NRC Research Press

ARTICLE

Estimating individual-tree aboveground biomass of tree species in the western U.S.A.

Krishna P. Poudel, Hailemariam Temesgen, Philip J. Radtke, and Andrew N. Gray

Abstract: Using a large dataset compiled from studies over the years covering 23 tree species, we developed methods to estimate total and components (stem, bark, branch, and foliage) of aboveground live tree biomass. Missing components in the dataset were imputed using species-specific or generalized (species combined into softwood and hardwood groups) Dirichlet imputation. Geometric means of the imputed stem wood proportions were 8% and 9% higher than the observed geometric mean of stem wood proportions in softwood and hardwood species, respectively. For other components, the differences were within 1%. On average, the component ratio method (CRM), used for the official United States forest carbon inventories, underestimated the aboveground biomass (AGB, kg) predictions by 3.7% with a very wide range (–70.3% to 31.6%). Compared with the CRM approach, equations developed in this study reduced RMSE of AGB by as much as 145.0%. On average, new equations reduced RMSE in predicting individual-tree AGB by 15.5% compared with the CRM approach and by 3.9% compared with a calibration of CRM AGB. Predicting AGB as a function of stem volume was not as accurate as using direct AGB equations. Generalized component ratio equations may be suitable for the stem wood component but were highly biased for other components.

Key words: biomass, western United States, Dirichlet imputation, component ratio method, carbon.

Résumé : À l'aide d'un important jeu de données compilées à partir d'études couvrant plusieurs années et 23 espèces d'arbres, nous avons élaboré des méthodes pour estimer la biomasse aérienne totale et ses composantes (fût, écorce, branches et feuillage) chez les arbres vivants. Les composantes manquantes dans le jeu de données ont été imputées en utilisant la méthode de Dirichlet individuellement pour chaque espèce ou collectivement en formant un groupe pour les espèces résineuses et un autre pour les espèces feuillues. Les moyennes géométriques imputées des proportions de bois de fût étaient respectivement 8 et 9 % plus élevées que les moyennes géométriques observées des proportions de bois de fût chez les espèces résineuses et feuillues. Dans le cas des autres composantes, les différences étaient en deçà de 1 %. En moyenne, la méthode des ratios des composantes (MRC), utilisée pour les inventaires officiels du carbone forestier aux États-Unis, sous-estimait les prédictions de biomasse aérienne (BA, kg) de 3,7 % avec un très grand écart (–70,3 à 31,6 %). Comparativement à la MRC, les équations élaborées dans cette étude ont réduit l'erreur quadratique moyenne (EQM) de la BA jusqu'à 145,0 %. En moyenne, les nouvelles équations pour prédire la BA d'arbres individuels ont réduit l'EQM de 15,5 % comparativement à la MRC, et de 3,9 % comparativement à une calibration de la BA par la MRC. La prédiction de la BA en fonction du volume du fût n'était pas aussi précise que l'utilisation directe des équations de la BA. Les équations généralisées du ratio des composantes peuvent convenir pour la composante du bois de fût mais elles étaient fortement biaisées pour les autres composantes. [Traduit par la Rédaction]

Mots-clés : biomasse, ouest des États-Unis, imputation par la méthode de Dirichlet, méthode des ratios des composantes, carbone.

Introduction

Forest biomass is important commercially for fuelwood and fiber production and scientifically for studies of ecosystem productivity (Parresol 1999). It is also useful in quantifying carbon stocks and sequestration rates, assessing potential impacts due to climate change, locating bio-energy processing plants, and mapping and planning fuel treatments (Temesgen et al. 2015). The official United States (US) forest carbon inventories rely on tree aboveground biomass (AGB) predictions from sample tree measurements and forest area estimates of the USDA Forest Service, Forest Inventory and Analysis (FIA) program (Heath et al. 2009).

The amount of biomass as living vegetation or dead wood and debris is an important factor that determines forest function in regulating atmospheric carbon (Brown 2002). An estimated 296 Gt of carbon is contained in above- and below-ground forest biomass (Food and Agriculture Organization of the United Nations (FAO) 2016). However, growth rates, harvest activities, natural disturbances such as wildfire, and loss of forest cover due to land-use changes can alter the carbon stocks and carbon-fixing ability of a forest. Thus, the potential contribution of forest ecosystems in meeting a variety of human goals requires good information on current and future biomass predictions.

Estimates of component biomass are important because of the different roles that they play in forest management. Crown biomass estimates are useful in assessing fuel load, formulating fire management strategies, and developing wildfire models. Knowledge of biomass present in small branches, leaves, and bark is useful when assessing the feasibility of bioenergy plant establish-

Received 22 August 2018. Accepted 12 December 2018.

K.P. Poudel. Department of Forestry, Mississippi State University, Box 9681, Mississippi State, MS 39762, USA.
H. Temesgen. Department of Forest Engineering, Resources, and Management, College of Forestry, Oregon State University, 280 Peavy Hall, Corvallis, OR 97331, USA.
P.J. Radtke. Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute, Blacksburg, VA 24061, USA.
A.N. Gray. USDA Forest Service, Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331, USA.
Corresponding author: Krishna P. Poudel (email: kpp70@msstate.edu).

 Published at www.nrcresearchpress.com/cjfr on 19 December 2018.

ment. Because direct measurement of biomass involves destructive sampling of the subject tree, allometric equations that estimate the biomass of tree components based on standing tree dendrometric variables are used. One of the desirable properties in fitting component biomass equations is that the sum of the predicted component biomass is the same as the total biomass obtained from the equation for AGB. In recent decades, this property was attained by fitting a constrained simultaneous system of equations using seemingly unrelated regression (e.g., Parresol 1999). Recently, component biomass has been predicted as the product of predicted total biomass and the component proportion has been predicted from generalized linear models such as Dirichlet regression (Poudel and Temesgen 2016*a*; Zhao et al. 2016; Dahlhausen et al. 2017), multinomial logistic regression (Poudel and Temesgen 2016*a*), or a fractional multinomial logit regression (Zhao et al. 2016).

The type and amount of data and the number of equations required to accurately quantify biomass is unknown. When predicting biomass at the local scale, site- and species-specific equations might be preferred to regional equations in their current form. Additionally, locally developed species-specific biomass equations may provide biomass estimates that are considerably different from the estimates obtained by using regional biomass equations without local calibration. On the other hand, generalized equations calibrated with location-specific best unbiased linear predictors based on small sample sizes can perform better than fitting location-specific equations based on much larger samples (de-Miguel et al. 2014). Another alternative is to directly include stand variables that are correlated with difference between local equations (e.g., Wirth et al. 2004; Forrester et al. 2017).

There have been multiple efforts to develop national-scale biomass estimators in the US and other countries. Jenkins et al. (2003) developed national-scale biomass equations based on meta-analysis of the compiled diameter-based equations. The resulting equations use diameter at breast height (DBH) as the only predictor of total AGB and were the basis of the US forest carbon inventories until 2009. Using the allometric equation form adopted by Jenkins et al. (2003), biomass increases with DBH (Heath et al. 2009) without any accounting for variation in stem form or height (Zhou and Hemstrom 2009). In 2009, the FIA updated its biomass prediction models by switching to a method known as the component ratio method (CRM) to estimate biomass of medium and large trees (Woodall et al. 2011). The CRM was proposed for consistent national biomass estimation based on FIA volume estimates with conversion and expansion factors for determining whole-tree and component biomass. One drawback of the CRM is its use of regional volume equations, which results in different volume and biomass estimates for trees of the same species, diameter, and height where they exist across regional boundaries (Heath et al. 2009). In addition, Domke et al. (2012) found that this attempt to balance accuracy with consistency resulted in an average decrease of 16% in national carbon stock estimates for the 20 most abundant tree species in the 48 conterminous US states.

The literature on aboveground biomass estimation is voluminous. Early biomass studies were focused on fitting small-scale site- and species-specific allometric equations, while recent studies have focused more on fitting techniques and application of remote sensing techniques. Use of remote sensing technology to monitor forest resources over time has also gained substantial momentum in recent years due to its ability to cover larger spatial extents compared with ground-based measurements (e.g., Jucker et al. 2017); however, the accuracy of such methods is still uncertain in comparison with the ground-based measurements, making them unsuitable for some applications (Poudel et al. 2018). Large-scale biomass studies based on remotely sensed data usually use estimates from one of the prevailing methods and equations such as the CRM or those developed by Jenkins et al. (2003) applied to ground-truth plots as the observed true value of biomass. The accuracy of the remotely sensed estimates is seldom tested against observed biomass data obtained from destructive sampling. Instead, the estimates obtained from these methods have been compared with one another.

Recently, Radtke et al. (2017) performed a large-scale biomass study covering 33 states in the eastern US with a collection of tree measurement data compiled from studies conducted over the past 115 years. They compared carbon and biomass estimates obtained from the current CRM approach with the estimates obtained from six different modifications of CRM, as well as the estimates obtained from a new fit of the AGB equations. The CRM in its present formulation underestimated AGB in eastern forests, with the range of underestimation varying from 6.2% to 17% depending on the species tested. An approach that replaced stem volume, sapling AGB, and component ratio equations resulted in the greatest reduction in estimation bias and root mean squared error (RMSE) in their study.

Two main issues surface when combining data obtained from different biomass studies: (*i*) component definition — components of AGB may be defined differently in different studies; and (*ii*) missing components — one or more biomass components could be missing in some studies. While there is not much that one can do with inconsistent variable definitions unless additional information is also reported in the past studies, there are methods for addressing missing components. Missing data are usually handled by complete-case analysis, i.e., by using trees that have data for all components; however, this can result in severe loss of information in the multivariate case (Hron et al. 2010).

Imputation methods have been successfully demonstrated as a way to fill in missing values to make better use of available information (Hron et al. 2010). Both univariate and multivariate imputations, as well as single and multiple imputations, have been used in forestry. Vonderach et al. (2018) used multiple imputation to fill in the missing biomass components to develop consistent sets of additive biomass equations for eight tree species in Germany; however, the imputed values from multiple imputation do not ensure the desired additive property unless it is assumed that the sum of the imputed component biomasses is the implied total biomass. Biomass components are compositional data as they are positive real numbers with a unit sum. Because the compositional data are represented in the simplex sample space rather than in the Euclidean space, the loss of information by using complete case analysis is even more severe (Hron et al. 2010). In other words, their sample space is the sample space of *D*-part compositions defined as

$$S^d = \left\{ (y_1, y_2, \ldots, y_D)^T \middle| y_i \geq 0,\ \sum_{i=1}^{D} y_i = 1 \right\}$$

where D represents the number of components and $d = D - 1$.

With data similar to that of Radtke et al. (2017), we assessed the accuracy of the CRM approach for 23 tree species from 10 states in the western US. New sets of volume, AGB, and component biomass equations were developed. Missing values of the components of AGB were recovered by using an imputation method. The goal was to identify whether the calibration of existing model predictions or fitting new models of AGB and component biomass would improve on the accuracy of existing methods currently in use. Specifically, the objectives of this study were to (*i*) assess the performance of Dirichlet imputation in predicting missing component biomass, (*ii*) fit equations to predict total stem cubic volume (cubic volume total stem (CVTS), m^3), (*iii*) evaluate the error produced by the CRM in estimating AGB, and (*iv*) develop new approaches to estimate AGB and component biomass and evaluate their performance against the existing methods.

Table 1. Summary statistics of diameter at breast height (DBH) and total height (HT) by species used to fit biomass equations in this study.

