NRC Research Press

ARTICLE

Additive biomass equations for slash pine trees: comparing three modeling approaches

Dehai Zhao, James Westfall, John W. Coulston, Thomas B. Lynch, Bronson P. Bullock, and Cristian R. Montes

Abstract: Both aggregative and disaggregative strategies were used to develop additive nonlinear biomass equations for slash pine (*Pinus elliottii* Engelm. var. *elliottii*) trees in the southeastern United States. In the aggregative approach, the total tree biomass equation was specified by aggregating the expectations of component biomass models, and their parameters were estimated by jointly fitting all component and total biomass equations using weighted nonlinear seemingly unrelated regression (NSUR) (SUR1) or by jointly fitting component biomass equations using weighted NSUR (SUR2). In an alternative disaggregative approach (DRM), the biomass component proportions were modeled using Dirichlet regression, and the estimated total tree biomass was disaggregated into biomass components based on their estimated proportions. There was no single system to predict biomass that was best for all components and total tree biomass. The ranking of the three systems based on an array of fit statistics followed the order of SUR2 > SUR1 > DRM. All three systems provided more accurate biomass predictions than previously published equations.

Key words: nonlinear seemingly unrelated regression, Dirichlet regression, biomass additivity, heteroscedasticity.

Résumé : Des stratégies d'agrégation et de désagrégation ont été utilisées pour développer des équations de biomasse non linéaires additives pour le pin d'Elliott typique (*Pinus elliottii* Engelm. var. *elliottii*) dans le sud-est des États-Unis. Dans l'approche agrégative, l'équation de la biomasse totale des arbres a été établie en agrégeant les prévisions des modèles des composantes de la biomasse, et leurs paramètres ont été estimés en ajustant conjointement toutes les équations de la biomasse totale et de ses composantes à l'aide de régressions non linéaires pondérées apparemment non reliées (NSUR) (SUR1) ou par ajustement conjoint des équations des composantes de la biomasse à l'aide du NSUR pondéré (SUR2). Dans une autre approche, par désagrégation (DRM), les proportions des composantes de la biomasse ont été modélisées à l'aide de la régression de Dirichlet et la biomasse totale estimée des arbres a été désagrégée en composantes de la biomasse en fonction de leurs proportions estimées. Aucun des systèmes permettant de prédire la biomasse s'est avéré le meilleur pour toutes les composantes et la biomasse totale des arbres. Le classement des trois systèmes, sur la base d'un ensemble de statistiques d'ajustement, suivait l'ordre suivant : SUR2 > SUR1 > DRM. Les trois systèmes ont fourni des prévisions de biomasse plus précises que les équations publiées précédemment. [Traduit par la Rédaction]

Mots-clés : régressions non linéaires apparemment non reliées, régression de Dirichlet, additivité de la biomasse, hétéroscédasticité.

Introduction

Slash pine (*Pinus elliottii* Engelm. var. *elliottii*) is an important commercial timber species in the southeastern United States (US). Slash pine has been planted on more than 4.2 million ha, covering a wide range from eastern Texas to southern North Carolina to south-central Florida, with 79% of the planted slash pine occurring in Florida and Georgia (Barnett and Sheffield 2005). To assess plantation productivity, nutrient cycling, energy flows, and carbon storage and sequestration, forest researchers, managers, and policymakers often need biomass equations of individual trees to accurately estimate different components of tree biomass. Stem biomass or variable-top stem dry mass prediction equations have been developed for slash pine (Parresol and Thomas 1989; Bailey and Fang 2000), but few studies have been conducted in component biomass models for slash pine, with the exception of Gonzalez-Benecke et al. (2014).

A desirable property of tree biomass component equations is that the predictions for the components sum to the prediction from the total tree equation (Kozak 1970; Parresol 2001; Bi et al. 2004; Zhao et al. 2015). Gonzalez-Benecke et al. (2014) developed separate biomass component equations and total tree biomass equations. This approach resulted in inconsistent biomass estimates in terms of additivity: sum of the component estimates did not equate to that given by the total equation (Kozak 1970). Separately fitting the biomass equations also ignores the inherent correlation among the biomass components measured on the same sample trees. Taking into account this correlation when estimating a system of additive biomass equations has greater statistical efficiency (Parresol 1999, 2001).

Parresol (2001) proposed an aggregation approach that has become a standard method that ensures additivity of individual component biomass estimates. In Parresol's approach, a nonlinear model is specified for each of M tree biomass components, and then these component models are aggregated to the total tree biomass. These aggregative models were usually estimated by jointly fitting all $M + 1$ equations using weighted nonlinear seemingly unrelated regression (NSUR) (e.g., Bi et al. 2004; Zhao et al.

Received 8 June 2018. Accepted 7 September 2018.

D. Zhao, B.P. Bullock, and C.R. Montes. Warnell School of Forestry and Natural Resources, The University of Georgia, Athens, GA 30602, USA.
J. Westfall. U.S. Forest Service, Northern Research Station, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073, USA.
J.W. Coulston. U.S. Forest Service, Southern Research Station, 1710 Research Center Drive, Blacksburg, VA 24060, USA.
T.B. Lynch. Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, OK 74078, USA.
Corresponding author: Dehai Zhao (email: zhaod@uga.edu).

2015; Wang et al. 2018; Dong et al. 2018). In this estimation method, hereafter referred to as SUR1, a constant $(M + 1) \times (M + 1)$ matrix is assumed for the inherent correlations among all biomass equations (biomass component equations and total biomass equation), and heteroscedasticity is addressed by having a unique weighting function for each equation.

Taking the fundamental additivity of biomass components themselves into consideration, Affleck and Diéguez-Aranda (2016) proposed to jointly fit M biomass component equations, rather than all $M + 1$ equations, using maximum likelihood estimation. For comparison purposes, here, the aggregative models were also estimated by jointly fitting M component equations using weighted NSUR, hereafter referred to as SUR2. In the SUR2, a constant $M \times M$ matrix is assumed for the cross-correlations among component equations only, and heteroscedasticity in component equations is addressed by different weighting functions. The SUR2 has seldom been used to estimate parameters of additive biomass equations, except in Affleck and Diéguez-Aranda (2016). When the SUR1 and SUR2 were used to fit aggregative models, they actually resulted in two different systems that might have similar model forms. Their difference in biomass predictions remains to be explored.

An alternative disaggregation approach is based on the development of component biomass fraction equations (Tang et al. 2000; Dong et al. 2015). In this approach, a total biomass model is first developed, and the estimated total tree biomass is disaggregated into tree components based on their proportions in the total. Two-stage nonlinear error-in-variable models (TSEM) is used to jointly estimate the coefficients of tree component models that guarantee additivity of the estimates of biomass components and totals (Tang et al. 2000). Zhao et al. (2016) recently evaluated Dirichlet, fractional multinomial logit, and log-ratio regression approaches for modeling biomass component proportions and found that the Dirichlet regression was superior to the other two methods. Unlike the approach of Tang et al. (2000), component biomass proportions could be directly modeled using the Dirichlet regression model (DRM) (Hijazi and Jernigan 2009; Zhao et al. 2016), and the total biomass model is separately developed. The estimated total tree biomass is disaggregated into tree components based on their estimated proportions. DRM guarantees that all of the estimated component proportions sum to 1. So this disaggregation approach, hereafter referred to as DRM, also guarantees the property of additivity among the components of tree biomass and total tree biomass.

In this study, three systems of additive biomass equations have been developed for slash pine in the southeastern US using SUR1, SUR2, and DRM, respectively. Their predictive performances were evaluated and compared with the biomass equations developed by Gonzalez-Benecke et al. (2014).

Materials and methods

Data description

Data used in this study are from two datasets of measurements on slash pine trees sampled from slash pine plantation stands in the coastal plain of Georgia and northern Florida. The first dataset consisted of measurements on 96 trees from destructive biomass sampling in 2016 and included diameter at breast height (DBH), total tree height, taper measurements, green mass of cut bolts, green mass of branches with foliage, green mass of disks, subsampled branches with foliage, and dry masses of disk wood, bark, branch, and foliage. The biomass sampling including field and laboratory measurements followed similar protocols as Zhao et al. (2015). Dry masses of stem wood, stem bark, branch, and foliage were calculated for each sampled tree using the method of Zhao et al. (2015). The second dataset consisted of 210 trees from legacy biomass database including DBH, total height, and dry mass of stem wood, stem bark, branch, and foliage. Tree total aboveground biomass is the sum of dry masses of tree components (stem wood, stem bark, branch, and foliage). Proportions of component biomass in tree total aboveground biomass were calculated as the ratio of component biomass to total biomass.

Table 1. Summary statistics of diameter at breast height (DBH, cm), total height (HT, m), stem wood, stem bark, branch, foliage, and total tree aboveground biomass (kg) for the sampled slash pine (*Pinus elliottii* var. *elliottii*) trees.

Variable	*N*	Mean	Minimum	Maximum	SD
DBH	306	18.4	3.0	53.3	9.5
HT	306	16.7	2.9	30.2	6.3
Stem wood	306	152.25	0.37	1391.71	220.81
Stem bark	306	22.51	0.36	137.96	23.61
Branch	306	19.90	0.04	325.77	38.01
Foliage	306	7.16	0.07	76.24	9.75
Total	306	201.83	0.98	1861.90	288.31

Note: *N*, number sampled; SD, standard deviation.

Summary statistics for DBH and total height of all trees used for developing biomass equations are shown in Table 1. Stem wood, stem bark, branch, foliage biomass, and tree total aboveground biomass of all trees and their relationships with tree DBH and total height are shown in Fig. 1. Stem wood, stem bark, branch, and foliage biomass proportions in tree total biomass and their relationships with tree DBH and total height are shown in Fig. 2.

Model specification and estimation

Aggregation approach — SUR1

Following the model structure specified in Parresol (2001), the biomass models for tree biomass components were constrained to equal the total tree biomass as follows:

$$
\begin{aligned}
&\boldsymbol{y}_1 = f_1(\boldsymbol{X}_1, \boldsymbol{\beta}_1) + \boldsymbol{\varepsilon}_1 \\
&\boldsymbol{y}_2 = f_2(\boldsymbol{X}_2, \boldsymbol{\beta}_2) + \boldsymbol{\varepsilon}_2 \\
&\qquad \ldots \\
&\boldsymbol{y}_M = f_M(\boldsymbol{X}_M, \boldsymbol{\beta}_M) + \boldsymbol{\varepsilon}_M \\
&\boldsymbol{y}_{M+1} = f_1(\boldsymbol{X}_1, \boldsymbol{\beta}_1) + f_2(\boldsymbol{X}_2, \boldsymbol{\beta}_2) + \ldots + f_M(\boldsymbol{X}_M, \boldsymbol{\beta}_M) + \boldsymbol{\varepsilon}_{M+1}
\end{aligned} \tag{1}
$$

where $\boldsymbol{y}_1$ to $\boldsymbol{y}_M$ represent the vector of biomass components (e.g., $M = 4$ for stem wood, stem bark, branch, and foliage), respectively; $\boldsymbol{y}_{M+1}$ is the vector of tree total aboveground biomass; $f_m(\boldsymbol{X}_m, \boldsymbol{\beta}_m)$ is a nonlinear function for biomass component m ($m = 1, 2, \ldots, M$); $\boldsymbol{\varepsilon}_i$ is the $N \times 1$ vector of residuals for equation i ($i = 1, \ldots, M + 1$); and N is the number of trees. The expectation of $\boldsymbol{\varepsilon}_i$ is $E(\boldsymbol{\varepsilon}_i) = \boldsymbol{0}$, then $E(\boldsymbol{y}_i) = \hat{\boldsymbol{y}}_i$, and thus $E(\boldsymbol{y}_{M+1}) = \sum_{m=1}^{M} \hat{\boldsymbol{y}}_m = \sum_{m=1}^{M} f_m(\boldsymbol{X}_m, \hat{\boldsymbol{\beta}}_m)$.

