
A single-tree compatible biomass model of *Quercus variabilis* Blume forests in North China*

Conghui Zheng · Euan G. Mason · Liming Jia · Songpo Wei · Caowen Sun · Jie Duan

Trees: Structure and Function. 01/2015

Abstract

Key message Fitting and comparing three sets of compatible biomass models for prediction of biomass or carbon stocks of natural and planted *Quercus variabilis* Blume forests.

Abstract To make the sum of estimated values from biomass models of various components of a tree equal to estimated tree total biomass for *Quercus variabilis* Blume (cork oak) forests in North China, single-tree compatible biomass models were developed. 100 trees from 100 plots in North China were felled to obtain biomass of aboveground components, and roots of 19 of those trees were extracted for measurement of root biomass. After Box-Cox transformations of variables, two sets of independent component biomass models with a dummy variable in models to define stand origin were separately built using linear mixed effects analyses (one set of models with site as a random factor; the other set without any random factor). Then three methods were compared to force compatibility of the biomass models: sums of linear mixed effects models, sums of linear models, and simultaneous equation fits based on linear models. Model parameters were estimated by ordinary least squares (OLS) or seeming unrelated regression procedures (SUR). Coefficients of determination (r^2), root mean square error (*Rmse*), residuals plots and histograms of residuals indicated that models fitted with sums of linear mixed effects models were the least biased and most precise at estimating total above-ground biomass. Further testing for the linear mixed effects models with jackknife validation and prediction sum of squares (*PRESS*) statistics indicated that the compatible biomass models can be used to predict biomass or carbon stocks of cork oak forests in North China within specific tree diameter at breast height and height ranges.

Keywords *Quercus variabilis* Blume · dummy variable · Box-Cox transformation · linear mixed effects models · seeming unrelated regression procedures · jackknife validation

Introduction

Forests play an important role in the mitigation of global climate change by sequestering CO₂. Estimating forest biomass with models facilitates indepth studies of ecosystem energy flows and carbon cycling. Monitoring of forest biomass and carbon stocks and establishment of biomass models suitable for larger areas are therefore increasingly important (Case and Hall, 2008; Gurdak et al., 2014; Jenkins et al., 2003; Muukkonen, 2007; Návar, 2009; Ritchie et al., 2013; Snorrason and Einarsson, 2006; Vallet et al., 2006). Both linear and nonlinear models are commonly used for biomass modelling

* C. Zheng and E. G. Mason have equally contributed to this work and should be regarded as co-first authors.

C. Zheng · L. Jia · S. Wei · C. Sun · J. Duan

Ministry of Education Key Laboratory of Silviculture and Conservation, Beijing Forestry University, Beijing, China

E. G. Mason

School of Forestry, University of Canterbury, Christchurch, New Zealand

L. Jia

Corresponding author e-mail: jlj@bjfu.edu.cn tel.: +86 10 62337055 fax: +86 10 62337055

(Durkaya et al., 2009; Sajdak et al., 2014; Schmidt et al., 2009; Torres Vélez and Del Valle, 2007; Zeng et al., 2011; Zianis et al., 2004). Moreover, compatible biomass modelling systems have been proposed and applied to ensure compatibility between total biomass models and component biomass models. Researchers have developed two principal ways of creating compatible biomass models: (1) the total biomass regression function is defined as the sum of separately calculated best regression functions of component biomasses; (2) joint-generalized least squares regression, like seemingly unrelated regressions (SUR), two-stage least-squares methods (2SLS), three-stage least-squares methods (3SLS) and generalized method of moments (GMM) (António et al., 2007; Balboa-Murias et al., 2006; Bi et al., 2010; Brandeis et al., 2006; Castedo-Dorado et al., 2012; Gómez-García et al., 2013; Gómez-García et al., 2014; Hasenauer et al., 1998; Carvalho and Parresol, 2003; Li and Zhao, 2013; Menéndez-Miguélez et al., 2013; Parresol, 1999, 2001; Ruiz-Peinado et al., 2011; Shaiek et al., 2011).

Oak forest accounts for 15.89% of forest stand area and 13.30% of forest stand volume in China (Luo et al., 2009). *Quercus variabilis* Blume (Cork oak) is used for timber, cork, soil and water conservation, and is mainly distributed in the areas of Taihang Mountain and Yan Mountain in North China. A large number of biomass models of oak have been created in Europe, North America, Portugal, Spain, Greece, and China (Bao et al., 1984; Canadell et al., 1988; Zianis et al., 2005; Ter-Michaelian et al., 1997; Paulo and Tomé, 2010; Fonseca and Parresol, 2001; Tian et al., 1997; Zianis et al., 2011). In most models tree height (H), diameter at breast height (DBH) or their combination (d^2h) are independent variables. Until now, *Quercus variabilis* Blume compatible biomass models in North China have not been developed. The objective of the study described here was therefore to establish and compare three sets of compatible biomass models applied to natural and planted *Quercus variabilis* Blume forests in North China: (1) A total biomass regression function defined as the sum of separately calculated best linear mixed effects models of the component biomasses (sums of linear mixed effects models); (2) A total biomass regression function defined as the sum of separately calculated best linear models of the component biomasses (sums of linear models); and (3) simultaneous equations of linear models of biomass based on seemingly unrelated regressions (SUR).

Methods

Measurements

100 cork oak plots in 13 stands with different site conditions and age distributions were created in the following locations: Gao-Luo forestry station, Qi-Jiahe forestry station, Bei-Tan forestry station and Heng-He forestry station of Zhong-tiaoshan region in Shanxi province, collective forests of Da-Geliao village in Xingtai city of Hebei province, Si-Zuolou forestry station and Xi-Shan forestry station in Beijing. 67 plots in naturally regenerated forest and 33 plots in plantations were included and the area of each plot was 20m × 20m (0.04ha). Measured variables were as follows: (1) Single tree variables including diameter at breast height, tree height, crown width, crown length; (2) Biomass data of various components obtained by fully weighing individual tree components in the mentioned plots including aboveground biomass of each component of 100 trees closest to the average within each plot (67 naturally regenerated trees and 33 planted trees) and root biomass of 19 of those 100 trees (12 naturally regenerated trees and 7 planted trees). Additionally, another 13 sets of root biomass data can be found in two related reports (Bao et al., 1984; Tian et al., 1997). For the biomass data, subsamples were also selected and weighed to get fresh mass in the field. Then all the subsamples were oven-dried at 85 °C in a laboratory until a constant mass was reached. Using the ratio of dry mass to fresh mass, biomass of

each component was computed including stem woody biomass, stem bark biomass, branch biomass, foliage biomass, total stem biomass (including bark), total crown biomass and total aboveground biomass for all the 100 trees. For 32 trees, root biomass and total tree biomass were obtained as well. Summaries of plots and weighed trees are shown in Table 1.

The following notations will be used hereafter. Other notations in some particular equations will be listed with those equations:

DBH is diameter at breast height over bark (1.3 m above the ground, cm)

H is total tree height (m)

d^2h is equal to $DBH * DBH * H$

Wi (model) is biomass (kg) for each component (i can be wood, bark, branch, foliage, root, stem, crown, total aboveground and total tree)

tWi (model) is transformed biomass for each component by the Box-Cox method (i can be wood, bark, branch, foliage, root, stem, crown, total aboveground and total tree)

p -value in the following tables is probability of type I error in Shapiro Wilks test.

“lme” is a linear mixed effects model

“lm” is a linear model

“sums of lme” is a method of sums of linear mixed effects model to force additivity of biomass models

“sums of lm” is a method of sums of linear model to force additivity of biomass models

SUR is a method of seemingly unrelated regressions to force additivity of biomass models

Table 1 Summary of plots and weighed trees

Items	Natural forests (67 trees)		Plantations (33 trees)	
	Type or interval	Number of trees	Type or interval	Number of trees
Site type	Sunny slope with medium or thick soil in the low-elevation area	25	Sunny slope with medium or thick soil in the low-elevation hilly	9
	Sunny slope with thin soil in the low-elevation area	9	Sunny slope with thin soil in the low-elevation hilly	11
	Shady slope with medium or thick soil in the low-elevation area	18	Shady slope with medium or thick soil in the low-elevation hilly	10
	Shady slope with thin soil in the low-elevation area	7	Shady slope with thin soil in the low-elevation hilly	3
	Flat ground with thick soil in the low-elevation area	3		
	Sunny slope with medium or thick soil in the low-and middle- elevation area	1		
	Shady slope with thin soil in the low-and middle- elevation area	4		
Age group (years)	Small age-aged stands (<40)	38	Small age-aged stands (<20)	4
	Middle-aged stands (41-60)	26	Middle-aged stands (21-40)	20
	Nearly mature stands (61-80)	2	Nearly mature stands (41-50)	9
	Mature stands (>80)	1		
DBH	[5,7)	1	[3,5)	2

class	[7,9)	8	[5,7)	2
(cm)	[9,11)	19	[7,9)	2
	[11,13)	20	[9,11)	3
	[13,15)	7	[11,13)	14
	[15,17)	5	[13,15)	5
	[17,19)	1	[15,17)	5
	[19,21)	2		
	[21,23)	2		
	[23,25)	2		

Independent biomass model construction for each component

Variable transformations: Before modeling, some explanatory and response variables (V) were transformed to new variables (tV) by Box-Cox transformation to make frequency distributions of those variables (tV) as close to normal distributions as possible. Equation (1) expresses a Box-Cox transformation (Sakia, 1992):

$$tV = \begin{cases} \frac{V^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log_e V & \lambda = 0 \end{cases} \quad (1)$$

Where, V was the original response or explanation variable (e.g., D , H , d^2h and Wi), tV was the response or explanatory variable after Box-Cox transformation (e.g., tD , tH , td^2h and tWi), λ was the parameter in the Box-Cox transformation.

In this study, only variables in bark biomass models were transformed by the logarithmic transformation in above formula, while variables in other models were transformed by Box-Cox transformations with a value of λ not equal to 0.

Model structure and selection: Linear mixed effects models (with site as a random factor) and linear models were used to establish independent component biomass models. Variables tD , tH , td^2h , D_i were chosen as possible explanatory variables where dummy variable D_i was introduced in a model to define forest origin ($D_i=1$ for natural forests; $D_i=0$ for plantation). Variables tWi were response variables. Model forms were (Peng and Lu, 2012; Su et al., 2012):

$$Y = X\beta + Zu + e \quad (2)$$

$$Y = X\beta + e \quad (3)$$

Where, formula (2) is a linear mixed effects model; formula (3) is a linear model ; Y is the vector of the response variable (e.g., tWi); X is the vector of fixed-effect regressors (e.g., tD , tH , td^2h , D_i); Z is the vector of random-effect regressors (e.g., site); β is the vector of fixed effect coefficients (e.g., a, b, c, e); u is the vector of the random- effect coefficients; e is the vector of errors.

