Methods for estimating aboveground biomass and its components for Douglas-fir and lodgepole pine trees

K.P. Poudel and H. Temesgen

Abstract: Estimating aboveground biomass and its components requires sound statistical formulation and evaluation. Using data collected from 55 destructively sampled trees in different parts of Oregon, we evaluated the performance of three groups of methods to estimate total aboveground biomass and (or) its components based on the bias and root mean squared error (RMSE) that they produced. The first group of methods used an analytical approach to estimate total and component biomass using existing equations and produced biased estimates for our dataset. The second group of methods used a system of equations fitted with seemingly unrelated regression (SUR) and were superior to the first group of methods in terms of bias and RMSE. The third group of methods predicted the proportion of biomass in each component using beta regression, Dirichlet regression, and multinomial log-linear regression. The predicted proportions were then applied to the total aboveground biomass to obtain the amount of biomass in each component. The multinomial log-linear regression approach consistently produced smaller RMSEs compared with both SUR methods. The beta and Dirichlet regressions were superior to both SUR methods except for Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) branch biomass, for which the simple SUR method produced smaller RMSE compared with the beta and Dirichlet regressions.

Key words: component ratio method, seemingly unrelated regression, multinomial log-linear regression, beta regression, Dirichlet regression.

Résumé : L'estimation de la biomasse aérienne et de ses composantes requiert une formulation et une évaluation statistiques fiables. À l'aide de données récoltées à partir d'un échantillonnage destructif de 55 arbres distribués dans différentes parties de l'Oregon, nous avons évalué la performance de trois groupes de méthodes pour estimer la biomasse aérienne totale ou ses composantes sur la base du biais et de l'erreur quadratique moyenne (EQM) qu'elles ont produits. Le premier groupe de méthodes utilisait une approche analytique pour estimer la biomasse totale et ses composantes à l'aide d'équations existantes, ce qui a produit des estimations biaisées pour notre fichier de données. Le deuxième groupe utilisait un système d'équations ajustées par régression apparemment indépendante (RAI) et a produit des résultats supérieurs à ceux du premier groupe en termes de biais et d'EQM. Le troisième groupe de méthodes estimait la proportion de la biomasse de chaque composante en utilisant des régressions bêta, Dirichlet et multinomiale log-linéaire (RML). Les proportions prédites ont ensuite été appliquées à la biomasse aérienne totale pour obtenir la biomasse de chaque composante. L'approche RML a systématiquement produit de plus petites valeurs d'EQM que celles des deux approches RAI. Les régressions bêta et Dirichlet étaient supérieures aux deux méthodes RAI à l'exception de la biomasse des branches de douglas de Menzies (*Pseudotsuga menziesii* (Mirb.) Franco) pour laquelle la RAI simple a produit une plus petite valeur d'EQM que celles des régressions bêta et Dirichlet. [Traduit par la Rédaction]

Mots-clés : méthode du rapport des composantes, régression apparemment indépendante, régression multinomiale log-linéaire, régression bêta, régression Dirichlet.

Introduction

Forests play a dual role in the global carbon cycle as an important carbon sink by removing carbon dioxide through photosynthesis and converting that photosynthate to forest biomass and as a carbon source by releasing carbon dioxide through respiration, wildfires, and decomposition. If forests are properly managed and timber is used for long-term products such as buildings, then forest management could result in a net reduction of atmospheric carbon, whereas burning of wood for residential and commercial uses increases carbon emission rates. Thus there is great concern about the uncertainty over whether forests will be a sink or a source of carbon in the future. Forest growth rates, harvest activities, natural disturbances such as wildfire, and loss of forest cover due to land-use changes are key factors that alter the carbon stock and absorption ability of a forest. It is essential to have knowledge of carbon stocks and fluxes to understand the current state and future course of the carbon cycle in response to changing land uses and climatic conditions (Hollinger 2008). The amount of biomass as living vegetation or dead wood and debris is an important factor that relates forestry function and atmospheric carbon regulation (Brown 2002).

Total carbon stocks in forest ecosystems can be divided into two main parts: aboveground pools and fluxes and belowground pools and fluxes (Hoover 2008). Aboveground biomass constitutes the major portion of carbon pools in forest ecosystems (Vashum and Jayakumar 2012). Xiao et al. (2003) found a 0.14 ratio of belowground biomass to aboveground biomass in a 73-year-old Scots pine forest. Czapowskyj et al. (1985) found 80% biomass in the aboveground components and 20% biomass in the belowground components of black spruce (*Picea mariana* B.S.P. (Mill.)) in Maine. Aboveground biomass estimation has received significant atten-

K.P. Poudel and H. Temesgen. Oregon State University, Department of Forest Engineering, Resources and Management, 204 Peavy Hall, Corvallis, OR 97331, USA.

Corresponding author: H. Temesgen (e-mail: hailemariam.temesgen@oregonstate.edu).

tion in recent years because of the fact that the change in aboveg-round biomass is associated with the components of climate change (Lu et al. 2002). It depicts the forest's potential to sequester and store carbon dioxide. Therefore, estimation of aboveground biomass is central in quantifying and monitoring of the amount of carbon stored in forest ecosystems.

The amount of biomass contained in a tree is influenced by various site factors such as stand density, site productivity, soil characteristics such as texture and moisture content, and tree characteristics such as species and age. Generally, aboveground biomass is estimated through the use of allometric equations that relate easily measurable attributes such as diameter at breast height (DBH, 1.3 m), height, etc., to total tree biomass or compo-nent biomass. It is critical to note, however, that different regions and species link these attributes to biomass through different functions such as logarithmic, linear, or quadratic forms (Zhou and Hemstrom 2009). Thus, site- and species-specific models might be preferred to regional models in their current form when predicting biomass at the local scale. de-Miguel et al. (2014a) found that the generalized mixed-effects metamodels calibrated with a location-specific small sample using a best unbiased linear predic-tor performed better than fitting location-specific equations based on much larger samples.

Aboveground biomass is commonly divided into three major components: bole (main stem), stem bark, and crown (branch wood and foliage). The component biomass models are useful to account for the variability within the tree. Moreover, different tree components are used for different purposes that require sep-arate estimates of component biomasses. Primarily, the bole is used for timber production, crown biomass can provide informa-tion on fuel load and wildfire assessment, and small branches and leaves are useful in bioenergy production. Thus the information on biomass in each component is required to determine the asso-ciated cost of transportation and processing biomass. However, on the other hand, if the objective is to process whole trees through cull removal or removal of small diameter trees for whole-tree chipping at the bioenergy plant, then the component biomass estimates may not be as essential.

The amount of biomass allocated to different tree components varies by species and by their geographic locations (Luo et al. 2013). In an effort to quantify aboveground biomass and net pri-mary production of pine, oak, and mixed pine–oak forests, Nunes et al. (2013) found that in pine stands, the greater distribution of biomass was from the stem component, whereas in oak and mixed stands, the foliage was the component with the greater distribution of biomass. Repola (2006) found that wood density in pine decreased from the butt to the top and that the gradient in wood density was steep at the butt but decreased in the upper part of the stem. Even though the vertical dependence was similar in birch (a hardwood species), the density gradient was much smaller. Because mass is a function of volume and density, it also affects the amount of biomass contained in trees. In addition, the allocation of biomass to different components is influenced by differences in management practices (Tumwebaze et al. 2013; de-Miguel et al. 2014b), age, and stand structure (de-Miguel et al. 2014b). Therefore, locally derived, species-specific biomass equa-tions may yield biomass estimates that are considerably different from the estimates obtained by using more generalized regional biomass equations without local calibration. Weighing trees in the field would be the most accurate method for estimating tree biomass. However, it is time consuming and very expensive, mak-ing the use of allometric equations inevitable. The type and amount of data and the number of equations required to accu-rately quantify biomass is unknown. Cautionary measures should be taken in developing and evaluating the methods for estimating aboveground biomass and its components.

When component models are fitted, the strength of relation-ship exhibited by bark, branch, and foliage models is nowhere

near that for the stem wood (Boudewyn et al. 2007). However, it is desired that the prediction from component regression equations add to the prediction from the total tree regression model (Parresol 2001). The relationships between component mass and easily measurable tree attributes differ considerably in conifers and hardwoods. Component biomass modeling in hardwood re-quires a more innovative approach due to decurrent form (Westfall et al. 2012).