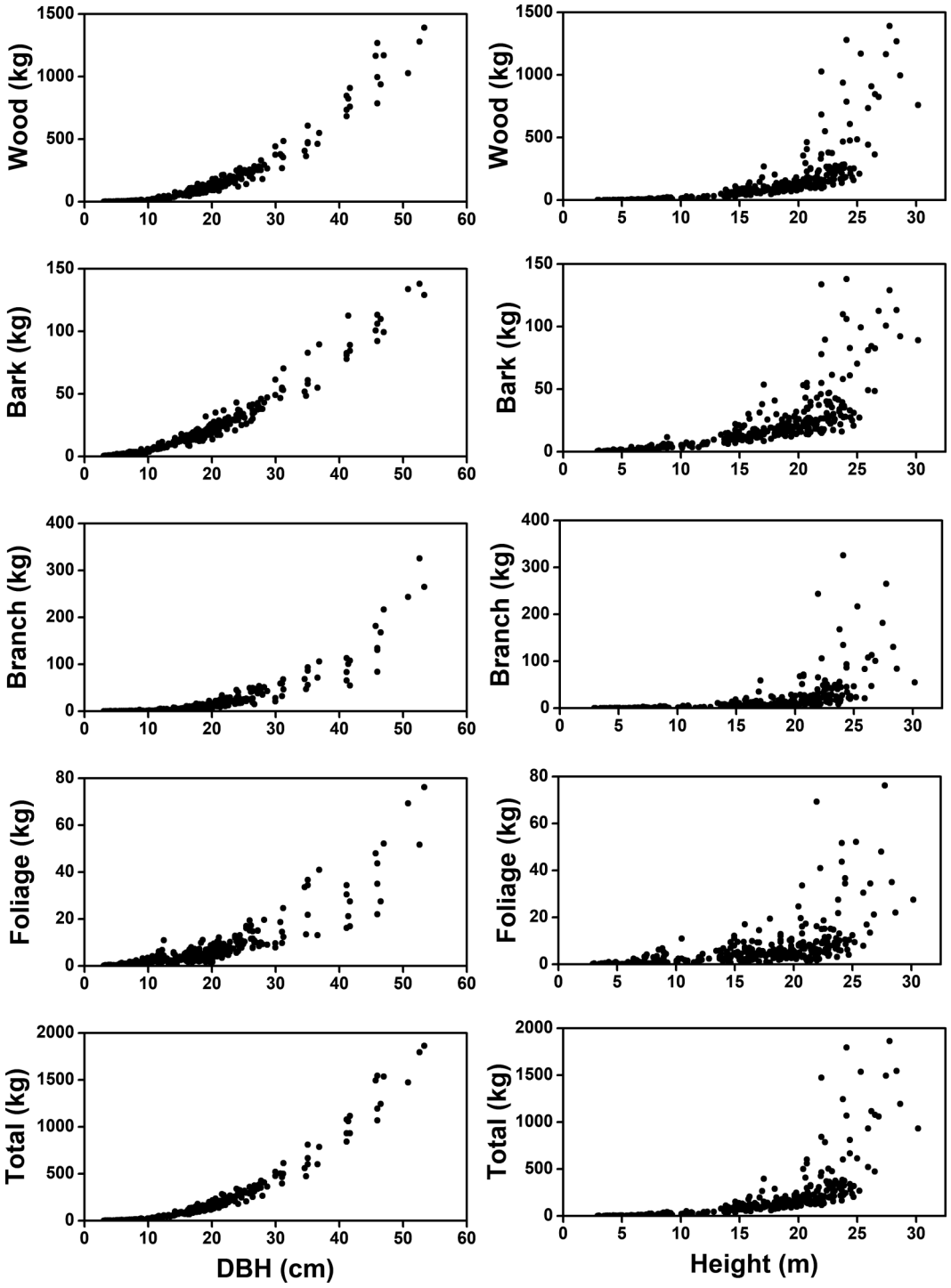
In the SUR1, all $M + 1$ biomass equations were jointly fit using weighted NSUR (Parresol 2001; Bi et al. 2004; Zhao et al. 2015). The total biomass equation was treated as component equations. For all $\boldsymbol{\varepsilon}_i$ ($i = 1, 2, \ldots, M + 1$), assume $\boldsymbol{\varepsilon}_i : iid(\boldsymbol{0}, \sigma_i^2 \boldsymbol{\psi}_i)$, where $\boldsymbol{\psi}_i$ is an $(N \times N)$ diagonal matrix. Heteroscedasticity in equation i is described by weighting function $\boldsymbol{\psi}_i$ of which the diagonal elements are not all identical.

Let $\mathbf{V}_i = \sigma_i \sqrt{\boldsymbol{\psi}_i^{-1}}$ and

$$
\mathbf{V}_{(M+1)N \times (M+1)N} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \ldots & \mathbf{0} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{V}_{M+1} \end{bmatrix} \tag{2}
$$

The inherent correlations among biomass components and total biomass measured on the same tree are described by the correlation matrix among all $(M + 1)$ biomass equations:

Fig. 1. Relationships between stem wood, stem bark, branch, foliage, and total biomass of the sampled slash pine trees and tree DBH and height.



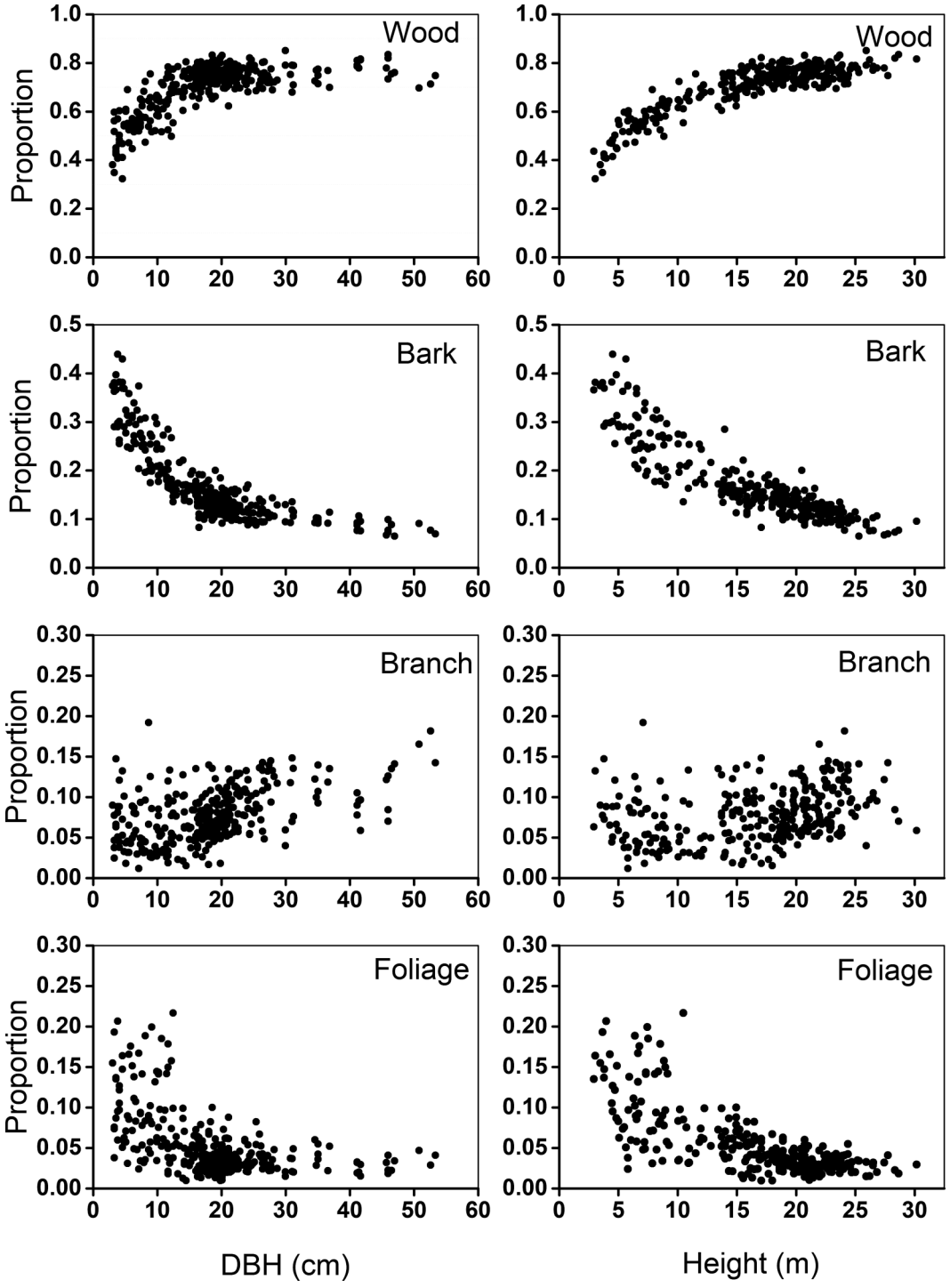
(3) $$\boldsymbol{\rho}_{(M+1)\times(M+1)} = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1(M+1)} \\ \rho_{21} & 1 & \dots & \rho_{2(M+1)} \\ \dots & \dots & \dots & \dots \\ \rho_{(M+1)1} & \rho_{(M+1)2} & \dots & 1 \end{bmatrix}$$

Let $\mathbf{C} = \boldsymbol{\rho} \otimes \mathbf{I}_N$, where $\otimes$ is the Kronecker product; $\mathbf{R} = \mathbf{VCV}$. Then the distribution for the residual terms $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \dots, \boldsymbol{\varepsilon}_{M+1}')'$ is assumed to follow $\boldsymbol{\varepsilon} : iid(\mathbf{0}, \mathbf{R})$ in the SUR1.

Aggregation approach — SUR2

In most cases, a total biomass is obtained by summing together the component biomass observations. For a given tree, tree total biomass $y_{M+1} = \sum_{m=1}^{M} y_m$. Due to this fundamental additivity of biomass data, the observed totals do not provide more information than component biomass observations. Therefore, Affleck and Diéguez-Aranda (2016) have estimated the parameters by jointly fitting M biomass component equations with maximum likelihood (ML) approach. That is, assume that $E(\boldsymbol{\varepsilon}_i) = \mathbf{0}$, and

Fig. 2. Relationships between proportions of stem wood, stem bark, branch, and foliage components in tree total aboveground biomass and tree size (DBH and height).



$E(\boldsymbol{y}_m) = \hat{\boldsymbol{y}}_m\ (m = 1, \ldots, M)$ for biomass components. After fitting, total biomass can be estimated as $\hat{\boldsymbol{y}}_{M+1} = \sum_{m=1}^{M} \hat{\boldsymbol{y}}_m = \sum_{m=1}^{M} f_m(\mathbf{X}_m, \hat{\boldsymbol{\beta}}_m)$.

In this study, biomass component equations were jointly fit using NSUR. Assume the error terms in biomass component biomass equations follow $\boldsymbol{\varepsilon}_m : iid(\mathbf{0}, \sigma_m^2 \boldsymbol{\psi}_m)\ (m = 1, 2, \ldots, M)$.