For variable selection of independent component models, an overall merit-based method was used. All the regression equations in different combinations of tD , tH , td^2h , D_i were compared by examining coefficients of determination (r^2) and root mean square errors ($Rmse$). Residual distribution homogeneity and model bias were visually checked by residual plots with loess regression lines overlaid on the plots. For an unbiased model, a loess line should be flat and located at the zero value on

the vertical axis in the residual plot (Jacoby, 2000). Normality of residuals was checked with histograms of residuals and by using a Shapiro Wilk test (probabilities of type I error or *p values* below 0.05 indicate a departure from a normal distribution). For each biomass component an "optimal" regression equation of each component (tWi) was selected. The optimal Wi model was also obtained by back-transforming the tWi model and then residuals were further examined.

Compatible biomass model establishment

After obtaining independent biomass models of each component, three sets of compatible biomass models for cork oak natural forest and plantation in North China were established and compared including: (1) a total biomass regression function defined as the sum of the separately calculated best linear mixed effects models of the biomass of its components (sums of lme); (2) a total biomass regression function defined as the sum of the separately calculated best linear models of the biomass of its components (sums of lm); and (3) simultaneous equations of linear models based on seemingly unrelated regressions (SUR). The following relationship should be satisfied (Fig.1).

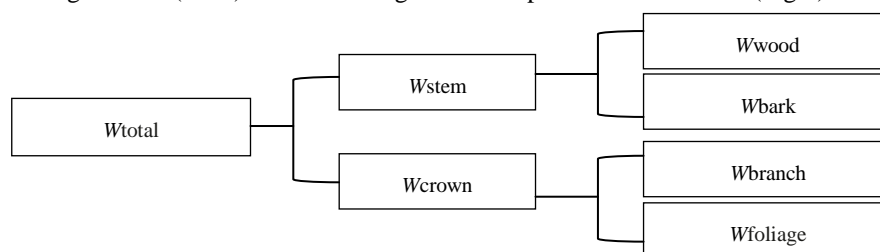


Fig. 1 Relationship between each component, where lines from right to left indicate summation

Methods (1) and (2), sums of lme and sums of lm, included stem wood biomass models, bark biomass models, branch biomass models, foliage biomass models and root biomass models. Then summing stem wood biomass models and bark biomass models to get stem biomass models, summing branch biomass models and foliage biomass models to get crown biomass models, and summing stem wood biomass models, bark biomass models, branch biomass models and foliage biomass models to get total aboveground biomass models for all the 100 trees. For 32 trees, tree total biomass models were also obtained by summing stem wood biomass models, bark biomass models, branch biomass models, foliage biomass models and root biomass models.

For method (3), simultaneous equations (based on linear models) were established with the estimated value of stem biomass of an independent linear stem biomass model (Xu, 1998). The remaining parameters in other models were jointly estimated by seemingly unrelated regressions (SUR). Given the limited number of sampled trees for root biomass, the root biomass model was not included in compatible models estimated using this method. Method (3) is described as formulae (4)-(9), where Wi is observed biomass value, Wi' is estimated biomass value, the $f_i(x)$ has the same form as that in the corresponding independent biomass model:

Group1:

$$W_{\text{total aboveground}} = W_{\text{stem}}' + f_{\text{crown}}(x) \quad (4)$$

$$W_{\text{crown}} = f_{\text{crown}}(x) \quad (5)$$

Group2:

$$W_{\text{wood}} = f_{\text{wood}}(x) \quad (6)$$

$$W_{\text{bark}} = W_{\text{stem}}' - f_{\text{wood}}(x) \quad (7)$$

Group3:

$$W_{\text{foliage}} = f_{\text{foliage}}(x) \quad (8)$$

$$W_{\text{branch}} = W_{\text{crown}} - f_{\text{foliage}}(x) \quad (9)$$

The parameters of the total aboveground biomass model and the crown biomass model were jointly estimated based on the estimated stem biomass using an independent stem biomass model (Group1). Then models of stem wood and bark, foliage and branch were separately jointly fitted based on the estimated stem biomass and crown biomass (Group2 and Group3). Whether or not to include any particular estimated parameter was decided by the significance of a t-test.

After establishing three sets of compatible biomass models using the three approaches, the values of r^2 and $Rmse$ were compared. Residual distribution homogeneity and model bias were visually checked by residual plots with loess regression lines overlaid on the plots; Normality of residuals was checked with Shapiro Wilks tests and by examining histograms of residuals. Then the most suitable set of compatible biomass models could be selected.

Since an independent validation data set was not available, the Leave-One-Out Jackknife method with *PRESS* (Prediction Sum of Squares) statistics was used to further test those most suitable models (Sánchez-González et al., 2007; Sánchez-González et al., 2005). *MPRESS* (the mean of the *PRESS* residuals see formula (10)) and *MPRESS* (the mean of the absolute values of the *PRESS* residuals) for each tW_i model were computed. Additionally, the residual ranges and prediction ranges of tW_i models and their corresponding jackknife tests were compared with residual plots.

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_{i-1})^2 \quad (10)$$

Where, y_i is the observed value of observation i , \hat{y}_{i-1} is the estimated value for observation i in a model fitted without this observation, n is the number of observations.

Results

Independent biomass models of each component

The dummy variable in this study specified stand origin. Fig.2 is a plot of the dependent variable (tW_{wood}) versus the independent variable ($tDBH$) with loess regression lines overlaid on it for stem wood biomass model and suggests that there was a difference between the relationships between tW_{wood} and $tDBH$ for natural forest (dotted line) and plantation (solid line). Similar differences were also found for other component biomass models. The dummy variable D_1 was significant ($P < 0.05$) in stem wood biomass models and foliage biomass models, which suggested that the two component models were indeed different between plantation and natural forest. While the coefficients of variable D_1 for other components models were not significant, so variable D_1 was not introduced to those models (see tW_i model in Tables 2 and 3).

The original root biomass dataset assembled 32 sets of data available in north China. However the biggest tree was an outlier (Fig.3). In order to establish a more accurate root biomass model for trees with more common sizes, the root biomass data of the biggest tree was excluded from modeling.

The best tW_i models of various components separately based on linear mixed effects models and linear models for individual biomass components of cork oak forest were selected. Summaries of those selected models are shown in Tables 2 and 3.

Most of the values of r^2 of the two sets of tWi models were above 0.9 (Tables 2 and 3). Most of the values of r^2 of Wi models were above 0.8 except for bark biomass model and foliage biomass model. All the values of $Rmse$ of Wi models were below 8 kg.

For “lme” tWi models, probabilities of type I errors for biomass models of branch and root were above 0.05 and those for stem wood, bark and foliage were below 0.05; For “lm” tWi models probabilities of type I errors for models of bark, branch and root were above 0.05 and those for stem wood and foliage were below 0.05. For “lme” Wi models and “lm” Wi models, all the probabilities of type I errors were below 0.01 (Tables 2 and 3).

As the loess curves of “lme” models were much closer to the baseline than those of “lm” models, only “lme” model residual plots and histograms of residuals for stem wood biomass model and root biomass model are shown in Fig.4 (Those for biomass models of bark, branch and foliage were quite similar and so they are not shown here). Residual ranges and predicted value ranges of “lme” models of various components are shown in Table 2. Heteroskedasticity of tWi models for “lme” was not obvious and for most of the models (Fig. 4a, c), and residuals approximately followed a normal distribution (Fig. 4b, d).

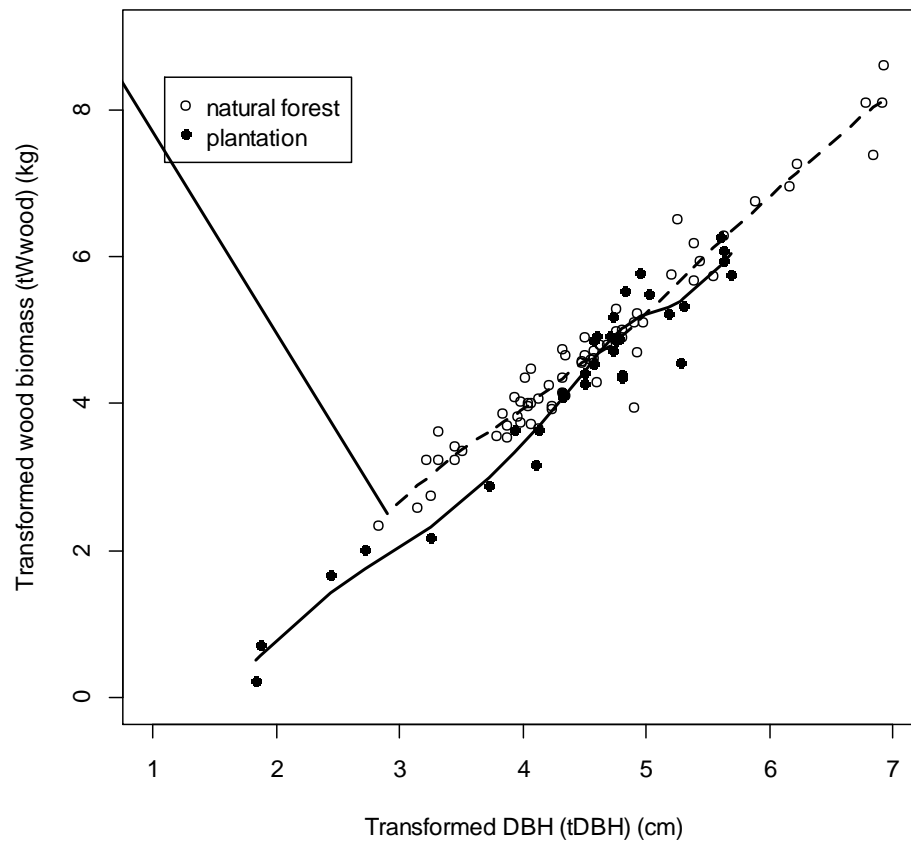


Fig. 2 Transformed stem wood biomass (tW_{wood}) versus transformed DBH ($tDBH$) for stem wood biomass model (The dotted line and solid line are respectively the loess regression lines for natural forest and plantation)

