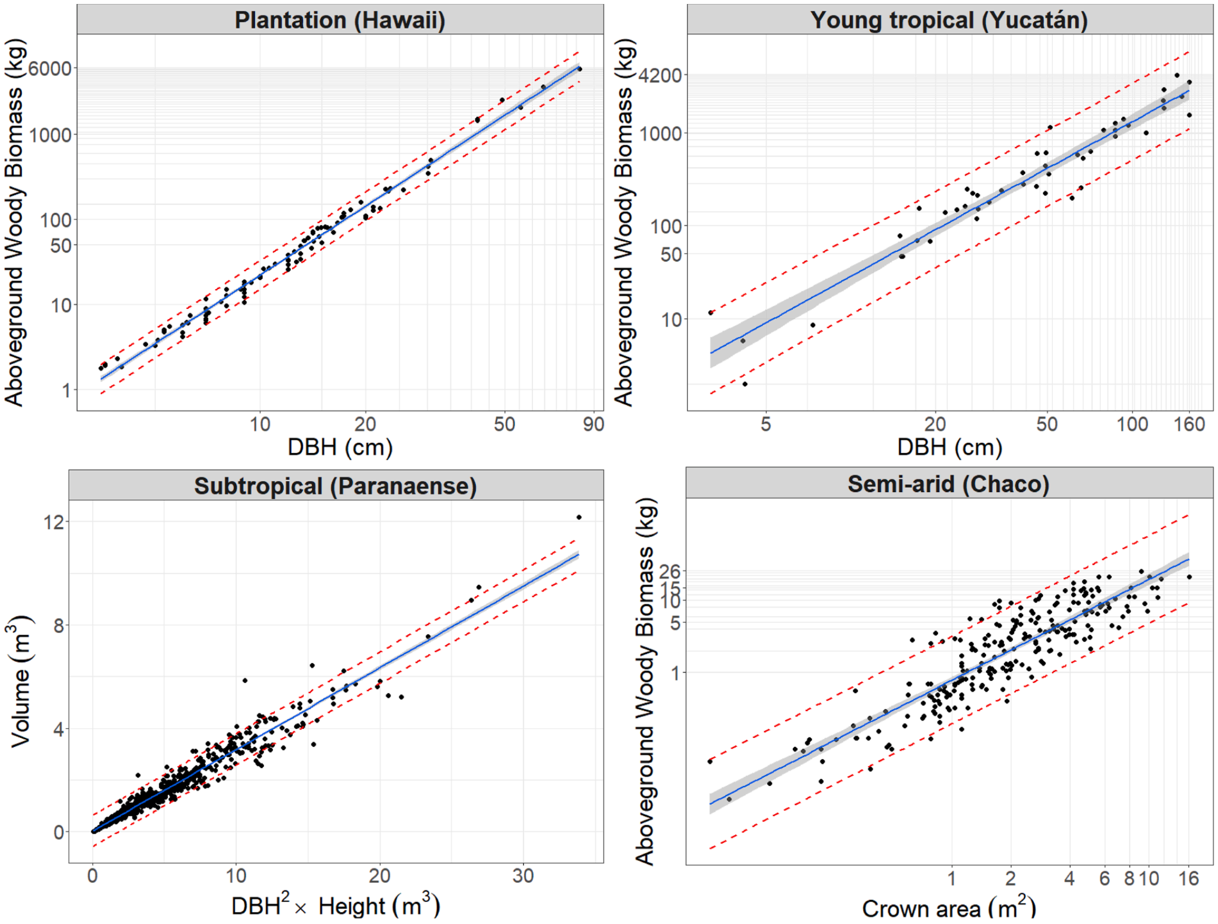


Fig. 3. Size distribution of harvested trees used for allometric models in the four study sites. The best fit regression lines are shown in blue. The grey area indicates the confidence in the model, while red dashed lines represent the confidence in predicting individuals.

Table 2
Allometric models for the four case studies. σ is the residual standard deviation.

Study site	Equation	R^2_{adj}	Number of trees	σ
Hawaii	$\log(AGB) = 3.07 + 2.68 \bullet \log(DBH)$	0.99	93	0.194
Yucatán	$\log(AGB) = -0.456 + 1.66 \bullet \log(DBH)$	0.92	48	0.461
Paranaense	$V = 0.0207 + 0.3161 \bullet DBH^2 \bullet Height$	0.95	655	0.307
Chaco	$\log(AGB) = -13.1 + 1.39 \bullet \log(CA)$	0.79	245	0.706

2.2. *Allometric models*

The allometric models predicting tree biomass or volume from various non-destructive measurements were obtained by linear fitting approaches (Fig. 3). In the case of the Selva Paranaense, the response variable was stem volume, and tree size was the predictor variable, represented by the stem diameter squared times tree height. In the other three sites, the response variable was the logarithm of aboveground biomass. For the Yucatán and the Hawaiian sites, the predictor variable was the logarithm of stem diameter; for the Chaco site, the predictor variable was the logarithm of crown area, calculated from the longest crown diameter and its perpendicular diameter.

To propagate uncertainty, the model residual variance, σ^2, is needed, along with the sum of squared deviations of the predictor variables. Since these statistics were not published in previous publications from our sites, we fit the equations to the original data (Table 2). Model fits were best for Hawaii, followed by Paranaense, Yucatán, and then Chaco (Table 2).