			DBH (cm)				HT (m)				
Species symbol	Common name	Species group	Minimum	Mean	Maximum	SD	Minimum	Mean	Maximum	SD	n
ABAM	Pacific silver fir	TF	4.5	23.5	90.4	20.4	3.1	14.3	40.3	9.2	37
ABCO	White fir	TF	28.4	42.6	66.5	11.7	17.7	28.2	39.7	7.1	10
ABGR	Grand fir	TF	12.4	43.9	84.3	22.6	8.7	28.1	44.4	11.7	17
ABLA	Subalpine fir	TF	3.5	17.3	44.4	9.6	2.2	11.5	27.9	6.9	98
ABMA	California red fir	TF	38.9	44.5	50.0	7.9	22.3	22.7	23.1	0.5	2
CHNO	Alaska yellow cedar	CL	9.3	30.1	60.2	20.2	6.8	18.3	26.6	8.1	5
JUOS	Utah juniper	WO	4.1	19.5	50.5	11.4	2.7	5.4	9.4	1.5	33
LAOC	Western larch	CL	16.0	38.4	60.7	15.1	18.9	25.7	34.2	5.3	7
PIEN	Engelmann spruce	SP	0.7	18.0	57.6	13.7	1.5	12.4	40.8	8.7	74
PIGL	White spruce	SP	3.6	17.6	57.6	12.8	2.2	13.0	37.5	9.1	108
PIMA	Black spruce	SP	3.5	16.3	38.4	5.9	2.9	13.9	30.1	4.3	67
PISI	Sitka spruce	SP	7.2	21.9	47.8	12.8	4.6	18.9	40.7	12.6	19
PICO	Lodgepole pine	PI	0.9	18.3	60.0	11.4	1.5	14.4	39.6	8.3	218
PIJE	Jeffrey pine	PI	14.5	29.4	43.4	10.0	8.4	15.7	21.0	4.0	9
PIMO	Western white pine	PI	6.0	7.8	9.3	1.7	4.3	5.2	6.9	1.5	3
PIPO	Ponderosa pine	PI	0.8	28.1	63.0	15.1	1.4	15.8	37.9	8.6	158
PIMO	Singleleaf pinyon	PI	0.4	25.1	115.6	18.0	1.4	6.4	14.0	2.5	75
PSME	Douglas-fir	DF	1.8	29.2	114.0	21.3	2.4	21.1	64.0	12.6	424
SESE	Redwood	CL	8.3	283.0	681.0	195.3	9.4	76.0	115.6	31.9	141
SEGI	Giant sequoia	CL	15.9	390.7	850.7	237.7	8.3	69.7	96.3	20.4	43
THPL	Western redcedar	CL	3.8	19.2	54.2	11.0	3.4	12.5	32.4	6.6	50
TSHE	Western hemlock	TF	3.1	28.8	77.8	16.6	3.5	22.3	53.0	11.9	115
TSME	Mountain hemlock	TF	13.5	40.6	76.2	25.1	6.1	21.0	37.2	12.7	9
ALRU	Red alder	AA	9.3	20.6	51.8	10.7	7.4	18.5	31.6	7.4	25
BEPA	Paper birch	MB	4.7	14.5	30.6	6.6	5.6	13.5	21.9	4.4	23
POTR	Quaking aspen	AA	6.0	18.4	34.8	8.3	5.4	16.0	25.6	5.7	26
POBA	Black cottonwood	AA	7.4	16.7	30.6	7.5	6.9	13.6	23.5	4.7	19
Total											1815

Note: Species groups are the group as identified in Jenkins et al. (2003): TF, true fir–hemlock; CL, cedar–larch; WO, woodland species; SP, spruce; PI, pine; DF, Douglas-fir; AA, alder–aspen–cottonwood–willow; MB, soft maple–birch. SD, standard deviation.

Materials and methods

Data

Radtke et al. (2015) compiled a large collection of taper, volume, biomass, and wood density data from North American trees destructively sampled over the past century. The collection included stem volume records from 15 485 trees and AGB records from 1815 trees from the western US (Table 1). Data used in this study came from 10 western US states (Alaska, Washington, Oregon, California, Nevada, Arizona, Utah, Idaho, Montana, and Colorado) from 23 tree species. Figure 1 shows the approximate locations where the sample trees were harvested. Many of the AGB records were separated into four major aboveground components — stem wood, stem bark, branches (wood and bark), and foliage — although some component measurements were unaccounted for in a number of sample trees (Fig. 2). Species groupings following Jenkins et al. (2003) were adopted to accommodate species with very small sample sizes (n < 9). Note that the choice of species grouping is completely arbitrary and that there are many other possibilities such as grouping by species functional traits, grouping by taxa, and specific gravity as done by Chojnacky et al. (2014).

Missing components were imputed using Dirichlet imputation by species or species groups. The Euclidean distance, squared Euclidean distance, and Mahalanobis distance are the commonly used distance metrics in the univariate imputations, but they are not appropriate for compositional data such as biomass components because the compositional data are represented in the simplex sample space rather than in the Euclidean space (Hron et al. 2010). Dirichlet imputation uses the Aitchison distance to impute missing components. The Aitchison distance for two compositions $\boldsymbol{x} = (x_1, x_2, \ldots, x_D)^{\mathrm{T}}$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_D)^{\mathrm{T}}$ is defined as

$$(1)\qquad d_A(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\frac{1}{D}\sum_{i=1}^{D-1}\sum_{j=i+1}^{D}\left(\ln\frac{x_i}{x_j} - \ln\frac{y_i}{y_j}\right)^2}$$

Dirichlet imputation requires at least some individuals with all components measured. For species missing any component observations on all trees, the imputation was performed at the species-group level. Dirichlet imputation was implemented with R function impCoda in library robCompositions (Templ et al. 2011). Dirichlet imputation results in different values than the actual values for the observed components; however, the original values can be recovered with a simple calculation:

$$(2)\qquad \hat{p}_{cj} = p_{cj} \times \frac{S_{\mathrm{O}j}}{S_{\mathrm{I}j}}$$

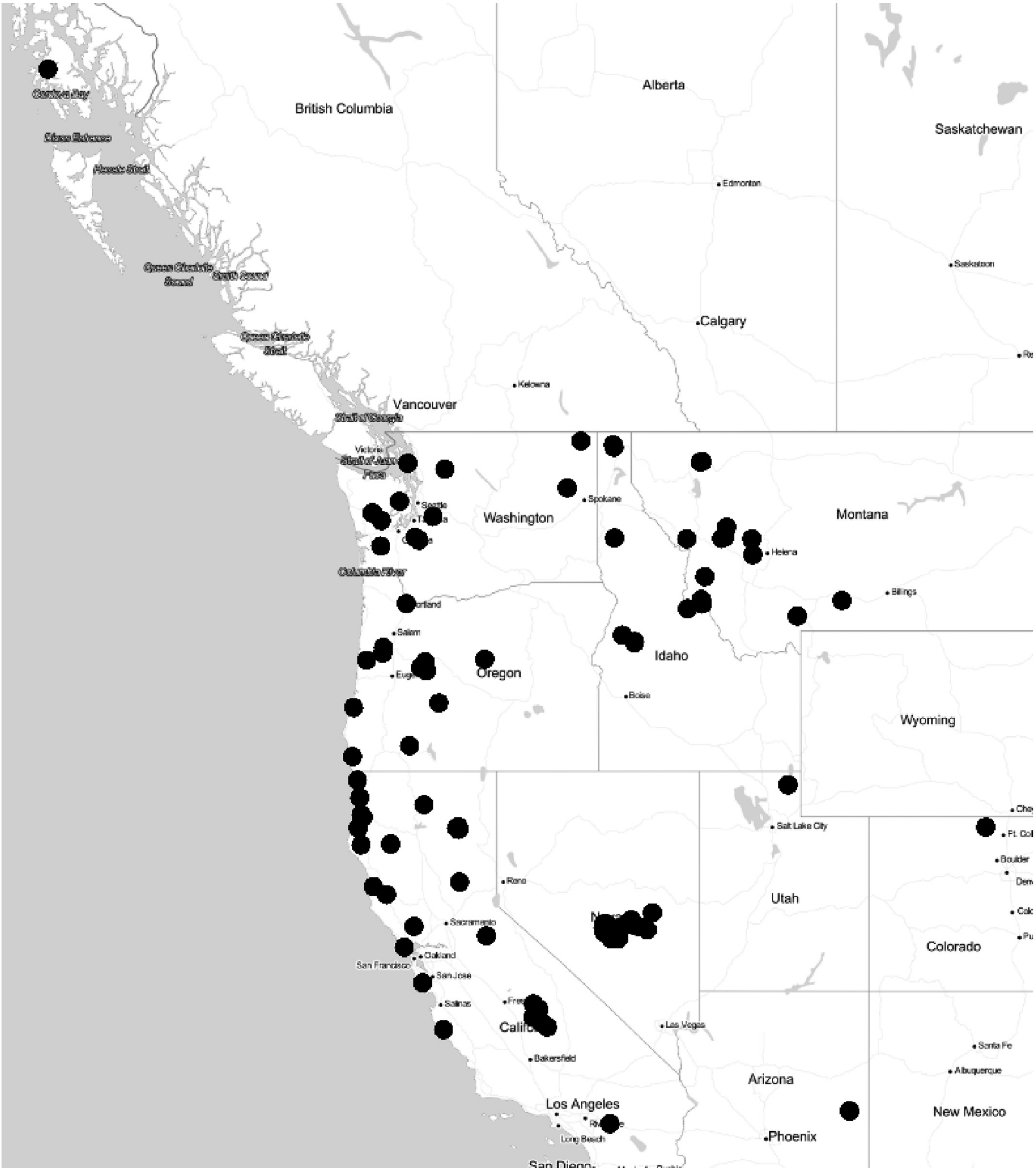
where $\hat{p}_{cj}$ is the recovered proportion of component c on tree j; p_{cj} is the observed or imputed proportion of component c on tree j; $S_{\mathrm{O}j}$ is the sum of observed proportions of non-missing components for tree j; and $S_{\mathrm{I}j}$ is the sum of imputed proportions of non-missing components for tree j. For a tree with no missing components, $S_{\mathrm{O}j} = S_{\mathrm{I}j} = 1$ and therefore $\hat{p}_{cj} = p_{cj}$.

Methods

Component ratio method

CRM is a multistep process that involves converting sound volume of wood in the merchantable stem (30 cm from the ground to a 10.2 cm inside-bark diameter top) using a compiled set of wood specific gravities. Biomass in stem bark is calculated using a set of percent bark and bark specific gravities compiled by Miles and Smith (2009). Biomass of stump wood and bark is based on volume

Fig. 1. Geographic distribution of sample trees used in this study. At least one tree per stand and up to 56 trees from a given forest were sampled at each site.