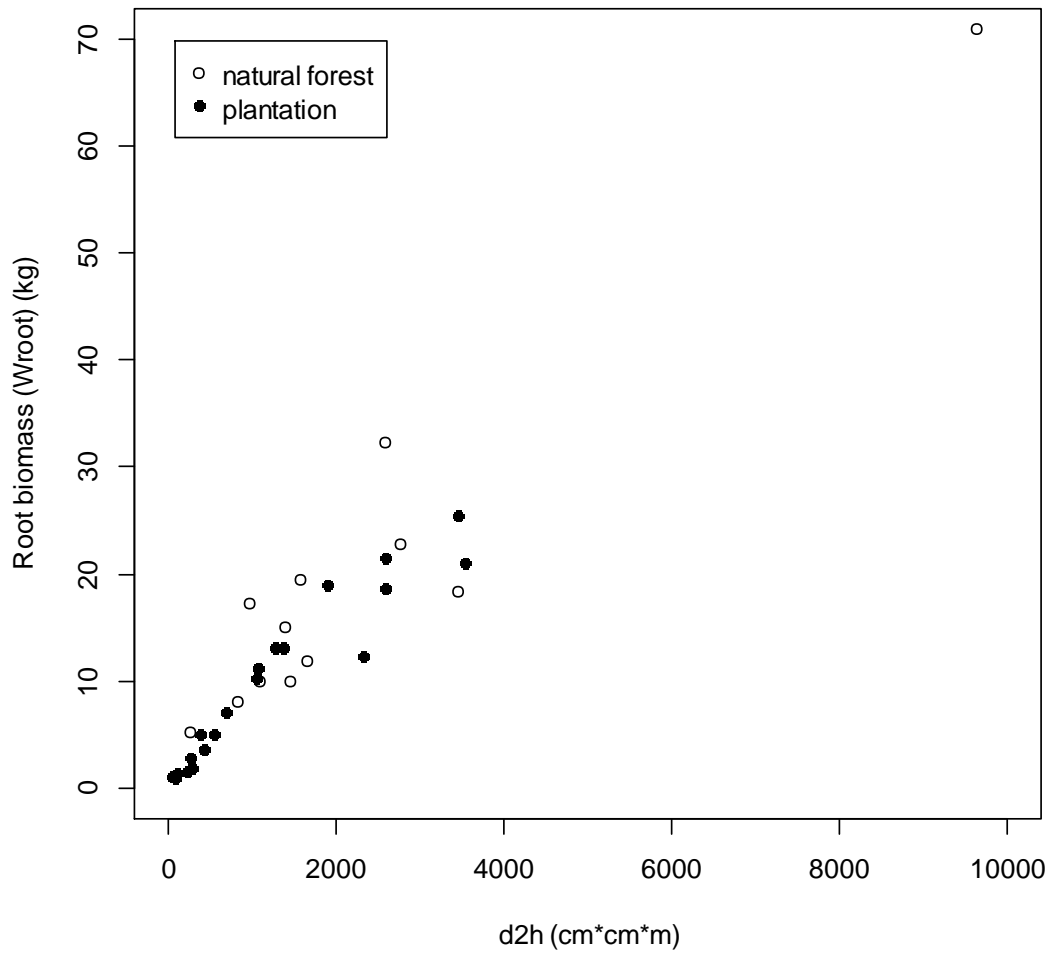


Fig. 3 Root biomass versus d^2h for 32 sets of data

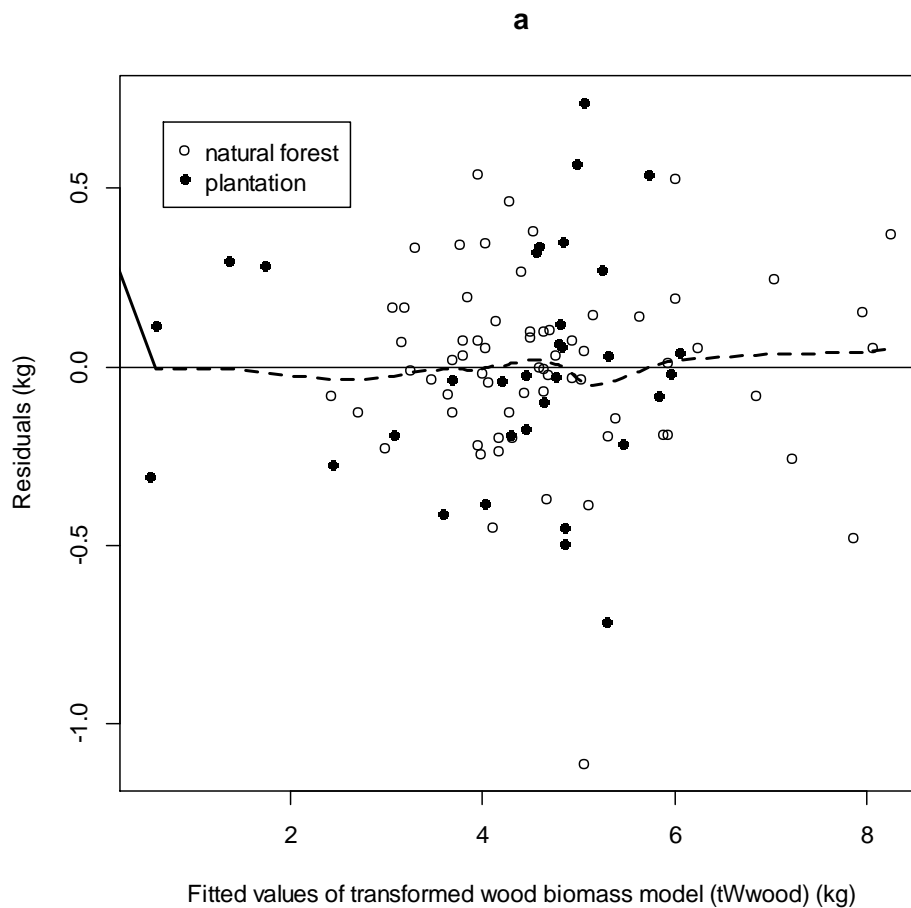
Table 2 Summary for independent biomass linear mixed effects models (lme)

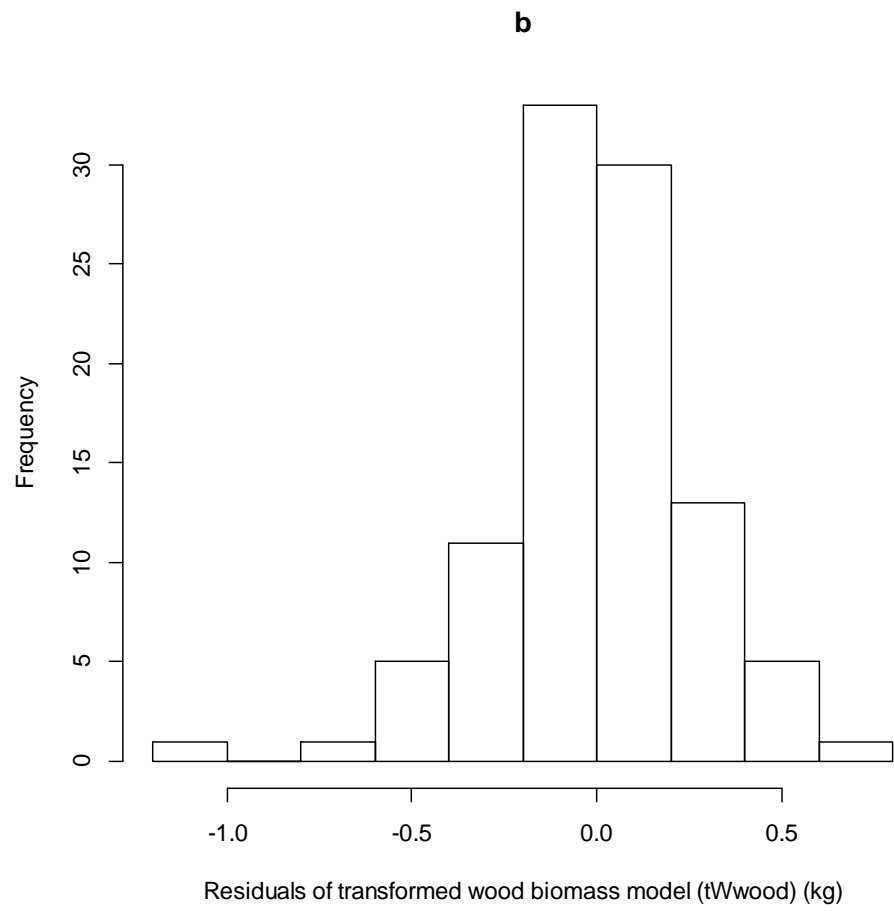
tW_i model	p -value (tW_i)	r^2 (tW_i)	$Rmse$ (tW_i)	p -value (W_i)	r^2 (W_i)	$Rmse$ (W_i)	predicted value range	Residual range
$tW_{wood}=a+b*tDBH$ $+c*D_1$	0.020	0.959	0.283	1.677e-08	0.958	7.645	(0.2,8.2)	(-1.5,0.7)
$tW_{bark}=a+b*tDBH$	0.002	0.903	0.223	< 2.2e-16	0.706	5.577	(-0.1,3.8)	(-1.0,0.9)
$tW_{branch}=a+b*d^2h+$ $c*H$	0.842	0.902	0.360	4.433e-11	0.898	4.876	(-0.3,5.5)	(-1.0, 1.0)
$tW_{foliage}=a+b*H+$ $c*d2h+e*D_1$	0.011	0.780	0.361	1.315e-05	0.748	1.105	(-0.6,2.7)	(-1.0,0.8)
$tW_{root}=a+b*d2h$	0.097	0.884	0.696	0.005232	0.839	1.824	(-0.8,11.0)	(-2.2,3.6)