With a substantial increase in the demand of forest biomass information in recent years, considerable efforts have been made to estimate aboveground biomass and its components. Stem wood biomass accounts for the major portion of aboveground biomass. In comparing different approaches of aboveground biomass esti-mation, Zhou and Hemstrom (2009) found that the proportion of softwood merchantable biomass ranged from 72% to 83% of the total aboveground biomass, with little variation among species. Their methods of obtaining total and component aboveground biomass include the U.S. Department of Agriculture (USDA) Forest Service's Forest Inventory and Analysis component ratio method (FIA-CRM), the regional approach for the Pacific Northwest (FIA-PNW), and the equations developed by Jenkins et al. (2003) (here-after, the Jenkins method). Using LiDAR data from a *Picea crassifolia* Kom. (a coniferous tree) stand, He et al. (2013) found that, on average, 72% of the total aboveground biomass of a tree is con-tained within the stem. The branches, foliage, and fruits in their study accounted 11%, 13%, and 4% of total aboveground biomass, respectively. However, Kuznetsova et al. (2011) found only 31% and 27% of aboveground biomass in the main stems of 8-year-old Scots pine (*Pinus sylvestris* L.) and lodgepole pine *Pinus contorta* var. *latifo-lia* Engelm. respectively, in the oil shale postmining landscapes in Estonia. They found 32% and 28% in shoots and 37% and 45% in the needles of those 8-year-old Scots pine and lodgepole pine, respec-tively. The stem biomass for 20-, 40-, and 60-year-old eastern Med-iterranean *Pinus brutia* Ten. estimated using intrinsically linear models was found by de-Miguel et al. (2014b) to be 79.8%, 80.5%, and 80.6% of total aboveground biomass, whereas the crown bio-mass was 20.2%, 19.5%, and 19.4%, respectively. They found the proportion of stem and crown biomass for uneven-aged stands to be 61.8% and 38.1%, respectively.

These differences in proportion of biomass in different compo-nents among species warrant for species-specific component mod-els. More than 90% of aboveground live tree biomass in Oregon is contributed by softwood species (Zhou and Hemstrom 2009). Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) by itself contrib-utes about 51% of total aboveground biomass of trees with a DBH > 12.5 cm (5 in.), and western hemlock (*Tsuga heterophylla* (Raf.) Sarg.) and ponderosa pine (*Pinus ponderosa* Douglas ex P. Lawson & C. Lawson) contribute about 8% each (Zhou and Hemstrom 2009). Additionally, Douglas-fir, western hemlock, grand fir (*Abies grandis* (Douglas ex D. Don) Lindl.), red alder (*Alnus rubra* Bong.), and lodgepole pine account for approximately 50% of the gross live volume in the Pacific Northwest. In this study, we developed and evaluated different approaches for estimating aboveground biomass and its components for Douglas-fir and lodgepole pine trees using the data collected destructively from sampled trees. The ideal method for estimating biomass for a population would be to randomly select sample plots, harvest all trees in that plot to estimate plot biomass, and then extrapolate to the population of interest. However, the design-based biomass estimation of a population is prohibitively expensive. Therefore, the sample tree measurement and forest area estimates are commonly obtained from independent forest inventory. The individual-tree biomass equations are then applied to the inven-tory data to obtain the biomass of the target population.

Approaches in biomass estimation depends on the scale of anal-ysis, need for detail, user group interests, and purpose of estima-tion (Zhou and Hemstrom 2009). Although there is a need for consistent methods of biomass estimation, there is no strong

Fig. 1. Locations of study sites. Data collection was carried out in the summers of 2012 and 2013.
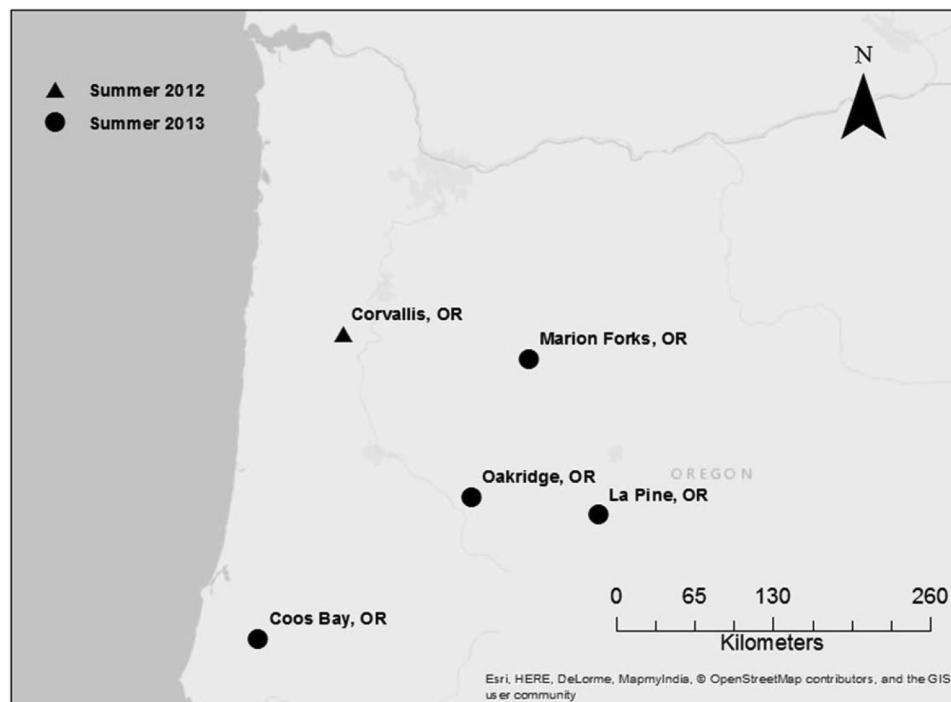
rationale to justify one method of estimation as being superior to another. In this study, we compared different methods that belonged to three major groups. The first group of methods (group I methods) uses an analytical approach to estimate total and component biomass using existing equations and are the FIA-CRM, FIA-PNW, and Jenkins methods. The second group of methods (group II methods) is a regression-based approach that used a system of equations fitted with seemingly unrelated regression (SUR). The dependent variables in the system of equations were component biomass and total biomass, and the independent variables were DBH, DBH and total height, or DBH and crown length. These two groups of methods give the amount of biomass contained in each component and also the total aboveground biomass. The third group of methods (group III methods) predicted the proportions of biomass in each component using beta, multinomial log-linear, and Dirichlet regressions. Predicted proportions are then applied to the observed total aboveground biomass to obtain the amount of biomass in each component.

Materials and methods

Data and study area

A detailed biomass data collection was carried out by destructively sampling 22 Douglas-fir and 33 lodgepole pine trees in different forests within the state of Oregon (Fig. 1). Efforts were made to select trees to give an approximately equal representation across a range of size classes while avoiding the trees with severe defects and that were close to the stand edges. Trees that were forked below breast height and trees with damaged tops were also not included in sampling. The field work was carried out between the first week of July and third week of September 2012 and 2013. The following tree-level attributes were recorded: DBH, total height, crown base height (height to the base of the first live branch), crown width, and main stem diameter at 0.15 m, 0.76 m, 1.37 m, and 2.4 m above ground and every 1.22 m afterwards. Diameter measurements were made on a total of 2975 and 4458 branches from Douglas-fir and lodgepole pine, respectively. The mean DBHs were 54.9 cm and 24.6 cm and mean heights were

33 m and 17 m for Douglas-fir and lodgepole pine trees, respectively (Table 1).

The crown of the sample tree was divided into three equal-length strata. For all first-order branches, i.e., for the branches that are directly attached to the main bole, height to branch base and diameter at the branch base were measured, and for the first branch and every third branch in each stratum, branch length and green mass of both live and dead branches were recorded. Four, three, and two branches from lower, middle, and upper stratum, respectively, were randomly selected for weighing with and without foliage. The needles were removed in the field to obtain a separate green mass of branch wood and foliage. These branches were then brought to the lab, keeping branch wood and foliage in separate paper bags, for drying. Disks that were 3–5 cm thick were removed from the top of the stump and every 5.18 m. Green mass of disks with and without bark, as well as four measurements 90° apart along the disk edge for thickness, was recorded in the field. A bark sample that was 5–10 cm long was removed from each disk. The width, thickness, and length of the bark samples were also recorded in the field.

The sample branches were chipped into small pieces to expedite the drying process and placed in a kiln for drying at 105 °C. Oven-dried mass was recorded by tracking the mass lost by each sample (disk, branch wood, needle, and bark) until no further mass was lost.

Total tree biomass computation involved additional steps. Once on the ground, the bole of the sampled tree was sectioned into 5.18 m long sections. Disks that were 3–5 cm thick were removed from the top of the stump and every 5.18 m; diameter inside bark at both ends of the 5.18 m sections were also obtained. The inside-bark cubic volume of these sections was calculated using the Smalian's formula, which was then converted to inside-bark section biomass by multiplying it by the mean density of the disks taken from two ends. Inside-bark volume of the stump was calculated as a cylinder, whereas the volume of the top section was calculated as a cone. Both the stump volume and volume of the top section were multiplied by the density of the lowest and top

Table 1. Summary statistics for tree data sampled.