2.3. *Approaches to error propagation*

Monte Carlo error propagation involves multiple iterations of a calculation, in which each iteration uses a random sample of the inputs. We used 10,000 Monte Carlo iterations for each of five approaches to error propagation (four for confidence in the model and one for prediction of individuals) applied to each of four forests with each of four inventory sample sizes, as detailed below. We characterized the uncertainty in the outputs by the coefficient of variation (CV), which is the standard deviation divided by the mean. Code and data to reproduce the analysis are publicly available (Lin et al., 2023)

2.3.1. *Confidence in the model: Analytical approach*

Uncertainty in the model fit can be described analytically using Equation (1) (Snedecor and Cochran 1989, p. 164):

$$\sigma\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x}^2)}{\sum(x_i-\bar{x}^2)}} \tag{1}$$

where σ is the standard deviation of the model residuals; n is the sample size used to develop the regression model; x_0 is the value of the explanatory variable of the individual for which the prediction is to be made; and $\bar{x}$ is the mean of the explanatory variables of all individuals x_i in the allometric data set used to develop the regression model. This equation assumes homoscedasticity in the residuals. We used Monte Carlo simulation to propagate uncertainty in the model, randomly sampling one value of σ for each iteration to be applied to all the trees in each iteration.

2.3.2. *Confidence in the model: Slope-intercept sampling*

Alternatively, uncertainty in the model can be described by propagation of uncertainty in the model coefficients. Model coefficients in allometric equations cannot be assumed to be independent. In the case of linear equations, the intercept and slope are negatively correlated (Zapata-Cuartas et al., 2012). We represented the covariance between the slope and intercept using the variance–covariance $\sum$ matrix derived from regression (Lande et al., 2003):$\sum = \sigma^2(X^{'}X)^{-1}$ where X denotes a matrix of model explanatory variables with a preceding column of $1'$s representing the intercept term. By assuming a multivariate normal distribution $N(\beta, \sum)$ using the derived $\sum$ matrix and means β equal to the original vector of model coefficients, we randomly obtained 10,000 sets of model coefficients for each allometric model. We then applied the 10,000 sets of model coefficients to the inventory trees to calculate 10,000 model outputs, and the CV of the output was reported.

2.3.3. *Confidence in the model: Bootstrapping*

To represent confidence in the model using bootstrapping, we resampled the entire number of allometric trees in each site with replacement, and then refitted a regression equation to the sample, from which we obtained one set of allometric model coefficients. We repeated the process to obtain 10,000 sets of model coefficients, which we then applied to the inventory trees to calculate 10,000 model outputs, from which the output CVs were estimated.

2.3.4. *Confidence in the model: Bayesian approach*

This method uses probability distributions to determine the uncertainty in the model parameters. A Markov Chain Monte Carlo algorithm was used to generate marginal posterior distributions for each of the model parameters using Metropolis-Hastings sampling (Metropolis et al., 1953, Hastings 1970). We calculated a maximum likelihood estimator of allometric model fit by using uninformative priors, which were distributed normally for the slope and intercept but uniformly for sigma to avoid initial negative values, following Réjou-Méchain et al. (2017). We then repeated the following stepwise procedure to obtain 10,000 sets of model parameters. (1) We assumed that the slope, intercept, and sigma followed normal distributions (because the likelihood of sigma being negative was now negligible) with means equal to the values from the previous step and standard deviations of 0.01 for slope, 0.002 for intercept, and 0.01 for sigma, and we obtained one random set of parameter values and calculated a new maximum likelihood estimator of the allometric model fit. (2) If the difference between the maximum likelihood estimators was greater than a random value between 0 and 1, then we accepted the new estimator. (3) We repeated this process using the new parameter values. Because we used uninformative priors, we discarded the first 10% of iterations as a burn-in period given the potentially poor fit. For the remaining parameter sets, we used 1 of every 1000 iterations to avoid autocorrelation in the data because each iteration was built based on the previous iteration. The details about the uninformative priors and the reasoning behind the default standard deviation values are presented in the R package “BIOMASS” (Réjou-Méchain et al., 2017).

2.3.5. *Uncertainty in the prediction of individuals*

The uncertainty in prediction of individuals can be described analytically using Equation 3 (Snedecor and Cochran 1989, p. 166), under the assumptions of homoscedasticity, as in the case of the uncertainty in model fit:

$$\sigma\sqrt{1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X_i-\bar{X})^2}} \tag{2}$$

with variables defined as in Equation (1). The uncertainty of prediction of individuals was then propagated using Monte Carlo simulation. For each tree, we added a random error with a mean of 0 and a standard deviation defined by Equation (2). We repeated the calculation 10,000 times to generate 10,000 model predictions, from which the output CVs were estimated.

2.4. Applying tree allometry to forest inventory data

For each study site, we randomly sampled 10, 100, 1000, and 10,000 trees from the population represented by the inventory data. We sampled with replacement, because we were limited by the number of trees in the inventory (1,850 trees for Hawaii, 6,732 for Selva Paranaense, 11,019 for Yucatán, and 1,070 for Chaco). Except in Hawaii, nested sampling was used, with small trees measured in smaller plots than large trees (as described above). To represent the population, we repeated the small trees by the ratio of expansion factors for the various size classes (10 or 40 for Chaco, 10 or 25 for Selva Paranaense, 18 or 25 or 125 for Yucatán). To replicate the diameter distribution of the population, we sampled from 10 strata defined by tree diameter with equal numbers of trees in each stratum. This stratification was important for small sample sizes, where the uncertainty varied depending on which trees were selected. In the case of only 10 trees, the biomass estimate was sensitive to the biggest tree sampled, and therefore we had to be careful to select a sample that gave a similar total biomass to the original data set. We intended the four data sets to illustrate the effect of differences in the forests, and this approach reduced the effect of random sampling of trees within each forest.

3. Results

3.1. Confidence in the model

There was close agreement in the four approaches to characterizing uncertainty in model fit: the analytical approach, random sampling of the slope and intercept, bootstrap refitting of the model, and the Bayesian approach (Figs. 4 and 5). The four sites, however, differed dramatically in uncertainty in model fit and were generally lowest where more trees were used in the construction of the allometric equations and where the model fit was good. Specifically, Selva Paranaense had the lowest uncertainty, with $R^2_{adj} = 0.95$ and 655 trees, followed by Hawaii, which had only 93 trees but an extremely high R^2_{adj} (0.99), because the trees in this monoculture plantation were more uniform in their allometry (Fig. 3). The other two sites had larger uncertainties: Chaco had a poor model fit ($R^2_{adj} = 0.79$) because the tree allometry was extremely variable and the model was based on crown diameters. Yucatán had only 48 trees to fit the allometric model (Table 2).