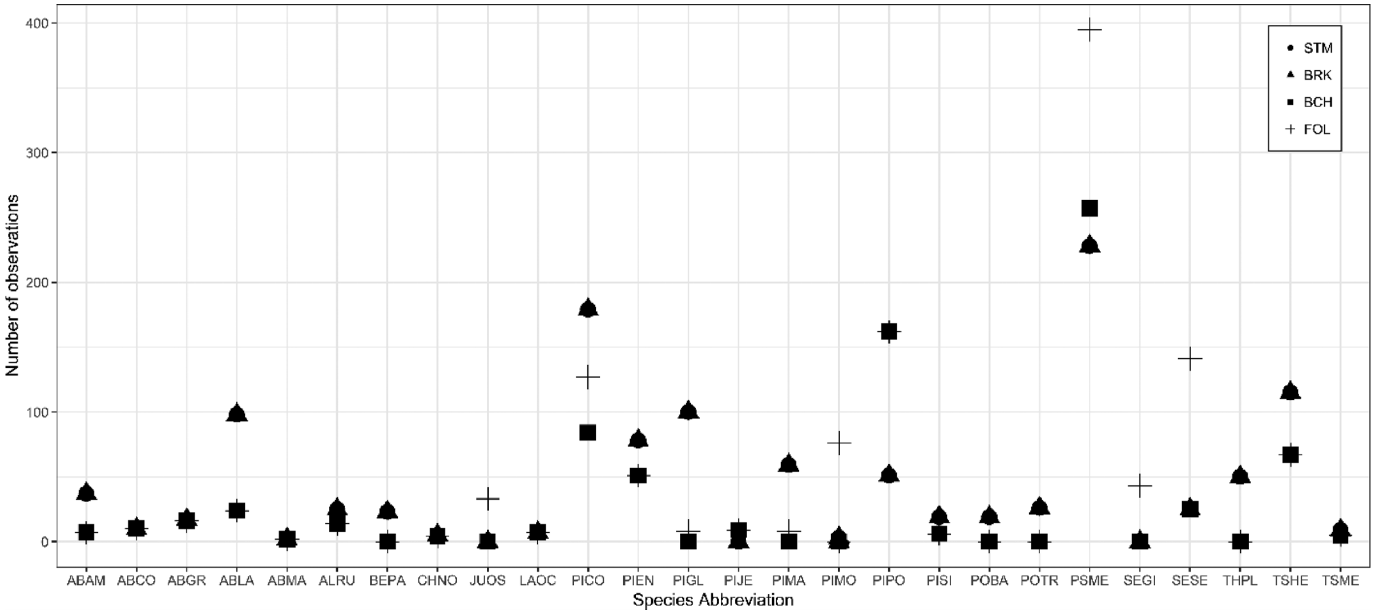


equations in Raile (1982) together with a compiled set of wood and bark specific gravities and bark percent coefficients. Biomass of non-merchantable top and limbs is calculated by subtracting biomass of merchantable stem and foliage obtained using Jenkins et al. (2003) equations and CRM estimated stump biomass from the total AGB obtained using Jenkins et al. (2003) equations and multiplying by the CRM adjustment factor. Total AGB is obtained by summing these component masses.

Volume equations

The FIA program of the USDA Forest Service has four regional units: Northern, Southern, Interior West, and Pacific Northwest.

Fig. 2. Number of observations available by aboveground biomass component and by species. Components are stem wood (STM), bark (BRK), branch (BCH), and foliage (FOL).



The CRM uses species-specific merchantable volume equations, some of which vary within and between FIA unit and subunit (Woodall et al. 2011). The merchantable stem volumes serve as the starting point for biomass conversion and expansion to AGB and other components. As a result, biomass estimates for trees of identical species, diameter, and height may differ based on their geographic locales.

The same regional volume equations provide estimates of wood volume from ground to tip. Exploratory data analysis showed a nonlinear relationship between DBH and CVTS, as well as between total tree height (HT) and CVTS. Therefore, a nonlinear equation in the form of eq. 3 was fitted by species without regard to geographic location:

$$(3) \qquad y = \exp[\beta_0 + \beta_1 \ln(\mathrm{DBH}) + \beta_2 \ln(\mathrm{HT})] + \varepsilon$$

where y is the CVTS, β_0, β_1, and β_2 are the regression parameters to be estimated from the data, and ε is the random error.

New AGB equations

Different linear and nonlinear models with varying predictor variables were examined to predict AGB. The predictors of AGB included DBH, HT, and their transformations. When a model was fitted with the log-transformation of the dependent variable, i.e., ln(AGB), the predicted values were obtained by multiplying the log-bias correction factor (CF) introduced by Baskerville (1972):

$$(4) \qquad \mathrm{CF} = \exp\left(\frac{\mathrm{MSE}}{2}\right)$$

where MSE is the mean squared error obtained by the least-squares regression.

Calibration of CRM estimated AGB

Poudel and Temesgen (2016*b*) used regression-based calibration factors to calibrate regional volume and component biomass equations. Poudel et al. (2018) used a similar approach to calibrate volume and biomass change estimated using multi-temporal lidar sampling and re-measured field inventory data. In this study, calibrated CRM estimates of AGB were developed from the CRM-estimated AGB using calibration coefficients obtained from fitting a simple linear regression model of the following form:

$$(5) \qquad \ln(\mathrm{AGB}_i) = \alpha + \beta \ln(\mathrm{AGB\_CRM}_i) + \varepsilon_i$$

where AGB_i is the observed AGB of tree i from destructive sampling or legacy studies, $\mathrm{AGB\_CRM}_i$ is the CRM-estimated AGB of tree i, α and β are the parameters in the linear regression, and ε_i is the random error. Then, with the least squares estimators a and b of α and β, respectively, the corresponding calibration equation is

$$(6) \qquad \widehat{\mathrm{AGB}}_i = \mathrm{CF} \times \exp[a + b \ln(\mathrm{AGB}_{\mathrm{CRM}_i})]$$

Thus, for the future observations, if one obtains $\mathrm{AGB\_CRM}_i$ as the regional predicted value of biomass, the unknown value AGB_i for the new tree can be estimated using eq. 6. Once again, CF is the multiplicative log-bias correction factor as defined in eq. 4. Note that the calibration coefficients, as well as the evaluation statistics, were obtained based on a 500-case resampling.

Volume-to-biomass conversion

The volume-to-biomass conversion factor for merchantable stem wood in the CRM is specific gravity, i.e., merchantable stem wood biomass = sg × V_{merch}, where sg is the wood specific gravity and V_{merch} is the merchantable wood volume. Radtke et al. (2017) achieved greatest reduction in estimation bias and RMSE of AGB by using whole-stem volume instead of merchantable volume. They converted the total stem cubic volume (CVTS) to whole-stem biomass by applying either species-specific specific gravity compiled by Miles and Smith (2009) or the observed specific gravity depending on the modeling scenario used.

In this study, we fitted simple linear regression models that converted whole-stem volume to total AGB. To add flexibility in converting stem volume to AGB, we included both slope and intercept in the regression model.

$$(7) \qquad \ln(\mathrm{AGB}_i) = \beta_0 + \beta_1 \ln(\widehat{\mathrm{CVTS}}) + \varepsilon_i$$

where $\widehat{\mathrm{CVTS}}$ is the cubic volume including top and stump obtained by using eq. 3 fitted in this study.

Component biomass estimation

Biomass in different components can be estimated in different ways, e.g., as a ratio of other components (e.g., the CRM approach), by fitting independent equations for component proportions (e.g., beta regression in Poudel and Temesgen 2016*a*), fitting a simultaneous system of component biomass equations (e.g., Parresol 1999), or fitting a simultaneous system of equations for component proportions (e.g., Dirichlet regression in Poudel and Temesgen 2016*a*). We used the Dirichlet regression to estimate the proportion of biomass in different aboveground components. The Dirichlet regression treats the aboveground biomass as compositional data and follows a Dirichlet distribution. Details of the Dirichlet regression are given in Poudel and Temesgen (2016*a*).

Evaluation

Accuracy of the CRM, as well as the new methods in estimating AGB, were compared for the study dataset using bias, bias percent, RMSE, and RMSE percent defined as follows:

$$\text{(8)} \qquad \text{Bias} = \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})}{n}$$

$$\text{(9)} \qquad \text{Bias percent} = \frac{\text{Bias}}{\bar{y}} \times 100$$

$$\text{(10)} \qquad \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{n}}$$

$$\text{(11)} \qquad \text{RMSE percent} = \frac{\text{RMSE}}{\bar{y}} \times 100$$

where y_i and $\widehat{y_i}$ are observed and predicted biomass in tree i, respectively, and $\bar{y}$ is the average biomass of n trees. These evaluation statistics for AGB estimation were obtained at each iteration of a 500-case resampling and averaged to obtain the final evaluation statistics.

Bias reflects how close the functional form of the model can get to the true relationship between the dependent and independent variables (Kuhn and Johnson 2013). RMSE gives more weight to the large but infrequent errors than the mean. Results from the model with high RMSE will significantly change the model fit with small deviations in the data (Kuhn and Johnson 2013). In addition, the RMSE penalizes models with large prediction errors. In general, bias in estimating equations should be avoided as much as possible. RMSE can be high due to natural variation in the population, but prediction equations with lower RMSE are more precise and thus more desirable. All statistical procedures were performed in R statistical software (version 3.4.4; R Core Team 2018).

We used these statistics to evaluate four approaches to estimate AGB and two approaches to estimate components of AGB. Approaches to estimate AGB are (*i*) current CRM, (*ii*) use of direct AGB equations fitted in this study, (*iii*) calibration of CRM-estimated AGB, and (*iv*) conversion of predicted CVTS to AGB using volume-to-biomass conversion equations. Approaches to estimate components of AGB are (*i*) generalized component ratio equations for hardwood and softwood species and (*ii*) species-specific component ratio equations obtained with Dirichlet regression. We would like to note that the grouping into hardwood and softwood to obtain generalized biomass equations may not be the best grouping (e.g., Ducey 2012; Forrester et al. 2017), but we used this grouping as it is used in the current CRM approach and the national-scale estimation technique developed by Jenkins et al. (2003).

Results

Imputation and CVTS modeling

The greatest differences (up to 9.2%) in the imputed and observed geometric means of component proportions were observed in imputing the stem wood component. The number of observations available for component biomass for each species is shown in Fig. 2, and geometric means of the observed and imputed component biomass proportions obtained using species-specific imputation are shown in Fig. 3. The maximum difference between observed and imputed geometric means was 9.2% (western hemlock (*Tsuga heterophylla* (Raf.) Sarg.) stem wood component).

Of the 27 species in the original dataset, only 16 species had information on all four biomass components, two species had information on three components, six species had information on two components, and three species had only one component. As mentioned before, the Dirichlet imputation requires all components to have some observed values, but some species in our dataset had one or more components with no observed values, e.g., no crown (branch and foliage) biomass information was available for western redcedar (*Thuja plicata* Donn ex D. Don (THPL); Fig. 2). Missing components in such cases were imputed using combined species data for softwood and hardwood species groups. The imputed geometric means are compared with the observed geometric means in Fig. 4. Similar to the species-specific imputations, the highest discrepancy was observed in stem wood biomass. The imputed geometric means were 8% and 9% higher than the observed geometric mean of stem wood proportions in softwood and hardwood species, respectively. For other components, the differences were within 1%.

Species-specific simple nonlinear equations with DBH and HT as predictor variables (eq. 3) were suitable for predicting CVTS inside bark (nonlinear $R^2 \geq 0.96$). Parameter estimates of the nonlinear inside-bark CVTS equation fitted in this study are given in Table 2. Parameters of all of the models were statistically significant at the 5% level of significance (p values <0.05). These volume equations were used to obtain CVTS-based estimates of AGB.

AGB estimation

Applying the current CRM approach to estimate AGB for the data used in this study indicated substantial bias and RMSE (Table 3). Percent bias ranged from –70.3% to 31.6%, with a positive overall bias across species averaging 3.7%. Overall average bias, however, might be misleading as the range of bias is so high (–70.3% to 31.6%). This is reflected in the RMSE, which ranged from 12.6% to 178.4%, with an overall average of 36.9%. In addition, the standard deviations of bias and RMSE were 22.4 and 33.0, respectively. The CRM produced the largest bias (–70.3%) and RMSE (178.4%) for Utah juniper (*Juniperus osteosperma* (Torr.) Little).

