Improved Allometric Equation for Aboveground Biomass: A Case Study of Four Tree Species in China

JIA WANG[1,2*], CHUNDONG LV[1,2], HE GAO[1,2], LONGYU ZHANG[1,2], FANGFEI ZHANG[1,2] AND ZHONGKE FENG[1,2]

1 Beijing Key Laboratory of Precision Forestry, Beijing Forestry University ,No. 35 Qinghua East Road, Haidian District, Box 111, Beijing100083, China

2 College of Forestry , Beijing Forestry University ,No. 35 Qinghua East Road, Haidian District, Box 111, Beijing100083, China

**Tel.: +86-13810664533 E-mail:wangjia2002_0@163.com*

Wang, J., Lv, Ch., Gao, H., Zhang, L., Zhang, F. and Feng, Zh. 2017. Improved Allometric Equation for Aboveground Biomass: A Case Study of Four Tree Species in China. *Baltic Forestry* 23(3): 636–643.

Abstract

Allometric equation is the most extensively applied method for estimating aboveground biomass (AGB). However, most studies focus only on an equation establishment and its precision and ignore errors in coefficient. Hence, the current paper aims to address the heteroscedasticity limitations of linear regression, which is the most extensively used method for solving the coefficient of allometric equations. In this paper, we proposed weighed linear regression with new weighed factor for estimating AGB. At meantime, we compared this method with other two common methods, which are simple linear regression and nonlinear regression, through solving their coefficients, accuracy assessment, and biomass estimate which forward compared with the biomass of 71 trees associated with 4 species at Xiaolong Mountain in the Gansu province of China. The results showed that the precision of coefficients of power progress of weighted linear regression and nonlinear least square regression is very close, and higher than the linear regression.

Keywords: Above ground biomass (AGP), allometric equation, linear regression, weighted linear regression, nonlinear regression

Introduction

Biomass estimation is very important in net primary productivity, energy conversion, nutrient and carbon cycling, biomass estimation (Saint-Andre2 et al. 2005, Delitti et al. 2006, Saglan et al. 2008). The most precise way to calculate biomass is through cutting down trees and measuring the weight of shoot system, which includes the trunk, branches, and leaves (Saglan et al .2008, Van et al. 2000, Norris et al. 2001, Brown et al. 2004, Wadham-Gagnon et al. 2006). Unfortunately, this destructive method will bring permanent and irreversible damages, including ecological system disasters, and human financial resource wastes (Delitti et al. 2006, Kale et al. 2004). Therefore, non-destructive techniques have recently become very common. These non-destructive techniques are based on the regression models that establish the relationship between the biomass growth, and its parameters (Brown et al. 2004, Lott et al. 2000, Claesson et al. 2001, Saatchi et al. 2007). Throughout the world, it is common to use the two biomass growth parameters, which are tree height (H), and the diameter at breast height (DBH) to develop the allometric equations needed to estimate the biomass.

Allometric equation is the common method for estimating AGB which show no significant difference when compared with actual biomass (Mary et al. 2001). The error of estimated biomass is determined through four factors; tree measurement, allometric model selection, single plot size, and landscape-scale representatively. Moreover, the allometric model itself is a source of error. For example, at present, almost all references utilize logarithmic conversion to linearize the data when using allometric equations for regression analysis (Chave et al. 2005). Subsequently, standard linear regression model is utilized to estimate the coefficients “*a*” and “*b*.” This is process called linear regression. Furthermore, the power function of linear regression is widely recommended in probability and statistics materials at present. However, this method results in serious heteroscedasticity.