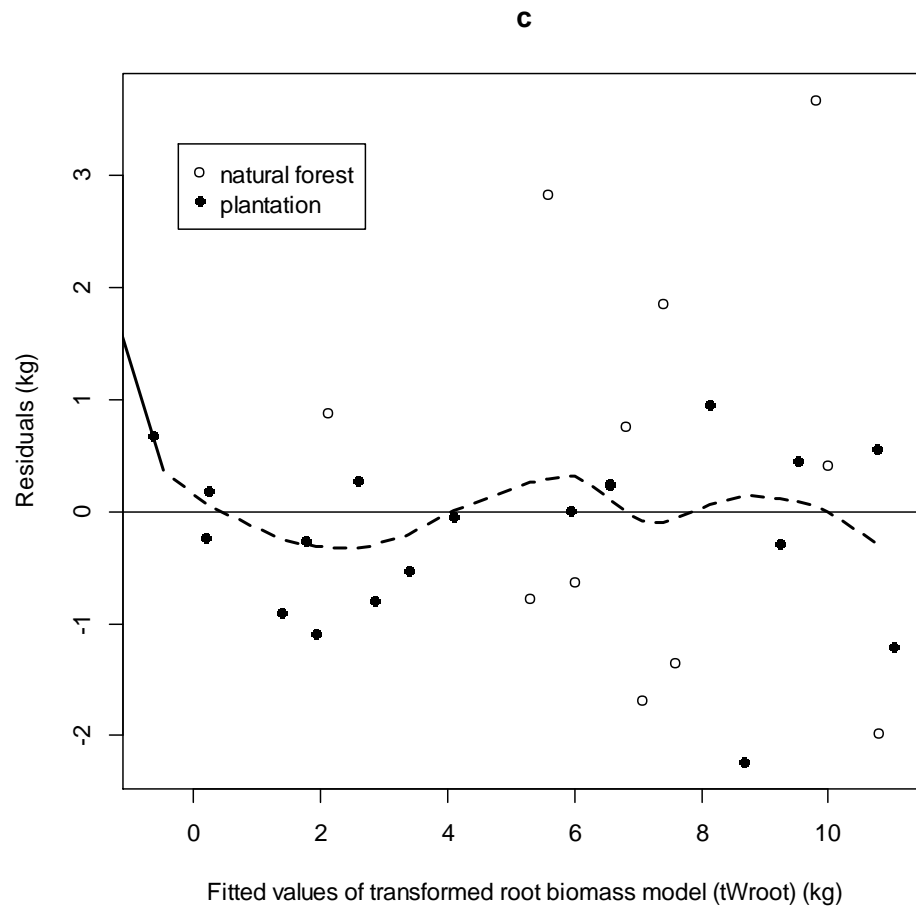
Table 3 Summary for independent biomass linear models (lm)

tW_i model	p -value (tW_i)	r^2 (tW_i)	$Rmse$ (tW_i)	p -value (W_i)	r^2 (W_i)	$Rmse$ (W_i)
--------------	--------------------------	------------------	----------------------	-------------------------	-----------------	---------------------

$tW_{wood}=a+b*tDBH+c*D_1$	0.022	0.944	0.332	8.697e-10	0.936	9.381
$tW_{bark}=a+b*tDBH$	0.126	0.875	0.253	1.048e-15	0.689	5.739
$tW_{branch}=a+b*tD2h+c*H$	0.889	0.874	0.409	4.544e-12	0.861	5.689
$tW_{foliage}=a+b*H+c*tD^2h+e*D_1$	0.009	0.768	0.371	1.116e-05	0.732	1.139
$tW_{root}=a+b*tD2h$	0.079	0.864	0.756	0.005085	0.811	1.979







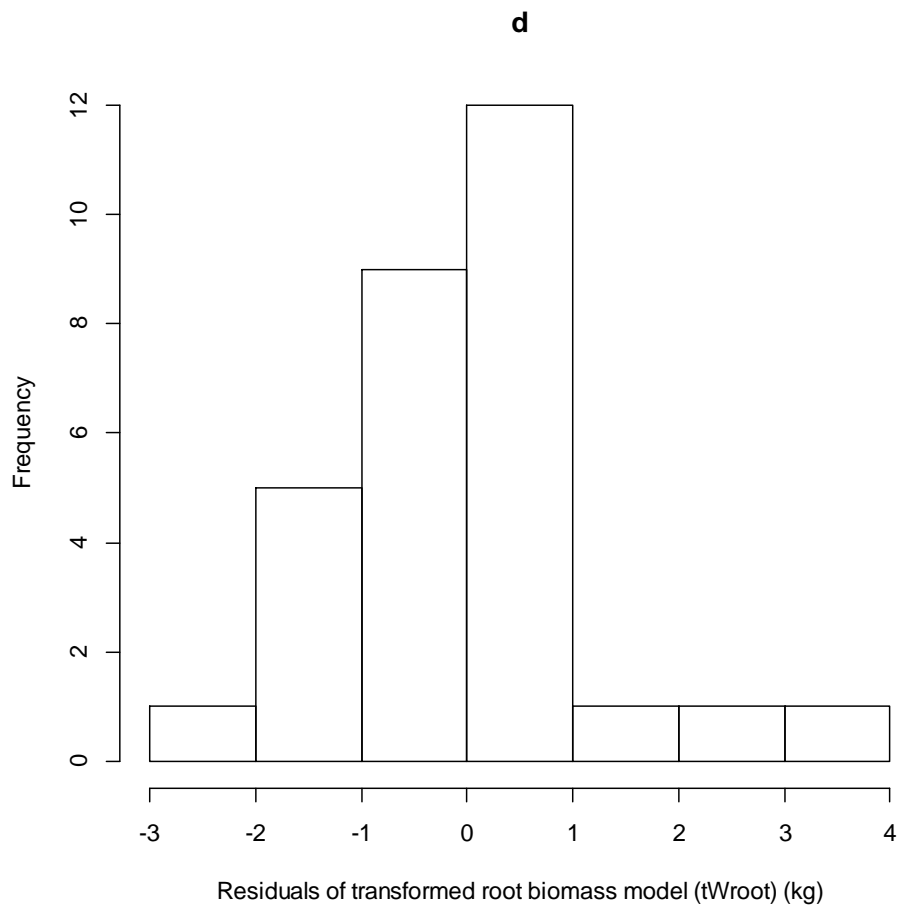


Fig.4 Loess residual plots and histograms of residuals for independent biomass linear mixed effects models (lme) (For Fig. 4a, c, the solid horizontal lines are baselines and the dotted lines are loess curves)

Compatible biomass models

Coefficients and standard errors of the coefficients for the three sets of compatible biomass models ((1) sums of lme; (2) sums of lm; and (3) SUR) are listed in Table 4.

The corresponding probabilities of type I errors in Shapiro Wilk tests (*p-value*) for “sums of lme” and “sums of lm” are in Table 5. The corresponding values for SUR are in Table 6.

Table 4 Coefficients and standard errors of the coefficients (*se*) for “sums of lme”, “sums of lm” and SUR (*Wi*)

Components (equation form and $\lambda(i)$)	Methods	Coefficients ($\pm se$)			
		$a (\pm se)$	$b (\pm se)$	$c (\pm se)$	$e (\pm se)$
stem	SUR	-1.776 ***	0.481 ***		
		(± 0.121)	(± 0.009)		
stem wood ($tW_{wood} = a + b * tDBH + c * D_j$; $\lambda(W_{wood}) = 0.16$,	Sums of	-1.710 ***	1.3449 ***	0.3909 **	
	lme	(± 0.229)	(± 0.0459)	(± 0.1369)	
$\lambda(DBH) = 0.45$)	Sums of	-1.995 ***	1.418 ***	0.249 ***	
	lm	(± 0.167)	(± 0.036)	(± 0.071)	
bark ($tW_{bark} = a + b * tDBH$;	Sums of	-2.687 ***	2.000 ***		
	lme	(± 0.230)	(± 0.092)		

$\lambda(Wbark)=0,$	Sums of	-2.865 ***	2.065 ***		
$\lambda(DBH) = 0)$	lm	(± 0.194)	(± 0.079)		
	SUR	-3.558 ***	2.336 ***		
		(± 0.277)	(± 0.097)		
crown	SUR	-2.300 ***	-0.086 **	0.453 ***	
		(± 0.200)	(± 0.029)	(± 0.031)	
branch	Sums of	-2.988 ***	0.563 ***	-0.194 ***	
($tWbranch=a+b*t2h+c*H;$	lme	(± 0.286)	(± 0.039)	(± 0.041)	
$\lambda(Wbranch) = 0.11,$	Sums of	-2.911 ***	0.560 ***	-0.197 ***	
$\lambda(d^2h) = 0.16)$	lm	(± 0.221)	(± 0.039)	(± 0.043)	
	Sums of	-2.126 ***	-0.122 **	0.349 ***	-0.311 **
foliage	lme	(± 0.220)	(± 0.039)	(± 0.036)	(± 0.096)
($tWfoliage=a+b*H+c*t2h+e*D_i;$	Sums of	-2.091 ***	-0.126 **	0.350 ***	-0.289 ***
$\lambda(Wfoliage) = 0.18,$	lm	(± 0.204)	(± 0.039)	(± 0.035)	(± 0.079)
$\lambda(d^2h) = 0.16)$	SUR	-2.077 ***	-0.086 **	0.318 ***	-0.252 ***
		(± 0.201)	(± 0.033)	(± 0.033)	(± 0.067)
root	Sums of	-2.058 **	0.185 ***		
($tWroot=a+b*t2h;$	lme	(± 0.763)	(± 0.016)		
$\lambda(Wroot) = 0.66,$	Sums of	-2.287 **	0.190 ***		
$\lambda(d^2h) = 0.42)$	lm	(± 0.643)	(± 0.014)		

Significant codes: *** 0.001, ** 0.01, * 0.05. The type III sums of squares were used in “sums of lme” and “sums of lm”. The t-tests were used in SUR.