Variable	Minimum	Mean	Maximum	CV
Douglas-fir				
Foliage (kg)	10.0	56.91	169.00	0.82
Branch (kg)	19.0	221.7	791.0	1.03
Bark (kg)	6.0	105.6	372.0	1.03
Bole (kg)	123.0	1916.9	6515.0	1.05
Total (kg)	187.0	2301.2	7578.0	1.04
DBH (cm)	19.3	54.9	114.0	0.54
Total height (m)	16.6	33.0	48.8	0.31
Lodgepole pine				
Foliage (kg)	5.0	17.61	54.00	0.70
Branch (kg)	10.0	47.6	178.0	0.90
Bark (kg)	1.0	11.9	49.0	0.97
Bole (kg)	24.0	220.2	798.0	0.97
Total (kg)	49.0	297.5	899.0	0.87
DBH (cm)	13.5	24.6	42.9	0.35
Total height (m)	9.2	17.0	31.9	0.35

Note: CV, coefficient of variation; DBH, diameter at breast height (1.3 m).

disks to obtain the stump and top biomass, respectively. Total inside-bark bole biomass was obtained by summing the stump, sections, and top masses. Outside-bark bole volume was obtained in the same way as to obtain total inside-bark stem volume, using outside-bark diameters in place of inside-bark diameters. Total stem bark volume was then obtained by subtracting inside-bark stem volume from the outside-bark stem volume. The density of bark samples was obtained as the ratio of dry mass to its volume calculated assuming a rectangular shape. Then, the method of bark biomass calculation was same as the method for inside-bark bole biomass calculation except that the bark sample densities were used instead of the disk density.

Individual branch wood and foliage biomass was obtained by fitting a species-specific log-linear model of the following form:

$$(1) \qquad \ln(y_{ij}) = \beta_{0i} + \beta_{1i} \ln(BD_{ij}) + \varepsilon_{ij}$$

where y_{ij} and BD_{ij} are oven-dried mass (kg) of branch wood (or foliage) and branch diameter (cm) at base of the jth branch on ith tree, respectively; β_{ij}s are regression parameters to be estimated from the data; $\ln(\cdot)$ is the natural logarithm; and ε_{ij}s are the random errors. The scatterplot of the logarithm of branch wood and foliage biomasses against the logarithm of branch basal diameter showed that a linear model was appropriate for modeling these relationships. Similar model form has been used by Temesgen et al. (2011) to model foliar biomass for Douglas-fir and ponderosa pine and by Poudel et al. (2015) to estimate crown biomass in Douglas-fir. When logarithmic regressions are used, there is inherent negative bias due to the fact that the dependent variable is transformed prior to estimation (Snowdon 1991). The commonly used remedy to this is to multiply the back-transformed results by a correction factor [exp(MSE/2)], proposed by Baskerville (1972), where MSE is the mean squared error obtained by the least-squares regression. However, there are conflicting remarks about the correction factor itself (e.g., Beaucham and Olson 1973; Flewelling and Pienaar 1981), and the effects of adjustment might be negligible (e.g., Harrison et al. 2009). Therefore, we did not use the correction factor in this study. The correction factors for Douglas-fir branch wood and foliage biomass models were 1.06 and 1.14, respectively, and for lodgepole pine, the correction factors were 1.12 and 1.16 for branch wood and foliage biomass models, respectively. Total branch wood and foliage biomass in each tree was obtained by summing these fitted values. The distribution of aboveground biomass in different components differed among species, with the majority of the aboveground biomass being present in the main bole (Fig. 2).

Methods for estimating aboveground biomass

Group I methods

There are many different methods available for the calculation of aboveground biomass and its components. The Forest Inventory and Analysis (FIA) of the USDA Forest Service uses the component ratio method (CRM) described in Heath et al. (2008) to produce national-level biomass and carbon estimates. The FIA-CRM method is a multistep process that involves converting sound volume of wood in the merchantable bole to bole biomass using a compiled set of wood-specific gravities. The biomass in bole bark is calculated using a compiled set of percent bark and bark-specific gravities. The set of wood- and bark-specific gravities and percent bark are compiled in Miles and Smith (2009). The biomass of tops and limbs is calculated as a proportion of the bole biomass, based on component proportions from Jenkins et al. (2003). Biomass in stump wood and bark is based on volume equations in Raile (1982), as well as the compiled set of wood- and bark-specific gravities. Total aboveground biomass is obtained by summing these component masses.

The FIA program of the USDA Forest Service has four regional units: Northern, Southern, Interior West, and Pacific Northwest (FIA-PNW). The FIA-PNW collects and maintains the data on plots in coastal Alaska, California, Hawaii, Oregon, Washington, and United Sates affiliated Pacific Islands. To calculate aboveground biomass, the FIA-PNW uses its specific set of equations. Tree stem biomass is calculated from the cubic volume and wood density factor. The specific equations used by FIA-PNW for volume and aboveground biomass in the Pacific Northwest can be found in Zhou and Hemstrom (2010).

The Jenkins et al. (2003) equations were derived by fitting regressions on pseudo data generated from previously published equations. The Jenkins et al. (2003) equations for total aboveground biomass are single-entry equations that only use DBH as a predictor variable. Biomass of the components is predicted as the proportions of total aboveground biomass using an exponential function of DBH, predicted differently for hardwood and softwood species groups.

Group II methods

Aboveground biomass equations are used to convert forest inventory data to biomass estimates. These equations combine biomass data obtained from destructive sampling with the dendrometric information through regression. Regression models for total and component aboveground biomass have been fitted using different fitting approaches with linear and nonlinear functions. The sets of component biomass equations can be fitted independently or as a system of equations. When a system of equations is fitted simultaneously, the residuals are correlated because the component biomasses come from the same tree. Therefore, the SUR method that allows the inclusion of dependencies among the error terms of the component biomass equations is commonly used to estimate component and total aboveground biomass (e.g., Parresol 1999; Lambert et al. 2005; Ritchie et al. 2013). The SUR models can be constrained such that the prediction of component equations sum to the prediction of total tree regression.

The single-entry DBH-based allometric models for component biomass are much more complex than for total aboveground biomass (Jenkins et al. 2003). Zhang et al. (2004) found that diameter at the base of the live crown was a better predictor of crown foliage biomass than DBH. For mixed-species Atlantic forest stands, a transformed nonlinear biomass equation that used squared DBH, total tree height, and wood density as predictor variables was the most accurate model for aboveground biomass (Nogueira Junior et al. 2014). Therefore, we extended the single-entry DBH-based SUR models to include other predictor variables. Both single- and multiple-entry SUR models were constrained such that the prediction of component equations sum to the pre-

Fig. 2. Aboveground biomass distribution in different tree components (percentage of total aboveground biomass) in different species (DF, Douglas-fir; LP, lodgepole pine). Note the differences in scale.
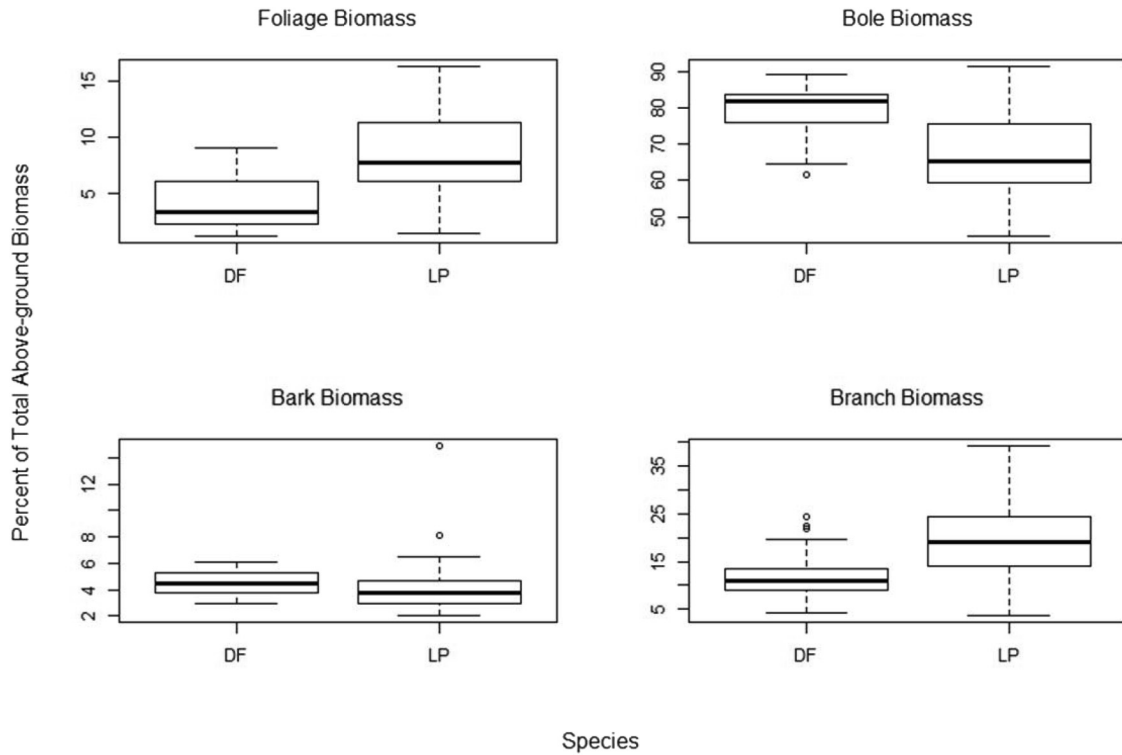
diction of total tree regression. The models for component and total biomass in the simple and extended SUR methods were in the following form:

$$(2) \quad \text{Bole} = \exp(a_{11} + a_{12}X_1 + a_{13}X_2 + a_{14}X_3)$$

$$(3) \quad \text{Bark} = \exp(a_{21} + a_{22}X_1 + a_{23}X_2 + a_{24}X_3)$$

$$(4) \quad \text{Branch} = \exp(a_{31} + a_{32}X_1 + a_{33}X_2 + a_{34}X_3)$$

$$(5) \quad \text{Foliage} = \exp(a_{41} + a_{42}X_1 + a_{43}X_2 + a_{44}X_3)$$

$$(6) \quad \text{Total} = \sum_{i,j} \exp(a_{ij} + a_{ij}X_1 + a_{ij}X_2 + a_{ij}X_3)$$

where a_{ij} ($i = 1,2,3,4$, and $j = 1,2,3,4$) are parameters to be estimated from the data and X_1, X_2, and X_3 are natural logarithms of DBH, total tree height, and crown length, respectively. Note that for the simple SUR method, only X_1 was used as explanatory variable. The dependent variables in these equations were component or total biomass (kg) in the original scale (not in the logarithmic scale). The choice of model form was primarily guided by past literature, and the diagnostic plots did not show a severe problem with the model form. The studentized residuals for the extended SUR models for bole biomass estimation in Douglas-fir and lodgepole pine trees are shown in Fig. 3. SAS procedure PROC MODEL (SAS Institute Inc. 2013) was used to fit both simple and extended SUR models. We made the necessary computational adjustments to match the component definitions in all methods and applied these methods to estimate component and total aboveground biomass of all five species.

Group III methods

The component biomass can be estimated as the proportions of total aboveground biomass. A proportion is bounded between 0 and 1, and therefore, the effect of explanatory variables tends to

be nonlinear, and the variance tends to decrease when the mean get closer to one of the boundaries. In this study, we evaluated three different methods for estimating proportions, namely the beta regression, Dirichlet regression, and the multinomial log-linear regression (MLR).

Beta regression

The beta regression model was introduced by Ferrari and Cribari-Neto (2004) and is useful when the variable of interest is continuous, restricted to the interval (0, 1) such as percentages, proportions, and fractions or rates, and related to other variables through a regression structure. Since then, it has been used in many fields including medicine (Hubben et al. 2008), economics (De Paola et al. 2010), education (Smithson and Verkuilen 2006), and forestry (Korhonen et al. 2007; Eskelson et al. 2011). Korhonen et al. (2007) used this technique for modeling percent canopy cover in a conifer-dominated study area in central Finland. Eskelson et al. (2011) used beta regression to estimate riparian percent shrub cover. This method produced smaller mean squared prediction error and absolute bias compared with the ordinary least squares and generalized least squares regression models used in their study. The usual beta distribution is of the following form:

$$(7) \quad f(y; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1 - y)^{\beta-1} \quad \text{for } 0 < y < 1$$

where α and $\beta > 0$ are two shape parameters and $\Gamma(\cdot)$ is the gamma function.

To obtain a regression structure for the mean of the response and a precision (dispersion) parameter, Ferrari and Cribari-Neto (2004) used a different parameterization of the beta distribution. With mean and precision parameters defined as $\mu = \alpha/(\alpha + \beta)$ and $\phi = (\alpha + \beta)$, respectively, the beta density function has the following form under new parameterization:

Fig. 3. Studentized residuals for the extended seemingly unrelated regression (SUR) models for bole biomass estimation in (*a*) Douglas-fir and (*b*) lodgepole pine trees.

a)

b)
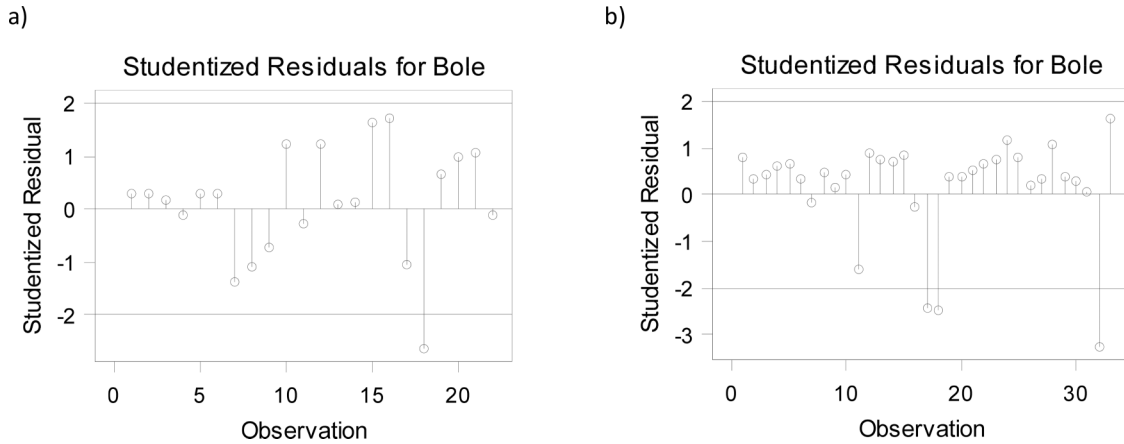
$$(8) \qquad f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma[(1-\mu)\phi]} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \quad \text{for } 0 < y < 1$$

where $0 < \mu < 1$ and $\phi > 0$.

With this parameterization, the beta regression model can be written as

$$(9) \qquad g(\mu_i) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} = \boldsymbol{\eta}_i$$

where $g(\cdot)$ is strictly increasing and is a double differentiable link function that maps $(0, 1)$ on to the real line $\mathbb{R}$, $\mathbf{x}_i = (x_{i1}, \ldots\ldots, x_{ik})^{\mathrm{T}}$ is a vector of k explanatory variables, $\boldsymbol{\beta} = (\beta_1, \ldots\ldots, \beta_k)^{\mathrm{T}}$ is a vector of k unknown regression parameters ($k < n$), and $\boldsymbol{\eta}_i$ is a linear predictor (i.e., $\boldsymbol{\eta}_i = \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$, usually $x_{i1} = 1$ for all i so that the model has an intercept) (Cribari-Neto and Zeileis 2010).

We used various combinations of explanatory variables to predict proportion of aboveground biomass present in different components. The logit link function $g(\mu) = \log[\mu/(1-\mu)]$ was used, thus the predicted proportions are obtained as $\mu_i = [\exp(\boldsymbol{\eta}_i)/(1 + \exp(\boldsymbol{\eta}_i))]$. The beta regression was performed in R version 3.1.2 (R Core Team 2014) with function betareg in package betareg (Cribari-Neto and Zeileis 2010). The predictor variables in the beta regression were DBH and total tree height for both species.

Dirichlet regression

The Dirichlet distribution is a multivariate generalization of the beta distribution and takes the following form:

$$(10) \qquad f(\mathbf{y}; \boldsymbol{\alpha}) = \frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_{c=1}^{C} y_c^{(\alpha_c-1)}$$

where α_c are the shape parameters for each variable, $\alpha_c > 0$ for all c (component ID, ranges from 1 to C), $y_c \in (0,1)$, $\sum_{c=1}^{C} y_c = 1$ for all c, and C is the number of variables. $\mathrm{B}(\alpha_c) = \left[(\prod_{c=1}^{C}\Gamma(\alpha_c))/\Gamma(\sum_{c=1}^{C}\alpha_c)\right]$ is the multinomial beta function. If $C = 2$, then the Dirichlet distribution reduces to the beta distribution. Maier (2014a) used the generalization of Ferrari and Cribari-Neto (2004) and reparameterized the Dirichlet distribution with mean and precision parameters $\mu_c = (\alpha_c/\phi)$ and $\phi = \alpha_0 = \sum_{c=1}^{C}\alpha_c$, respectively. Then, the Dirichlet density has the following form:

$$(11) \qquad f(\mathbf{y}; \boldsymbol{\mu}, \phi) = \frac{1}{\mathrm{B}(\boldsymbol{\mu}\phi)} \prod_{c=1}^{C} y_c^{(\mu_c\phi-1)}$$

where $0 < \mu < 1$ and $\phi > 0$.