Equation forms along with their parameter estimates for the AGB equations fitted in this study are presented in Table 4. Even though the form of the prediction equation is same for all species, estimation of parameters was done either through linear or nonlinear least squares approach. Species for which the nonlinear model was used have a log-bias correction factor (CF) equal to 1.0000 in Table 4. The log-bias correction factor for logarithmic models ranged from 1.0089 to 1.0356. Bias in predicting total AGB by fitting new equations ranged from –2.7% for singleleaf pinyon (*Pinus monophylla* Torr. & Frém.) to 7.2% for giant sequoia (*Sequoiadendron giganteum* (Lindl.) J. Buchholz), with an overall mean across species of 0.9% (SD = 2.7%; Table 4). RMSE ranged from 7.1% for grand fir (*Abies grandis* (Douglas ex D. Don) Lindl.) to 40.3%, for giant sequoia, with a mean of 21% (SD = 8.8%; Table 4).

Fig. 3. Geometric mean of the observed and imputed component biomass proportions. Imputed values are obtained using Dirichlet imputation of compositional data by species. BCH, FOL, BRK, and STM represent bare branch, foliage, stem bark, and stem wood components, respectively.

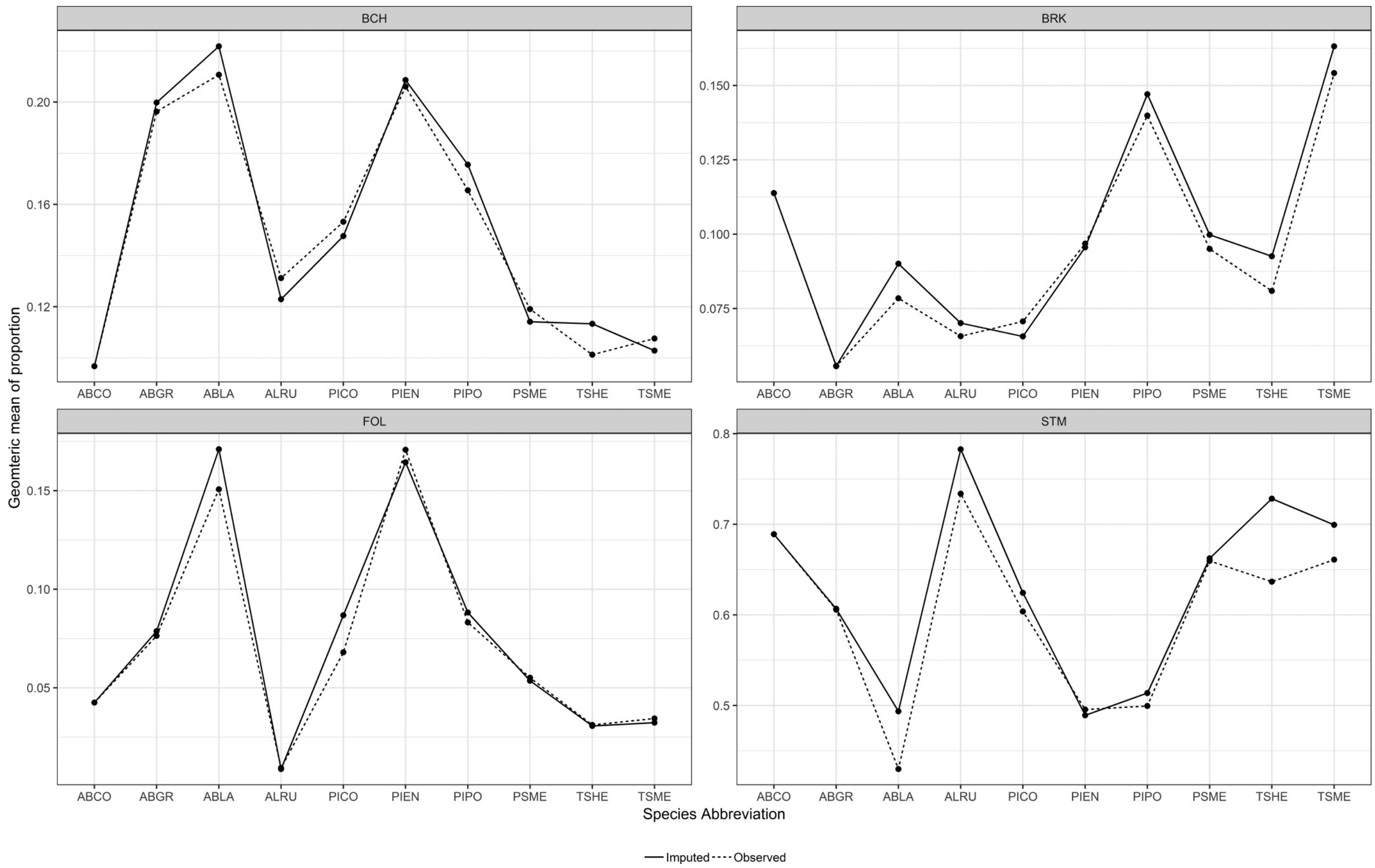
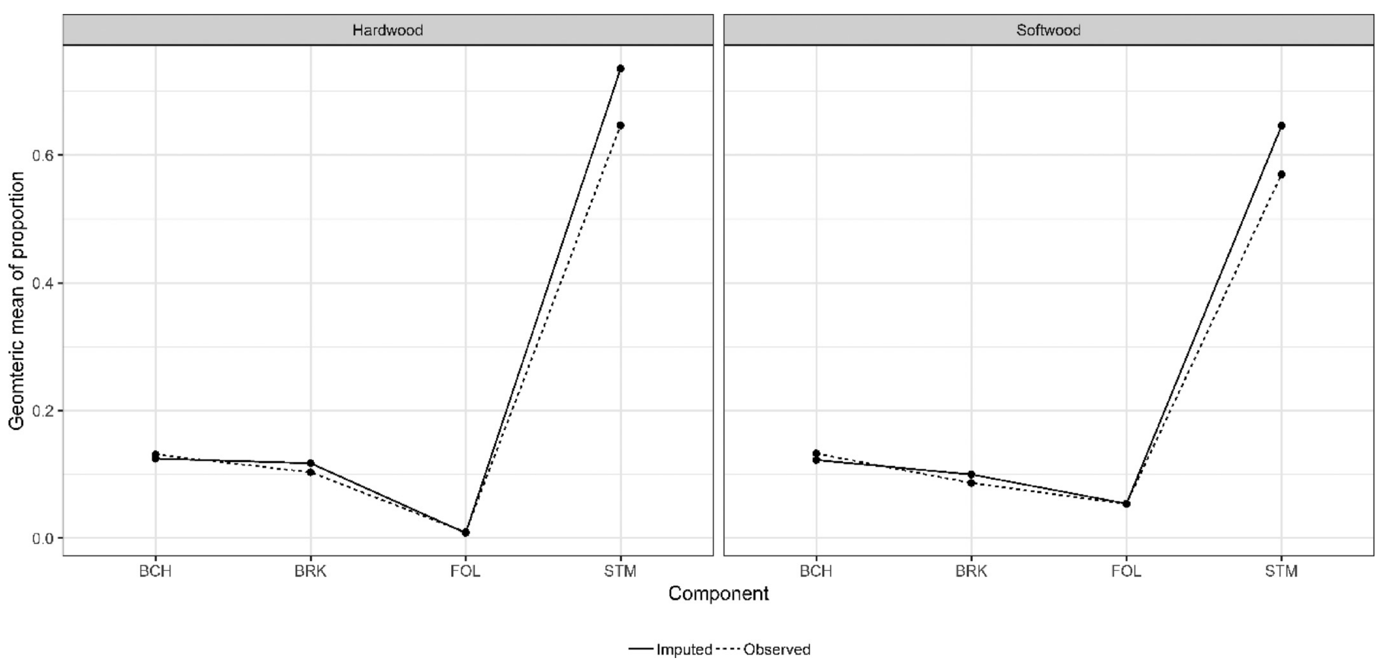


Fig. 4. Comparison of geometric mean of the original data and imputed data obtained from combined species imputation of component proportions.



Calibration coefficients along with the bias and RMSE percent obtained by calibrating CRM-estimated AGB with the data in this study are presented in Table 5. Cross-validation bias in AGB estimation using the calibration ranged from −8.4% to 7.0%, and cross-validation RMSE ranged from 6.0% to 51%. Mean bias and RMSE were 1% (SD = 3.7%) and 25.3% (SD = 10.7%), respectively.

Evaluation statistics obtained from CRM, calibration of CRM, and new fit of the AGB equations were compared for their accu-

Table 2. Parameter estimates and standard errors (in parentheses) of inside-bark CVTS equations fitted in this study.

Species symbol	Common name	*n*	*a*	*b*	*c*
ABAM	White fir	1417	–10.0535 (0.05)	1.5960 (0.02)	1.3454 (0.03)
ABCO	Grand fir	94	–9.1826 (0.24)	1.5459 (0.07)	1.1730 (0.09)
ABGR	Subalpine fir	169	–9.8457 (0.11)	1.7032 (0.04)	1.1489 (0.05)
ABLA	California red fir	278	–10.2670 (0.10)	1.6937 (0.03)	1.3084 (0.04)
ABMA	Alaska yellow cedar	673	–9.6022 (0.08)	1.8398 (0.02)	0.9508 (0.02)
CHNO	Western larch	29	–11.3275 (0.34)	1.9347 (0.11)	1.3343 (0.20)
CADE	Incense-cedar	171	–9.6396 (0.07)	1.6648 (0.03)	1.0801 (0.04)
PICEA	Spruce spp.	9	–10.3812 (0.47)	1.8241 (0.13)	1.1885 (0.20)
PIEN	Engelmann spruce	509	–10.3011 (0.08)	1.6880 (0.02)	1.3210 (0.03)
PIGL	White spruce	21	–10.5658 (0.30)	2.1248 (0.10)	0.8943 (0.10)
PISI	Sitka spruce	244	–10.1366 (0.12)	1.8640 (0.03)	1.0835 (0.05)
PICO	Lodgepole pine	264	–10.0258 (0.11)	1.8580 (0.04)	1.0636 (0.04)
PIFL	Limber pine	16	–11.4217 (0.50)	2.1371 (0.09)	1.1979 (0.09)
PIJE	Jeffrey pine	39	–10.4643 (0.29)	2.0332 (0.12)	0.9415 (0.13)
PILA	Sugar pine	253	–10.6985 (0.24)	2.4396 (0.07)	0.4890 (0.08)
PIMO	Western white pine	29	–9.5997 (0.36)	1.4028 (0.09)	1.4115 (0.13)
PIPO	Ponderosa pine	3554	–10.5808 (0.04)	2.1110 (0.01)	0.9126 (0.01)
PSME	Douglas-fir	5536	–9.5504 (0.02)	1.7420 (0.01)	1.0212 (0.01)
SESE	Redwood	187	–10.7638 (0.19)	2.0497 (0.04)	0.9477 (0.05)
THPL	Western redcedar	605	–9.5468 (0.11)	2.0363 (0.2)	0.6601 (0.05)
TSUGA	Hemlock spp.	25	–9.4684 (0.33)	1.8321 (0.05)	0.9369 (0.10)
TSHE	Western hemlock	821	–9.9763 (0.07)	1.9583 (0.01)	0.9254 (0.02)
TSME	Mountain hemlock	19	–10.9806 (0.44)	1.8803 (0.14)	1.3024 (0.14)
ALRU	Red alder	404	–10.3462 (0.09)	2.0513 (0.03)	0.9448 (0.04)
POTR	Quaking aspen	119	–10.7487 (0.18)	1.8981 (0.05)	1.2408 (0.07)

Note: The model form was $CVTS_{ib} = \exp[a + b \times \ln(DBH) + c \times \ln(HT)]$, where $CVTS_{ib}$ is the inside-bark cubic volume from ground to the tip of the tree, DBH is the diameter at breast height (cm), and HT is the total tree height (m). Parameter estimates were statistically significant at the 5% level of significance (p values <0.05).