Table 5 *p*-value in Shapiro Wilks test for sums of lme and sums of lm

Wi model	sums of lme	sums of lm
$tWstem=tWwood+tWbark$	2.643e-12	8.645e-13
$tWcrown =tWbranch+tWfoliage$	2.503e-11	7.41e-12
$tWtotal\ aboveground=tWwood+tWbark+$ $tWbranch+tWfoliage$	1.552e-13	7.998e-14
$tWtotaltree=tWwood+tWbark+tWbranch+tWfoliage+tWroot$	0.04626	0.08583

Table 6 r^2 , *Rmse* and *p*-value in Shapiro Wilks tests for simultaneous equations (SUR)

Simultaneous equations based on lm		r^2	<i>Rmse</i>	<i>p</i> -value (Wi)
Group1	<i>Wstem</i>	0.954	9.990	8.349e-13
	<i>Wcrown</i>	0.893	5.560	3.88e-07
	$Wtotal\ aboveground = Wstem + Wcrown$	0.951	14.007	8.187e-13
Group2	$Wwood = Wstem - Wbark$	0.964	7.020	7.566e-05
	<i>Wbark</i>	0.719	5.453	2.354e-14
Group3	$Wbranch=Wcrown-Wfoliage$	0.894	4.998	4.986e-07
	<i>Wfoliage</i>	0.739	1.125	0.0002

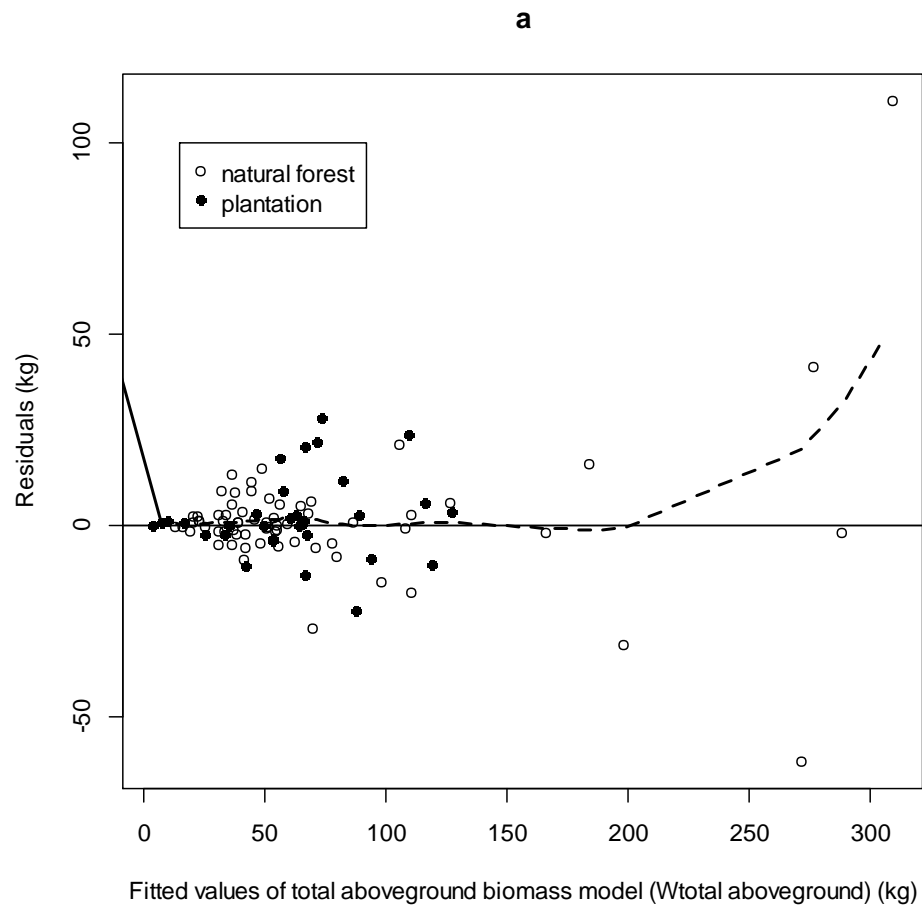
All the coefficients for the three sets of compatible biomass models were significant at $\alpha=0.01$ confidence level (Table 4). Generally speaking, standard errors for the three sets of models were very

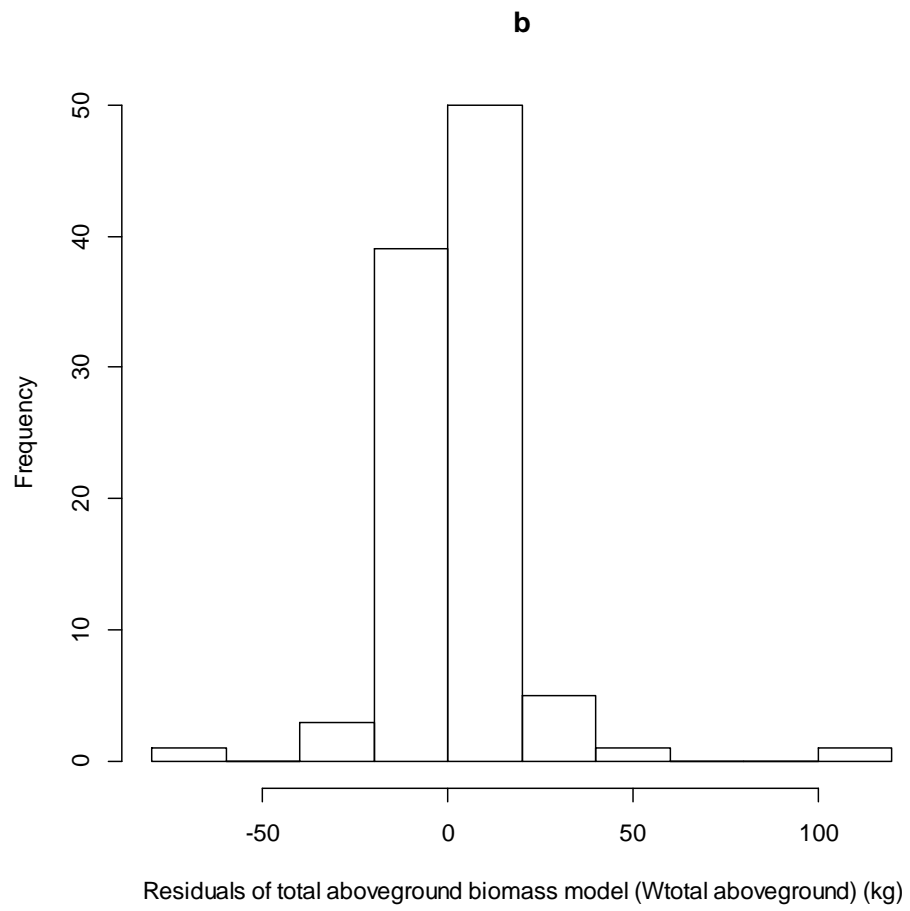
low compared to the magnitudes of predicted values. For sums of lme models, the values ranged from 0.016 to 0.763 (Table 4).

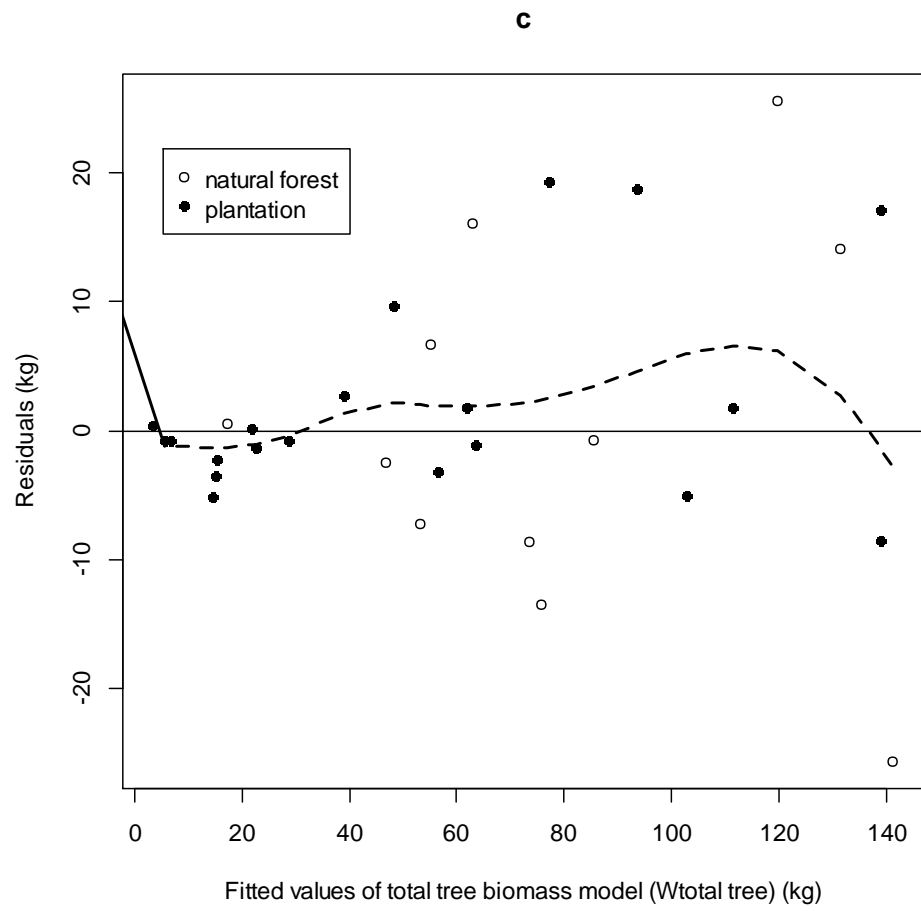
Most of the values of r^2 of Wi models in SUR were above 0.8 except for the bark biomass model and the foliage biomass model, and values of r^2 in SUR were a little larger than those of “sums of lm” models for stem wood biomass, bark biomass, branch biomass and foliage biomass (Table 3, Table 6). Most of the values of $Rmse$ in SUR were below 10 kg except for the total aboveground biomass model and values of $Rmse$ in SUR were a little smaller than those in “sums of lm” models for stem wood biomass, bark biomass, branch biomass, foliage biomass (Table 3, Table 6).

The loess curves of “sums of lme” models were closest to the x-axis, followed by “sums of lm”, and SUR with the biggest deviation from the x-axis, and only “sums of lme” models residual plots and histograms of residuals for total aboveground biomass and total tree biomass are shown in Fig.5 with the same symbol meanings as Fig. 4. The loess lines of each “sums of lme” model generally produced a flat line located at the baseline except for few big trees with DBH values above 19 cm or H values above 15m (Fig. 5a, c). For most of the “sums of lme” Wi models, the residuals did not follow normal distribution and heteroskedasticity was obvious (Fig. 5b, d).

To sum up, Values of r^2 were a little higher and values of $Rmse$ were a little lower in SUR than those in “sums of lm”, but SUR had higher bias than “sums of lm” for most components, and the “sums of lm” had higher bias than the “sums of lme” whose loess lines of each model generally produced a flat line located at the baseline except for few big trees (Fig. 5a, c), therefore the "sums of lme" in the three above methods was the most suitable way to force additivity of different components in this study.