The Dirichlet regression is useful for modeling data that represent the components as percentage of the total. With the usual parameterization, the regression model can be formulated as

$$(12) \qquad g(\alpha_c) = \boldsymbol{\eta}_c = \mathbf{X}^{[c]}\boldsymbol{\beta}^{[c]}$$

Where $g(\cdot)$ is the link function, which is $\log(\cdot)$ for the model with usual parameterization (Maier 2014a), the superscript $[c]$ represents the predicted proportion of component c. The predicted values are obtained as $\mu_c = \exp(\boldsymbol{\eta}_c)$. The Dirichlet regression was performed in R version 3.1.2 (R Core Team 2014) with function DirichReg in package DirichletReg (Maier 2014b). The predictor variables used in Dirichlet regression were also the DBH and total tree height for both species.

Multinomial log-linear regression

In this method, four components (bole, bole bark, branch, and foliage) were set to four nominal values. The models to predict proportions of total tree biomass found in bole wood, stem bark, foliage, and branch were fit simultaneously using a multinomial logit model. The models for component proportions were

$$(13) \qquad p_{\text{Bole}} = \frac{1}{1 + e^{(a_1+a_2X_1+a_3X_2)} + e^{(b_1+b_2X_1+b_3X_2)} + e^{(c_1+c_2X_1+c_3X_2)}}$$

$$(14) \qquad p_{\text{Bark}} = \frac{e^{(a_1+a_2X_1+a_3X_2)}}{1 + e^{(a_1+a_2X_1+a_3X_2)} + e^{(b_1+b_2X_1+b_3X_2)} + e^{(c_1+c_2X_1+c_3X_2)}}$$

$$(15) \qquad p_{\text{Foliage}} = \frac{e^{(b_1+b_2X_1+b_3X_2)}}{1 + e^{(a_1+a_2X_1+a_3X_2)} + e^{(b_1+b_2X_1+b_3X_2)} + e^{(c_1+c_2X_1+c_3X_2)}}$$

$$(16) \qquad p_{\text{Branch}} = \frac{e^{(c_1+c_2X_1+c_3X_2)}}{1 + e^{(a_1+a_2X_1+a_3X_2)} + e^{(b_1+b_2X_1+b_3X_2)} + e^{(c_1+c_2X_1+c_3X_2)}}$$

where p_{Bole}, p_{Bark}, p_{Foliage}, and p_{Branch} are proportions of total aboveground biomass in bole, bark, foliage, and branch, respectively; $X_1 = \text{DBH}$; $X_2 = \text{total tree height}$; and a_i, b_i, and c_i ($i = 1,2,3$) are model parameters. The multinomial logit fit provides the "probability" of observing these components and can be considered as the proportion of biomass in each component and estimated by model parameters (Boudewyn et al. 2007). The MLR was performed in R version 3.1.2 (R Core Team 2014) with function multinom in package nnet (Venables and Ripley 2002). The biomass present in each component was used as the frequency weight, and the bole component was used as the reference group.

Table 2. Mean bias and root mean squared error (RMSE) for component and total aboveground biomass produced by the FIA-CRM, FIA-PNW, and Jenkins methods.

Species	Bias (kg)					RMSE (kg)				
	Foliage	Bark	Branch	Bole	Total	Foliage	Bark	Branch	Bole	Total
FIA-CRM method										
DF	−79.6	−189.6	−73.3	−109.7	−32.5	126.0	281.1	136.8	289.8	234.5
LP	3.2	−4.8	19.4	25.1	72.4	12.7	12.5	37.9	70.5	101.8
FIA-PNW method										
DF	−14.9	−143.2	−161.2	182.8	−136.6	37.7	232.5	280.3	311.7	376.1
LP	6.6	−3.9	22.7	38.8	64.2	12.0	9.4	40.6	78.7	99.7
Jenkins method										
DF	−102.1	−255.5	−134.8	−93.8	−586.2	176.0	418.7	246.5	597.6	1327.7
LP	3.5	−18.7	29.7	49.1	63.6	9.9	26.7	44.6	102.5	104.2

Note: DF, Douglas-fir; LP, lodgepole pine.

The methods used to estimate aboveground biomass and (or) its components in this study were applied to Douglas-fir and lodgepole pine trees. Performance of all the methods was evaluated based on the bias and RMSE produced by each method. In statistics, bias is defined as the difference between the true value of an unknown parameter and the expected value of its estimator. Bias in this study is defined as the mean difference between the measured or observed value and the predicted value of the variable of interest.

Results and discussion

The FIA-CRM, FIA-PNW, and the Jenkins methods were biased and produced the highest values for RMSE. The mean bias and RMSE produced by these methods are given in Table 2. These methods produced similar estimates for total aboveground biomass for lodgepole pine trees, but for the Douglas fir trees, these methods provided differing estimates. The Jenkins method for Douglas fir produced total aboveground biomass that was 18.4% and 23.7% higher than the estimates provided by the FIA-PNW and FIA-CRM methods, respectively. These methods provided the component biomass estimates for lodgepole pine that were closer to each other than the component biomass estimates for Douglas-fir. However, none of these methods were consistent in over- or under-predicting the component masses (Fig. 4). Previous studies have also shown great discrepancies in biomass estimates obtained from the FIA-CRM, FIA-PNW, and Jenkins methods. For example, Nay and Bormann (2014) found the Jenkins equation to predict up to 190% of the site-specific foliage biomass equation for Douglas-fir.

The lodgepole pine trees in our sample were the trees with smaller DBHs (mean DBH, 24.6 cm) compared with the DBHs of the Douglas-fir trees (mean DBH, 54.9 cm). Indeed, the Jenkins, FIA-CRM, and FIA-PNW methods were more sensitive to tree size compared with other methods. For example, the RMSE percent for total aboveground biomass using the Jenkins equations for Douglas fir dropped from 57.7% to 11.1% when this approach was applied to the trees that are < 94 cm in DBH. With the FIA-CRM method, the RMSE percent for total aboveground biomass of Douglas fir decreased from 10.2% to 7.1% for trees that are < 94 cm in DBH. Similarly, the RMSE percent for total aboveground biomass of Douglas fir decreased from 16.3% to 8.5% in using the FIA-PNW method for trees that are < 94 cm in DBH.

The parameter estimates and their approximate standard errors of the simple and extended SUR models are presented in Table 3. The mean bias and RMSE produced by simple and extended SUR approaches are presented in Table 4. These locally fitted SUR models (both simple and extended) provided component and total aboveground biomass estimates that were more accurate (lower RMSE) than the estimates produced by the FIA-CRM, FIA-PNW, and Jenkins methods. Including additional explanatory variables than just DBH in the SUR models resulted in the decrease in RMSE percent from 10.7% to 8.3% for Douglas fir and from 22.8% to 20.5% for lodgepole pine total aboveground biomass. The RMSE for bole biomass estimation was reduced by 2.3% and 6.9% for Douglas fir and lodgepole pine, respectively, by using the extended SUR model instead of the simple SUR models. It is logical because one would, for example, expect differences, at least, in the bole biomass for a same DBH tree with a different height, which would not be accounted for by DBH-only models.

However, it should be noted that even though the RMSE for total aboveground biomass is decreased by using the extended SUR model, the RMSE for some component biomass slightly increased (Fig. 5). This could have been avoided by not constraining the extended SUR models, i.e., fitting independent component models rather than fitting a system of equations, which, in turn, would have affected the additivity of the component models.

The beta regression, Dirichlet regression, and MLR provided the predicted proportions of each component biomass. The predicted proportions were then applied to observed total aboveground biomass to obtain predicted biomass estimates in different components ($\hat{B}_c$), i.e., $\hat{B}_c = \hat{p}\text{AGB}$, where $\hat{p}$ and AGB are predicted proportions and observed total aboveground biomass (kg), respectively. In practice, it is unreasonable to assume that information on actual or measured total AGB would be available. In that case, the AGB should be replaced with $\widehat{\text{AGB}}$, i.e., the predicted total aboveground biomass. The model given in eq. 6 can be used to obtain $\widehat{\text{AGB}}$. The parameter estimates for these models are available on request. The bias and RMSE produced by these methods are given in Table 5. These methods unbiasedly predicted component proportions for all species. These methods also provided component biomass estimates that were more accurate (lower RMSE) than the estimates produced by the FIA-CRM, FIA-PNW, and Jenkins methods.