Let $\mathbf{V}_m = \sigma_m \sqrt{\boldsymbol{\psi}_m^{-1}}\ (m = 1, 2, \ldots, M)$ and

$$\mathbf{V}_{MN\times MN} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \ldots & \mathbf{0} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{V}_M \end{bmatrix} \quad (4)$$

The correlation matrix among component biomass equations is

$$(5)\qquad \boldsymbol{\rho}_{M\times M}=\begin{bmatrix}1 & \rho_{12} & \dots & \rho_{1M}\\ \rho_{21} & 1 & \dots & \rho_{2M}\\ \dots & \dots & \dots & \dots\\ \rho_{M1} & \rho_{M2} & \dots & 1\end{bmatrix}$$

From the viewpoint of model fitting, the difference between the SUR1 and SUR2 is that in the SUR1 one additional variance function $\mathrm{Var}(\boldsymbol{y}_{M+1})=\sigma^2_{M+1}\boldsymbol{\psi}_{M+1}$ (i.e., $\mathbf{V}_{M+1}=\sigma_{M+1}\sqrt{\boldsymbol{\psi}^{-1}_{M+1}}$) is specified for the total biomass equation, and additional constant correlations $\rho_{(M+1)m}$ ($m=1, 2, \dots, M$) are specified for cross-correlations between the total biomass equation and component biomass equations.

In the SUR2, however, the variance of tree total biomass $\boldsymbol{y}_{M+1}=\sum_{m=1}^{M}y_m$ can be estimated by

$$(6)\qquad \mathrm{Var}(y_{M+1})=\sum_{m=1}^{M}\hat{\sigma}^2_m\hat{\psi}_m+\sum\sum_{m\neq m'}\hat{\rho}_{mm'}\hat{\sigma}_m\sqrt{\hat{\psi}_m}\hat{\sigma}_{m'}\sqrt{\hat{\psi}_{m'}}$$

and the cross-covariance between the total biomass and biomass component m is

$$(7)\qquad \mathrm{Cov}[y_{M+1},y_m]=E\left[\left(\sum_{i=1}^{M}\varepsilon_i\right)\varepsilon_m\right]=\hat{\sigma}^2_m\hat{\psi}_m+\hat{\sigma}_m\sqrt{\hat{\psi}_m}\left(\sum_{i\neq m}^{M}\hat{\rho}_{im}\hat{\sigma}_i\sqrt{\hat{\psi}_i}\right)$$

The cross-correlation between the total biomass and biomass component m is

$$(8)\qquad \mathrm{Corr}[y_{M+1},y_m]=\frac{\mathrm{cov}[y_{M+1},y_m]}{\sqrt{\mathrm{var}(y_{M+1})}\sqrt{\mathrm{var}(y_m)}}=\frac{\hat{\sigma}^2_m\hat{\psi}_m+\hat{\sigma}_m\sqrt{\hat{\psi}_m}\left(\sum_{i\neq m}\hat{\rho}_{im}\hat{\sigma}_i\sqrt{\hat{\psi}_m}\right)}{\hat{\sigma}_m\sqrt{\hat{\psi}_m}\sqrt{\sum_{i=1}^{M}\hat{\sigma}^2_i\hat{\psi}_i+\sum\sum_{i\neq j}\hat{\rho}_{ij}\hat{\sigma}_i\sqrt{\hat{\psi}_i}\hat{\sigma}_j\sqrt{\hat{\psi}_j}}}$$

If homoscedasticity exists in each biomass component equation, i.e., $\psi_m=1$ ($m=1, 2, \dots, M$), or if heteroscedasticity exits in biomass component equations but can be addressed by the same weighting function, i.e., $\psi_m=\psi$ ($m=1, 2, \dots, M$), then the cross-correlation between the total biomass and component m can be reduced to a constant parameter:

$$(9)\qquad \rho_{(M+1)m}=\frac{\hat{\sigma}_m+\left(\sum_{i\neq m}\hat{\rho}_{im}\hat{\sigma}_i\right)}{\sqrt{\sum_{i=1}^{M}\hat{\sigma}^2_i+\sum\sum_{i\neq j}\hat{\rho}_{ij}\hat{\sigma}_i\hat{\sigma}_j}}\qquad (m=1, 2, \dots, M)$$

In reality, however, heteroscedasticity almost always exits in component biomass equations. It is practical to address the heteroscedasticity problem by having a unique weighting function for each biomass equation (Parresol 2001; Zhao et al. 2015). Because $\hat{\psi}_i\neq\hat{\psi}_j$ ($i\neq j$; $i, j=1, 2, \dots, M$) and they are functions of tree dimensions, the cross-correlations between the total biomass and biomass components (eq. 8) cannot be reduced to constant parameters (eq. 9) using the SUR2 (also see Affleck and Diéguez-Aranda 2016).

Tree biomass components were modeled as a power function of tree dimensions as

$$(10)\qquad f_m(\boldsymbol{X}_m,\boldsymbol{\beta}_m)=\beta_{m0}\mathrm{DBH}^{\beta_{m1}}\mathrm{HT}^{\beta_{m2}}$$

where $\boldsymbol{X}_m$ is tree diameter at breast height (DBH, cm) and tree total height (HT, m), and $\boldsymbol{\beta}_m=(\beta_{m0}, \beta_{m1}, \beta_{m3})$ to be estimated for biomass component m. Here, $m=1, 2, 3, 4$ for stem wood, stem bark, branch, and foliage, respectively. Each component equation can contain its own independent variables.

The following equation was fit using stepwise regression for each biomass equation in the system:

$$(11)\qquad \ln\left(\hat{\varepsilon}^2_i\right)=\ln\left(\sigma^2_i\right)+\gamma_{i1}\ln(\mathrm{DBH})+\gamma_{i2}\ln(\mathrm{HT})$$

where ln stands for natural logarithm, $\hat{e}_i$ represents the estimated errors for equation i in the system fitted using the unweighted NSUR. The resulting significant parameters $\hat{\gamma}_{i1}$ and $\hat{\gamma}_{i2}$ at $\alpha=0.05$ form the weighting functions $\hat{\psi}_i=\mathrm{DBH}^{\hat{\gamma}_{i1}}\mathrm{HT}^{\hat{\gamma}_{i2}}$ ($i=1, 2, \dots, M+1$) in the SUR1 or ($i=1, 2, \dots, M$) in the SUR2.

The aggregative models were fitted using the four-step fitting method (Zhao et al. 2015) using the weighted NSUR with nonlinear ordinary least squares (OLS) estimation (SUR1 or SUR2) and using the SAS/ETS® MODEL procedure (SAS Institute Inc. 2011). To verify the efficiency of parameter estimates obtained using the weighted NSUR, the primary functions were also estimated using weighted OLS estimation with the same weighting functions used in NSUR.

Disaggregation approach — DRM

Tang et al. (2000) initially developed a disaggregation strategy (also see Dong et al. 2015). In their strategy, a total biomass model $y_T=f_T(\boldsymbol{X}_T,\boldsymbol{\beta}_T)$ needs to be first developed. The biomass component models are defined as $y_m=f_m(\boldsymbol{X}_m,\boldsymbol{\beta}_m)$ ($m=1, 2, \dots, M$) and the component m fraction in tree total biomass is derived as $p_m=f_m(\boldsymbol{X}_m,\boldsymbol{\beta}_m)/\sum_{i=1}^{M}f_i(\boldsymbol{X}_i,\boldsymbol{\beta}_i)$. Then the estimated total biomass $\hat{y}_T$ is disaggregated into the component biomass: $y_m=p_m\times\hat{y}_T+\varepsilon_m$ ($m=1, 2, \dots, M$), that is, $\hat{y}_m=p_m\times\hat{y}_T$. The parameters associated with biomass component models are estimated through jointly fitting these component models with two-stage nonlinear error-in-variable models (TSEM) (Tang et al. 2001; Tang and Wang 2002) or NSUR (Dong et al. 2015).

In the current study, we used another disaggregation approach. The total biomass model was separately developed, and biomass component proportions were directly modeled using the Dirichlet regression model (DRM). Specifically, let $\{y_1, \dots, y_M\}$ represent the M individual biomass components, total biomass $y_T=\sum_{m=1}^{M}y_m$, and the component biomass proportions $p_m=y_m/y_T$. Thus, the vector of component proportions is $\boldsymbol{p}=(p_1, \dots, p_M)'$ with constraints $p_m\in(0, 1)$ and $\sum_{m=1}^{M}p_m=1$. Assume that the fractional components $\boldsymbol{p}$ follow the Dirichlet distribution with parameters $\alpha_1, \dots, \alpha_M>0$:

$$(12)\qquad \mathrm{Dir}(M,\boldsymbol{\alpha})=f(\boldsymbol{p}|\boldsymbol{\alpha})=\frac{1}{B(\boldsymbol{\alpha})}\prod_{m=1}^{M}p_m^{\alpha_m-1}$$

where

$$B(\boldsymbol{\alpha})=\frac{\prod_{m=1}^{M}\Gamma(\alpha_m)}{\Gamma\left(\sum_{m=1}^{M}\alpha_m\right)}=\frac{\prod_{m=1}^{M}\Gamma(\alpha_m)}{\Gamma(\alpha_0)},$$

$$\boldsymbol{\alpha}=(\alpha_1, \dots, \alpha_M)\ \text{and}\ \alpha_0=\sum_{m=1}^{M}\alpha_m$$

Table 2. Parameter estimates and their standard errors (SE) and *p* values for aggregative biomass equations fitted with SUR1 and SUR2.

			SUR1			SUR2		
Biomass component	Variable	Parameter	Estimate	SE	*p* value	Estimate	SE	*p* value
Stem wood		$\hat{\beta}_{10}$	0.0127	0.0006	<0.0001	0.0121	0.0007	<0.0001
	DBH	$\hat{\beta}_{11}$	2.0432	0.0251	<0.0001	2.0545	0.0315	<0.0001
	HT	$\hat{\beta}_{12}$	1.0387	0.0359	<0.0001	1.0398	0.0402	<0.0001
Stem bark		$\hat{\beta}_{20}$	0.0429	0.0024	<0.0001	0.0435	0.0031	<0.0001
	DBH	$\hat{\beta}_{21}$	1.7285	0.0371	<0.0001	1.8014	0.0437	<0.0001
	HT	$\hat{\beta}_{22}$	0.3557	0.0469	<0.0001	0.2732	0.0551	<0.0001
Branch		$\hat{\beta}_{30}$	0.0027	0.0003	<0.0001	0.0015	0.0002	<0.0001
	DBH	$\hat{\beta}_{31}$	3.1120	0.1088	<0.0001	3.0256	0.0446	<0.0001
	HT	$\hat{\beta}_{32}$	−0.2909	0.1255	0.0211			
Foliage		$\hat{\beta}_{40}$	0.0283	0.0041	<0.0001	0.0306	0.0048	<0.0001
	DBH	$\hat{\beta}_{41}$	2.7373	0.0936	<0.0001	2.8935	0.0972	<0.0001
	HT	$\hat{\beta}_{42}$	−0.9643	0.1109	<0.0001	−1.1482	0.1182	<0.0001

Note: DBH, diameter at breast height; HT, total height.