Heteroscedastic studies are more common in the field of economics, because economists use the linear regression model and hypothesis testing to test heteroscedasticity, and reduce the heteroscedastic influence on the regression by using the weighted approach. Wang and Feng (Wang and Feng 2006) carried out studies on the current and widely used logarithm conversion method to address the power function problem, especially heteroscedasticity. Heteroscedasticity was corrected by weighted regression by many authors applying to every observation a weighting equal to the inverse of the variance of the residuals. Xu indicated that determining the weight function is one of the key problems in estimation methods for weighted regression (Xu 2003). Different weighted factors may be tested depending on the independent variables of model (Schlaegel 1982, Parresol 1999, Parresol 2001, Cunia 1987, Tang et al.2001). The optimization methodology of the value of the exponent k was proposed by Harvey. Estimation of weights when fitting nonlinear equations was described also between others by Williams and Gregoire (1993). The other way for solving the heteroscedastic problem is the nonlinear regression which was wildly used in western countries (Williams and Gregoire 1993, Reed and Green 1985, Tang and Wang. 2002, José de Jesús Návar Cháidez et al. 2004, Li and Zhao 2013).

These studies show that both the weighted regression and the nonlinear regression had similar results. In addition, their standard deviation and correlations indicate that they are clearly superior to that of traditional logistic conversions. Arevalo et al. (2007) also applied the traditional logarithm transformation, and the weighted and nonlinear regression methods to study the AGB of four asexually reproducing trees in New York. Arevalo et al. (2007) used mathematical software to subject the sample tree data, wherein, they concluded that the accuracy of the former two functions are nearly the same, and that the latter has a high standard deviation and index correlation. Feng (1999) cited the statistical data from Baskerville (1972), and pointed out that the allometric equation coefficient calculated using logarithms reduces the biomass by 10% to 20%. It is important to note that this error can be reduced, and that the current development of computer technology now allows obtaining nonlinear parameters easier. In the study conducted by Cole and Ewel (2006) the calculations of four valuable tropical tree species pointed out that ignoring the heteroscedasticity of dependent variables often results in underestimating tree biomasses in the final equation.

Thus, this study was tried to solve the problem of heteroscedasticity on biomass calculation, and to develop allometric equations for biomass estimations of the different species in China.

Methods

The allometric equation, $W = a(DBH^2H)^b$, is used in this study for calculating AGB, because it is the most common model in calculating a single tree biomass in China (Shen Y.Z.2011, Du W.Z.2012). In this equation, the DBH is the diameter at breast height, the H is total tree height, and the W is the AGB. For estimating different methods for solving the coefficient a and b of allometric equation $W = a(DBH^2H)^b$, take Xi and Yi as variables for the observation value of DBH^2H and W at the ith. Once this is completed, the allometric equation $W = a(DBH^2H)^b$ can be written in the form of the power function $y_i = ax_i^b + \varepsilon_i, \quad (i = 1, 2, L\ , n)$, and after log transformed, the power function is be changed into a simple linear regression form: $\ln(y_i - \varepsilon_i) = \ln a + b \ln x_i$, where ε_i is the random error[30,31]. The most common method to estimate a and b of allometric equation of is the logarithmic transformation, which is usually applied to convert the nonlinear power function to a log-linear model.

Linear regression

The power function is as follows:

$$y_i = ax_i^b + \varepsilon_i, \quad (i = 1, 2, L\ , n) \tag{1}$$

Move ε_i to the left side of the equation and transform to logarithm form:

$$\ln(y_i - \varepsilon_i) = \ln a + b \ln x_i \tag{2}$$

That is

$$\ln(y_i - \varepsilon_i) = \ln[y_i(1 - \varepsilon_i / y_i)] = \ln y_i + \ln(1 - \varepsilon_i / y_i) = \ln a + b \ln x_i \tag{3}$$

make $k_i = \varepsilon_i / y_i$

Considering that k_i is very small, we use series expansion on $V_{(k+1)}$, and only for once remainder, we can get the following formula:

$$\ln y_i \approx \ln a + b \ln x_i + k_i \tag{4}$$

Make $y_i' = \ln y_i$

$a' = \ln a$

$x_i' = \ln x_i$

We can get:

$$y_i \approx a' + bx_i' + k_i \tag{5}$$

According to the least square criterion, make $\sum_{i=1}^{n} k_i^2 = \min$, $\hat{b}$ and $\hat{a}'$ can be calculated, considering $\hat{a}', \hat{b}$ is the estimated value of a and b, then $\hat{a} = e^{\hat{a}'}$, this is the approach that has been commonly used – the power function linear regression.