Table 3. Bias and root mean squared error (RMSE) percentages obtained from applying CRM to estimate AGB of the legacy tree data.

Species symbol	Common name	Bias %	RMSE %
ABAM	Pacific silver fir	–9.0	38.1
ABCO	White fir	11.4	18.0
ABGR	Grand fir	9.5	17.7
ABLA	Subalpine fir	27.0	49.4
JUOS	Utah juniper	–70.3	178.4
PIEN	Engelmann spruce	14.7	28.0
PIGL	White spruce	10.2	31.7
PIMA	Black spruce	19.1	27.5
PISI	Sitka spruce	6.2	23.5
PICO	Lodgepole pine	20.3	39.7
PIJE	Jeffrey pine	31.6	37.7
PIPO	Ponderosa pine	24.3	39.3
PIMO	Singleleaf pinyon	27.4	42.6
PSME	Douglas-fir	–0.3	26.8
SESE	Redwood	–32.4	57.6
SEGI	Giant sequoia	–17.4	44.7
THPL	Western redcedar	13.4	23.8
TSHE	Western hemlock	–10.8	32.4
TSME	Mountain hemlock	–1.8	12.6
ALRU	Red alder	1.7	23.9
BEPA	Paper birch	–4.5	22.8
POTR	Quaking aspen	0.6	12.8
POBA	Black cottonwood	14.5	20.8

racy. Calibration of the CRM using simple linear regression reduced RMSE by as much as 126.5% (for Utah juniper). On average, calibration of CRM reduced RMSE in estimating individual-tree AGB by 11.7% (SD = 26.6%). Compared with the CRM approach, new equations reduced RMSE for all species. RMSE in AGB was reduced by as much as 145.0% (for Utah juniper) by using new equations fitted in this study. The smallest reduction in RMSE was observed in Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco; 0.7% reduction). On average, new equations reduced RMSE in estimating individual-tree AGB by 15.5% (SD = 29.3%) compared with the CRM approach and by 3.9% (SD = 6.4%) compared with the calibration of CRM.

We fitted volume-to-biomass equations using predicted CVTS from volume equations fitted in this study as the predictor of the AGB (eq. 7). Coefficients for volume-based AGB prediction equations are given in Table 6. When there were no data for fitting species-specific volume-to-biomass equations, coefficients were obtained from species group specific volume-to-biomass equations. RMSEs were obtained only for species that had both volume and biomass data.

Component biomass estimation

Jenkins et al. (2003) fitted two sets of generalized component ratio equations for softwood and hardwood trees as a function of DBH. These equations are used to estimate the component ratios in the CRM approach, as well as the total AGB. We fitted generalized component ratio equations simultaneously with Dirichlet regression and used both DBH and height as the predictors. Parameter estimates and their standard errors for softwood and hardwood species groups obtained from Dirichlet regression fit are given in Table 7. These coefficients were applied to the total AGB estimated using equations fitted in this study to obtain biomass in different components.

All-species mean bias for stem wood was –1.4% (SD = 10.6%), with individual-species bias ranging from –20.2% to 28.6% (Table 8). Seven of 23 species had bias greater than 10% for the stem wood component. RMSE for stem wood ranged from 4.8% to 32.6%, with an all-species average of 16.6%. Bias and RMSE were higher for other components, with all-species average biases of –17.2% (range –72.8% to 27.7%), 1.2% (range –167.8% to 43.3%), and –19.0% (range –249.1% to 38.0%) for bark, branch, and foliage components, respectively, and all-species average RMSEs of 44.2% (range 19.5% to 94.6%), 47.6% (range 22.1% to 190.3%), and 67.7% (range 25.5% to 326.0%) for bark, branch, and foliage components, respectively.

Compared with the generalized component ratio equations for softwood and hardwood trees, bias and RMSE were reduced sub-

Table 4. Parameter estimates and standard errors (in parentheses) of species-specific biomass equations and evaluation statistics (bias and root mean squared error (RMSE) percentages) obtained from fitting regression models with the data used in this study.

Common name	*a*	*b*	*c*	*d*	CF	Bias %	RMSE %
Pacific silver fir	−2.7231 (0.22)	1.8546 (0.12)	0.7567 (0.18)		1.0000	−1.0	10.6
White fir	−1.5759 (0.62)	2.2315 (0.17)			1.0089	0.9	11.7
Grand fir	−2.2106 (0.30)	2.4297 (0.07)			1.0000	−0.4	7.1
Subalpine fir	−5.5175 (0.88)	2.6795 (0.11)	1.2805 (0.46)	−0.0759 (0.02)	1.0000	2.4	22.5
Utah Juniper	−0.7457 (0.50)	1.8999 (0.14)			1.0000	3.4	33.4
Engelmann spruce	−2.6483 (0.19)	1.4762 (0.09)	1.1357 (0.08)		1.0000	0.8	20.0
White spruce	−1.9682 (0.18)	1.9401 (0.09)	0.4210 (0.09)		1.0000	−1.7	22.3
Black Spruce	−2.6988 (0.15)	1.9463 (0.11)	0.6183 (0.12)		1.0000	0.7	16.5
Sitka spruce	−2.5826 (0.16)	1.8196 (0.15)	0.7108 (0.13)		1.0091	−0.4	22.1
Lodgepole pine	−1.8641 (0.14)	1.8010 (0.03)	0.5617 (0.03)		1.0000	−1.7	22.0
Jeffrey pine	−3.1894 (0.75)	3.2244 (0.27)			1.0258	3.0	23.6
Ponderosa pine	−0.6616 (0.09)	0.8288 (0.09)	0.2127 (0.12)	0.4145 (0.08)	1.0246	−0.6	25.2
Singleleaf pinyon	−0.5048 (0.26)	1.4016 (0.04)	0.8522 (0.12)		1.0000	−2.7	29.5
Douglas-fir	−2.8246 (0.11)	1.6385 (0.03)	1.0474 (0.04)		1.0000	0.3	26.1
Redwood	−3.0491 (0.14)	1.9407 (0.05)	0.6876 (0.09)		1.0205	4.9	28.7
Giant sequoia	−3.8735 (0.42)	2.1251 (0.11)	0.601 (0.21)		1.0356	7.2	40.3
Western redcedar	−2.5443 (0.12)	1.5701 (0.12)	0.9162 (0.13)		1.0151	2.3	22.3
Western hemlock	−2.7552 (0.10)	1.8521 (0.08)	0.7642 (0.08)		1.0209	6.6	30.6
Mountain hemlock	−0.9861 (0.58)	2.0975 (0.14)			1.0000	−2.3	8.8
Red alder	−3.8516 (0.33)	2.3174 (0.22)	0.6283 (0.25)		1.0281	−1.8	18.9
Paper birch	−2.5782 (0.22)	2.5299 (0.08)			1.0179	2.6	19.0
Quaking aspen	−3.0439 (0.42)	2.0345 (0.07)	0.6420 (0.12)		1.0000	−0.1	8.9
Black cottonwood	−3.8009 (0.41)	1.6300 (0.14)	1.3354 (0.16)		1.0000	−0.3	12.7

Note: Prediction equations were in the form of $\exp(a + bX1 + cX2)$, except for subalpine, which was $\exp(a + bX1 + cX2 + dX3)$, and ponderosa pine, which was $\exp(a + bX1 + cX1^2 + dX2)$; $X1 = \ln(\text{DBH})$; $X2 = \ln(\text{HT})$; $X3 = \text{HT}$. CF is the Baskerville (1972) log-bias correction factor, DBH is the diameter at breast height (cm), and HT is the total tree height (m). CF = 1.0000 indicates that the dependent variable (AGB) was not log-transformed.

Table 5. Calibration coefficients and standard errors (in parentheses) and bias and root mean squared error (RMSE) percentages obtained from calibrating CRM-estimated AGB using simple linear regression.

Common name	*a*	*b*	Bias %	RMSE %	CF
Pacific silver fir	0.9116 (0.07)	0.8699 (0.02)	5.9	28.0	1.0148
White fir	1.3292 (0.34)	0.8429 (0.05)	0.3	11.3	1.0052
Grand fir	0.7777 (0.23)	0.9168 (0.03)	0.9	15.0	1.0215
Subalpine fir	1.0475 (0.07)	0.8720 (0.02)	0.8	31.0	1.0409
Utah juniper	2.9692 (0.20)	0.4795 (0.04)	1.8	51.9	1.1077
Engelmann spruce	1.1920 (0.04)	0.8446 (0.01)	1.5	23.5	1.0147
White spruce	0.9368 (0.05)	0.8820 (0.01)	−3.0	27.0	1.0274
Black spruce	1.1011 (0.09)	0.8374 (0.02)	−1.1	18.9	1.0201
Sitka spruce	1.0507 (0.10)	0.8615 (0.02)	−2.2	19.1	1.0134
Lodgepole pine	1.0353 (0.06)	0.8637 (0.01)	−0.3	28.7	1.0788
Jeffrey pine	1.0447 (0.23)	0.8995 (0.04)	−1.1	12.5	1.0066
Ponderosa pine	1.7960 (0.06)	0.7777 (0.01)	2.6	25.4	1.0481
Singleleaf pinyon	1.9756 (0.10)	0.7480 (0.02)	4.9	33.4	1.0643
Douglas-fir	1.0408 (0.03)	0.8652 (0.004)	4.6	38.1	1.0274
Redwood	0.9150 (0.07)	0.9021 (0.01)	6.0	30.6	1.0248
Giant sequoia	−0.0464 (0.22)	0.9833 (0.02)	7.0	41.4	1.0366
Western redcedar	0.3483 (0.12)	0.9746 (0.02)	−8.4	29.2	1.0430
Western hemlock	0.8680 (0.06)	0.8679 (0.03)	5.9	35.6	1.0240
Mountain hemlock	1.1898 (0.16	0.8579 (0.01)	−0.7	6.0	1.0122
Red alder	0.0852 (0.15)	0.9885 (0.03)	−1.4	20.1	1.0202
Paper birch	0.0267 (0.14)	0.9859 (0.03)	−0.9	20.3	1.0174
Quaking aspen	0.4224 (0.14)	0.9347 (0.03)	−3.8	14.7	1.0236
Black cottonwood	1.3979 (0.11)	0.7769 (0.03)	2.7	19.7	1.0142

Note: Calibration coefficients, as well as evaluation statistics, were obtained from 500 case resampling. The calibration equation is $\text{AGB} = \text{CF} \times \exp[a + b \times \ln(\text{AGB\_CRM})]$. CF is the log-bias correction factor.

stantially by using species-specific component ratio equations, as was expected (Table 9). Bias in stem wood was less than 2% for all species, with an average of 0.5% (range 0.0% to 1.7%) and average RMSE of 11.4% (range 2.3% to 18.5%). Similar to the generalized component ratio equations, bias and RMSE were comparatively higher for other components. Average biases for bark, branch, and foliage were −3.9% (range −17.5% to 2.5%), 0.9% (range −1.7% to 7.9%), and −1.8% (range −17.5% to 3.6%), respectively, and average RMSEs were 25.0% (range 6.5% to 59.4%), 27.4% (range 3.4% to 48.2%), and 39% (range 13.5% to 82.4%) for bark, branch, and foliage components, respectively.