The expected value, variance, covariance, and correlations of the component proportions, respectively, are

$$\mathrm{E}[p_m] = \frac{\alpha_m}{\alpha_0}$$

$$\mathrm{Var}[p_m] = \frac{\alpha_m(\alpha_0 - \alpha_m)}{\alpha_0^2(\alpha_0 + 1)}$$

$$\mathrm{Cov}[p_m, p_{m'}] = \frac{-\alpha_m\alpha_{m'}}{\alpha_0^2(\alpha_0 + 1)} \qquad (m \neq m')$$

$$\mathrm{Cor}[p_m, p_{m'}] = -\sqrt{\frac{\alpha_m\alpha_{m'}}{(\alpha_0 - \alpha_m)(\alpha_0 - \alpha_{m'})}} \qquad (m \neq m')$$

The log link functions for the shape parameter of each component can be related to tree dimension variables as

$$\log(\alpha_m) = g_m(\boldsymbol{X}_m, \boldsymbol{\beta}_m) \qquad (m = 1, \ldots, M) \tag{13}$$

The maximum likelihood estimates of β parameters are obtained with the full log-likelihood of the Dirichlet distribution defined in eq. 14:

$$l(\boldsymbol{p}|\boldsymbol{\alpha}) = \log\Gamma\left(\sum\nolimits_{m=1}^{M}\alpha_m\right) - \sum\nolimits_{m=1}^{M}\log\Gamma(\alpha_m) + \sum\nolimits_{m=1}^{M}(\alpha_m - 1)\log(p_m) \tag{14}$$

Then

$$\hat{\alpha}_m = \exp\{g_m(\boldsymbol{X}_m, \hat{\boldsymbol{\beta}}_m)\}$$

$$\hat{p}_m = \hat{\alpha}_m/\hat{\alpha}_0$$

$$\hat{y}_m = \hat{p}_m \times \hat{y}_T \qquad (m = 1, \ldots, M)$$

Owing to $\sum_{m=1}^{M}\hat{p}_m = 1$, the DRM can also guarantee additivity of estimates of biomass components and the total: $\sum_{m=1}^{M}\hat{y}_m = \hat{y}_T$.

In this study, the following log link functions were used for biomass component proportions:

$$\log(\alpha_m) = \beta_{0i} + \beta_{1m}\log(\mathrm{DBH}) + \beta_{2m}\log(\mathrm{HT}) \qquad (m = 1, \ldots, M) \tag{15}$$

The DRMs for biomass component proportions were fitted using R Package DirichletReg (Maier 2014). The total biomass model was estimated using weighted nonlinear least squares (WNLS):

$$y_T = \beta_0\mathrm{DBH}^{\beta_1}\mathrm{HT}^{\beta_2} + \varepsilon_T \tag{16}$$

Model assessment and evaluation

For each modeling approach, the models were fitted to the entire dataset (sample size N = 306). Percent mean error (E), percent mean absolute error (MABE), percent root mean squared error (RMSE), and pseudo R^2 were used to evaluate the predictive performances of the aggregative models (fitted using SUR1 or SUR2) and the disaggregative models (DRM), based on the biomass predictions of M components and the total:

$$E_i\,(\%) = \frac{100}{N}\sum_{j=1}^{N}\frac{y_{ij} - \hat{y}_{ij}}{y_{ij}} \tag{17}$$

$$\mathrm{MABE}_i\,(\%) = \frac{100}{N}\sum_{j=1}^{N}\frac{|y_{ij} - \hat{y}_{ij}|}{y_{ij}} \tag{18}$$

$$\mathrm{RMSE}_i\,(\%) = 100\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\frac{y_{ij} - \hat{y}_{ij}}{y_{ij}}\right)^2} \tag{19}$$

$$R_i^2 = 1 - \frac{\sum_{j=1}^{N}(y_{ij} - \hat{y}_{ij})^2}{\sum_{j=1}^{N}(y_{ij} - \bar{y}_i)^2} \tag{20}$$

where y_{ij} and $\hat{y}_{ij}$ are the jth observed biomass and jth predicted biomass, respectively, for component i or total, and $\bar{y}_i$ is the mean of N observed biomass for the same component or total.

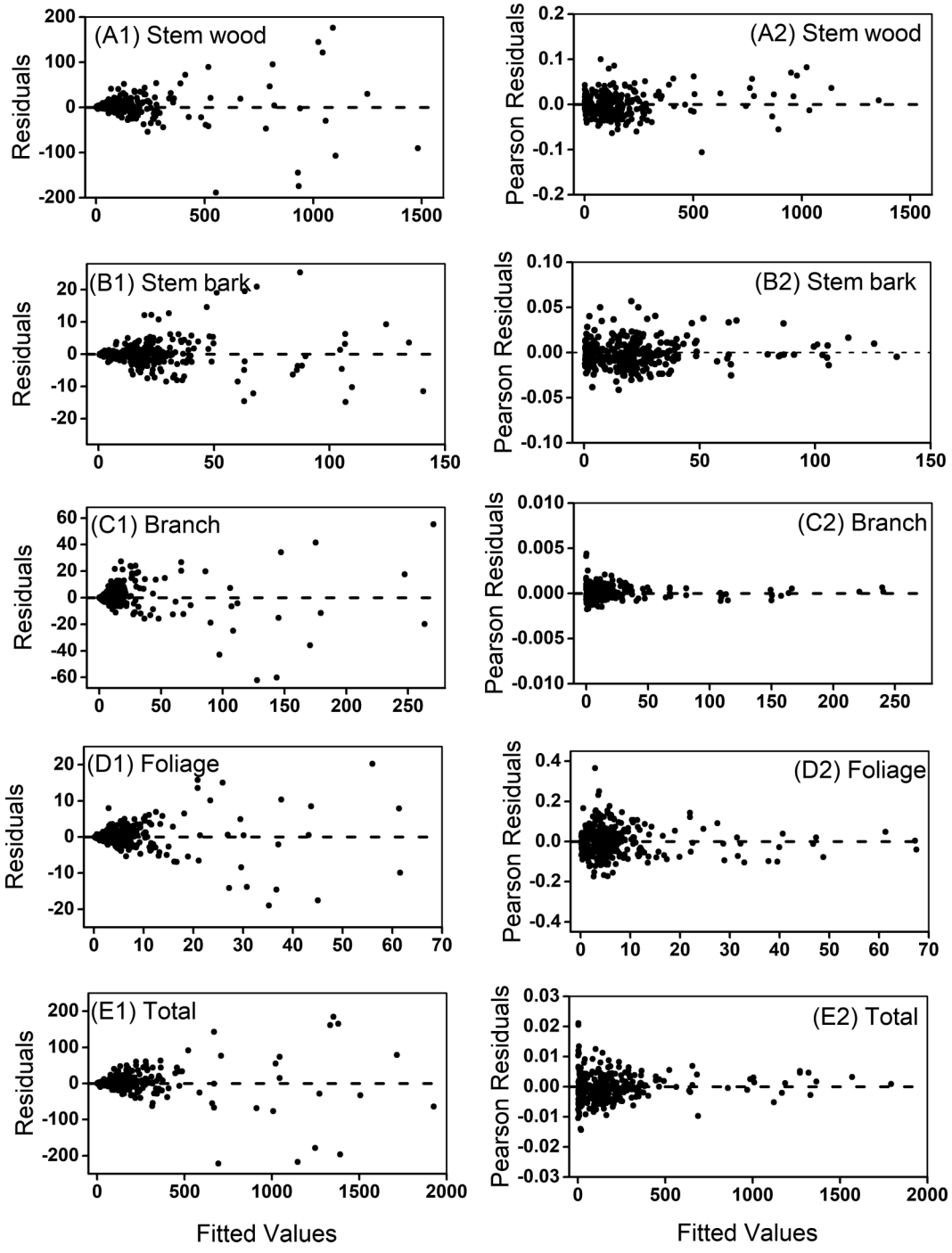
In this study, we did not conduct cross-validation using techniques such as data splitting or leave-one-out validation because they do not provide any additional information on model performance compared with the statistics obtained for model fit to the entire dataset (Kozak and Kozak 2003; Zhao et al. 2015; Gonzalez-Benecke et al. 2018). However, the predictive performance of the three systems was compared with previously published equations of Gonzalez-Benecke et al. (2014). We also refitted the functions of Gonzalez-Benecke et al. (2014) to our data used in this study and tried to keep the comparisons as equitable as possible.

Results and discussion

Aggregative biomass equations with SUR1

The component biomass equations (stem wood, stem bark, branch, and foliage) and total aboveground biomass equation

Fig. 3. Residual plots for each biomass component in the aggregative equations fitted with SUR1 without weigh functions (left: A1–E1) and Pearson residual plots for each biomass component in the aggregative equations fitted with SUR1 with different weighting functions (right: A2–E2).



were fitted jointly with weighted NSUR. The fitted additive biomass equations are shown below. The parameters acting as powers of DBH and HT were highly significant in each biomass equation (Table 2).

$$
\begin{aligned}
&\text{Wood:} && \hat{y}_1 = \hat{\beta}_{10}\mathrm{DBH}^{\hat{\beta}_{11}}\mathrm{HT}^{\hat{\beta}_{12}} \\
&\text{Bark:} && \hat{y}_2 = \hat{\beta}_{20}\mathrm{DBH}^{\hat{\beta}_{21}}\mathrm{HT}^{\hat{\beta}_{22}} \\
&\text{Branch:} && \hat{y}_3 = \hat{\beta}_{30}\mathrm{DBH}^{\hat{\beta}_{31}}\mathrm{HT}^{\hat{\beta}_{32}} \\
&\text{Foliage:} && \hat{y}_4 = \hat{\beta}_{40}\mathrm{DBH}^{\hat{\beta}_{41}}\mathrm{HT}^{\hat{\beta}_{42}} \\
&\text{Total:} && \hat{y}_5 = \hat{y}_1 + \hat{y}_2 + \hat{y}_3 + \hat{y}_4
\end{aligned}
\tag{21}
$$

The powers of DBH were positive in each biomass component, while the powers of HT were positive in stem wood and bark components and negative in branch and foliage components (Table 2). The positive powers of DBH and HT suggested their positive relationship with stem wood and bark biomass. This implies that, for the same DBH, tree stem wood and bark biomass increased with increasing tree height. The positive powers of DBH but negative powers of HT in branch and foliage components implied that their biomass increased with increasing DBH, but for the same DBH, branch and foliage biomass decreased with increasing tree height.

The error term of the total biomass equation of model 1 is actually a linear combination of other error terms: $\varepsilon_{M+1} = \sum_{m=1}^{M} \varepsilon_m$. When all $M + 1$ equations are jointly estimated using the seemingly unrelated regression, a singular across-equation variance and covariance matrix may occur (Bi et al. 2004). This happens if no weighting function or the same weighting function is used for all system equations. However, the SUR1 in the current study fit all $M + 1$ equations using weighted NSUR and had a unique weighting function for each equation. The weighting functions $DBH^{4.178}$, $DBH^{3.602}$, $DBH^{5.943}$, $DBH^{6.593}HT^{-4.444}$, and $DBH^{5.639}$ were used for stem wood, stem bark, branch, foliage, and total biomass equations, respectively. The different weighting functions not only stabilized the residual variances (Fig. 3), but also removed the singularity problem in across-equation variance and covariance matrix.

The singularity problem is overcome in SAS PROC MODEL by computing a generalized inverse of the variance and covariance matrix through setting the part of the matrix for the total tree biomass to zero. The parameter estimates are equivalent to fitting the first M component equations while leaving the total tree biomass equation out of the system. This is a special case of the SUR2, fitting M component equations using NSUR without the weighting or with the same weighting function for each component equation.

In the SUR1, a constant $(M + 1) \times (M + 1)$ matrix was assumed for cross-correlations among all equations, including the total biomass equation. In the current study, the following constant 5 × 5 matrix was estimated for correlations among four biomass components (stem wood, stem bark, branch, and foliage) and total biomass across all trees:

(22)
$$\begin{array}{c} \\ \text{Wood} \\ \text{Bark} \\ \text{Branch} \\ \text{Foliage} \\ \text{Total} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Wood} & \text{Bark} & \text{Branch} & \text{Foliage} & \text{Total} \end{array} \\ \begin{pmatrix} 1 & 0.328 & 0.175 & 0.199 & 0.783 \\ & 1 & -0.052 & 0.024 & 0.434 \\ & & 1 & 0.367 & 0.459 \\ & & & 1 & 0.416 \\ & & & & 1 \end{pmatrix} \end{array}$$

There were high correlations between total biomass and biomass components, between stem wood and stem bark components, and between branch and foliage components.