Obviously, because the power function linear regression only merely satisfies the need of minimizing the quadratic sum of relative error of dependent variables, instead of absolute error, coefficients $\hat{a}', \hat{b}$ obtained by $\sum_{i=1}^{n} k_i^2 = \min$,not $\sum_{i=1}^{n} \varepsilon_i^2 = \min$, therefore, difference certainly be contained.

The $\hat{a}', \hat{b}$ which are derived from linear regression are the estimated value of a and b. Make $v_i = \hat{a}x_i^{\hat{b}} - y_i$, the estimated value of σ^2 can be obtained through the following formula:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} v_i^2 \Big/ (n-2) \qquad (6)$$

The Co-factor Matrix obtained by cofactor propagation law as the follow:

$$\begin{pmatrix} Q_{\hat{a}'\hat{a}'} & Q_{\hat{a}'\hat{b}} \\ Q_{\hat{a}'\hat{b}} & Q_{\hat{b}\hat{b}} \end{pmatrix} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Q}_{YY'}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1} \qquad (7)$$

Where $\mathbf{A}^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1' & x_2' & \cdots & x_n' \end{pmatrix}$

Furthermore, $\hat{\sigma}_{a_1}$ $\hat{\sigma}_{b_1}$ (estimation of Variance of $\hat{a}$ $\hat{b}$) and $\hat{\sigma}_{a_1b_1}$ (estimation of Covariance of $\hat{a}$ $\hat{b}$) can be calculated:

$$\hat{\sigma}_{aa}^2 = \hat{\sigma}^2\hat{a}^2 Q_{\hat{a}'\hat{a}'} \qquad (8)$$

$$\hat{\sigma}_{ab} = \hat{\sigma}^2\hat{a} Q_{\hat{a}'\hat{b}'} \qquad (9)$$

$$\hat{\sigma}_{bb}^2 = \hat{\sigma}^2 Q_{\hat{b}\hat{b}} \qquad (10)$$

Weighed linear regression

Multiply y_i on both sides of the equation:

$$y_i' \approx a' + bx_i' + k_i$$

and considering $k_i = \varepsilon_i / y_i$, we can get a new observational equation

$$y_i y_i' \approx y_i a' + y_i x_i' b + \varepsilon_i \qquad (11)$$

Under the condition that $\sum_{i=1}^{n} \varepsilon_i^2 = \min$, we can get coefficients $\hat{\hat{a}}$、$\hat{\hat{b}}$ (to distinguish linear regression, we suppose that the estimated value of a, b is $\hat{\hat{a}}$、$\hat{\hat{b}}$) and error function :

$$v_i = y_i\hat{\hat{a}}' + y_i x'\hat{\hat{b}} - y_i y_i' \qquad (12)$$

the unbiased estimated value of σ^2 can be calculated using the following formula:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} v_i^2 \Big/ (n-2) \qquad (13)$$

Here, the weight of $\ln y_i$ is y_i^2, and it makes $\sum_{i=1}^{n} \varepsilon_i^2 = \min$. It is more reasonable than the linear regression. The Co-factor Matrix obtained by cofactor propagation law as follow:

$$\begin{pmatrix} Q_{\hat{\hat{a}}'\hat{\hat{a}}'} & Q_{\hat{\hat{a}}'\hat{\hat{b}}} \\ Q_{\hat{\hat{a}}'\hat{\hat{b}}} & Q_{\hat{\hat{b}}\hat{\hat{b}}} \end{pmatrix} = (B^T B)^{-1} \qquad (14)$$

where

$$B = \begin{pmatrix} y_1 & y_1 x_1' \\ y_2 & y_2 x_2' \\ \vdots & \vdots \\ y_n & y_n x_n' \end{pmatrix}$$

estimation of the Variance and estimation of Covariance of $\hat{\hat{a}}$、$\hat{\hat{b}}$ obtained by formulas (8), (9) and (10).

Nonlinear least square regression

Although weighed linear regression is a simple process, it is only an approximation method. In this part, we present a more precise method as the nonlinear least square regression.