Although the approaches to quantifying uncertainty in the model were in general agreement, some differences deserve attention. Bootstrapping in the Yucatán data set departed from the other approaches in giving a higher uncertainty (Fig. 5). Because the sample size was small (48 trees), bootstrap sampling was sensitive to outliers, which might be represented repeatedly or not at all in each bootstrap sample (Fig. 6).

We did not expect confidence in the model to depend on the number of trees to which it was applied. In fact, increasing the numbers of trees in the inventory sample from 10, 100, 1000 to 10,000 did not affect the overall uncertainty in the application of the models (see CVs in Fig. 4 and flat lines in Fig. 5).

3.2. Confidence in the prediction of individuals

In addition to describing the uncertainty of forest biomass using four different methods, we also described the uncertainty in predicting biomass of individual trees (Figs. 4 and 5). The uncertainty declined with the number of trees in the inventory sample, with the estimate of the uncertainty becoming negligible when many trees were measured. Clearly, when the number of inventoried trees is large, the uncertainty due to predicting individuals is insufficient to represent the uncertainty in the models correctly.

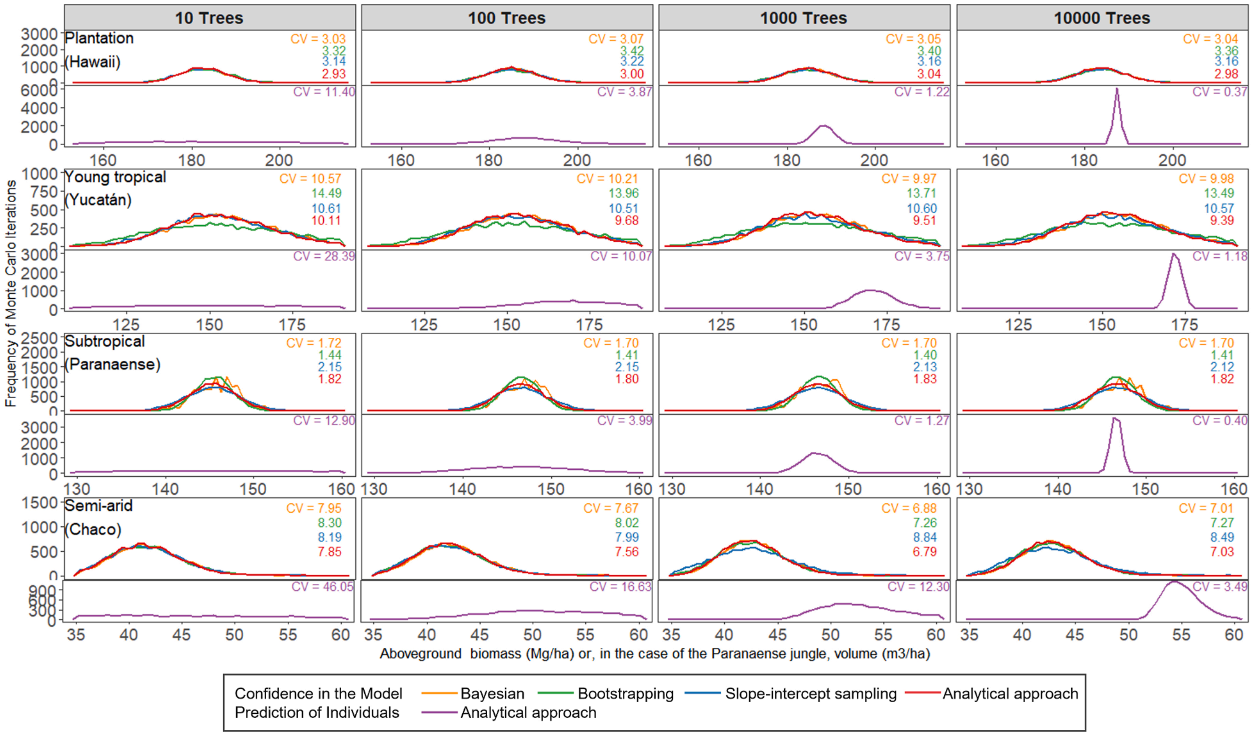


Fig. 4. Frequency distributions of estimates of forest biomass (or volume, in the Selva Paranaense) per hectare based on 10,000 Monte Carlo iterations for 4 sites, 4 scenarios of inventory sample sizes, and 5 approaches to characterizing uncertainty in the estimates. CV indicates coefficients of variation (%).