The beta regression produced smaller RMSEs compared with the simple SUR models, except for Douglas-fir branch biomass, whereas it produced smaller RMSEs than the extended SUR models for both the Douglas-fir and lodgepole pine component biomasses. The beta regression produced a 2.8% higher RMSE for Douglas-fir branch biomass, but for foliage, bark, and bole biomass, it decreased the RMSE by 5.4%, 7.8%, and 8.8%, respectively, compared with the simple SUR models. In case of lodgepole pine, the beta regression decreased RMSEs by 16.5%, 4.2%, 13.7%, and 24.6% for foliage, bark, branch, and bole biomass, respectively, compared with the simple SUR models. Comparing the performance of beta regression against the extended SUR models, it decreased RMSEs by 8.1%, 11.8%, 0.2%, and 6.5% for Douglas-fir and by 13.6%, 4.2%, 8.4%, and 17.7% for lodgepole pine foliage, bark, branch, and bole biomasses, respectively. This was expected because the measurement of proportion takes values on the open interval (0, 1), and the influence of explanatory variables on continuous responses bounded between 0 and 1 can be investigated with the beta regression (Ferrari and Cribari-Neto 2004).

Fig. 4. Mean component biomass estimates produced by the FIA-CRM, FIA-PNW, and Jenkins methods in Douglas-fir and lodgepole pine trees.
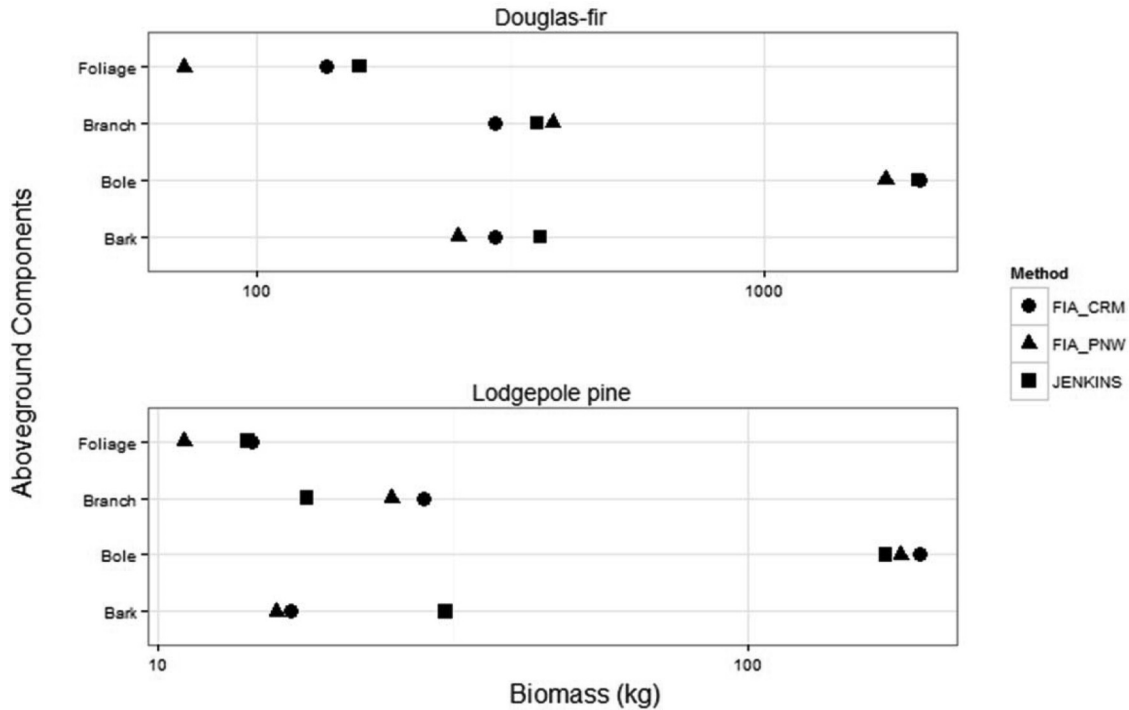

Table 3. Parameter estimates and their approximate standard errors of the simple and extended seemingly unrelated regression (SUR) models for Douglas-fir and lodgepole pine.

	Douglas-fir		Lodgepole pine	
Parameter	Estimate	Approximate standard error	Estimate	Approximate standard error
Simple SUR model				
a_{11}	−0.349	0.096	−2.374	0.334
a_{12}	1.925	0.021	2.373	0.093
a_{21}	−3.247	0.701	−4.401	1.150
a_{22}	1.919	0.154	2.105	0.324
a_{31}	−2.604	0.822	−3.384	1.137
a_{32}	1.942	0.180	2.202	0.318
a_{41}	−3.242	0.896	−2.472	0.855
a_{42}	1.758	0.197	1.637	0.245
Extended SUR model				
a_{11}	−1.993	0.062	−1.622	0.174
a_{12}	1.597	0.012	1.067	0.069
a_{13}	0.827	0.009	1.234	0.077
a_{21}	−4.459	0.806	−4.602	1.165
a_{22}	2.169	0.176	1.842	0.643
a_{23}	—	—	0.361	0.606
a_{31}	−3.856	0.454	−6.345	0.921
a_{32}	1.795	0.093	4.678	0.318
a_{33}	0.611	0.046	−1.865	0.283
a_{41}	−4.517	0.564	−5.134	0.776
a_{42}	1.497	0.147	3.843	0.324
a_{43}	0.779	0.171	−1.647	0.325

Similarly, the Dirichlet regression also produced a smaller RMSE compared with the simple SUR methods, except for Douglas-fir branch biomass, but it produced smaller RMSEs than the extended SUR models for both Douglas-fir and lodgepole pine component biomasses. The Dirichlet regression produced a 2.1% higher RMSE for Douglas-fir branch biomass; however, for foliage, bark, and bole biomass, it decreased RMSE by 1.6%, 3.7%, and 8.7% compared with the simple SUR models. In the case of lodgepole pine, the Dirichlet regression decreased RMSEs by 7.4%, 0.8%, 14.7%, and 23.2% for foliage, bark, branch, and bole biomass, respectively, compared with the simple SUR models. Comparing the performance of Dirichlet regression with the performance of the extended SUR models, it decreased RMSEs by 4.2%, 7.8%, 0.9%, and 6.3% for Douglas-fir and by 4.5%, 0.8%, 9.5%, and 16.3% for lodgepole pine foliage, bark, branch, and bole biomasses, respectively. One advantage of using Dirichlet regression over beta regression is that the Dirichlet regression allows simultaneous fitting of the component proportions, and therefore, the predicted proportions sum to 1. The better performance of Dirichlet regression models could be attributed to their capability to capture the variance in proportions that sum to a constant exhibiting skewness and heteroscedasticity.

The MLR consistently produced a smaller RMSE compared with both SUR methods for both species and all components. The MLR decreased RMSEs by 2.1%, 4.7%, 1.1%, and 9.1% for Douglas-fir and by 10.2%, 5.0%, 14.9%, and 23.5% for lodgepole pine foliage, bark, branch, and bole biomasses, respectively, compared with the simple SUR models. The MLR produced RMSEs that were 4.7%, 8.8%, 4.2%, and 6.8% smaller for Douglas-fir and 7.4%, 5.0%, 9.7%, and 16.6% smaller for lodgepole pine foliage, bark, branch, and bole biomasses, respectively, compared with the RMSEs produced by the extended SUR models.

There was no single method to predict proportion that was best for all both species and component biomass. The beta regression produced the smallest RMSEs for foliage and bark biomass estimation, whereas the MLR produced the smallest RMSEs for branch and bole biomass estimation for Douglas-fir. In the case of lodgepole pine, the beta regression produced the smallest RMSEs for foliage and bole biomass estimation, but the MLR produced the smallest RMSEs for bark and branch biomass estimation. It should be noted, however, that the RMSE percent produced by these methods were within 4.2% of each other, except for the lodgepole pine foliage estimation; in which case, the Dirichlet regression produced a 9.1% higher RMSE compared with the beta regression.

Table 4. Mean bias and root mean squared error (RMSE) for component and total aboveground biomass produced by the simple and extended seemingly unrelated regression (SUR) approaches.

Species	Bias (kg)					RMSE (kg)				
	Foliage	Bark	Branch	Bole	Total	Foliage	Bark	Branch	Bole	Total
Simple SUR method										
DF	4.0	0.4	0.3	−44.1	−39.3	18.4	28.2	66.2	249.2	245.8
LP	0.7	0.1	2.3	−2.7	0.5	8.8	7.1	28.7	79.7	67.9
Extended SUR method										
DF	−0.7	12.7	−16.6	−16.3	−20.9	19.9	32.5	72.9	204.3	190.9
LP	2.5	0.2	6.8	−10.9	−1.2	8.3	7.1	26.2	64.4	60.9

Note: DF, Douglas-fir; LP, lodgepole pine.