Discussion

Forest ecosystems contribute substantially to global climate change mitigation by sequestering and storing carbon. Forest carbon inventories are obtained by using tree and area measure-

Table 6. Parameter estimates and standard errors (in parentheses) of the volume-to-biomass equations fitted in this study.

Common name	*a*	*b*	CF	Basis	RMSE %
Pacific silver fir	6.3389 (0.02)	0.8690 (0.01)	1.0333	GRP	—
White fir	6.4429 (0.06)	0.8396 (0.09)	1.0165	SPP	19.9
Grand fir	6.3170 (0.06)	1.0150 (0.04)	1.0271	SPP	15.2
Subalpine fir	6.2418 (0.05)	0.8175 (0.02)	1.0382	SPP	40.2
Utah juniper	—	—	—	—	—
Engelmann spruce	6.1041 (0.05)	0.7423 (0.01)	1.0469	SPP	44.3
White spruce	6.3052 (0.04)	0.7890 (0.01)	1.0265	SPP	25.8
Black spruce	6.1913 (0.03)	0.7655 (0.01)	1.0373	GRP	—
Sitka spruce	6.1467 (0.04)	0.8489 (0.02)	1.0101	SPP	22.0
Lodgepole pine	6.2043 (0.03)	0.8057 (0.01)	1.0531	SPP	32.8
Jeffrey pine	6.6600 (0.06)	0.9500 (0.04)	1.0071	SPP	13.8
Ponderosa pine	6.4661 (0.03)	0.7567 (0.01)	1.0482	SPP	26.6
Singleleaf pinyon	6.3278 (0.03)	0.7959 (0.01)	1.0706	GRP	26.2
Douglas-fir	6.4313 (0.01)	0.9589 (0.005)	1.0242	SPP	29.8
Redwood	6.4452 (0.03)	0.9011 (0.01)	1.0209	SPP	23.9
Giant sequoia	6.2848 (0.02)	0.9323 (0.004)	1.0247	GRP	—
Western redcedar	6.1092 (0.05)	0.9114 (0.02)	1.0167	SPP	—
Western hemlock	6.3161 (0.02)	0.9079 (0.01)	1.0209	SPP	29.7
Mountain hemlock	6.5393 (0.06)	0.8812 (0.03)	1.0136	SPP	8.7
Red alder	6.3749 (0.08)	1.0025 (0.04)	1.0288	SPP	18.9
Paper birch	6.3147 (0.05)	0.9525 (0.02)	1.0209	GRP	—
Quaking aspen	6.2622 (0.05)	0.9157 (0.02)	1.0106	SPP	13.5
Black cottonwood	6.3147 (0.05)	0.9525 (0.02)	1.0209	GRP	—

Note: Volume-based prediction equation is $\widehat{\mathrm{AGB}} = \mathrm{CF} \times \exp[a + b\ \ln(\widehat{\mathrm{CVTS}})]$, where CF is the log-bias correction factor and $\widehat{\mathrm{CVTS}}$ is the whole-stem inside-bark cubic volume using equations fitted in this study (Table 2). "Basis" represents whether the coefficients were obtained by fitting species-specific (SPP) or species group specific (GRP) volume to biomass equations. Utah juniper did not have volume data. Root mean squared error (RMSE) percentages are obtained only for species that had both volume and biomass data.

Table 7. Parameter estimates and standard errors (in parentheses) of the general component ratio equations for softwood and hardwood tree species obtained by fitting Dirichlet regression.

Component	*a*	*b*	*c*
Softwood			
Stem	−1.7196 (0.1993)	−0.6652 (0.0753)	1.5372 (0.0896)
Bark	−1.4062 (0.1749)	−0.1796 (0.0607)	0.7537 (0.0731)
Foliage	1.9428 (0.1501)	−0.2280 (0.0493)	−0.1462 (0.0628)
Branch	0.8862 (0.1492)	−0.3819 (0.0530)	0.3627 (0.0653)
Hardwood			
Stem	−3.5960 (1.3917)	−1.9231 (0.4472)	2.7287 (0.5462)
Bark	−2.7325 (1.3676)	−1.5205 (0.4359)	1.8614 (0.5316)
Foliage	4.6466 (1.3446)	0.0086 (0.4336)	−0.9938 (0.5142)
Branch	3.6937 (1.4698)	−0.1730 (0.4446)	−0.3566 (0.5650)

Note: Form of the prediction equation for component i is $\hat{p}_i = \dfrac{\exp[a_i + b_i \ln(\mathrm{DBH}) + c_i \ln(\mathrm{HT})]}{\sum_{\forall i}\exp[a_i + b_i \ln(\mathrm{DBH}) + c_i \ln(\mathrm{HT})]}$, where DBH is the diameter at breast height (cm) and HT is the total tree height (m).

ments along with biomass equations. Therefore, large-scale forest biomass estimation is important to assess the role of the forestry sector in mitigating climate change impacts. In this study, we used "legacy" data (Radtke et al. 2015) compiled from different studies over the years to estimate total aboveground live tree biomass, as well as the biomasses of different tree components.

Accuracy of imputing component proportions was assessed by comparing observed and imputed geometric means because the closed geometric mean is the best linear unbiased estimator with respect to the geometry of the simplex of the center of the distribution of a random composition (Pawlowsky-Glahn and Egozcue 2002), and the easiest way to impute missing values in compositional data is to replace the missing values of a component by the geometric mean of available data for that component (Hron et al. 2010). Some other problems in biomass studies include purposeful selection of trees, lack of measurement of the basic variables such as height and crown width, and different field and lab procedures used in different studies, as they lead to different errors in measurements of AGB and CVTS from different felled tree studies.

The CRM underestimated AGB for trees used in this study dataset. This result is consistent with the results reported in Radtke et al. (2017), who found a 12.2% underestimation by the current CRM approach for eastern trees. Predictor variables for AGB in newly fitted models included DBH and HT, with logarithmic and quadratic terms used as transformations that varied by species. Even though the form of the prediction equation is same for all species, estimation of parameters was done through either a linear or nonlinear least squares approach. Of special interest were subalpine fir (*Abies lasiocarpa* (Hook.) Nutt.) and ponderosa pine (*Pinus ponderosa* Douglas ex P. Lawson & C. Lawson). In the case of subalpine fir, HT, not the logarithm of HT, was statistically significant when it appeared with the logarithm of DBH in the model. For ponderosa pine, the residual vs. fitted values showed a curved pattern when the independent variables were logarithms of DBH and HT (Fig. 5, left), suggesting a nonlinear relationship between the dependent and independent variables. The diagnostic plot was much improved when the squared logarithm of DBH, i.e., $[\ln(\mathrm{DBH})]^2$, was added to the model (Fig. 5, right). We investigated different model forms and weighting schemes before settling on this model form. Residual analysis did not show any problems with the model fits for other species.

Both positive and negative biases were produced by the new AGB equations developed in this study. The higher bias and RMSE produced for giant sequoia could potentially be attributed to the larger values of DBH, HT, and AGB for this species (Table 1). Average DBH, HT, and AGB for giant sequoia were 390.7 cm, 69.7 m, and 145 619.2 kg, respectively, whereas average DBH, HT, and AGB for other species were 44.3 cm, 21.0 m, and 7929.0 kg, respectively. Additionally, relatively higher bias (4.9%) was observed for redwood trees (*Sequoia sempervirens* (Lamb. ex D. Don) Endl.), which also have larger DBH, HT, and AGB compared with other species. However, this was not the case for Utah juniper (JUOS) and west-

Table 8. Bias and root mean squared error (RMSE) percentages in component biomass proportions obtained from fitting general Dirichlet regression by softwood and hardwood species.

	STM		BRK		BCH		FOL	
Common name	Bias %	RMSE %	Bias %	RMSE %	Bias %	RMSE %	Bias %	RMSE %
Pacific silver fir	–8.4	16.1	–29.8	47.5	31.6	45.3	–14.7	50.1
White fir	–3.2	13.8	4.3	25.5	12.3	59.6	3.8	45.8
Grand fir	–12.8	22.3	–70.1	94.6	43.3	50.6	38.0	66.3
Subalpine fir	–17.7	25.7	–36.8	46.0	19.7	30.2	28.1	38.3
Utah juniper	3.9	21.8	–8.9	21.0	–24.0	26.3	16.5	27.8
Engelmann spruce	–4.4	18.6	–0.1	52.7	–0.8	40.5	13.7	39.4
White spruce	–16.1	24.1	–30.3	43.9	33.4	45.7	0.8	40.5
Black spruce	0.6	16.7	–12.6	22.3	11.9	46.9	–17.4	53.1
Sitka spruce	6.2	10.3	–42.3	48.7	–7.1	24.2	0.5	25.5
Lodgepole pine	12.5	21.1	–40.3	70.9	–28.7	63.2	–9.9	57.0
Jeffrey pine	–20.2	24.7	15.0	44.0	17.5	37.6	31.6	37.6
Ponderosa pine	–5.2	21.1	27.7	46.9	–0.4	52.2	–14.2	53.0
Singleleaf pinyon	28.6	32.6	–28.0	45.5	–29.4	31.2	–30.8	42.4
Douglas-fir	–1.3	13.2	–8.2	30.7	6.4	48.4	6.5	78.5
Redwood	1.3	4.8	25.8	28.2	–167.8	190.3	–61.8	87.6
Giant sequoia	9.0	12.1	–72.8	79.6	2.9	35.1	–47.9	96.8
Western redcedar	–2.8	14.3	–40.6	49.6	27.5	37.1	–25.6	50.3
Western hemlock	–1.3	11.5	–23.4	43.8	29.0	52.8	–59.4	95.4
Mountain hemlock	2.9	15.3	19.5	29.5	1.2	22.1	–93.2	131.6
Red alder	5.8	9.2	–62.0	67.4	18.0	35.7	–249.1	326.0
Paper birch	–0.7	9.5	–9.8	26.1	2.3	34.0	15.7	45.2
Quaking aspen	0.5	6.9	25.8	32.4	–40.4	56.2	9.5	35.6
Black cottonwood	–10.0	15.3	2.0	19.5	14.2	30.0	22.2	33.1

Note: Components are stem wood (STM), bark (BRK), branch (BCH), and foliage (FOL). Overall average biases for STM, BRK, BCH, and FOL were –1.5%, –18.5%, –1.1%, and –12.4% for softwood and –1.1%, –11.0%, –1.5%, and –50.4% for hardwood species, respectively. Overall RMSEs for STM, BRK, BCH, and FOL were 17.9%, 45.8%, 49.4%, and 58.8% for softwood and 10.2%, 36.4%, 39.0%, and 110.0% for hardwood species, respectively.

Table 9. Bias and root mean squared error (RMSE) percentages in component biomass proportions obtained from fitting species-specific Dirichlet regression.