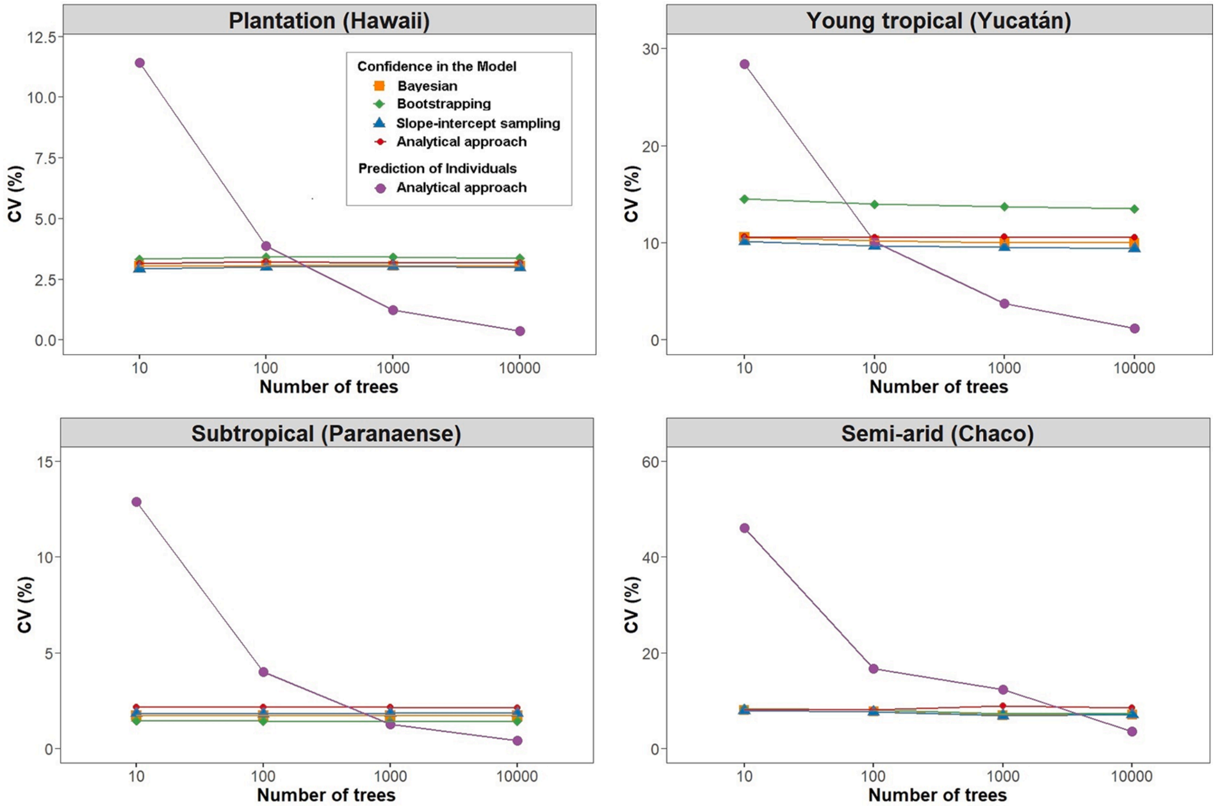


Fig. 5. Uncertainty in biomass estimates, indicated by the coefficients of variation (CV) across 4 sites, 4 scenarios of tree numbers, and 5 approaches to characterizing uncertainty in the estimates.

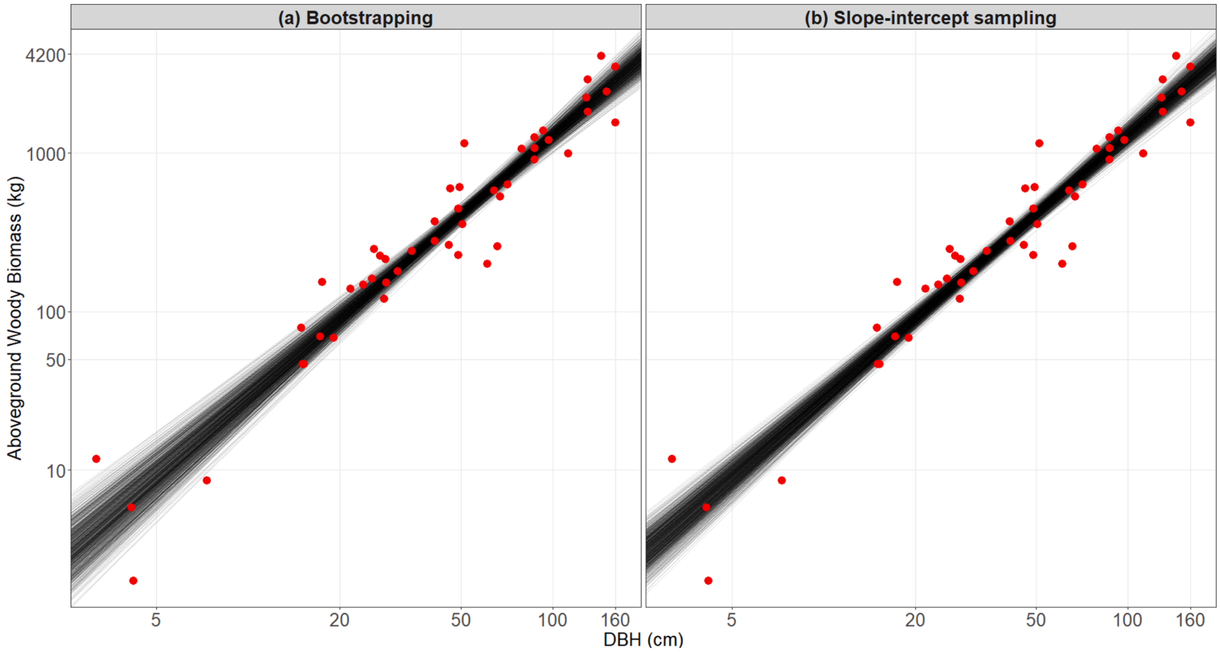


Fig. 6. The sampled trees from the Yucatán site are shown in red along with 10,000 possible regression lines based on (a) bootstrap fitting, based on 48 data points, randomly sampled with replacement and (b) random samples of the slope and intercept based on their variance and covariance. Uncertainty characterized by bootstrapping (a) is higher than by slope-intercept sampling (b; Figs. 4 and 5).

Forest biomass calculated by predicting individual trees had a higher mean than that calculated accounting for uncertainty in the model fit, except in the Selva Paranaense, where the allometric model was not based on log-transformed values. This bias arises from the application of the linear fit in log–log space, due to a lack of proper weighting of residuals: high values tend to inflate the estimated mean. This bias can be avoided by using non-linear fitting approaches that account for the dependence of the variance of the residuals upon tree diameter.