Fig. 5. Root mean squared error (RMSE, kg) produced by simple and extended seemingly unrelated regression (SUR) models in estimating component biomass in Douglas-fir and lodgepole pine trees.
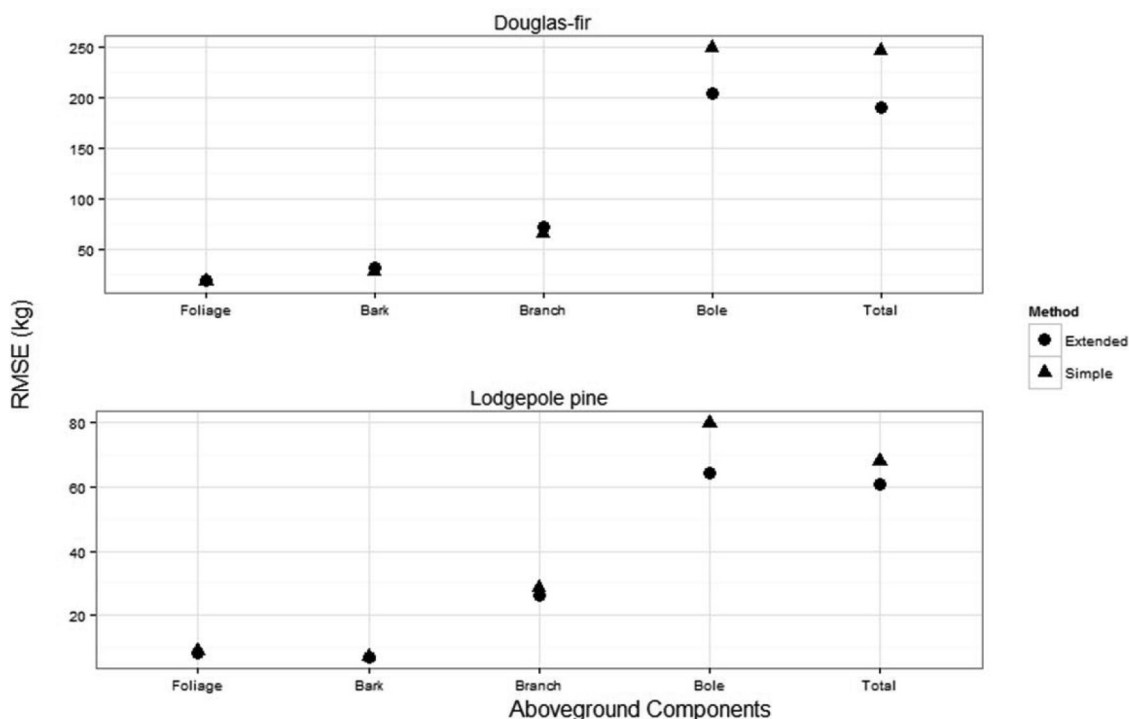

Table 5. Mean bias and root mean squared error (RMSE) of component biomass produced by the beta, multinomial loglinear regression (MLR), and Dirichlet regression methods.

Species	Bias (kg)				RMSE (kg)			
	Foliage	Bark	Branch	Bole	Foliage	Bark	Branch	Bole
Beta regression method								
DF	−0.218	−2.619	−1.221	2.994	15.3	20.0	72.5	79.8
LP	−0.122	0.105	−0.111	−0.766	5.9	6.6	22.2	25.5
MLR method								
DF	0.001	−0.005	−0.004	−0.083	17.2	23.2	63.7	74.3
LP	−0.007	−0.006	−0.025	−0.114	7.0	6.5	21.6	27.9
Dirichlet regression method								
DF	−1.948	−2.971	6.807	−1.979	17.5	24.3	70.8	83.0
LP	−1.515	−1.356	0.729	1.990	7.5	7.0	21.7	28.6

Note: DF, Douglas-fir; LP, lodgepole pine. Predicted component biomass was obtained by applying predicted proportions to the observed total aboveground biomass.

Summary and conclusion

The methods to estimate aboveground biomass and (or) its components used in this study provided differing estimates for total aboveground biomass and its components. Both simple and ex-tended SUR methods performed better than the FIA-CRM, the FIA-PNW, and the Jenkins methods for our dataset. The FIA-CRM, FIA-PNW, and Jenkins methods provided results comparable with both SUR methods for smaller trees (DBH, <94 cm), but they were highly biased for bigger trees. The methods for estimating proportions were clearly superior to the FIA-CRM, FIA-PNW, and Jenkins methods in terms of bias and RMSE. These methods were also superior to both simple and extended SUR methods with some exceptions. However, none of these methods was clearly superior to the other.

The Jenkins method for our Douglas fir trees produced total aboveground biomass that was 18.4% and 23.7% higher than the estimates provided by the FIA-PNW and FIA-CRM methods, respectively. Zhou and Hemstrom (2009) reported similar differences (17% and 20%, respectively) for the major softwood species in Oregon. The FIA-CRM, FIA-PNW, and Jenkins methods produced up to 3.4 and 2.6 times higher mean bark biomass estimates for Douglas-fir and lodgepole pine trees. The bole biomass estimates were within 10% for Douglas fir and within 22% for lodgepole pine. These methods were very inconsistent for branch and foliage biomass, overestimating for one species–component combination and underestimating for other combinations. The FIA-CRM and Jenkins methods were developed for larger scale biomass estima-

tion; therefore, the estimates obtained from these methods are inconsistent at the smaller scale.

In fitting the SUR models, the use of explanatory variables other than DBH was justified by the reduction in RMSE produced by the extended SUR method in estimating total aboveground biomass. This also reduced the RMSE for bole biomass. However, it did not improve prediction for some components. This could be because the added variable provided more information for the bole and total biomass but not for other components. We also found that, for Douglas fir, the addition of crown length rather than total height to the DBH-only model was useful in estimating foliage and branch biomass. By using crown length instead of total height, the adjusted R^2 of the foliage and branch models increased from 0.44 and 0.55 to 0.81 and 0.89, respectively.

The knowledge of biomass distributions in different tree components is essential to determine which portion of the tree can provide what amount of biomass for different purposes. The proportion of component biomass can be predicted using the beta regression, Dirichlet regression, and MLR. These methods were not only unbiased, but also produced very small RMSE values compared with the FIA-CRM, FIA-PNW, and Jenkins methods. They also provided better results compared with both SUR methods for most of the species–component combinations. One desired property in the component biomass estimation is the property of additivity, which can be attained by simultaneous fitting of component proportions in the Dirichlet regression and MLR. These methods ensure that the predicted component proportions sum to 1.

Even though the methods or models that are capable of predicting forest biomass at the large scale are desired, the use of such models without local calibration could lead to serious bias. When the large-scale models are calibrated, these models could perform even better than the locally fitted equations (e.g., de-Miguel et al. 2014a). With increasing interest in the utilization of component biomass for bioenergy, the importance of component models is also increasing. The differences in the scale of development and application of model could result in higher bias for the component biomass estimation. The findings of this study should provide information on the efficiency of selected methods in quantifying component and total aboveground biomass. Application of the methods to predict component proportions for other species and locations and with a larger dataset would further validate their accuracy.

References

Baskerville, G. 1972. Use of logarithmic regression in the estimation of plant biomass. Can. J. For. Res. 2(1): 49–53. doi:10.1139/x72-009.

Beaucham, J.J., and Olson, J.S. 1973. Corrections for bias in regression estimates after logarithmic transformation. Ecology, 54(6): 1403–1407. doi:10.2307/1934208.

Boudewyn, P., Song, X., Magnussen, S., and Gillis, M. 2007. Model-based, volume-to-biomass conversion for forested and vegetated land in Canada. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Information Report BC-X-411.

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.

Cribari-Neto, F., and Zeileis, A. 2010. Beta regression in R. J. Stat. Softw. 34(2): 1–24.

Czapowskyj, M.M., Robison, D.J., Briggs, R.D., and White, E.H. 1985. Component biomass equations for Black Spruce in Maine. USDA Forest Service, Northeastern Research Station, Broomall, Pennsylvania, Research Paper NE-564.