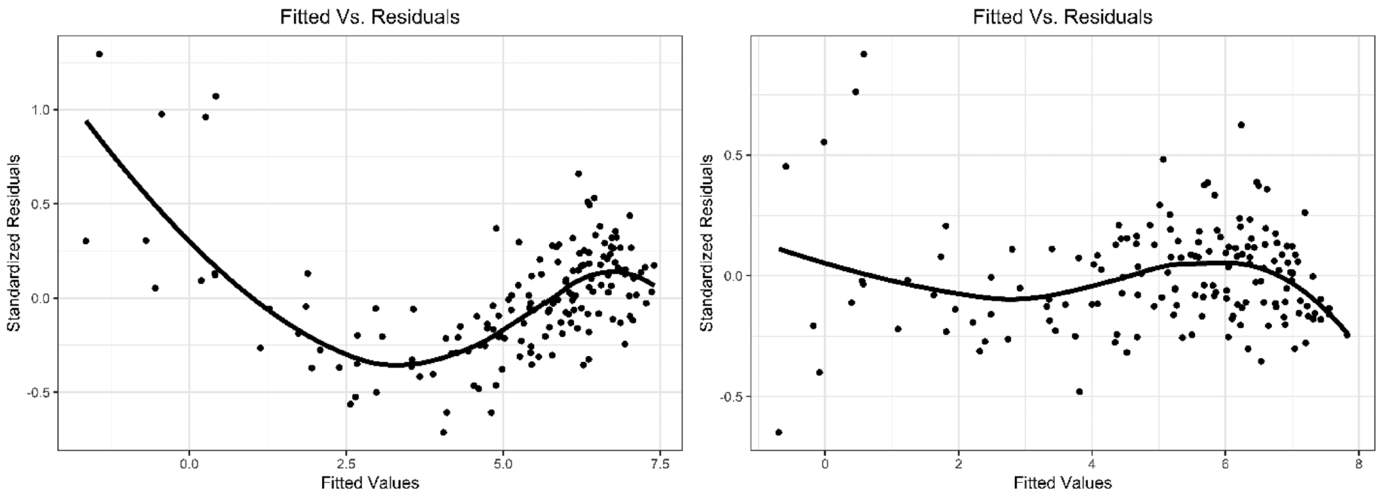
	STM		BRK		BCH		FOL	
Common name	Bias %	RMSE %	Bias %	RMSE %	Bias %	RMSE %	Bias %	RMSE %
Pacific silver fir	0.9	13.2	–4.1	26.2	0.2	25.7	–1.4	32.0
White fir	0.0	5.8	–1.3	22.4	1.5	31.7	–0.2	39.4
Grand fir	1.7	18.5	–5.7	59.4	–1.7	29.5	–3.5	54.5
Subalpine fir	0.7	16.5	–4.8	24.6	0.0	20.2	0.3	23.1
Utah juniper	0.2	10.5	–0.5	14.3	–0.6	3.4	0.4	13.5
Engelmann spruce	0.3	13.7	–2.4	34.3	–0.1	25.2	0.7	29.9
White spruce	0.7	16.7	–8.1	26.4	1.2	27.6	–0.4	38.9
Black spruce	1.1	15.9	–9.4	18.7	1.5	39.9	–1.2	53.5
Sitka spruce	0.0	5.1	–0.4	22.5	0.1	13.3	0.3	14.5
Lodgepole pine	1.4	17.3	–17.5	41.1	–0.3	44.1	3.6	60.3
Jeffrey pine	0.3	12.1	2.5	30.6	–1.1	12.8	–2.0	13.9
Ponderosa pine	0.7	17.3	–2.2	30.2	–0.8	34.0	0.8	52.3
Singleleaf pinyon	0.5	12.2	–1.3	21.3	–1.3	7.5	1.0	25.0
Douglas-fir	0.6	13.4	–12.8	30.2	7.9	45.6	–4.6	82.4
Redwood	0.0	2.3	–0.2	6.5	0.8	35.5	0.0	52.6
Giant sequoia	0.0	2.8	0.2	15.0	–0.3	14.5	–1.0	38.8
Western redcedar	0.6	13.3	–5.1	24.6	1.0	23.8	–1.7	33.1
Western hemlock	0.4	10.8	–2.4	35.7	1.6	32.8	–7.8	39.7
Mountain hemlock	0.1	6.1	–0.9	9.7	1.5	23.3	–2.8	19.8
Red alder	0.1	5.7	–2.5	21.4	2.1	25.7	–17.5	43.4
Paper birch	0.4	11.5	–4.4	22.8	2.7	39.7	–4.0	49.7
Quaking aspen	0.2	11.2	–3.0	19.5	2.3	48.2	1.1	62.2
Black cottonwood	0.2	11.3	–2.7	18.2	1.7	25.3	–2.1	24.4

Note: Components are stem wood (STM), bark (BRK), branch (BCH), and foliage (FOL). Overall average biases for STM, BRK, BCH, and FOL were 0.5%, –4.0%, 0.6%, and –1.0% for softwood and 0.2%, –3.2%, 2.2%, and –5.6% for hardwood species, respectively. Overall RMSEs for STM, BRK, BCH, and FOL were 11.8%, 26.0%, 25.8%, and 37.7% for softwood and 9.9%, 20.5%, 34.7%, and 44.9% for hardwood species, respectively.

ern hemlock (TSHE), which had greater bias and RMSE than redwood (SESE).

We calibrated the CRM AGB estimates by fitting a calibration equation based on simple linear regression (eq. 5). Poudel and Temesgen (2016*b*) found substantial reduction in bias and RMSE by adjusting volume and biomass obtained from the regional equations with the calibration coefficients obtained from simple linear regression. It was expected that the new fit would provide the best results followed by the calibrated and uncalibrated CRM, respectively. For some species, the calibration did not reduce bias

Fig. 5. Scatterplot of fitted values vs. studentized residuals obtained by fitting logarithmic AGB equation without (left) and with (right) squared logarithm of DBH, i.e., [ln(DBH)]², in the model for ponderosa pine trees.



and RMSE, e.g., it increased RMSE by 11.3% for Douglas-fir trees. This could be because, in the CRM system, regional variation in Douglas-fir AGB might have been accounted for by using regional merchantable volume equations while the calibration was not optimized to account for regional variation in Douglas-fir biomass.

Many national forest inventories are primarily designed for the estimation of growing stock volume, with estimates of biomass and carbon obtained as secondary products (Radtke et al. 2017). Sample data for development of volume and taper equations are much more abundant than biomass data. Additionally, estimation of the stem wood component in CRM is based on merchantable stem wood volume. Volume-based AGB prediction was not as accurate as the prediction of AGB using direct AGB equations for most of the species, which was expected because of the multistep nature of this approach and the error associated in the volume prediction. Volume-based AGB estimation can be viewed as the conversion of CVTS to AGB and was expected to perform worse than the calibration of CRM-estimated AGB; however, it performed better than the calibration of CRM AGB for 7 out of 16 species (Tables 5 and 6).

Accuracy of the generalized component ratios was generally quite low and highly variable for individual components of biomass (Table 8). Large values of bias and RMSE were observed in two cases: (*i*) when the number of observations is less than 20 (e.g., grand fir) and (*ii*) when the component biomass is proportionally very small (e.g., giant sequoia, redwood, and red alder (*Alnus rubra* Bong.)). Compared with the generalized component ratio equations for softwood and hardwood trees, bias and RMSE were reduced substantially by using species-specific component ratio equations, as was expected (Table 9). One possible reason for poor performance of the generalized component ratio models could be inherent on how the species are grouped. We graphically investigated potential grouping of species by plotting standardized coefficients of the species-specific biomass equations; however, we did not observe any discernible and consistent pattern in them (Fig. 6).

Three possibilities were tested for obtaining AGB: (*i*) as a function of predicted CVTS, i.e., predict CVTS using eq. 2 and coefficients in Table 2 and convert it to AGB using eq. 7 and coefficients in Table 6; (*ii*) as a function of CRM AGB, i.e., predict AGB using eq. 6 and coefficients in Table 5; and (*iii*) using direct AGB equations fitted in this study, i.e., using equations in Table 4. We found that the use of direct biomass equations is most appropriate when the objective is to obtain accurate estimate of AGB. This makes sense from a statistical perspective as well because this approach minimizes the error in our variable of interest — AGB. Component biomass can be obtained by multiplying the AGB obtained from one of the three approaches by the generalized component ratio equations (Table 7) or the species-specific component ratio equations (Supplementary Table 1[1]). When possible, the use of species-specific equations is suggested as they are notably more accurate than the generalized component ratio equations.