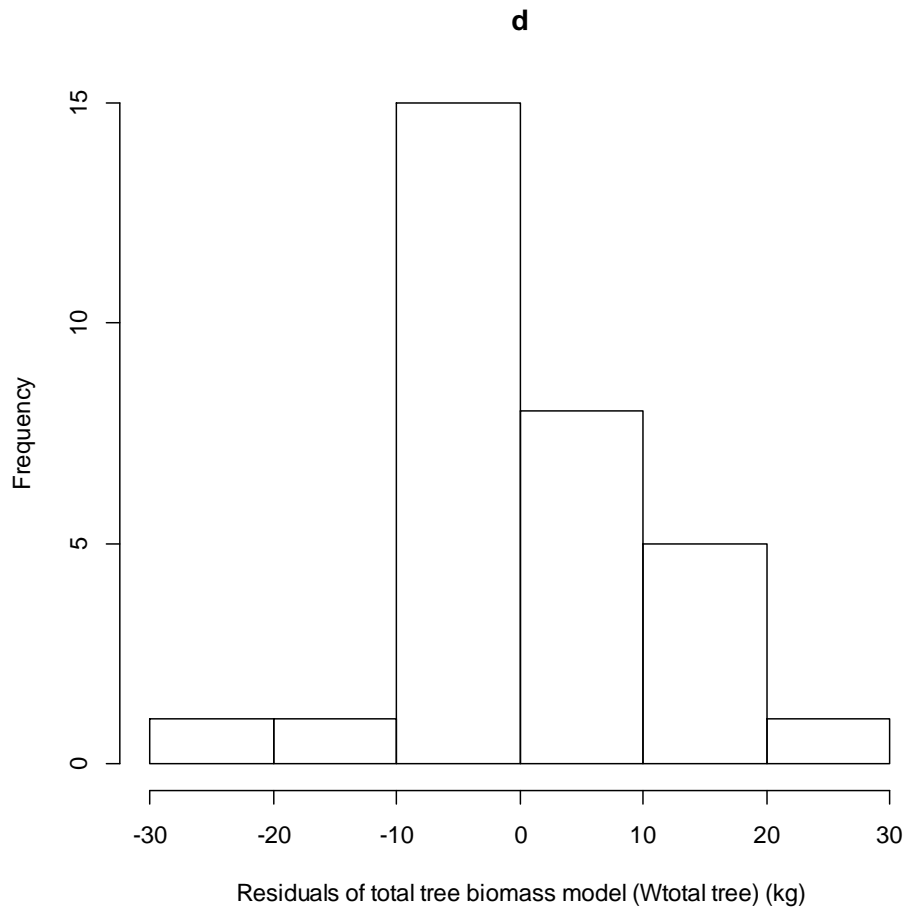


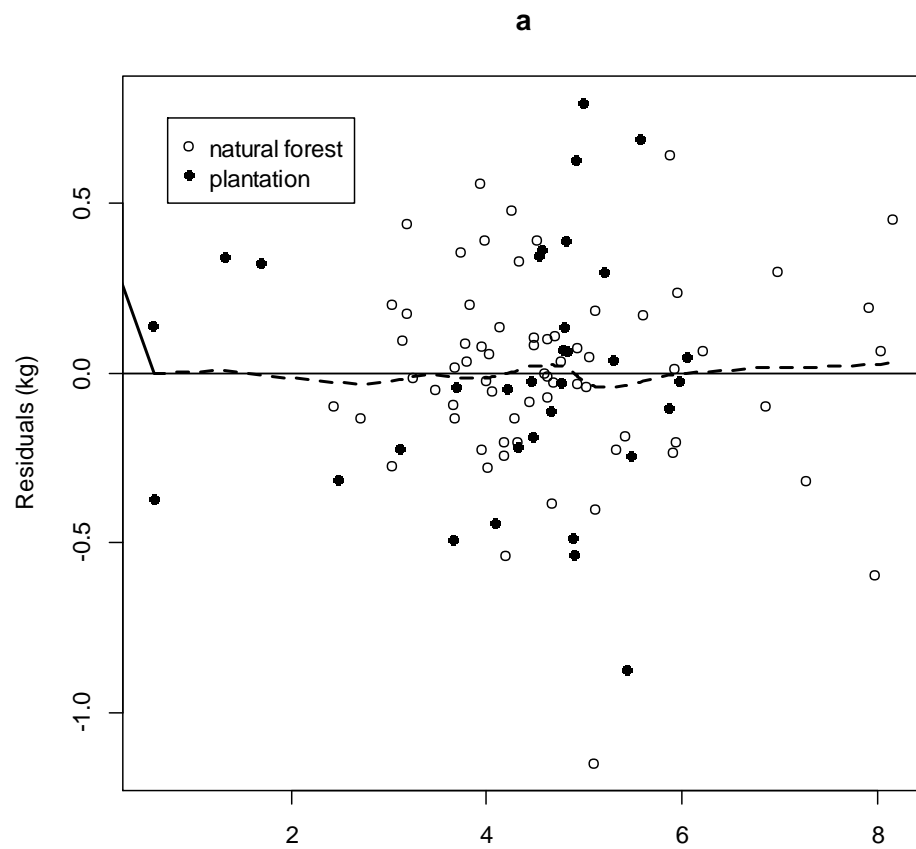
Fig. 5 Loess residual plots and histograms of residual for compatible linear mixed effects models (sums of lme)

Jackknife test and *PRESS* statistics of linear mixed effects models

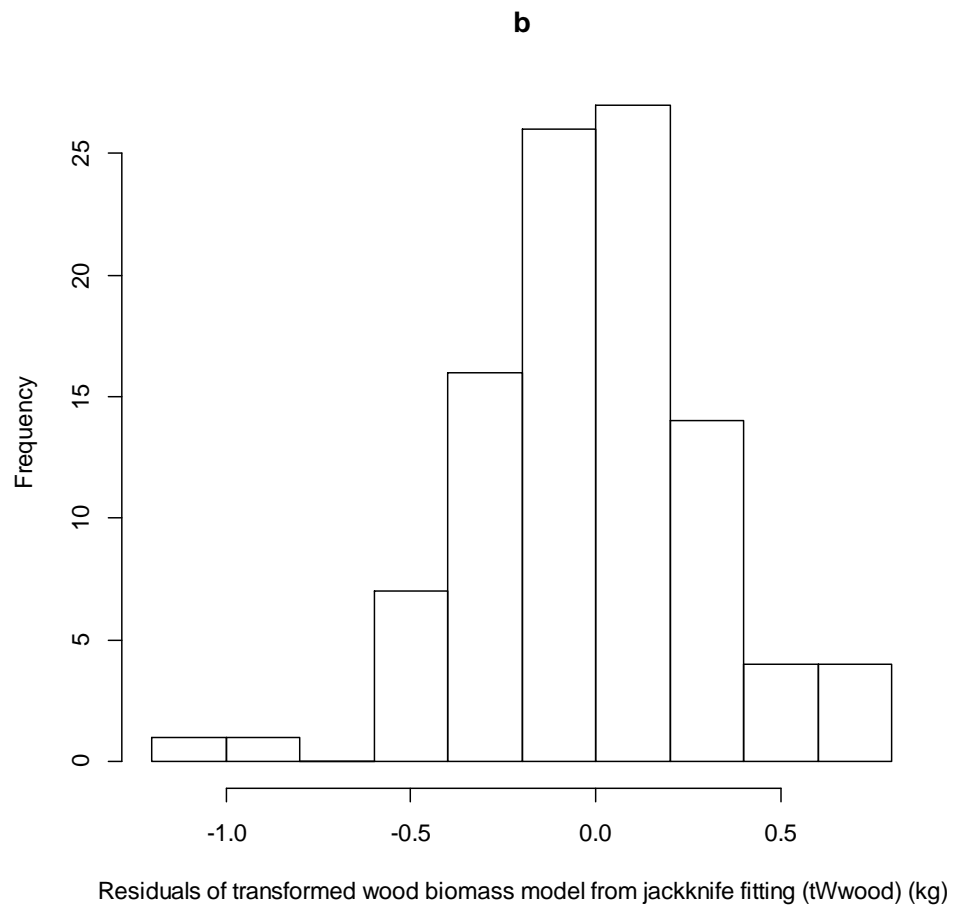
Compatible biomass models based on the method of “sums of lme” were selected as the best models and so they were validated with jackknife methods and *PRESS* (Prediction Sum of Squares) statistics. The *MPRESS* and *MAPRESS* statistics for each *tWi* model are shown in Table 7. Residual plots and histograms of residuals for the jackknife testing of stem wood biomass model and root biomass model (*tWi*) are shown in Fig.6 with the same symbol meanings as Fig.4. Residual ranges and predicted value ranges for the jackknife tests of various components are shown in Table 7.

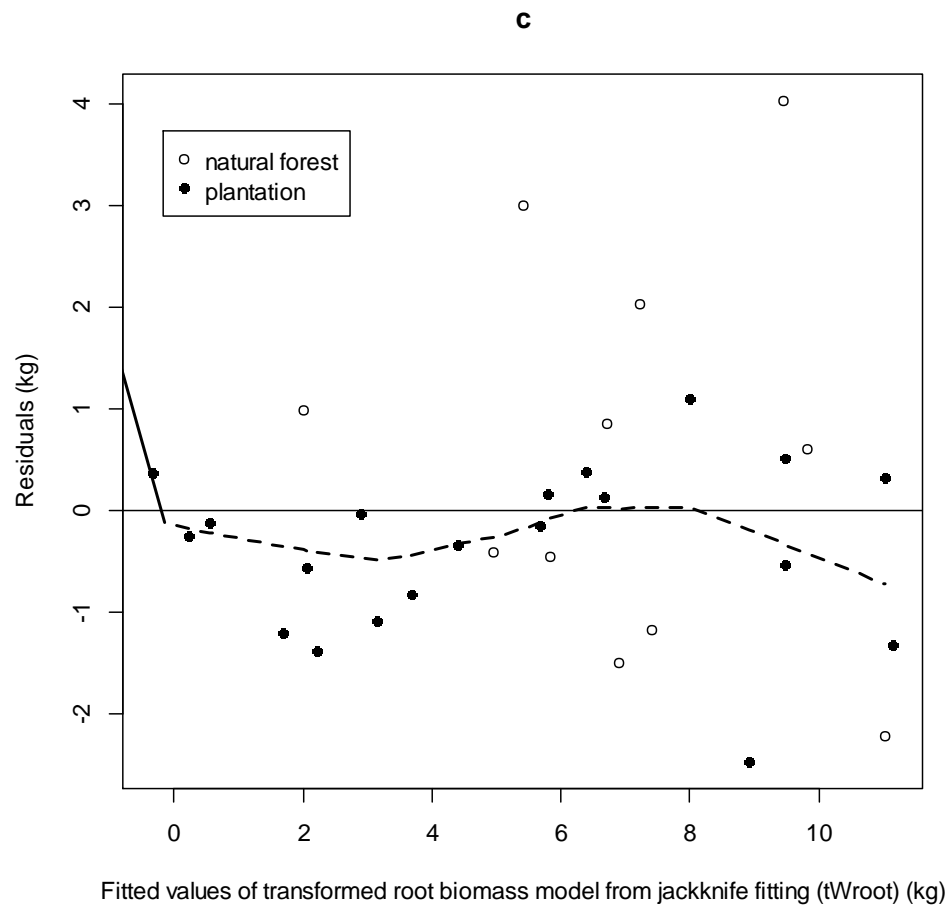
Table7 Summary for each “sums of lme” *tWi* model from jackknife fitting