De Paola, M., Scoppa, V., and Lombardo, R. 2010. Can gender quotas break down negative stereotypes? Evidence from changes in electoral rules. J. Pub. Econ. 94(5): 344–353. doi:10.1016/j.jpubeco.2010.02.004.

de-Miguel, S., Mehtatalo, L., and Durkaya, A. 2014a. 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.

de-Miguel, S., Pukkala, T., Assaf, N., and Shater, Z. 2014b. Intra-specific differences in allometric equations for aboveground biomass of eastern Mediterranean *Pinus brutia*. Ann. For. Sci. 71(1): 101–112. doi:10.1007/s13595-013-0334-4.

Eskelson, B.N., Madsen, L., Hagar, J.C., and Temesgen, H. 2011. Estimating riparian understory vegetation cover with beta regression and copula models. For. Sci. 57(3): 212–221.

Ferrari, S., and Cribari-Neto, F. 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31(7): 799–815. doi:10.1080/0266476042000214501.

Flewelling, J.W., and Pienaar, L.V. 1981. Multiplicative regressions with lognormal errors. For. Sci. 27(2): 281–289.

Harrison, R.B., Terry, T.A., Licata, C.W., Flaming, B.L., Meade, R., Guerrini, I.A., Strahm, B.D., Xue, D., Lolley, M.R., Sidell, A.R., Wagoner, G.L., Briggs, D., and Turnblom, E.C. 2009. Biomass and stand characteristics of a highly productive mixed Douglas-fir and western hemlock plantation in coastal Washington. West. J. Appl. For. 24(4): 180–186.

He, Q.S., Chen, E.X., An, R., and Li, Y. 2013. Above-ground biomass and biomass components estimation using LiDAR data in a coniferous forest. Forests, 4(4): 984–1002. doi:10.3390/f4040984.

Heath, L.S., Hansen, M., Smith, J.E., Miles, P.D., and Smith, B.W. 2008. 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; October 21–23, 2008; Park City, UT. Edited by W.H. McWilliams, G.G. Moisen, and R.L. Czaplewski. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, Proc. RMRS-P-56CD.

Hollinger, D.Y. 2008. Defining a landscape-scale monitoring tier for the North American carbon program. In Field measurements for forest carbon monitoring. Edited by C. Hoover. Springer Netherlands. pp. 3–16.

Hoover, C.M. 2008. Field measurements for forest carbon monitoring: a landscape-scale approach. Springer, New York.

Hubben, G.A.A., Bishai, D., Pechlivanoglou, P., Cattelan, A.M., Grisetti, R., Facchin, C., Compostella, F., Bos, J., Postma, M., and Tramarin, A. 2008. The societal burden of HIV/AIDS in northern Italy: an analysis of costs and quality of life. AIDS Care, 20(4): 449–455. doi:10.1080/09540120701867107.

Jenkins, J.C., 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.

Korhonen, L., Korhonen, K.T., Stenberg, P., Maltamo, M., and Rautiainen, M. 2007. Local models for forest canopy cover with beta regression. Silva Fenn. 41(4): 671. doi:10.14214/sf.275.

Kuznetsova, T., Tilk, M., Parn, H., Lukjanova, A., and Mandre, M. 2011. Growth, aboveground biomass, and nutrient concentration of young Scots pine and lodgepole pine in oil shale post-mining landscapes in Estonia. Environ. Monit. Assess. 183(1–4): 341–350. doi:10.1007/s10661-011-1925-x.

Lambert, M.C., Ung, C.H., and Raulier, F. 2005. Canadian national tree aboveground biomass equations. Can. J. For. Res. 35(8): 1996–2018. doi:10.1139/x05-112.

Lu, D., Mausel, P., Brondizio, E., and Moran, E. 2002. Above-ground biomass estimation of successional and mature forests using TM images in the Amazon Basin. In Advances in spatial data handling. Edited by D.E. Richardson and P. van Oosterom. Springer. pp. 183–196.

Luo, Y., Wang, X., Zhang, X., Ren, Y., and Poorter, H. 2013. Variation in biomass expansion factors for China's forests in relation to forest type, climate, and stand development. Ann. For. Sci. 70(6): 589–599. doi:10.1007/s13595-013-0296-6.

Maier, M.J. 2014a. DirichletReg: Dirichlet regression for compositional data in R. Research Report Series. Department of Statistics and Mathematics, 125. WU Vienna University of Economics and Business, Vienna. Available from http://epub.wu.ac.at/4077/.

Maier, M.J. 2014b. DirichletReg: Dirichlet regression in R. R package version 0.5-2. Available from http://cran.revolution-computing.com/2014-11-25/web/packages/DirichletReg/DirichletReg.pdf.

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, Newtown Square, Pennsylvania, Research Note NRS-38 .

Nay, S.M., and Bormann, B.T. 2014. Site-specific Douglas-fir biomass equations from the Siskiyou Mountains, Oregon, compared with others from the Pacific Northwest. For. Sci. 60(6): 1140–1147. doi:10.5849/forsci.13-084.

Nogueira Junior, L.R., Engel, V.L., Parrotta, J.A., Galvão de Melo, A.C., and Ré, D.S. 2014. Allometric equations for estimating tree biomass in restored mixed-species Atlantic forest stands. Biota Neotrop. 14(2): 1–9. doi:10.1590/1676-06032013008413.

Nunes, L., Lopes, D., Rego, F.C., and Gower, S.T. 2013. Aboveground biomass and net primary production of pine, oak and mixed pine–oak forests on the Vila Real district, Portugal. For. Ecol. Manage. 305: 38–47. doi:10.1016/j.foreco.2013.05.034.

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

Parresol, B.R. 2001. Additivity of nonlinear biomass equations. Can. J. For. Res. 31(5): 865–878. doi:10.1139/x00-202.

Poudel, K.P., Temesgen, H., and Gray, A.N. 2015. Evaluation of sampling strategies to estimate crown biomass. For. Ecosyst. 2: 1. doi:10.1186/s40663-014-0025-0.

R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.r-project.org/.

Raile, G.K. 1982. Estimating stump volume. USDA Forest Service, St Paul, Minnesota, Research Paper NC-224.

Repola, J. 2006. Models for vertical wood density of Scots pine, Norway spruce and birch stems, and their application to determine average wood density. Silva Fenn. 40(4): 673. doi:10.14214/sf.322.

Ritchie, M.W., Zhang, J., and Hamilton, T.A. 2013. Aboveground tree biomass for *Pinus ponderosa* in northeastern California. Forests, **4**(1): 179–196. doi:10.3390/f4010179.

SAS Institute Inc. 2013. Base SAS 9.4 procedures guide: statistical procedures. 2nd edition. SAS Institute Inc, Cary, North Carolina. Available from http://support.sas.com/documentation/cdl/en/procstat/66703/PDF/default/procstat.pdf.

Smithson, M., and Verkuilen, J. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological Methods, **11**(1): 54. doi:10.1037/1082-989X.11.1.54.

Snowdon, P. 1991. A ratio estimator for bias correction in logarithmic regressions. Can. J. For. Res. **21**(5): 720–724. doi:10.1139/x91-101.

Temesgen, H., Monleon, V., Weiskittel, A., and Wilson, D. 2011. Sampling strategies of efficient estimation of tree foliage biomass. For. Sci. **57**(2): 153–163.

Tumwebaze, S.B., Bevilacqua, E., Briggs, R., and Volk, T. 2013. Allometric biomass equations for tree species used in agroforestry systems in Uganda. Agroforestry Systems, **87**(4): 781–795. doi:10.1007/s10457-013-9596-y.

Vashum, K.T., and Jayakumar, S. 2012. Methods to estimate above-ground biomass and carbon stock in natural forests — a review. J. Ecosyst. Ecogr. **2**(4): 1–7. doi:10.4172/2157-7625.1000116.

Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S. 4th edition. Springer, New York.

Westfall, J.A., MacFarlane, D.W., and Weiskittel, A.R. 2012. Biomass measurement and modeling challenges for hardwood species in the northern region. *In* Moving from status to trends: Forest Inventory and Analysis (FIA) symposium 2012, December 4–6, 2012, Baltimore, Maryland.

Xiao, C.W., Yuste, J.C., Janssens, I.A., Roskams, P., Nachtergale, L., Carrara, A., Sanchez, B.Y., and Ceulemans, R. 2003. Above- and belowground biomass and net primary production in a 73-year-old Scots pine forest. Tree Physiol. **23**(8): 505–516. doi:10.1093/treephys/23.8.505.

Zhang, Y.J., Borders, B.E., Will, R.E., and Posadas, H.D. 2004. A model for foliage and branch biomass prediction for intensively managed fast growing loblolly pine. For. Sci. **50**(1): 65–80.

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, Portland, Oregon, Research Paper PNW-RP-584.

Zhou, X., and Hemstrom, M.A. 2010. Timber volume and aboveground live tree biomass estimations for landscape analyses in the Pacific Northwest. USDA Forest Service, Portland, Oregon, General Technical Report PNW-GTR-819.