Aggregative biomass equations with SUR2

Stem wood, stem bark, branch, and foliage biomass component equations were fitted jointly with weighted NSUR, as follows:

(23)
$$\begin{aligned} &\text{Wood:} && \hat{y}_1 = \hat{\beta}_{10}\text{DBH}^{\hat{\beta}_{11}}\text{HT}^{\hat{\beta}_{12}} \\ &\text{Bark:} && \hat{y}_2 = \hat{\beta}_{20}\text{DBH}^{\hat{\beta}_{21}}\text{HT}^{\hat{\beta}_{22}} \\ &\text{Branch:} && \hat{y}_3 = \hat{\beta}_{30}\text{DBH}^{\hat{\beta}_{31}} \\ &\text{Foliage:} && \hat{y}_4 = \hat{\beta}_{40}\text{DBH}^{\hat{\beta}_{41}}\text{HT}^{\hat{\beta}_{42}} \end{aligned}$$

After fitting these component equations, the total biomass was estimated by summing the estimated component biomass: $\hat{y}_5 = \hat{y}_1 + \hat{y}_2 + \hat{y}_3 + \hat{y}_4$. The parameters acting as powers of DBH were highly significant in each biomass equation; the parameters acting as powers of HT were highly significant in stem wood, stem bark, and foliage equations, but not significant in the branch equation (Table 2). There were small differences in parameter estimates between SUR1 and SUR2, and most of the standard errors for SUR2 were larger than those of SUR1 (Table 2).

The weighting functions $\hat{\psi}_1$, $\hat{\psi}_2$, $\hat{\psi}_3$, and $\hat{\psi}_4$ were $DBH^{4.168}$, $DBH^{3.610}$, $DBH^{5.044}$, and $DBH^{6.471}HT^{-3.905}$, respectively, for stem wood, stem bark, branch, and foliage components. These weighting functions stabilized the residual variances for biomass components (not shown).

Table 3. Parameter estimates and their standard errors (SE) and p values for biomass component proportions fitted using DRM, and total biomass equation fitted using WNLS.

Component	Parameter	Estimate	SE	p value
DRM				
Stem wood: $\log(\alpha_1)$	Intercept	1.1301	0.1369	<0.0001
	log(DBH)	−0.9406	0.0681	<0.0001
	log(HT)	2.2182	0.0860	<0.0001
Stem bark: $\log(\alpha_2)$	Intercept	2.344	0.1415	<0.0001
	log(DBH)	−1.1568	0.0840	<0.0001
	log(HT)	1.4415	0.1022	<0.0001
Branch: $\log(\alpha_3)$	Intercept	−1.0547	0.2005	<0.0001
	log(HT)	1.2355	0.0698	<0.0001
Foliage: $\log(\alpha_4)$	Intercept	1.8557	0.0487	<0.0001
WNLS				
Total biomass	$\hat{\beta}_0$	0.0328	0.0023	<0.0001
	$\hat{\beta}_1$	2.1731	0.0291	<0.0001
	$\hat{\beta}_2$	0.6829	0.0420	<0.0001

Note: DBH, diameter at breast height; HT, total height.

In the SUR2, we assume a constant $M \times M$ matrix only for cross-correlations among M biomass component equations. In this study, a constant 4 × 4 matrix was estimated for four biomass components as follows:

(24)
$$\begin{array}{c} \\ \text{Wood} \\ \text{Bark} \\ \text{Branch} \\ \text{Foliage} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Wood} & \text{Bark} & \text{Branch} & \text{Foliage} \end{array} \\ \begin{pmatrix} 1 & 0.327 & 0.191 & 0.211 \\ & 1 & -0.032 & 0.018 \\ & & 1 & 0.426 \\ & & & 1 \end{pmatrix} \end{array}$$

There were high correlations between stem wood and stem bark components and between branch and foliage components. In the SUR2, the cross-correlation between the total biomass and biomass component is a function of all weighting functions for biomass components (see eq. 8). As in the SUR1, we used a unique weighting function for each biomass component equation; therefore, the correlation between total biomass and component biomass was not a constant across all trees.

When no weighting function is employed, or the same weighting function is used for all biomass component equations, $\hat{\psi}_m = \hat{\psi}$ $(m = 1, 2, \ldots, M)$ in the SUR2, the correlations between total biomass and biomass components can be reduced to a constant across all trees and can be calculated from the variance–covariance matrix of component errors using eq. 9.

Disaggregative biomass equations with DRM

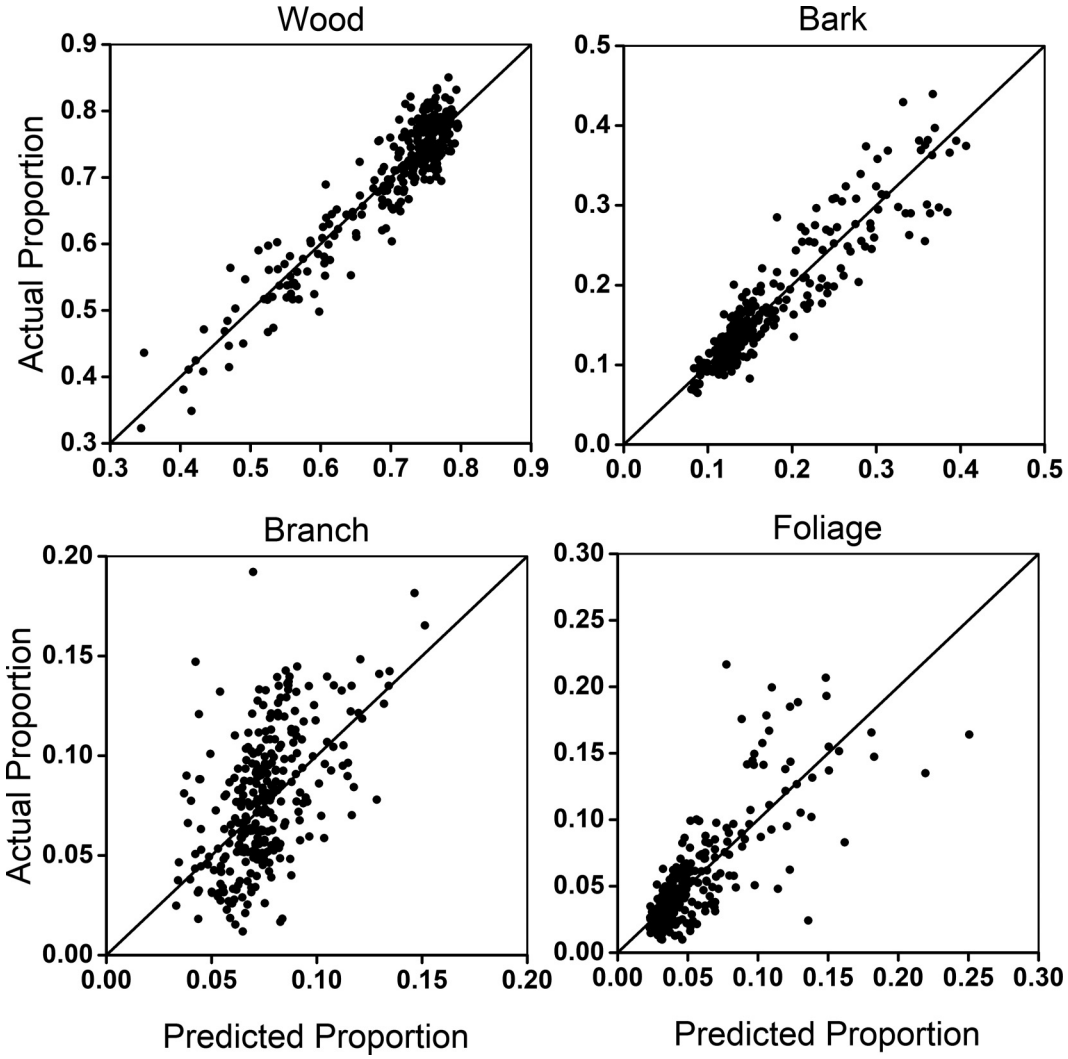
The total tree aboveground biomass equation was fitted with WNLS, and the estimated parameters are shown in Table 3.

(25)
$$\hat{y}_t = \hat{\beta}_0\text{DBH}^{\hat{\beta}_1}\text{HT}^{\hat{\beta}_2}$$

The proportions of stem wood, stem bark, branch, and foliage components in tree total aboveground biomass changed with increasing tree DBH and HT, with highly varied branch and foliage proportions (Fig. 2). The proportions were fitted to tree DBH and HT using the DRM. Both DBH and HT were significantly related to α_1 and α_2, only HT was significantly related to α_3, and neither was related to α_4 (Table 3).

Biomass component proportions were estimated by the following model:

Fig. 4. Comparisons of fractional components in tree total aboveground biomass with the predictions from the DRM.



$$
\begin{aligned}
\hat{\alpha}_0 &= 6.3962 + 3.0960\mathrm{DBH}^{-0.9406}\mathrm{HT}^{2.2182} \\
&\quad + 10.4228\mathrm{DBH}^{-1.1568}\mathrm{HT}^{1.4415} + 0.3483\mathrm{HT}^{1.2355} \\
\hat{p}_{\mathrm{wood}} &= 3.0960\mathrm{DBH}^{-0.9406}\mathrm{HT}^{2.2182}/\hat{\alpha}_0 \\
\hat{p}_{\mathrm{bark}} &= 10.4228\mathrm{DBH}^{-1.1568}\mathrm{HT}^{1.4415}/\hat{\alpha}_0 \\
\hat{p}_{\mathrm{branch}} &= 0.3483\mathrm{HT}^{1.2355}/\hat{\alpha}_0 \\
\hat{p}_{\mathrm{foliage}} &= 6.3962/\hat{\alpha}_0
\end{aligned} \tag{26}
$$

Due to the high variances in component proportions, R^2 values for predicting the component proportions were usually not large, e.g., about 0.5–0.7 for loblolly pine (*Pinus taeda* L.) (Zhao et al. 2016). In the current study, the Dirichlet model for slash pine component biomass proportions resulted in R^2 values of 0.878, 0.858, 0.280, and 0.634 for stem wood, stem bark, branch, and foliage component proportions, respectively. The DRM model with DBH and HT fitted well for slash pine component proportions, except for branch proportion (Fig. 4).

The predicted proportions were then applied to the estimated total aboveground biomass (eq. 25) to obtain the estimates of different component biomass:

$$
\begin{aligned}
&\text{Wood:} && \hat{y}_1 = \hat{y}_t \times \hat{p}_{\mathrm{wood}} \\
&\text{Bark:} && \hat{y}_2 = \hat{y}_t \times \hat{p}_{\mathrm{bark}} \\
&\text{Branch:} && \hat{y}_3 = \hat{y}_t \times \hat{p}_{\mathrm{branch}} \\
&\text{Foliage:} && \hat{y}_4 = \hat{y}_t \times \hat{p}_{\mathrm{foliage}}
\end{aligned} \tag{27}
$$

Comparison of modeling approaches

All systems slightly overestimated stem wood, stem bark, and total tree biomass (<4%) and largely overestimated branch and foliage biomass (16%–24%) (Table 4 and Fig. 5). For larger trees (DBH > 26 cm), it is obvious that the DRM predicted much more foliage biomass than the SUR2, and the SUR1 predicted the least foliage (Fig. 5).