4. Discussion

4.1. Prediction of individuals vs. Confidence in the model

Calculating uncertainty in estimates of forest biomass is complicated, involving sampling error, allometric models, and sometimes wood density and carbon fraction (Lu et al 2012, Temesgen et al 2015, Yanai et al. 2020). Correctly representing the uncertainty in tree allometry is especially challenging. The confidence interval is the standard measurement of error when reporting forest resources (Birigazzi et al., 2019), but there has been some confusion about when to propagate the confidence in the prediction of individuals versus the confidence in the model.

When an allometric model is applied to a small number of trees, the uncertainty of predicting individuals is important, because the confidence interval around predicting the mass of an individual tree is always larger than the confidence interval around the best-fit model (Fig. 2; Picard et al., 2012). In our case studies, for an extremely small inventory of only 10 trees, the uncertainty in predicting individuals was always more important than the uncertainty in the model fit, and in most cases this was true for 100 trees (Fig. 5). Only in the case of the Chaco forest, with the greatest allometric uncertainty, was the uncertainty in predicting individuals important for an inventory of 1000 trees. Thus, it is important to address the uncertainty in predicting individuals when small numbers of trees are involved, as is sometimes the case in expensive manipulation studies. The precise number of trees required to make this source of uncertainty negligible depends on specifics of the case.

In applications such as national forest inventories, involving very large numbers of trees, it is more important to represent the uncertainty in the model; individual uncertainties become unimportant due to the law of large numbers, where $CV \sim n^{-\frac{1}{2}}$ (Fig. 4). In cases such as these, it is incorrect and potentially very misleading to propagate the uncertainty in individuals and ignore the uncertainty in model fit. For example, when large numbers of trees are sampled by remote sensing, this approach can produce highly certain estimates of forest carbon (95% < 1% error), despite using allometric regressions with considerable uncertainty (Gonzalez et al., 2010).

4.2. Cross-site differences in allometric uncertainty

We chose four case studies that differed in the nature of the allometric relationship. The allometry of trees of consistent form and size (e.g., *Eucalyptus* plantation in Hawaii) can be characterized with high confidence with a small number of trees. Trees in the Paranaense forest are not very uniform, but a very large number of trees were used to construct the allometric equation. These two case studies were characterized by allometries with high R^2_{adj} values and low residual standard deviations (Table 2) and correspondingly low biomass uncertainty (CV's near 3%; Fig. 5). The other two case studies had allometries with lower R^2_{adj} values and higher residual standard deviations (Yucatán and Chaco; Table 2) and correspondingly high biomass uncertainty (CV's near 10%; Fig. 5).

The semi-arid Chaco site was interesting, given that the multi-stemmed nature of the trees precluded the use of a dbh-based allometry. The allometry based on crown area was inherently more variable than the dbh-allometries used at the other sites, such that the uncertainty associated with the mean value was higher than the Hawaiian and Paranaense sites and similar to the Yucatán site. The highly variable allometry of the Chaco site made the uncertainty associated with the predictions of individuals more important, even at relatively large inventory sizes of ~ 1000 trees. The Chaco allometry was able to capture the growth form of these multi-stemmed trees, but individual variation was still high and important relative to other sites.

4.3. Approaches to quantifying confidence in the model

In our case studies, all four approaches to propagating uncertainty in the allometric model gave generally similar results (Fig. 5). The biggest difference was for the young tropical forest in the Yucatán site, where bootstrapping differed from the other three approaches, producing higher estimates of uncertainty. The allometric model at this site was poorly constrained at the low end of tree diameter and biomass (Fig. 3); only 4 small trees were sampled (of 48 trees in total) and these 4 were in poor agreement. In situations such as this, adding observations to the allometric dataset may be useful. Bootstrapping allows outliers in a data set to be omitted by chance in some iterations and represented multiple times in others (Fig. 6). The other three approaches all use the allometric sample set to describe the model uncertainty, whereas bootstrapping represents the possibility that the allometric sample is not representative of the underlying population.

The other three approaches produced consistent estimates of uncertainty in model fit. Practical considerations might affect the choice of approach. Affordability and ownership of the analytical process are important to the sustainability of forest monitoring systems (FAO, 2020), thus simpler approaches may be preferred in countries with low capacities. The analytical approach is amenable to implementation in Excel (Yanai et al. 2010), while Bayesian and slope-intercept approaches require specialized software packages. Monte Carlo and bootstrapping approaches can be implemented in Excel (FCPF, 2021) but are easier to manage in a more powerful programming language.

4.4. Implications for policy and larger research communities

Allometric calculations of carbon sequestration from forest inventory assessments are increasingly used at national scales as part of international carbon emission trading agreements (e.g., REDD+). These agreements often require an assessment of uncertainty of reported emission reductions (Pelletier et al., 2013). A wide range of methods have been employed to quantify this uncertainty, some of them egregiously wrong (Yanai et al., 2020). We hope that this paper will contribute to more accurate estimates of uncertainty in allometric relationships (Picard et al., 2012) and their use when applied to large-scale forest inventories.

National forest inventories are common frameworks for forest monitoring and international reporting for climate mitigation and national forest policy formulation. It is crucial that reporting institutions understand when and how to report uncertainties. Our case-study based approach aims to support continuing efforts to address technical gaps through capacity development (Romijn et al. 2015, Neeff et al. 2017) and to highlight the need to appropriately address uncertainties given the wildly variable quality in uncertainty reporting to date.

In this study, we showed how the quality of the allometric data affects the importance of the uncertainty of the prediction of individuals when scaling up to whole inventories. In most cases, this source of uncertainty can be ignored at large scales. In contrast, confidence in the model should always be reported, in addition to other sources of error, of which sampling error is generally the most important (Phillips et al. 2000). We showed that there are multiple options for adequately accounting for uncertainty in model fit (analytical approach, slope-intercept sampling, bootstrapping, and Bayesian analysis), but that accounting for uncertainty based only on predicting individuals seriously underestimates allometric uncertainty when the number of trees is large, as is certainly the case in national forest inventories.