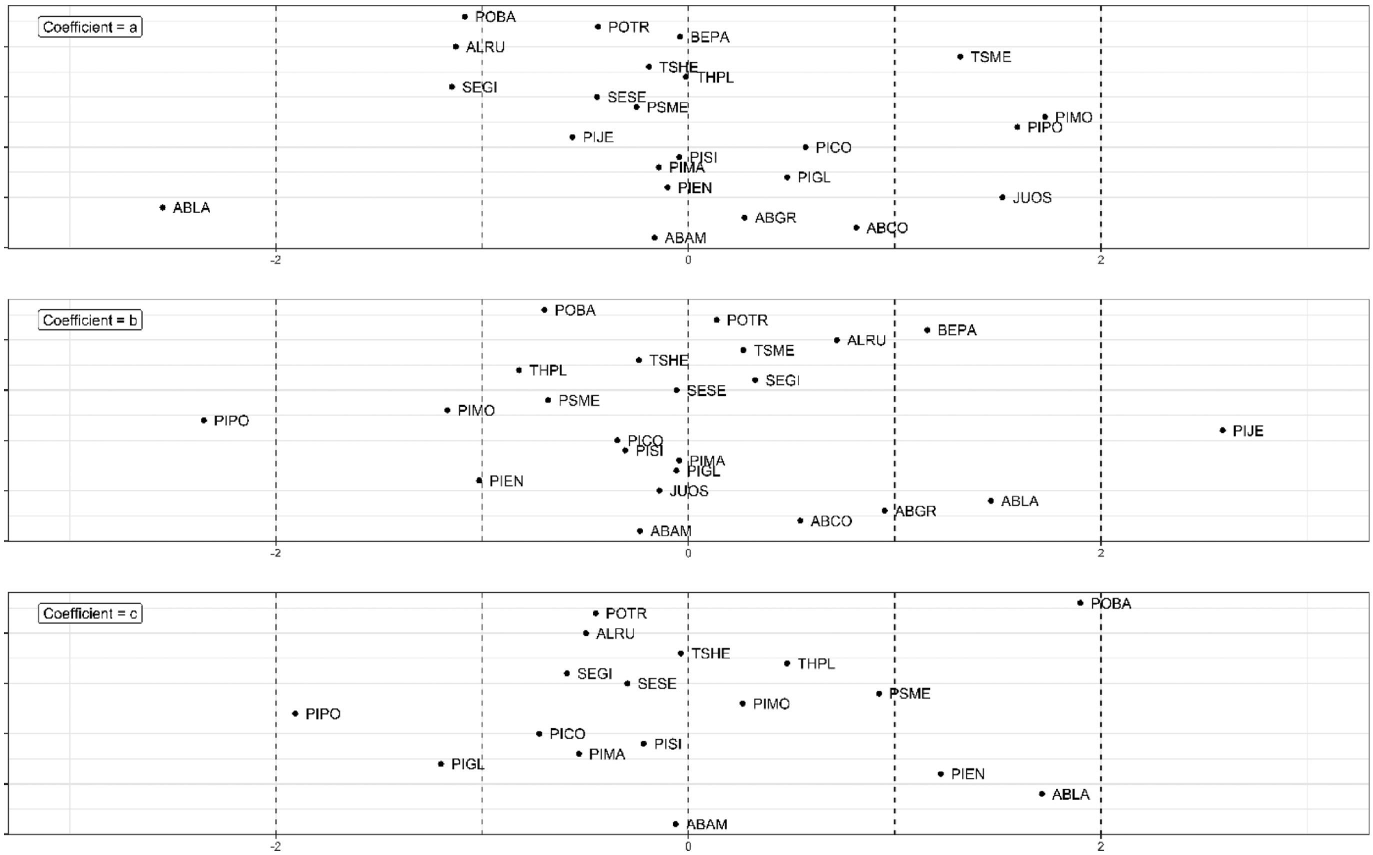
Conclusions

We used species-specific and generalized (when enough observations were not available in each component) Dirichlet imputation to fill in the missing biomass components. The imputed geometric means were up to 9% higher than the observed geometric means of stem wood proportion, but for other components, the differences were within 1%. The CRM underestimated the AGB with very wide range across species. Results obtained from the calibration of CRM using simple linear regression were intermediate between the CRM approach and new equations. Compared with the CRM approach, new equations reduced RMSE for all species. RMSE in AGB was reduced by as much as 145.0% by using new equations fitted in this study. The smallest reduction in RMSE was observed in Douglas-fir (0.7% reduction). As expected, the volume-based AGB prediction was not as accurate as the prediction of AGB using direct AGB equations for most of the species; however, it performed better than the calibration of CRM AGB for 7 out of 16 species. Our results showed that generalized component ratio equations may be suitable for the stem wood component but were highly biased for other components. In addition, large bias and RMSE were observed for species with a small number of observations and for components with very small proportions of the total. As expected, the species-specific component ratio equations were more accurate than the generalized component ratio equations.

Results suggest that calibration of existing biomass equations is not sufficient compared with the new fit. In addition, biomass components are better estimated on a species-specific basis rather than by arbitrary grouping. Graphical comparison of standard-

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfr-2018-0361.

Fig. 6. Standardized parameters of species-specific biomass equations. Standardization was done by subtracting mean and dividing by standard deviation (SD) of the parameters, e.g., $a_{\text{standardized}} = \dfrac{\alpha - \overline{\alpha}}{\text{SD}(\alpha)}$.



ized coefficients of species-specific aboveground biomass equations developed in this study did not provide a discernible pattern in them. However, a more biological rather than a purely statistical approach considering ecological strata and species functional traits or a spatially driven approach could be of interest to a larger audience. Therefore, we recommend that future studies in large-scale biomass estimation should consider species grouping by taxonomic association, functional traits, and spatial distribution to account for limited species-specific data. There may be opportunities in improving modeling approaches such as path evaluation, i.e., predicting total and disaggregating into components versus predicting components and aggregating to obtain total biomass, consideration of optimal weighting schemes, and cross-validation methods. A comprehensive analysis with the data from the entire US, including geographic and ecological parameters and species as random effects in the biomass equations, is underway.

References

Baskerville, G.L. 1972. Use of logarithmic regression in the estimation of plant biomass. Can. J. For. Res. **2**: 49–53. doi:10.1139/x72-009.

Brown, S. 2002. Measuring carbon in forests: current status and future challenges. Environ. Pollut. **116**(3): 363–372. doi:10.1016/S0269-7491(01)00212-3. PMID:11822714.

Chojnacky, D.C., Heath, L.S., and Jenkins, J.C. 2014. Updated generalized biomass equations for North American tree species. Forestry, **87**: 129–151. doi:10.1093/forestry/cpt053.

Dahlhausen, J., Uhl, E., Heym, M., Biber, P., Ventura, M., Panzacchi, P., Tonon, G., Horváth, T., and Pretzsch, H. 2017. Stand density sensitive biomass functions for young oak trees at four different European sites. Trees, **31**: 1–16.

de-Miguel, S., Mehtätlo, L., and Durkaya, A. 2014. Developing generalized, calibratable, mixed-effects meta-models for large-scale biomass prediction. Can. J. For. Res. **44**(6): 648–656. doi:10.1139/cjfr-2013-0385.

Domke, G.M., Woodall, C.W., Smith, J.E., Westfall, J.A., and McRoberts, R.E. 2012. Consequences of alternative tree-level biomass estimation procedures on U.S. forest carbon stock estimates. For. Ecol. Manage. **270**: 108–116. doi:10.1016/j.foreco.2012.01.022.

Ducey, M.J. 2012. Evergreenness and wood density predict height–diameter scaling in trees of the northeastern United States. For. Ecol. Manage. **279**: 21–26. doi:10.1016/j.foreco.2012.04.034.

Food and Agriculture Organization of the United Nations (FAO). 2016. Global forest resource assessment 2015. How are the world's forests changing? 2nd ed. Food and Agriculture Organization of the United Nations, Rome.

Forrester, D.I., Tachauer, I.H.H., Annighoefer, P., Barbeito, I., Pretzsch, H., Ruiz-Peinado, R., Stark, H., Vacchiano, G., Zlatanov, T., Chakraborty, T., Saha, S., and Sileshi, G.W. 2017. Generalized biomass and leaf area allometric equations for European tree species incorporating stand structure, tree age and climate. For. Ecol. Manage. **396**: 160–175. doi:10.1016/j.foreco.2017.04.011.

Heath, L.S., Hansen, M., Smith, J.E., Smith, W.B., and Miles, P.D. 2009. Investigation into calculating tree biomass and carbon in the FIADB using a biomass expansion factor approach. *In* Proceedings of the Forest Inventory and Analysis (FIA) Symposium 2008, Park City, Utah, 21–23 October 2008. *Edited by* W.H. McWilliams, G.G. Moisen, and R.L. Czaplewski. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Colo., For. Serv. Proc. RMRS-P-56CD.

Hron, K., Templ, M., and Filzmoser, P. 2010. Imputation of missing values for compositional data using classical and robust methods. Comput. Stat. Data Anal. **54**: 3095–3107. doi:10.1016/j.csda.2009.11.023.

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

Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., Dalponte, M., van Ewijk, K.Y., Forrester, D.I., Haeni, M., Higgins, S.I., Holdaway, R.J., Iida, Y., Lorimer, C., Marshall, P.L., Momo, S., Moncrieff, G.R., Ploton, P., Poorter, L., Rahman, K.A., Schlund, M., Sonké, B., Sterck, F.J., Trugman, A.T., Usoltsev, V.A., Vanderwel, M.C., Waldner, P., Wedeux, B.M.M., Wirth, C., Wöll, H., Woods, M., Xiang, W., Zimmermann, N.E., and Coomes, D.A. 2017. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. Glob. Change Biol. **23**(1): 177–190. doi:10.1111/gcb.13388.

Kuhn, M., and Johnson, K. 2013. Applied predictive modeling. Springer-Verlag, New York. doi:10.1007/978-1-4614-6849-3.

Miles, P.D., and Smith, W.B. 2009. Specific gravity and other properties of wood and bark for 156 tree species found in North America. USDA Forest Service, Northern Research Station, Newtown Square, Penn., For. Serv. Res. Note NRS-38. doi:10.2737/NRS-RN-38.

Parresol, B.R. 1999. Assessing tree and stand biomass: a review with examples and critical comparisons. For. Sci. **45**(4): 573–593.

Pawlowsky-Glahn, V., and Egozcue, J.J. 2002. BLU estimators and compositional data. Math. Geol. **34**(3): 259–274. doi:10.1023/A:1014890722372.

Poudel, K.P., and Temesgen, H. 2016*a*. Methods for estimating aboveground biomass and its components for Douglas-fir and lodgepole pine trees. Can. J. For. Res. **46**(1): 77–87. doi:10.1139/cjfr-2015-0256.

Poudel, K.P., and Temesgen, H. 2016*b*. Calibration of volume and component biomass equations for Douglas-fir and lodgepole pine in western Oregon forests. For. Chron. **92**(2): 172–182. doi:10.5558/tfc2016-036.

Poudel, K.P., Flewelling, J.W., and Temesgen, H. 2018. Predicting volume and biomass change from multi-temporal lidar sampling and remeasured field inventory data in Panther Creek Watershed, Oregon, USA. Forests, **9**: 28. doi:10.3390/f9010028.

R Core Team. 2018. R: a language and environment for statistical computing. Version 3.4.4. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org.

Radtke, P., Walker, D., Weiskittel, A., Frank, J., Coulston, J., and Westfall, J. 2015. Legacy tree data: a national database of detailed tree measurements for volume, weight, and physical properties. *In* Proceedings of Forest Inventory and Analysis (FIA) Symposium 2015, Portland, Ore., 8–10 December 2015. *Edited by* S.M. Stanton and G.A. Christensen. USDA Forest Service, Pacific Northwest Research Station, Portland, Ore., For. Serv. Gen. Tech. Rep. PNW-GTR-931. pp. 25–30.

Radtke, P., Walker, D., Frank, J., Weiskittel, A., DeYoung, C., MacFarlane, D., Domke, G., Woodall, C., Coulston, J., and Westfall, J. 2017. Improved accuracy of aboveground biomass and carbon estimates for live trees in forests of the eastern United States. Forestry, **90**: 32–46. doi:10.1093/forestry/cpw047.

Raile, G.K. 1982. Estimating stump volume. USDA Forest Service, North Central Forest Experiment Station, St. Paul, Minn., For. Serv. Res. Pap. NC-224.

Temesgen, H., Affleck, D., Poudel, K., Gray, A., and 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. doi:10.1080/02827581.2015.1012114.

Templ, M., Hron, K., and Filzmoser, P. 2011. robCompositions: an R-package for robust statistical analysis of compositional data. *In* Compositional data analysis: theory and applications. *Edited by* V. Pawlowsky-Glahn and A. Buccianti. John Wiley & Sons, Chichester, U.K. pp. 341–355. doi:10.1002/9781119976462.ch25.

Vonderach, C., Kändler, G., and Dormann, C.F. 2018. Consistent set of additive biomass functions for eight tree species in Germany fit by nonlinear seemingly unrelated regression. Ann. For. Sci. **75**: 49. doi:10.1007/s13595-018-0728-4.

Wirth, C., Schumacher, J., and Schulze, E.D. 2004. Generic biomass functions for Norway spruce in Central Europe — a meta-analysis approach toward prediction and uncertainty estimation. Tree Physiol. **24**: 121–139. doi:10.1093/treephys/24.2.121. PMID:14676030.

Woodall, C., Heath, L.S., Domke, G.M., and Nichols, M.C. 2011. Methods and equations for estimating aboveground volume, biomass, and carbon for trees in the U.S. forest inventory, 2010. USDA Forest Service, Northern Research Station, Newtown Square, Penn., For. Serv. Gen. Tech. Rep. NRS-88. doi:10.2737/NRS-GTR-88.

Zhou, X., and Hemstrom, M.A. 2009. Estimating aboveground tree biomass on forest land in the Pacific Northwest: a comparison of approaches. USDA Forest Service, Pacific Northwest Research Station, Portland, Ore., For. Serv. Res. Pap. PNW-RP-584. doi:10.2737/PNW-RP-584.

Zhao, D., Kane, M., Teskey, R., and Markewitz, D. 2016. Modeling aboveground biomass components and volume-to-weight conversion ratios for loblolly pine trees. For. Sci. **62**: 463–473.