As for $y = ax^b$, if $i(i = 1,2,..., n)$, the dependent variable is y_i, and the measurement error is ε_i, we can derive the following formula:

$y_i = a(x_i)^b + e_i$, $(i = 1, 2,...,n)$.

If $k(k = 0, 1, 2,..., m)$, the approximate value of a, b is $a^{(k)}$, $b^{(k)}$ we can derive the following formula:

$y_i^{(k)} = a^{(k)}(x_i)^{b^{(k)}}$.

The following formula was then derived using a Taylor series:

$$y_i = y_i^{(k)} + \frac{y_i^{(k)}}{a^{(k)}}\delta_a^{(k+1)} + y_i^{(k)}\ln(x_i)\delta_b^{(k+1)} + R_i^{(k)} + \varepsilon_i$$

$R_i^{(k)}$ is the second order remainder.

make

$$\underset{n\times2}{\mathbf{A}_{(k)}} = \begin{pmatrix} y_1^{(k)}/a^{(k)} & y_1^{(k)}\ln(x_1) \\ y_2^{(k)}/a^{(k)} & y_2^{(k)}\ln(x_2) \\ \vdots & \vdots \\ y_n^{(k)}/a^{(k)} & y_n^{(k)}\ln(x_n) \end{pmatrix}, \quad \underset{2\times1}{\mathbf{X}_{(k+1)}} = \begin{pmatrix} \delta_a^{(k+1)} \\ \delta_b^{(k+1)} \end{pmatrix},$$

$$\underset{n\times1}{\mathbf{L}_{(k)}} = \begin{pmatrix} y_1 - y_1^{(k)} \\ y_2 - y_2^{(k)} \\ \cdots \\ y_n - y_n^{(k)} \end{pmatrix}, \quad \underset{n\times1}{\mathbf{R}^{(k)}} = \begin{pmatrix} R_1^{(k)} \\ R_2^{(k)} \\ \cdots \\ R_n^{(k)} \end{pmatrix}, \quad \underset{n\times1}{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

The formula can be written as:

$$\mathbf{A}_{(k)}\mathbf{X}_{(k+1)} - \mathbf{L}_{(k)} + \mathbf{R}_{(k)} + \varepsilon = 0 \qquad (17)$$

In formula (17), take $\hat{\mathbf{X}}_{(k+1)}$ for the estimated value of $\mathbf{X}_{(k+1)}$, and V_i for the most probable error of y_i, then

$$\underset{n\times1}{\mathbf{V}_{(k)}} = \begin{pmatrix} V_1^{(k)} \\ V_2^{(k)} \\ \vdots \\ V_n^{(k)} \end{pmatrix}$$

and error equation of (16) is

$$\mathbf{V}_{(k+1)} = \mathbf{A}_{(k)}\hat{\mathbf{X}}_{(k+1)} - \mathbf{L}_{(k)}, \qquad (18)$$

in which $V_{(k+1)}$ contains the observation error and the second order remainders error.

Utilize least square criterion on (18) to solve $\hat{\mathbf{X}}_{(k+1)}$:

$$\hat{\mathbf{X}}_{(k+1)} = \begin{pmatrix} \hat{\delta}_a^{(k+1)} \\ \hat{\delta}_b^{(k+1)} \end{pmatrix} = (\mathbf{A}_{(k)}^{T}\mathbf{A}_{(k)})^{-1}\mathbf{A}_{(k)}^{T}\mathbf{L}_{(k)} \quad (19)$$

Furthermore, the k+1th estimated value of a, b is

$$\begin{pmatrix} \hat{a}^{(k+1)} \\ \hat{b}^{(k+1)} \end{pmatrix} = \begin{pmatrix} \hat{a}^{(k)} \\ \hat{b}^{(k)} \end{pmatrix} + \begin{pmatrix} \hat{\delta}_a^{(k)} \\ \hat{\delta}_b^{(k)} \end{pmatrix} \quad (20)$$

Calculation steps:

1) The input coefficients ($\hat{a}^{(0)}, \hat{b}^{(0)}$) used were obtained from the linear formula, then calculate $y_i^{(0)}$, $\mathbf{A}_{(0)}$, $\mathbf{L}_{(0)}$. Use the formula (19) to get $\hat{\mathbf{X}}_{(1)}$and $\hat{a}^{(1)}, \hat{b}^{(1)}$ obtained by the formula (20).