Component	<i>p</i> -value	<i>PRESS</i>	<i>MPRESS</i>	<i>MAPRESS</i>	predicted value range	Residual range
stem wood	0.057	9.779	0.0006	0.231	(0.2,8.2)	(-1.5,0.7)
bark	0.0004	6.333	-8.955e-05	0.174	(-0.1,3.8)	(-1.0,0.9)
branch	0.820	15.639	-0.003	0.318	(-0.3,5.5)	(-1.0, 1.0)
foliage	0.009	14.682	-0.002	0.284	(-0.6,2.7)	(-1.0,0.8)
root	0.089	60.481	-0.039	1.111	(-0.8,11.0)	(-2.1, 4.3)



Fitted values of transformed wood biomass model from jackknife fitting (tWwood) (kg)





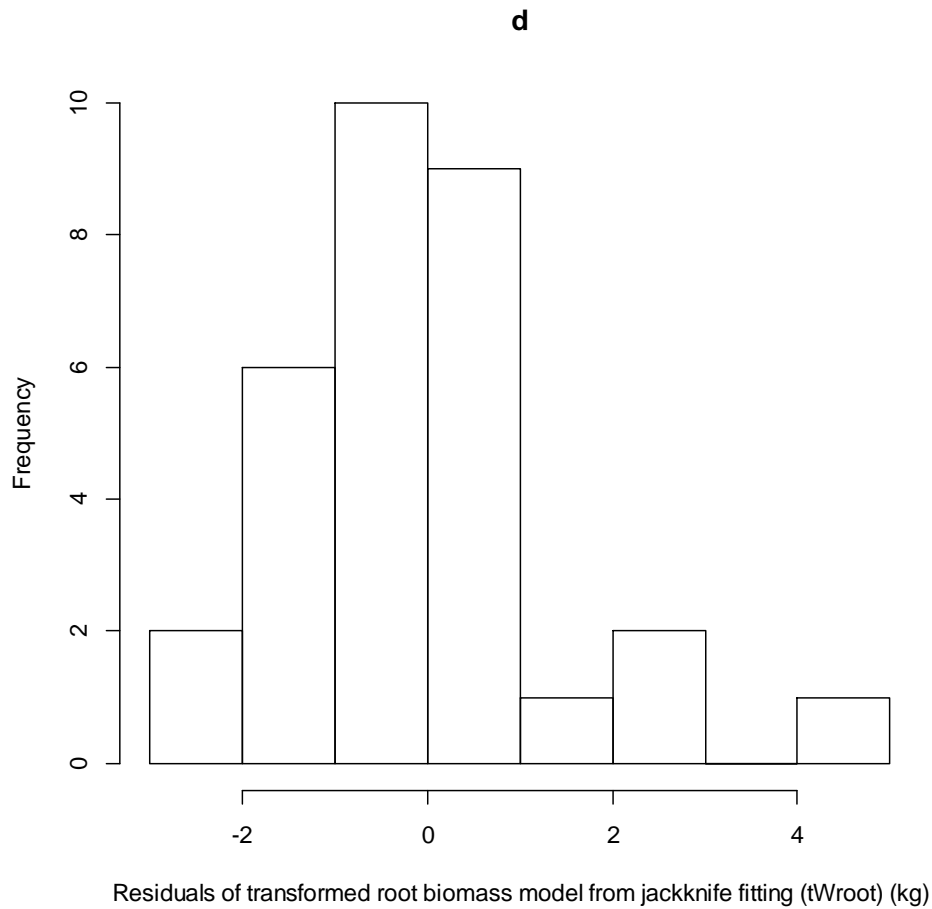


Fig.6 Residual plots and histograms of residuals for jackknife testing

Table 7 shows values of *PRESS*, *MPRESS* and *MAPRESS* for each tW_i model. Values of *MPRESS* for the stem wood biomass model, bark biomass model, branch biomass model and foliage biomass model were respectively 0.0006, $-8.955e-05$, -0.003 and -0.002 ; and values of *MAPRESS* for the four models were respectively 0.231, 0.174, 0.318 and 0.284. Values of *MPRESS* and *MAPRESS* for the root biomass model were respectively -0.039 and 1.111.

It can be seen from Figs. 4, 6 that for the tW_i model the scatter distribution in residual plots from fitting and those from the jackknife validation were similar and frequency distributions in histograms of residuals were similar as well. In Fig. 6a, c, heteroskedasticity was not obvious and the loess curves were close to the x-axes. In addition, residuals approximately followed normal distributions (Fig. 6 b, d).

Comparing the residual ranges of biomass models of various components in Tables 2 and 7, it can be seen that for stem wood biomass, bark biomass, branch biomass and foliage biomass, the residual ranges of tW_i models and jackknife tests were approximately equal, respectively were from -1.5 to 0.7 , from -1.0 to 0.9 , from -1.0 to 1.0 and from -1.0 to 0.8 . The predicted value ranges of above four tW_i models and their jackknife tests were also approximately equal, respectively were from 0.2 to 8.2 , from -0.1 to 3.8 , from -0.3 to 5.5 and from -0.6 to 2.7 . For root biomass, the prediction ranges of tW_i models and

jackknife tests were approximately equal, i.e., both were from -0.8 to 11.0; however, the residual ranges were from -2.2 to 3.6 for *tWroot* model and from -2.1 to 4.3 for its jackknife test.

Discussion

Selection of compatible models

Parresol (1999, 2001) proposed and compared two methods to force additivity of biomass models: (1) the total biomass regression function is defined as the sum of the separately calculated best regression functions of the biomass of its components; (2) joint-generalized least squares regression considering the contemporaneous correlations among equations, like seemingly unrelated regressions (SUR). By comparing standard errors of the coefficients of the two procedures and their own confidence and prediction intervals, the author pointed out that procedure (2) was recommended over procedure (1). Other researchers also reported that system estimating methods were more flexible and directly used simultaneous methods such as two-stage least squares (2SLS), three-stage least squares (3SLS), seemingly unrelated regressions (SUR), generalized method of moments (GMM). (António et al., 2007; Balboa-Murias et al., 2006; Bi et al., 2010; Brandeis et al., 2006; Gómez-García et al., 2013; Gómez-García et al., 2014; Hasenauer et al., 1998; Carvalho and Parresol, 2003; Li and Zhao, 2013; Menéndez-Miguélez et al., 2013; Ruiz-Peinado et al., 2011). Just few researchers compared the aboving procedure (1) and procedure (2) again before using system estimating methods to fit models, e.g., (Návar et al., 2002).

However some researchers fitted biomass models using procedure (1) rather than procedure (2) to make sure of the compatibility, i.e., fitting equations firstly and then evaluating additivity of predicted biomass components (Saint-André et al., 2005; Shaiek et al., 2011).

In our study, three methods including “sums of lme”, “sums of lm” and SUR were compared on r^2 , *Rmse* and residuals plots, and the "sums of lme" in the three above methods was finally selected as the most suitable way to force additivity of different components in this study. This study may have been a special case as Parresol suggested in his paper (Parresol, 2001) that

“...if one is faced with a small sample size and if contemporaneous correlations are small, then there is a loss in finite sample efficiency with estimated generalized NSUR, In this case, procedure (1) is best”.

Root biomass model and bark biomass model

The residual range for the root biomass model from fitting was a little wider than that from jackknife validation, but not overly (Figs.4, 6 c), so the root biomass model can be considered to have met the conditions for a jackknife validation. The bark biomass model used a logarithmic transformation and according some researchers (Baskerville, 1972; Wauters et al., 2008; Zianis et al., 2011), this may produce systematic underestimation of the dependent variable Y when converting the estimated $\log_e Y$ back to the original scale Y. A correction factor should be used to correct this bias according to those researchers. However, the correction factor was not used in this study as little bias was detected after back transformation.

Conclusions

Biomass models obtained using the “sums of lme” method were selected as the most suitable in this study because they were less biased than those fitted using “sums of lm” or SUR methods.

Forms of the “sums of lme” tWi models (with only fixed effects) were respectively $tW_{wood}=a+b*tDBH+c*D_1$, $tW_{bark}=a+b*tDBH$, $tW_{branch}=a+b*td2h+c*H$, $tW_{foliage}=a+b*H+c*td^2h+e*D_1$ and $tW_{root}=a+b*td2h$. Then the Wi models can be obtained by converting the estimated tWi models back to the original scale Wi .

The coefficients of determination (r^2) for the biomass models were 0.958 for the stem wood biomass model, 0.706 for the bark biomass model, 0.898 for the branch biomass model, 0.748 for the foliage biomass model and 0.839 for the root biomass model. Root mean square errors ($Rmse$) for those five models were respectively 7.645 kg, 5.577 kg, 4.876 kg, 1.105 kg and 1.824 kg. Standard errors of the coefficients (se) of those models ranged from 0.016 to 0.763. The residual ranges for the biomass models and their corresponding jackknife tests were approximately equal, respectively from -1.5 to 0.7 for stem wood, from -1.0 to 0.9 for bark, from -1.0 to 1.0 for branch and from -1.0 to 0.8 for foliage. The residual ranges for the root biomass model and its jackknife test were not so similar, respectively from -2.2 to 3.6 and from -2.1 to 4.3.

Within the specific ranges of DBH (4cm, 19cm) or H (5m, 15m) in the dataset used in this study, the biomass models fitted with the “sums of lme” method can be used for calculation and prediction of biomass or carbon stocks of the cork oak forests in North China.