CRediT authorship contribution statement

Jian Lin: Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization. **Javier G.P. Gamarra:** Conceptualization, Methodology, Formal analysis, Writing – original draft. **John E. Drake:** Conceptualization, Writing – original draft, Visualization. **Anibal**

Cuchietti: Resources, Writing – review & editing. **Ruth D. Yanai:** Conceptualization, Methodology, Resources, Writing – original draft, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data Availability statement: Lin, J., Gamarra, J.G.P, Drake, J.E., Cuchietti, A., Yanai, R.D. 2023. Github repository for the scaling of uncertainties in allometric models. https://github.com/jlin270/Scaling_up_uncertainties_in_tree_allometry.

Acknowledgements

The Yucatan site was studied under the MEX-SMIC network, an affiliation of academic institutions, government and NGO investigators, including the civil association Uyoolché AC in Felipe Carrillo Puerto, Quintana Roo. David López Merlín and Gonzalo Sánchez provided data and photographs from that site, and Dan Binkley provided data and photographs from the plantation forest in Hawaii. Craig Wayson and Marcela Olguín helped with the Yucatán data; Kenneth Fisher and two anonymous reviewers provided useful feedback on the manuscript. This publication is a product of QUEST (Quantifying Uncertainty in Ecosystem Studies), a working group dedicated to advancing uncertainty analysis in ecosystem studies (www.quantifyinguncertainty.org) and QUERCA (Quantifying Uncertainty Estimates and Risk for Carbon Accounting), which is funded by the US Department of State and US Agency for International Development through the SilvaCarbon Program.

References

Barker, A., Origo, N., Nightingale, J., 2015. Quality assurance framework for forest monitoring programmes. https://eprintspublications.npl.co.uk/6570/1/OP10.pdf. (accessed October 25, 2022).

Binkley, D., Senock, R., Bird, S., Cole, T.G., 2003. Twenty years of stand development in pure and mixed stands of Eucalyptus saligna and nitrogen-fixing Facaltaria moluccana. For. Ecol. Manage. 182, 93–102. https://doi.org/10.1016/S0378-1127(03)00028-8.

Birigazzi, L., Gregoire, T.G., Finegold, Y., Golec, R.D.C., Sandker, M., Donegan, E., Gamarra, J.G., 2019. Data quality reporting: good practice for transparent estimates from forest and land cover surveys. Environ. Sci. Policy 96, 85–94. https://doi.org/10.1016/j.envsci.2019.02.009.

Boyden, S., Binkley, D., Senock, R., 2005. Competition and Facilitation Between Eucalyptus and Nitrogen-Fixing Falcataria in Relation to Soil Fertility. Ecology 86, 992–1001. https://doi.org/10.1890/04-0430.

Baumann, M., Levers, C., Macchi, L., Bluhm, H., Waske, B., Gasparri, N.I., Kuemmerle, T., 2018. Mapping continuous fields of tree and shrub cover across the Gran Chaco using Landsat 8 and Sentinel-1 data. Remote Sens. Environ. 216, 201–211. https://doi.org/10.1016/j.rse.2018.06.044.

Cabido, M., Díaz, S., Acosta, A., 1992. La vegetación del Chaco Árido en el oeste de la provincia de Córdoba, Argentina. Doc. Phytosociol. 14, 447–456.

Cabrera, A., 1971. Fitogeografía de la República Argentina. Boletín de la Sociedad Argentina de Botánica 14, 1–42.

Cabrera, A. 1994. Regiones fitogeográficas argentinas. En Kugler WF (Ed.) Enciclopedia argentina de agricultura y jardinería. Tomo 2. 2a edición. 1a reimpresión. Acme. Buenos Aires. Argentina. Fascículo 1. pp. 1-85.

Campbell, J.L., Green, M.B., Yanai, R.D., Woodall, C.W., Fraver, S., Harmon, M.E., Domke, G.M., 2019. Estimating uncertainty in the volume and carbon storage of downed coarse woody debris. Ecol. Appl. 29 (2), e01844.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., Perez, R., 2004. Error propagation and scaling for tropical forest biomass estimates. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359 (1443), 409–420. https://doi.org/10.1098/rstb.2003.1425.

CONAFOR (Comisión Nacional Forestal), 2010. Inventario nacional forestal y de suelos México 2004-2009 (Reporte Técnico). Jalisco, México: Comisión Nacional Forestal. https://www.snieg.mx/Documentos/IIN/Acuerdo_3_X/Manual_y_Procedimientos_para_el_Muestreo_de_Campo_INFyS_2004-2009.pdf (accessed February 15, 2023.

Conti, G., Enrico, L., Casanoves, F., Díaz, S., 2013. Shrub biomass estimation in the semiarid Chaco forest: a contribution to the quantification of an underrated carbon stock. Ann. For. Sci. 70 (5), 515–524.

Cunia T., 1987. Error of forest inventory estimates: its main components. In: Wharton E. H., Cunia T. (eds) Estimating tree biomass regressions and their error Part E, USDA For. Serv., Northeast. For. Exp. Sta., Broomall, PA, USA, Gen. Tech. Rep. NE-117, pp 114.

Debell, D.S., Cole, T.G., Whitesell, C.D., 1997. Growth, Development, and Yield in Pure and Mixed Stands of Eucalyptus and Albizia. For. Sci. 43, 286–298. https://doi.org/10.1093/forestscience/43.2.286.

DeBell, D.S., Whitesell, C.D., Schubert, T.H., 1989. Using N2-Fixing Albizia to Increase Growth of Eucalyptus Plantations in Hawaii. For. Sci. 35, 64–75. https://doi.org/10.1093/forestscience/35.1.64.

FAO. 2020. Better data, better decisions – Towards impactful forest monitoring. Forestry Working Paper No. 16. Rome. 10.4060/cb0437en.