2) Taking $\hat{a}^{(1)}, \hat{b}^{(1)}$ as new estimated value of *a* and *b*, calculate $y_i^{(1)}$, $\mathbf{A}_{(1)}$, $\mathbf{L}_{(1)}$ and use formula (19) to solve $\hat{\mathbf{X}}_{(2)}$use formula (20) to solve $\hat{a}^{(2)}, \hat{b}^{(2)}$.

3) Take $\hat{a}^{(2)}, \hat{b}^{(2)}$ as new estimated value of *a, b* to get $\hat{a}^{(3)}, \hat{b}^{(3)}$, and so on, up to $\mathbf{V}_{(k)}^{T}\mathbf{V}_{(k)}$ reach minimum or $\hat{\delta}_a^{(k+1)}$, $\hat{\delta}_b^{(k+1)}$is enough small (at this time $\mathbf{R}^{(k)} \rightarrow 0$).

Assuming that the iteration ends at the k+1th step, the standard variance formula is:

$$\hat{\sigma}_0^2 = \mathrm{V}_{(k)}^{T}\mathrm{V}_{(k)} \,/\, (n-2) \quad (21)$$

Furthermore, $\hat{\sigma}_{a_1}$, $\hat{\sigma}_{b_1}$ (estimation of the variance of $\hat{a}, \hat{b}$) and $\hat{\sigma}_{a_1 b_1}$ (estimation of the covariance of $\hat{a}, \hat{b}$) can be calculated as follow:

$$\begin{pmatrix} \hat{\sigma}_a^2 & \hat{\sigma}_{ab} \\ \hat{\sigma}_{ab} & \hat{\sigma}_b^2 \end{pmatrix} = \hat{\sigma}_0^2(\mathrm{A}_{(k)}^{T}\mathrm{A}_{(k)})^{-1} \quad (22)$$

Model evaluation

In order to determine the best allometric model, we used the adjust coefficient of determination (R^2), the root mean square error (*RMSE*), Akaike Information Criterion (*AICc*) to analyse the statistics of each regression model. Coefficient of determination (R^2) is common indicator that can tell us what percentage of the total variation in a dependent variable is explained by the predictor variables and how well the model fits the data. The larger the value of adjusted R^2, the better the regression model is (Mann 2001). The adjusted R^2 will penalize you for adding independent variables (*K* in the equation) that do not fit the model. In regression analysis, it can be tempting to add more variables to the data as you think of them. Some of those variables will be significant, but you can't be sure that significance is just by chance. The adjusted R^2 will compensate for this by that penalizing you for those extra variables (Mendenhall et al 2006)

$$Adjusted \quad R^2 = 1 - \frac{(1-R^2)(N-1)}{N-p-1}, \quad (23)$$

where *N* is the sample size, *p* is the number of independent variables, and *N-p-1* is called the degrees of freedom (*df*) of the regression.

The root mean square error (*RMSE*) (also known as the Root Mean Square Deviation, *RMSD)* was used as a standard statistical metric to measure model performance in biomass research. It is the widely way used in model evaluations. While it has been used to assess model performance for many years, there is no consensus on the most appropriate metric for model errors. In the field of biomass, many present the *RMSE* as a standard metric for model errors, the *RMSE* penalizes variance as it gives errors with larger absolute values more weight than errors with smaller absolute values. The *RMSE* is calculated for the data set as

$$RMSE = \sqrt{\mathrm{SSE/n}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\ , \quad (24)$$

where y_i is the observed dry weight, y_i is the predicted dry weight, *n* is the sample size, accounts for the unexplained or "leftover" variance of the regression model (Mendenhall et al. 2006). The smaller the *RSME*, the better the regression model is (Arrest & Franklin 2007).Meanwhile, *AIC* has been chosen as a criterion for regression model, selection. *AIC* measures the amount of information lost in the specific model, hence, the "best" model is the one with minimum *AIC* value (Chave et al. 2014, Bumham and Anderson 2002, Basuki et al. 2009). Given n/k < 40 in this study, a bias adjustment was required for small sample size, expressed as *AICc* in Equation 25:

*AICc = n*ln (RSS/n) + 2*k + (2*k*(k+1))/(n-k-1)*, (25)

where *n* is the number of samples, *k* is the number of independent variables in the model, and *RSS* is the residual sums of the squares from the regression model.