Compared with the SUR2, the total biomass equation in the SUR1 was treated as component equations, specifying one additional variance function and additional constant cross-correlation parameters between total biomass and component equations. As mentioned above, the total tree biomass was obtained by adding component biomass together. Its variance should be a function of the biomass component variances and cross-correlations among component equations (eq. 6) (Affleck and Diéguez-Aranda 2016). Therefore, the residual variance of the total biomass estimated in the SUR1 could likely be biased. The analytical comparison of SUR1 and SUR2 in Appendix A also showed that the SUR2 should be more reasonable for estimating the aggregative models. With regard to biomass predictions, the SUR2 decreased E, MABE, and RMSE for stem wood, stem bark, branch, and total tree aboveground biomass; however, for foliage biomass, it increased E, MABE, and RMSE compared with the SUR1 models. The SUR1 and SUR2 models had a very close values of R^2 for predicting stem

Table 4. Statistics for predicting the component biomass of slash pine (*Pinus elliottii* var. *elliottii*) trees from the equations developed using the aggregative approach fitted with SUR1 and SUR2, the equations developed using the disaggregation strategy but fitted with DRM, previously published equations (GB14), and the newly fitted functions (GB_NEW) of Gonzalez-Benecke et al. (2014).

Method	Biomass	*E* (%)	MABE (%)	RMSE (%)	R^2
SUR1	Stem wood	−2.971	10.548	14.561	0.978
	Stem bark	−3.805	12.831	17.897	0.965
	Branch	−23.657	44.758	79.700	0.921
	Foliage	−18.685	39.726	65.541	0.849
	Total	−2.608	9.839	13.994	0.980
SUR2	Stem wood	−1.576	10.281	14.132	0.978
	Stem bark	−2.875	12.627	17.735	0.967
	Branch	−21.553	43.201	74.714	0.917
	Foliage	−20.103	40.342	67.313	0.819
	Total	−1.593	9.590	13.698	0.981
DRM	Stem wood	−0.736	11.042	14.831	0.977
	Stem bark	−3.017	13.360	18.275	0.966
	Branch	−16.688	41.999	71.142	0.914
	Foliage	−18.690	39.844	66.112	0.762
	Total	−0.404	10.318	14.259	0.981
GB14	Stem wood	9.400	17.549	24.469	0.973
	Stem bark	−0.631	29.057	35.709	0.853
	Branch	−28.177	50.499	94.467	0.901
	Foliage*	−113.353	117.815	200.849	—
	Total	5.746	14.410	20.128	0.980
GB_NEW	Stem wood	3.628	12.079	16.483	0.982
	Stem bark	1.194	15.299	19.448	0.959
	Branch	−137.213	150.617	295.590	0.934
	Foliage	−43.388	59.068	101.474	0.859
	Total	2.534	10.897	15.026	0.984

Note: *E*, mean error; MABE, mean absolute error; RMSE, root mean squared error.

*The original parameter estimates in slash pine foliage biomass model (F4) of Gonzalez-Benecke et al. (2014) were not correct. Here, the calculation was based on new parameters provided by Gonzalez-Benecke as requested, but the resultant R^2 still was not meaningful.

wood, stem bark, branch, and total tree aboveground biomass; the SUR2 had smaller R^2 for foliage biomass than the SUR1.

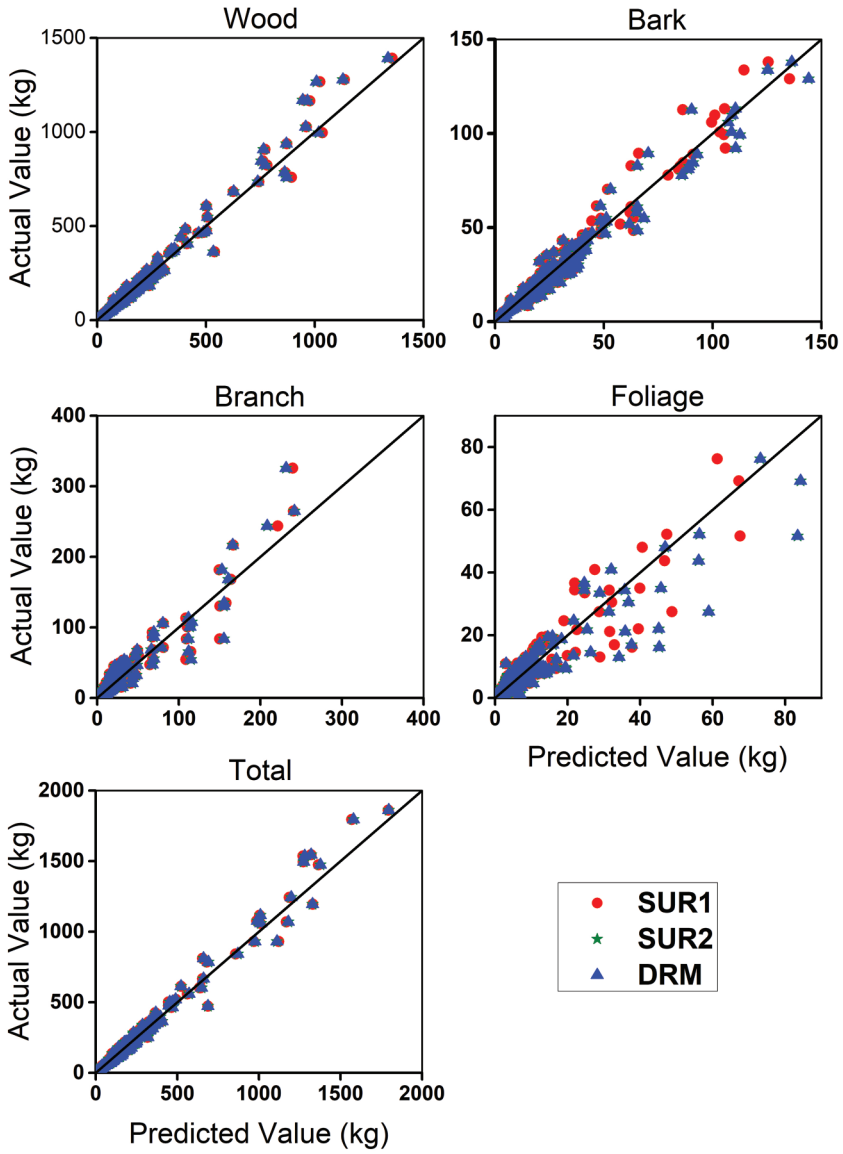
It is interesting that in the DRM system even though R^2 values for predicting component proportions were small (0.878, 0.858, 0.280, and 0.634 for stem wood, stem bark, branch, and foliage, respectively), R^2 values for predicting component biomass and total biomass were relatively large; even for predicting branch biomass, the R^2 value was 0.914 (Table 4). The DRM system produced a smaller R^2 value for foliage biomass compared with SUR1 and SUR2. In addition to biomass predictions, an advantage of the DRM approach is that it facilitates testing the effects of variables on biomass allocations (Zhao et al. 2016). In the current study, eq. 26 in the DRM system described how component proportions in tree total aboveground biomass changed with tree size (DBH and total tree height) for slash pines. Based on these proportion equations, we can estimate any ratio between two components or between one component and a subtotal. For example, the bark to stem over-bark biomass fraction (BFRAC) was estimated by $\hat{p}_{bark}/(\hat{p}_{bark} + \hat{p}_{wood}) = 1/(1 + 0.297\text{DBH}^{0.2162}\text{HT}^{0.7767})$. This equation provides more accurate BFRAC estimates than the BFARC equation of Gonzalez-Benecke et al. (2014): $R^2 = 0.91$ for the former and $R^2 = 0.84$ for the latter, when used in our biomass data. The BFARC was overestimated for larger trees and underestimated for smaller trees by the equation of Gonzalez-Benecke et al. (2014).

Total biomass, stem wood biomass, and stem bark biomass can usually be better predicted than crown biomass components (Bi et al. 2010; Zhao et al. 2015; Wang et al. 2018; Dong et al. 2018). Some crown size measurements could improve crown biomass prediction (Zhao et al. 2015; Dong et al. 2018). In the current study, our three systems were based on tree DBH and total height, which still predicted crown biomass well for slash pine, with $R^2 > 0.91$ for branch biomass and $R^2 > 0.76$ for foliage biomass (Table 4).

Based on *E*, MABE, RMSE, and R^2 for each biomass component and total biomass, the three systems of biomass equations (SUR1, SUR2, and DRM) could be ranked. The SUR1 was best for foliage estimation and DRM was best for branch estimation, while the SUR2 was best for stem wood, stem bark, and total tree aboveground biomass. The overall ranking of the three systems followed the order of SUR2 > SUR1 > DRM, and their sum of the ranks was 33, 40, and 41, respectively. It is surprising that SUR2 was actually better for predicting total biomass than SUR1 even though SUR1 actually uses total biomass as a dependent variable. This result tends to confirm Affleck and Diéguez-Aranda's perspective (Affleck and Diéguez-Aranda 2016). Our empirical findings would reduce the motivation for doing SUR1 if it were to be true on other datasets as well. Our analytical comparison also demonstrated that there appears to be no particular reason to prefer SUR1 over SUR2 (see Appendix A). Poudel and Temesgen (2016) used the SUR1 and DRM methods to develop biomass equations for Douglas-fir and lodgepole pine trees. They found that the DRM was superior to the SUR1, except for Douglas-fir branch biomass, for which the SUR1 was better than the DRM. However, our results indicated that the SUR1 and DRM had almost the same performance for stem wood, stem bark, and total biomass, and the SUR1 was better for foliage biomass, while the DRM was better for branch biomass (Table 4).

With the biomass data from this study, the previously published DBH–height based biomass equations of Gonzalez-Benecke et al. (2014) overestimated stem bark and foliage biomass (Fig. 6), even after using the correct coefficients provided by Gonzalez-Benecke (personal communication; the published parameters were not correct). Overestimated foliage biomass resulted in a meaningless (negative) R^2 in Table 4. The estimated foliage biomass for some large trees using their modified model did not make sense — having the estimation of 100–218 kg of foliage biomass (Fig. 6), more than the observed maximum value of 76.2 kg (Table 1). The data used in the study of Gonzalez-Benecke et al. (2014) had relatively narrow DBH and HT ranges (1.3–32.6 cm and 1.5–22.9 m, respectively) and small DBH and HT averages (9.9 cm and 8.1 m, respectively), while the data used in fitting the equation systems in this study had wider DBH and HT ranges (3.0–53.3 cm and 2.9–30.2 m, respectively) with larger average DBH (18.4 cm) and average HT (16.7 m). The biomass functions of Gonzalez-Benecke et al. (2014) were refit to our data used in this study, and parameter estimates are reported in Table B1 (see Appendix B). These newly fitted equations underestimated stem wood, stem bark, and total tree biomass and more largely overestimated branch and foliage biomass than SUR1, SUR2, and DRM. Furthermore, as mentioned earlier, the equations of Gonzalez-Benecke et al. (2014) were developed separately and thus did not have the additivity among tree biomass components and total biomass. For example, when their previously published equations were used for our biomass data, the difference of total biomass obtained by adding the estimated biomass components together and by the total biomass equation ranged from −96.2% to 2.9%, and on average, the former gave 5.3% more total biomass than the latter. Even for the newly fitted equations of Gonzalez-Benecke et al. (2014), the sum of the estimated biomass components still gave 7.4% more total biomass on average than that estimated from the total biomass equation (the range of the difference from −124.6% to 2.8%). In the current study, however, our three systems guarantee the additivity of the estimated biomass components and total. Overall, graphical examinations (Figs. 5 and 6) and the comparisons of the summary statistics (Table 4) suggested that all three new systems were better than the equations of Gonzalez-Benecke et al. (2014) for biomass predictions.