Fargione, J.E., Bassett, S., Boucher, T., Bridgham, S.D., Conant, R.T., Cook-Patton, S.C., Ellis, P.W., Falcucci, A., Fourqurean, J.W., Gopalakrishna, T., et al., 2018. Natural climate solutions for the United States, Science. Advances 4(11):eaat1869. https://doi.org/10.1126/sciadv.aat1869.

FCPF, 2021. Guidance note on estimating the uncertainty of emission reductions using Monte Carlo simulation. September 2021. World Bank, Washington DC.

Ferretti, M., König, N., Rautio, P., Sase, H., 2009. Quality assurance (QA) in international forest monitoring programmes: activity, problems and perspectives from East Asia and Europe. Ann. For. Sci. 66 (4), 1–12. https://doi.org/10.1051/forest/2009025.

Gonzalez, P., Asner, G.P., Battles, J.J., Lefsky, M.A., Waring, K.M., Palace, M., 2010. Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California. Remote Sens. Environ. 114 (7), 1561–1575. https://doi.org/10.1016/j.rse.2010.02.011.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov Chains and their applications. Biometrika 57, 97–109. https://doi.org/10.1093/biomet/57.1.97.

Hernández-Stefanoni, J.L., Dupuy, J.M., Johnson, K.D., Birdsey, R., Tun-Dzul, F., Peduzzi, A., López-Merlín, D., 2014. Improving species diversity and biomass estimates of tropical dry forests using airborne LiDAR. Remote Sens. (Basel) 6 (6), 4741–4763. https://doi.org/10.3390/rs6064741.

Hill, T.C., Williams, M., Bloom, A.A., Mitchard, E.T.A., Ryan, C.M., 2013. Are inventory based and remotely sensed above-ground biomass estimates consistent? PLoS One 8 (9), e74170.

IPCC. 2019. "Summary for Policymakers". In: 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. P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.- O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.). In press.https://www.ipcc.ch/srccl/chapter/summary-for-policymakers/.

Jandl, R., Bauhus, J., Bolte, A., Schindlbacher, A., Schüler, S., 2015. Effect of climate-adapted forest management on carbon pools and greenhouse gas emissions. Current Forestry Reports 1 (1), 1–7. https://doi.org/10.1007/s40725-015-0006-8.

Köhl, M., Lister, A., Scott, C.T., Baldauf, T., Plugge, D., 2011. Implications of sampling design and sample size for national carbon accounting systems. Carbon Balance Manag. 6 (1), 1–20. https://doi.org/10.1186/1750-0680-6-10.

Lande, R., Engen, S., Saether, B.E., 2003. Stochastic Population Dynamics in Ecology and Conservation. Oxford University Press, Oxford, UK.

Lin, J., Gamarra, J.G.P, Drake, J.E., Cuchietti, A., Yanai, R.D. 2023. Github repository for the scaling of uncertainties in allometric models. https://github.com/jlin270/Scaling_up_uncertainties_in_tree_allometry.

Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Vaglio Laurin, G., Saah, D., 2012. Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates. Int. J. For. Res. 2012 https://doi.org/10.1155/2012/436537.

McRoberts, R.E., Westfall, J.A., 2016. Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. Ann. For. Sci. 73 (3), 625–633. https://doi.org/10.1007/s13595-015-0473-x.

Melson, S.L., Harmon, M.E., Fried, J.S., Domingo, J., 2011. Estimates of live-tree carbon stores in the Pacific Northwest are sensitive to model selection. Carbon Balance Manag. 6, 2–16. https://doi.org/10.1186/1750-0680-6-2.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21 (6), 1087–1092. https://doi.org/10.1063/1.1699114.

Montagnini F. , Eibl B., Fernandez R., 2006. Rehabilitation of degraded lands in Misiones Argentina. Bois et Forets des Tropiques. 288 (2):51-65. 10.19182/bft2006.288.a20317.

Neeff, T., Somogyi, Z., Schultheis, C., Mertens, E., Rock, J., Brötz, J., Federici, S., 2017. Assessing progress in MRV capacity development: experience with a scorecard approach. Clim. Pol. 17 (2), 203–212.

Ochoa-Franco, A.D.P., Valdez-Lazalde, J.R., Santos-Posadas, H.M.D.L., Hernández-Stefanoni, J.L., Valdez-Hernández, J.I., Ángeles-Pérez, G., 2019. La densidad, riqueza y composición arbóreas definen parches detectados remotamente en una selva subperennifolia. Rev. Biol. Trop. 67 (4), 692–707. https://doi.org/10.15517/rbt.v67i4.34422.

Paré, D., Bernier, P., Lafleur, B., Titus, B.D., Thiffault, E., Maynard, D.G., Guo, X., 2013. Estimating stand-scale biomass, nutrient contents, and associated uncertainties for tree species of Canadian forests. Can. J. For. Res. 43 (7), 599–608. https://doi.org/10.1139/cjfr-2012-0454.

Pelletier, J., Martin, D., Potvin, C., 2013. REDD+ emissions estimation and reporting: Dealing with uncertainty. Environ. Res. Lett. 8 (3), 034009 https://doi.org/10.1088/1748-9326/8/3/034009.

Phillips, D.L., Brown, S.L., Schroeder, P.E., Birdsey, R.A., 2000. Toward error analysis of large-scale forest carbon budgets. Glob. Ecol. Biogeogr. 9 (4), 305–313. https://doi.org/10.1046/j.1365-2699.2000.00197.x.

Picard N., Saint-André L., Henry M., 2012. Manual for building tree volume and biomass allometric equations: from field measurement to prediction. Food and Agricultural Organization of the United Nations, Rome, and Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, 215 pp.

Picard, N., Boyemba Bosela, F., Rossi, V., 2015. Reducing the error in biomass estimates strongly depends on model selection. Ann. For. Sci. 72 (6), 811–823. https://doi.org/10.1007/s13595-014-0434-9.