Acknowledgements This research was jointly supported by China Scholarship Council, scientific and research base construction projects of Beijing Municipal Education Commission (SYSBL2009), forestry science promotion project of the State Forestry Bureau (2011-44), open fund project of Beijing Forestry University ‘985’ advantage subject innovation platform (000-1108003), and special fund project for forestry public service industry and research (201004021). We acknowledge the strong support from Zhong Tiaoshan National Forest Authority, Xingtai County Forestry Bureau, Si Zuolou forestry station and Xi Shan forestry station in Beijing.

Conflict of interest The authors declare that they have no conflict of interest.

References:

- António N, Tomé M, Tomé J, Soares P, Fontes L (2007) Effect of tree, stand, and site variables on the allometry of *Eucalyptus globulus* tree biomass. *Canadian Journal of Forest Research* 37: 895-906
- Balboa-Murias MÁ, Rodríguez-Soalleiro R, Merino A, Álvarez-González JG (2006) Temporal variations and distribution of carbon stocks in aboveground biomass of radiata pine and maritime pine pure stands under different silvicultural alternatives. *Forest Ecol Manag* 237: 29-38
- Bao X, Chen L, Chen Q, Ren J, Hu Y, Li Y (1984) Biomass of *Quercus variabilis* forest. *Acta Phytoecologica et Geobotanica Sinica*, 84: 313-320 (in Chinese).
- Baskerville GL (1972) Use of logarithmic regression in the estimation of plant biomass. *Canadian Journal of Forest Research* 2: 49-53
- Bi H, Long Y, Turner J, Lei Y, Snowdon P, Li Y, Harper R, Zerihun A, Ximenes F (2010) Additive prediction of aboveground

- biomass for *Pinus radiata* (D. Don) plantations. *Forest Ecol Manag* 259: 2301-2314
- Brandeis TJ, Delaney M, Parresol BR, Royer L (2006) Development of equations for predicting Puerto Rican subtropical dry forest biomass and volume. *Forest Ecol Manag* 233: 133-142
- Canadell J, Riba M, Andrés P (1988) Biomass Equations for *Quercus ilex* L. in the Montseny Massif, Northeastern Spain. *Forestry* 61: 137-147
- Carvalho JP, Parresol BR (2003) Additivity in tree biomass components of Pyrenean oak (*Quercus pyrenaica* Willd.). *Forest Ecol Manag* 179: 269-276
- Case BS, Hall RJ (2008) Assessing prediction errors of generalized tree biomass and volume equations for the boreal forest region of west-central Canada. *Canadian Journal of Forest Research* 38: 878-889
- Castedo-Dorado F, Gómez-García E, Diéguez-Aranda U, Barrio-Anta M, Crecente-Campo F (2012) Aboveground stand-level biomass estimation: a comparison of two methods for major forest species in northwest Spain. *Ann Forest Sci* 69: 735-746
- Durkaya A, Durkaya B, Ünsal A (2009) Predicting the above-ground biomass of calabrian pine (*Pinus brutia* Ten.) stands in Turkey. *African Journal of Biotechnology* 8: 2483-2488
- Gómez-García E, Crecente-Campo F, Diéguez-Aranda U (2013) Above-ground biomass equations for birch (*Betula pubescens* Ehrh.) and pedunculate oak (*Quercus robur* L.) in north western Spain. *Madera y Bosques* 19: 71-91(in Spanish)
- Gómez-García E, Crecente-Campo F, Tobin B, Hawkins M, Nieuwenhuis M, Diéguez-Aranda U (2014) A dynamic volume and biomass growth model system for even-aged downy birch stands in south-western Europe. *Forestry* 87: 165-176
- Gurdak DJ, Aragão LEOC, Rozas-Dávila A, Huasco WH, Cabrera KG, Doughty CE, Farfan-Rios W, Silva-Espejo JE, Metcalfe DB, Silman MR, Malhi Y (2014) Assessing above-ground woody debris dynamics along a gradient of elevation in Amazonian cloud forests in Peru: balancing above-ground inputs and respiration outputs. *Plant Ecology & Diversity* 7: 143-160
- Hasenauer H, Monserud RA, Gregoire TG (1998) Using simultaneous regression techniques with individual-tree growth models. *Forest Sci* 44: 87-95
- Jacoby WG (2000) Loess: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies* 19: 577-613
- Jenkins JC, Chojnacky DC, Heath LS, Birdsey RA (2003) National-scale biomass estimators for United States tree species. *Forest Sci* 49: 12-35
- Li H, Zhao P (2013) Improving the accuracy of tree-level aboveground biomass equations with height classification at a large regional scale. *Forest Ecol Manag* 289: 153-163
- Luo W, Zhang W, Huang Y (2009) Cork oak. China Forestry Publishing House, Beijing (in Chinese).
- Menéndez-Miguélez M, Canga E, Barrio-Anta M, Majada J, Álvarez-Álvarez P (2013) A three level system for estimating the biomass of *Castanea sativa* Mill. coppice stands in north-west Spain. *Forest Ecol Manag* 291: 417-426
- Ter-Mikaelian MT, Korzukhin MD (1997) Biomass equations for sixty-five North American tree species. *Forest Ecol Manag* 97: 1-24
- Muukkonen P (2007) Generalized allometric volume and biomass equations for some tree species in Europe. *European Journal of Forest Research* 126: 157-166
- Návar J, Méndez E, Dale V (2002) Estimating stand biomass in the Tamaulipan thornscrub of northeastern Mexico. *Ann Forest Sci* 59: 813-821
- Návar J (2009) Allometric equations for tree species and carbon stocks for forests of northwestern Mexico. *Forest Ecol Manag* 257: 427-434
- Parresol BR (1999) Assessing Tree and Stand Biomass: A Review with Examples and Critical Comparisons. *Forest Sci* 45: 573-593
- Parresol BR (2001) Additivity of nonlinear biomass equations. *Canadian Journal of Forest Research* 31: 865-878
- Paulo JA, Tomé M (2010) Predicting mature cork biomass with t years of growth from one measurement taken at any other age. *Forest Ecol Manag* 259: 1993-2005

- Peng H, Lu Y (2012) Model selection in linear mixed effect models. *J Multivariate Anal* 109: 109-129
- Ritchie M, Zhang J, Hamilton T (2013) Aboveground Tree Biomass for *Pinus ponderosa* in Northeastern California. *Forests* 4: 179-196
- Ruiz-Peinado R, Del Rio M, Montero G (2011) New models for estimating the carbon sink capacity of Spanish softwood species. *Forest Systems* 20: 176-188
- Saint-André L, M Bou AT, Mabiala A, Mouvondy W, Jourdan C, Roupsard O, Deleporte P, Hamel O, Nouvellon Y (2005) Age-related equations for above- and below-ground biomass of a Eucalyptus hybrid in Congo. *Forest Ecol Manag* 205: 199-214
- Sajdak M, Velázquez-Martí B, López-Cortés I, Fernández-Sarría A, Estornell J (2014) Prediction models for estimating pruned biomass obtained from *Platanus hispanica* Münchh. used for material surveys in urban forests. *Renew Energ* 66: 178-184
- Sakia RM (1992) The Box-Cox transformation technique: a review. *The statistician* 41: 169-178
- Sánchez-González M, Ca Ellas I, Montero González G (2007) Generalized height-diameter and crown diameter prediction models for cork oak forests in Spain. *Investigación agraria. Sistemas y recursos forestales* 16: 76-88
- Sánchez-González M, Tomé M, Montero G (2005) Modelling height and diameter growth of dominant cork oak trees in Spain. *Ann Forest Sci* 62: 633-643
- Schmidt A, Poulain M, Klein D, Krause K, Peña-Rojas K, Schmidt H, Schulte A (2009) Allometric above-belowground biomass equations for *Nothofagus pumilio* (Poep. & Endl.) natural regeneration in the Chilean Patagonia. *Ann Forest Sci* 66: 513
- Shaiek O, Loustau D, Trichet P, Meredieu C, Bachtobji B, Garchi S, Aouni MH (2011) Generalized biomass equations for the main aboveground biomass components of maritime pine across contrasting environments. *Ann Forest Sci* 68: 443-452
- Snorrason A, Einarsson SF (2006) Single-tree biomass and stem volume functions for eleven tree species used in Icelandic forestry. *Icelandic Agricultural Sciences* 19: 15-24
- Su X, Yan X, Tsai C (2012) Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics* 4: 275-294
- Fonseca TJF, Parresol BR (2001) A new model for cork weight estimation in Northern Portugal with methodology for construction of confidence intervals. *Forest Ecol Manag* 152: 131-139
- Tian Q, Zhou R, Zhang J (1997) Biomass of *Quercus variabilis* population. *Journal of Beijing Forestry University* 19 (SuPP.2): 113-117 (in Chinese).
- Torres Vélez DA, Del Valle JI (2007) Growth and yield modelling of *Acacia mangium* in Colombia. *New Forest* 34: 293-305
- Vallet P, Dhôte J, Mogueédec GL, Ravart M, Pignard G (2006) Development of total aboveground volume equations for seven important forest tree species in France. *Forest Ecol Manag* 229: 98-110
- Wauters JB, Coudert S, Grallien E, Jonard M, Ponette Q (2008) Carbon stock in rubber tree plantations in Western Ghana and Mato Grosso (Brazil). *Forest Ecol Manag* 255: 2347-2361
- Xu H (1998) Studies on standing tree biomass models and the corresponding parameter estimation. Dissertation, Beijing Forestry University (in Chinese).
- Zeng WS, Zhang HR, Tang SZ (2011) Using the dummy variable model approach to construct compatible single-tree biomass equations at different scales-a case study for Masson pine (*Pinus massoniana*) in southern China. *Canadian Journal of Forest Research* 41: 1547-1554
- Zianis D, Mencuccini M (2004) On simplifying allometric analyses of forest biomass. *Forest Ecol Manag* 187: 311-332
- Zianis D, Muukkonen P, Mäkipää R, Mencuccini, M (2005) Biomass and stem volume equations for tree species in Europe. *Silva Fennica Monographs* 4: 63
- Zianis D, Xanthopoulos G, Kalabokidis K, Kazakis G, Ghosn D, Roussou O (2011) Allometric equations for aboveground biomass estimation by size class for *Pinus brutia* Ten. trees growing in North and South Aegean Islands, Greece. *European Journal of Forest Research* 130: 145-160