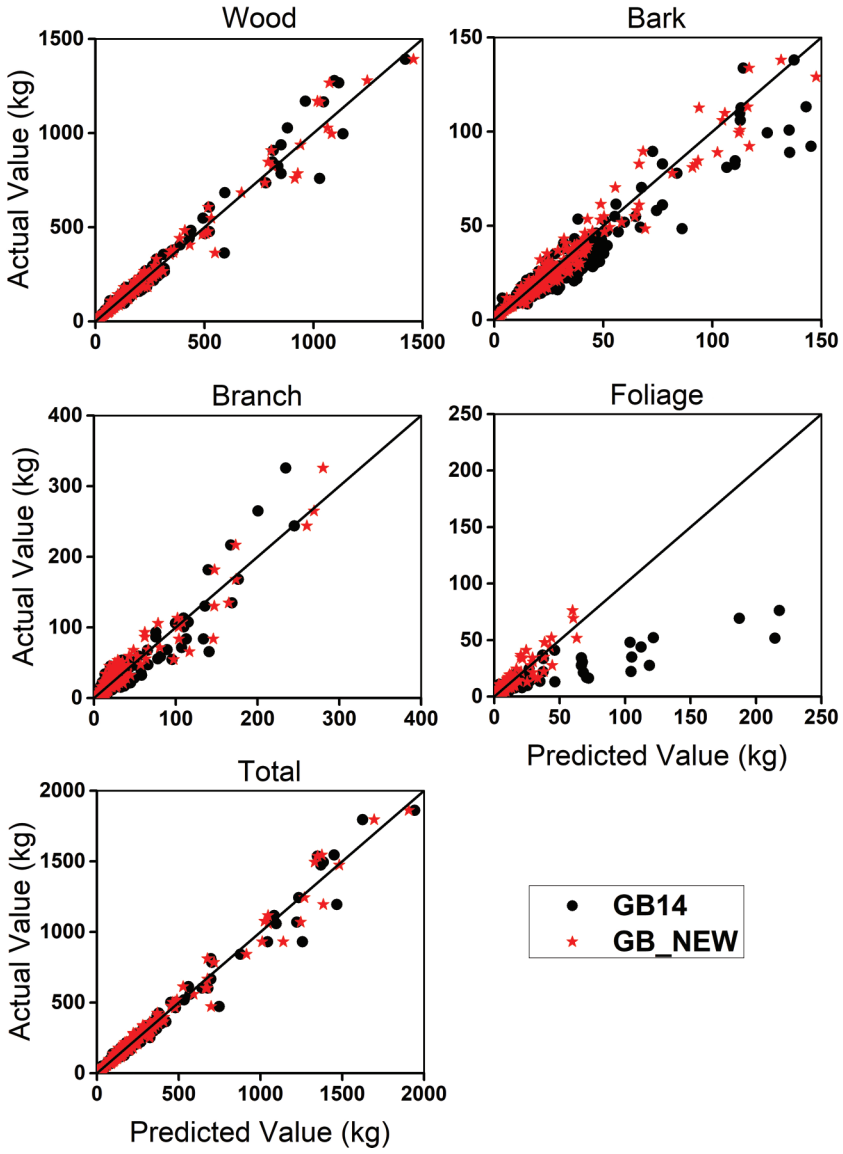
Fig. 5. Comparisons of stem wood, stem bark, branch, foliage, and total tree aboveground biomass predictions from the systems associated with SUR1, SUR2, and DRM.



All of the new systems of tree biomass equations are based on tree DBH and HT only, because in our legacy biomass dataset, there was no information about tree crown size measurements or stand-level factors such as age, stand density, and site quality. In addition to tree DBH and HT, including crown length, crown width, and diameter at the base of live crown could further improve crown biomass predication (Zhao et al. 2015), and incorporating stand-level factors may also improve predictive performance for other biomass components (Gonzalez-Benecke et al. 2018). Depending on data availability, we will integrate these factors into the systems of biomass equations to improve their predictive performance. It should be noted that the main purpose of this work is to compare three modeling approaches that ensure the additivity of nonlinear component biomass equations.

With the same weighting functions used in NSUR, aggregative equations of two systems were also fitted using weighted nonlinear ordinary least squares (OLS) estimation, and parameter estimates are given in Table B2 (in Appendix B). Our results clearly showed that NSUR led to a reduction in the standard error of estimates (Tables 2 and B2). This finding confirmed that the NSUR approach should achieve more efficient estimates compared with OLS in the presence of correlations between error terms of biomass equations (Binkley and Nelson 1988). Nord-Larsen et al. (2017), however, found that NSUR did not get lower standard error of estimates than OLS, which was contrary to our expectations. Their finding might result in the limited biomass data (from 679 trees including 13 different species) and computation. The NSUR method requires estimation of the cross-equation covariance matrix from the OLS residuals, which increases the variability of the estimation for small sample size. To realize the efficiency gain that NSRU has over OLS, you must have a reasonable amount of data (SAS Institute Inc. 2011). For the system of component equations only, estimating model parameters using SUR2 little improved prediction of all total and component biomass compared with OLS estimation (Tables 4 and B3). Estimating parameters of the system of component and total equations, SUR1 little increased prediction bias for stem wood, stem bark, and total biomass but largely decreased prediction bias for branch and foliage biomass. In summary, our results do not support the findings of Nord-Larsen et al. (2017) but do confirm that it is reasonable to prefer NSUR, especially SUR2 over OLS.

Fig. 6. Stem wood, stem bark, branch, foliage, and total tree aboveground biomass predictions from DBH- and height-based biomass equations developed by Gonzalez-Benecke et al. (2014) (GB14) and their functions refit to new data (GB_NEW).



Conclusions

We developed three systems of additive biomass equations for slash pine trees with a wide range of tree sizes and across a wide proportion of the species' geographic range in the US. The aggregative models were estimated either by jointly fitting all $M + 1$ biomass equations (M component biomass plus total biomass) with weighted NSUR (SUR1) or by jointly fitting M component biomass equations with weighted NSUR (SUR2). Due to different assumptions about the variance of the total biomass and cross-correlations between the total biomass and component equations, the SUR1 and SUR2 resulted in two aggregative systems, even though they had the same model form. In an alternative disaggregation strategy (DRM), component biomass proportions in tree aboveground biomass were modeled with the Dirichlet regression approach and the estimated total biomass was disaggregated into biomass components based on their estimated proportions. With regard to biomass predictions, there was no single system to predict biomass that was best for all components and total tree biomass. The overall ranking of the three systems followed the order of SUR2 > SUR1 > DRM based on numerous fit statistics. All three systems provided more accurate biomass predictions than those previously available in the literature. Whether the results of the three approaches are commonly true for other datasets such as hardwood species is worthwhile to study further.

Acknowledgements

This research was funded by the USDA Forest Service (Agreement 16-JV-11330145-075). This work was also partially supported by the USDA National Institute of Food and Agriculture, McIntire Stennis Project OKL0 2843, and the Division of Agricultural Sciences and Natural Resources at Oklahoma State University. We thank the Plantation Management Research Cooperative (PMRC) technical staff for their hard work in field sampling and data collection and the PMRC members for permitting access and destructive sampling on their lands.

References

Affleck, D.L.R., and Diéguez-Aranda, U. 2016. Additive nonlinear biomass equations: a likelihood-based approach. For. Sci. **62**(2): 129–140. doi:10.5849/forsci.15-126.

Bailey, R.L., and Fang, Z. 2000. Merchantable-stem green and dry weight prediction equations based on a new segmented-stem taper model. The University of Georgia, Athens, Ga., Plantation Management Research Cooperative (PMRC) Technical Report 2000-8.

Barnett, J.P., and Sheffield, R.M. 2005. Slash pine: characteristics, history, status and trends. *In* Slash Pine: Still Growing and Growing! Proceedings of the Slash Pine Symposium. *Edited by* E.D. Dickens, J.P. Barnett, W.G. Hubbard, and E.J. Jokela. USDA Forest Service, Southern Research Station, Asheville, N.C., Gen. Tech. Rep. SRS-76. pp. 1–6.

Bi, H., Turner, J., and Lambert, M.J. 2004. Additive biomass equations for native eucalypt forest trees of temperate Australia. Trees, **18**: 467–479. doi:10.1007/s00468-004-0333-z.

Bi, H., Long, Y., Turner, J., Lei, Y., Snowdon, P., Li, Y., Harper, R., Zerihun, A., and Ximenes, F. 2010. Additive prediction of aboveground biomass for *Pinus radiata* (D. Don) plantations. For. Ecol. Manage. **259**: 2301–2314. doi:10.1016/j.foreco.2010.03.003.

Binkley, J.K., and Nelson, C.H. 1988. A note on the efficiency of seemingly unrelated regression. Am. Stat. **42**(2): 137–139. doi:10.1080/00031305.1988.10475545.

Dong, L., Zhang, L., and Li, F. 2015. A three-step proportional weighting system of nonlinear biomass equations. For. Sci. **61**(1): 35–45. doi:10.5849/forsci.13-193.

Dong, L., Zhang, L., and Li, F. 2018. Additive biomass equations based on different dendrometric variables for two dominant species (*Larix gmelini* Rupr. and *Betula platyphylla* Suk.) in natural forests in the eastern Daxing'an Mountains, Northeast China. Forests, **9**: 261. doi:10.3390/f9050261.

Gonzalez-Benecke, C.A., Gezan, S.A., Albaugh, T.J., Allen, H.L., Burkhart, H.E., Fox, T.R., Jokela, E.J., Maier, C.A., Martin, T.A., Rubilar, R.A., and Samuelson, L.J. 2014. Local and general above-stump biomass functions for loblolly pine and slash pine trees. For. Ecol. Manage. **334**: 254–276. doi:10.1016/j.foreco.2014.09.002.

Gonzalez-Benecke, C.A., Zhao, D., Samuelson, L.J., Martin, T.A., Leduc, D.J., and Jack, S.B. 2018. Local and general above-ground biomass functions for *Pinus palustris* trees. Forests, **9**: 310. doi:10.3390/f9060310.

Hijazi, R.H., and Jernigan, R.W. 2009. Modeling compositional data using Dirichlet regression models. J. Appl. Probab. Stat. **4**(1): 77–91.

Kozak, A. 1970. Methods for ensuring additivity of biomass components by regression analysis. For. Chron. **46**: 402–405. doi:10.5558/tfc46402-5.

Kozak, A., and Kozak, R. 2003. Does cross validation provide additional information in the evaluation of regression models? Can. J. For. Res. **33**(6): 976–987. doi:10.1139/x03-022.

Maier, M.J. 2014. DirichletReg: Dirichlet regression for compositional data in R. WU Vienna University of Economics and Business, Vienna, Research Report Series/Department of Statistics and Mathematics, 125.

Nord-Larsen, T., Meilby, H., and Skovsgaard, J.P. 2017. Simultaneous estimation of biomass models for 13 tree species: effects of compatible additivity requirements. Can. J. For. Res. **47**(6): 765–776. doi:10.1139/cjfr-2016-0430.

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.

Parresol, B.R., and Thomas, C.E. 1989. A density-integral approach to estimating stem biomass. For. Ecol. Manage. **26**: 285–297. doi:10.1016/0378-1127(89)90089-3.