Pollard, J. E., Westfall, J. A., Patterson, P. A., Gartner, D. L., 2005. FIA quality assurance program: evaluation of a tree matching algorithm for paired forest inventory data. In: 2003 Proceedings of the Fifth Annual Forest Inventory and Analysis Symposium; 2003 November 18-20; New Orleans, LA. Gen. Tech. Rep. WO-69. Washington, DC: U.S. Department of Agriculture Forest Service. 222p.

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., Hérault, B., 2017. Biomass: an R package for estimating above-ground biomass and its uncertainty in tropical forests. Methods Ecol. Evol. 8 (9), 1163–1167. https://doi.org/10.1111/2041-210X.12753.

Romijn, E., Lantican, C.B., Herold, M., Lindquist, E., Ochieng, R., Wijaya, A., Verchot, L., 2015. Assessing change in national forest monitoring capacities of 99 tropical countries. For. Ecol. Manage. 352, 109–123. https://doi.org/10.1016/j.foreco.2015.06.003.

Secretaría de Ambiente y Desarrollo Sustentable. 2007. Primer Inventario Nacional de Bosques Nativos - Informe Regional Selva Misionera. 1^{a} ed. Buenos Aires. Secretaría de Ambiente y Desarrollo Sustentable. https://www.argentina.gob.ar/sites/default/files/primer_inventario_nacional_-_informe_regional_selva_misionera_0.pdf.

Secretaría de Desarrollo Sustentable y Política Ambiental. 2001. Primer Inventario Nacional de Bosques Nativos - Manual de análisis de fustes. Buenos Aires: Secretaría de Gobierno de Ambiente y Desarrollo Sustentable de la Nación. https://www.argentina.gob.ar/sites/default/files/manual_analisis_de_fustes_0.pdf.

Secretaría de Gobierno de Ambiente y Desarrollo Sustentable de la Nación. 2019. Segundo Inventario Nacional de Bosques Nativos: manual de campo. Buenos Aires: Secretaría de Gobierno de Ambiente y Desarrollo Sustentable de la Nación. https://www.argentina.gob.ar/sites/default/files/manual_de_campo_inbn2.pdf.

Snedecor, G.W., Cochran, W.G., 1989. Statistical Methods, 8th Edition. Iowa State University Press, Ames.

Temesgen, H., Affleck, D., Poudel, K., Gray, A., Sessions, J., 2015. A review of the challenges and opportunities in estimating above ground forest biomass using tree-level models. Scand. J. For. Res. 30 (4), 326–335. https://doi.org/10.1080/02827581.2015.1012114.

UNFCCC. 2011. Decision 2/CP.17. Outcome of the Work of the Ad Hoc Working Group on Long-term Cooperative Action under the Convention FCCC/CP/2011/9/Add.1; http://unfccc.int/resource/docs/2011/cop17/eng/09a01.pdf.

Wayson, C.A., Johnson, K.D., Cole, J.A., Olguín, M.I., Carrillo, O.I., Birdsey, R.A., 2015. Estimating uncertainty of allometric biomass equations with incomplete fit error information using a pseudo-data approach: methods. Ann. For. Sci. 72 (6), 825–834. https://doi.org/10.1007/s13595-014-0436-7.

Whitesell, C.D., Miyasaka, S.C., Strand, R.F., Schubert, T.H., McDuffie, K.E., 1988. Equations for Predicting Biomass in 2-to 6-Year-Old Eucalyptus saligna in Hawaii. USDA Forest Service Res. Note PSW - 402, 1–6. https://www.fs.usda.gov/psw/publications/documents/psw_rn402/psw_rn402.pdf.

Yanai, R.D., Battles, J.J., Richardson, A.D., Blodgett, C.A., Wood, D.M., Rastetter, E.B., 2010. Estimating uncertainty in ecosystem budget calculations. Ecosystems 13 (2), 239–248. https://doi.org/10.1007/s10021-010-9315-8.

Yanai, R.D., Wayson, C., Lee, D., Espejo, A.B., Campbell, J.L., Green, M.B., Zukswert, J. M., Yoffe, S.B., Aukema, J.E., Lister, A.J., Kirchner, J.W., Gamarra, J.G.P., 2020. Improving uncertainty in forest carbon accounting for REDD+ mitigation efforts. Environ. Res. Lett. 15 (12), 124002 https://doi.org/10.1088/1748-9326/abb96f.

Yanai, R.D., Young, A.R., Campbell, J.L., Westfall, J., Barnett, C., Dillon, G., Green, M.B., Woodall, C., 2022. FIA QC paper Measurement uncertainty in a national forest inventory: results from the Northern Region of the USA. Can. J. For. Res. 53 (3), 163–177. https://doi.org/10.1139/cjfr-2022-0062.

Yanai, R.D., J.E. Drake, H.L. Buckley, B.S. Case, P. Lilly, and R.C. Woollons. Propagating uncertainty in predicting individuals and means illustrated with foliar chemistry and forest biomass. Ecosystems, in review.

Zak, M.R., Cabido, M., Hodgson, J.G., 2004. Do subtropical seasonal forests in the Gran Chaco, Argentina, have a future? Biol. Conserv. 120 (4), 589–598.

Zapata-Cuartas, M., Sierra, C.A., Alleman, L., 2012. Probability distribution of allometric coefficients and Bayesian estimation of aboveground tree biomass. For. Ecol. Manage. 277, 173–179. https://doi.org/10.1016/j.foreco.2012.04.030.

Zuloaga, F.O., O. Morrone, and M.J. Belgrano., 2008. Catalogue of the vascular plants of the southern cone (Argentina, southern Brazil, Chile, Paraguay and Uruguay). Volume 2: Dicotyledoneae: Acanthaceae-Fabaceae (Abarema-Schizolobium) (pp. xx +-985). Missouri Botanical Garden Press, St Louis.