Poudel, K.P., and Temesgen, H. 2016. 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.

SAS Institute Inc. 2011. SAS/ETS® 9.3 User's Guide. SAS Institute Inc., Cary, N.C.

Tang, S., and Wang, Y. 2002. A parameter estimation program for the error-in-variable model. Ecol. Modell. **156**: 225–236. doi:10.1016/S0304-3800(02)00173-4.

Tang, S., Zhang, H., and Xu, H. 2000. Study on establish and estimate method of compatible biomass model. Sci. Silvae Sin. **36**(Suppl. 1): 19–27. [In Chinese with English abstract.]

Tang, S., Li, Y., and Wang, Y. 2001. Simultaneous equations, error-in-variable models, and model integration in systems ecology. Ecol. Modell. **142**: 285–294. doi:10.1016/S0304-3800(01)00326-X.

Wang, X., Zhao, D., Liu, G., Yang, C., and Teskey, R.O. 2018. Additive tree biomass equations for *Betula platyphylla* Suk. plantations in northeast China. Ann. For. Sci. **75**: 60. doi:10.1007/s13595-018-0738-2.

Zhao, D., Kane, M., Markewitz, D., Teskey, R., and Clutter, M. 2015. Additive tree biomass equations for midrotation loblolly pine plantations. For. Sci. **61**: 613–623. doi:10.5849/forsci.14-193.

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**(5): 463–473. doi:10.5849/forsci.15-129.

Appendix A

An analytical comparison of the SUR1 and SUR2 approaches

SUR2 estimates model parameters by minimizing

$$S_2 = \sum_{m=1}^{M}\sum_{i=1}^{N}\frac{e_{mi}^2}{\psi_{mi}} + \sum_{m}^{M}\sum_{m'\neq m}^{M}\sum_{i}^{N}\frac{e_{mi}e_{m'i}}{\sqrt{\psi_{mi}\psi_{m'i}}} \tag{A1}$$

where $e_{mi} = y_{mi} - f_m(\boldsymbol{X}_{mi}, \boldsymbol{\beta}_m)$, and ψ_{mi} is the weighting function for component m ($m = 1, 2, \ldots, M$) evaluated for observation i ($i = 1, 2, \ldots, N$).

SUR1 estimates model parameters by minimizing

$$S_1 = S_2 + \sum_{i=1}^{N}\frac{e_{Ti}^2}{\psi_{Ti}} + \sum_{m}^{M}\sum_{i}^{N}\frac{e_{mi}e_{Ti}}{\sqrt{\psi_{mi}\psi_{Ti}}} \tag{A2}$$

where $e_{Ti} = y_{Ti} - \sum_{m}^{M} f_m(\boldsymbol{X}_{mi}, \boldsymbol{\beta}_m)$, and ψ_{Ti} is the weighting function for the total biomass evaluated for observation i ($i = 1, 2, \ldots, N$).

According to the definition $e_{Ti} = \sum_{m}^{M} e_{mi}$, then $e_{Ti}^2 = \sum_{m=1}^{M} e_{mi}^2 + \sum_{m}\sum_{m'\neq m} e_{mi}e_{m'i}$ and $e_{Ti}e_{mi} = e_{mi}^2 + \sum_{m'\neq m}^{M} e_{mi}e_{m'i}$. S_1 can be written as

$$\begin{aligned} S_1 = {} & \sum_{m=1}^{M}\sum_{i=1}^{N} e_{mi}^2\left(\frac{1}{\psi_{mi}} + \frac{1}{\psi_{Ti}} + \frac{1}{\sqrt{\psi_{mi}\psi_{Ti}}}\right) \\ & + \sum_{m=1}^{M}\sum_{m'\neq m}^{M}\sum_{i=1}^{N} e_{mi}e_{m'i}\left(\frac{1}{\sqrt{\psi_{mi}\psi_{m'i}}} + \frac{1}{\psi_{Ti}} + \frac{1}{\sqrt{\psi_{mi}\psi_{Ti}}}\right) \end{aligned} \tag{A3}$$

From S_1, it is evident that SUR1 estimates could be obtained by a SUR2 procedure (that does not use the total biomass as a dependent variable), but with different (more complex) weighting functions. If the weighting functions used in S_2 do make sense to appropriately describe the variance and covariance structure of the comments, then the use of the more complex weights in S_1 must be less efficient, and so too must SUR1, as pointed out by a reviewer.

An analytical comparison of the OLS1 and OLS2 approaches

OLS2 approach: assuming no cross-equation correlations, then jointly fitting M component equations using weighted nonlinear least squares (OLS) estimation is to minimize

$$S_{OLS2} = \sum_{m=1}^{M}\sum_{i=1}^{N}\frac{e_{mi}^2}{\psi_{mi}} \tag{A4}$$

OLS1: assuming no cross-equation correlations, then jointly fitting $M + 1$ equations (M components plus the total) using weighted nonlinear least squares (OLS) estimation is to minimize

$$S_{OLS1} = S_{OLS2} + \sum_{i=1}^{N}\frac{e_{Ti}^2}{\psi_{Ti}} = \sum_{m=1}^{M}\sum_{i=1}^{N}\frac{e_{mi}^2}{\psi_{mi}} + \sum_{i=1}^{N}\frac{e_{Ti}^2}{\psi_{Ti}} \tag{A5}$$

Owing to $e_{Ti} = \sum_{m}^{M} e_{mi}$ and $e_{Ti}^2 = \sum_{m=1}^{M} e_{mi}^2 + \sum_{m}\sum_{m'\neq m} e_{mi}e_{m'i}$, then S_{OLS1} can be written as

$$S_{OLS1} = \sum_{m=1}^{M}\sum_{i=1}^{N} e_{mi}^2\left(\frac{1}{\psi_{mi}} + \frac{1}{\psi_{Ti}}\right) + \sum_{m=1}^{M}\sum_{m'\neq m}^{M}\sum_{i=1}^{N} e_{mi}e_{m'i}\left(\frac{1}{\psi_{Ti}}\right) \tag{A6}$$

It is evident that OLS1 estimates could be obtained using component biomass only as dependent variables, not necessarily using the total biomass as a dependent variable.

Appendix B

Table B1. Parameter estimates and their standard errors (SE) and *p* values for non-additive biomass equations of Gonzalez-Benecke et al. (2014) refitted to the same data set used in fitting additive equation systems in this study.

Biomass	Model	Parameter	Estimate	SE	*p* value
STEM	$= d_1 \cdot \mathrm{DBH}^{d_2} \cdot \mathrm{HT}^{d_3}$	$\hat{d}_1$	0.01788	0.00259	<0.0001
		$\hat{d}_2$	2.14514	0.02505	<0.0001
		$\hat{d}_3$	0.86523	0.05806	<0.0001
BFRAC	$= e^{[d_1 + d_2 \cdot \ln(\mathrm{DBH}^2 \cdot \mathrm{HT})]}$	$\hat{d}_1$	0.10972	0.02927	0.0002
		$\hat{d}_2$	−0.22138	0.00422	<0.0001
BRANCH	$= d_1 \cdot \mathrm{DBH}^{d_2} \cdot e^{d_3 \cdot \mathrm{DBH}} \cdot \mathrm{HT}^{d_4}$	$\hat{d}_1$	0.18365	0.14139	0.1950
		$\hat{d}_2$	1.85375	0.33436	<0.0001
		$\hat{d}_3$	0.04258	0.00906	<0.0001
		$\hat{d}_4$	−0.70825	0.13377	<0.0001
FOLIAGE	$= d_1 \cdot \mathrm{DBH}^{d_2} \cdot e^{d_3 \cdot DBH} \cdot \mathrm{HT}^{d_4}$	$\hat{d}_1$	0.13329	0.09308	0.15318
		$\hat{d}_2$	1.77433	0.33873	<0.0001
		$\hat{d}_3$	0.02340	0.00954	0.0147
		$\hat{d}_4$	−0.66092	0.16849	0.0001
TASB	$= d_1 \cdot \mathrm{DBH}^{d_2} \cdot \mathrm{HT}^{d_3}$	$\hat{d}_1$	0.02879	0.00422	<0.0001
		$\hat{d}_2$	2.28976	0.02562	<0.0001
		$\hat{d}_3$	0.60060	0.05833	<0.0001

Note: DBH, diameter at breast height; HT, total height; STEM, above stump stem over-bark biomass (stem wood + stem bark biomass); BFRAC, bark to stem over-bark biomass fraction; BRANCH, total branch biomass; FOLIAGE, total needle biomass; TASB, total above-stump biomass.

Table B2. Parameter estimates and their standard errors (SE) and *p* values for aggregative biomass equations fitted with weighted nonlinear ordinary least squares (OLS) estimation.

			OLS1			OLS2		
Biomass component	Variable	Parameter	Estimate	SE	*p* value	Estimate	SE	*p* value
Stem wood		$\hat{\beta}_{10}$	0.0108	0.0007	<0.0001	0.0122	0.0007	<0.0001
	DBH	$\hat{\beta}_{11}$	2.0768	0.0259	<0.0001	2.0524	0.0321	<0.0001
	HT	$\hat{\beta}_{12}$	1.0572	0.0374	<0.0001	1.0395	0.0411	<0.0001
Stem bark		$\hat{\beta}_{20}$	0.0425	0.0026	<0.0001	0.0438	0.0031	<0.0001
	DBH	$\hat{\beta}_{21}$	1.7825	0.0402	<0.0001	1.7902	0.0438	<0.0001
	HT	$\hat{\beta}_{22}$	0.3012	0.0499	<0.0001	0.2825	0.0552	<0.0001
Branch		$\hat{\beta}_{30}$	0.0054	0.0006	<0.0001	0.0016	0.0002	<0.0001
	DBH	$\hat{\beta}_{31}$	2.0963	0.1254	<0.0001	3.0087	0.0449	<0.0001
	HT	$\hat{\beta}_{32}$	0.4925	0.1443	0.0007			
Foliage		$\hat{\beta}_{40}$	0.0332	0.0056	<0.0001	0.0322	0.0052	<0.0001
	DBH	$\hat{\beta}_{41}$	2.8501	0.0980	<0.0001	2.8992	0.1042	<0.0001
	HT	$\hat{\beta}_{42}$	−1.1327	0.1227	<0.0001	−1.1716	0.1295	<0.0001

Note: OLS1, OLS used to fit all total and component biomass equations, and compared with SUR1; OLS2, OLS used to fit component equations only, and compared with SUR2. DBH, diameter at breast height; HT, total height.

Table B3. Statistics for predicting the component biomass of slash pine (*Pinus elliottii* var. *elliottii*) trees from the primary functions in additive systems developed with the aggregative approach in this study but were fitted with OLS.

Method	Biomass	*E* (%)	MABE (%)	RMSE (%)	R^2
OLS1	Stem wood	−0.607	10.374	14.031	0.980
	Stem bark	−2.743	12.596	17.591	0.966
	Branch	−33.485	58.234	99.328	0.632
	Foliage	−20.529	40.633	67.928	0.833
	Total	−1.191	9.676	13.630	0.973
OLS2	Stem wood	−1.608	10.281	14.141	0.977
	Stem bark	−2.887	12.612	17.728	0.966
	Branch	−22.0164	43.460	75.307	0.916
	Foliage	−20.427	40.549	67.831	0.820
	Total	−1.676	9.608	13.742	0.981

Note: OLS1, OLS used to fit all total and component biomass equations, and compared with SUR1; OLS2, OLS used to fit component equations only, and compared with SUR2. *E*, mean error; MABE, mean absolute error; RMSE, root mean squared error.