Herewith, adjusted R^2, *RMSE, AICc,* should provide sufficient justification for the strength of the allometric regression models.

Results

Data collection

The study site is located in the Xiaolong Mountains (33° 312 –34° 342 N, 104° 232 –106° 432 E) in the southeastern part of Gansu Province. It covers an area of 623,808.0 hectares. The area is warm temperate, with a semi-humid continental monsoon climate, an average annual temperature of 7 °C to 12 °C, and an annual rainfall of 460 mm to 800 mm.

Xiaolong Mountains has a forest area of 338,829.8 hm^2, a forest coverage rate of 62.46%, and a stocking volume of 25,536,825 m^3, 86.9% of which is young and middle-aged forest, and 13.1% is near mature, mature, and over-mature forest. The dominant species are mainly hardwoods, with oak trees as the majority. Sharp tooth oaks, Chinese pines, and oriental oaks make up 71.52% of the total area of the forest and 71.42% of the total volume. Mixed miscellaneous forests are also dominant in the area, which includes Northeast China ash, dogwoods, maple, linden, and cedar.

The dataset we used is a subset of the data collected in the Dang Chuan Forestry Centre of the Xiaolong Mountain Forestry Bureau. Table 1 gives the descriptive statistics of the sample trees: a total 71 felled sample trees were included in the experiments, 10 *Pinus armandii Franch*, 10 *Betulaalbo-sinensis Burk*, 19 *Pinus tabuliformis* and 32 *Quercus aliena var. acuteserrata*. DBH and height of all of the selected trees were measured after felling with a measuring tape, and the biomass of each was calculated after sampling and then after drying.

Table 1. The descriptive statistics of the sample trees

Species	Sample size	DBH* (cm)	Height (m)	Total biomass (kg)
Quercus aliena var.acuteserrata	32	4.3-47.7	4.9-24	2.5-1786.97
Pinus tabuliformis	19	4.7-40	6-17.5	3.42-538.03
Pinus armandii Franch	10	4.6-38.3	3.5-15.8	4.08-566.41
Betula albo-sinensis Burk	10	6.5-44.2	7.9-22.3	8.41-1185.49

*DBH is diameter at breast height

After sample trees' felling, the material was carefully separated into the following compartments, according to the procedure adopted by Soares and Schaeffer-Novelli (2005): leaves; branches (diameter at large end C2.5 cm); trunks; and barks. All of the compartments were weighted in the field (wet biomass). Then the compartments were oven-dried, and their dry biomass was weighted. The dry biomass of the tree was finally computed as the sum of all the compartments' dry biomass.

Allometric equations

Based on the observed biomass of the sample trees in the experimental plot, the power function of linear regression, the weighted linear regression and nonlinear regression were built, the coefficients estimates and estimated precision value of AGB are shown in Table 2.

It can be seen from Table 2 that usually the coefficients obtained using power function weighted linear regression and nonlinear least square regression are very close, but differ remarkably from the one obtained using linear regression. Also the absolute values of Variance and covariance of coefficients of weighted linear regression and nonlinear least square regression are similar, and lower than the linear regression. Therefore, the result is that the precision of coefficients of weighted linear regression and nonlinear least square regression is very close, and higher than the linear regression.

Table 2. Parameters of three allometric equations for 71 trees, 4 species at Xiaolong Mountains, China

Tree species	equations	a	b	$\hat{\sigma}_a$	$\hat{\sigma}_{ab}$	$\hat{\sigma}_b$	$\hat{\sigma}_0$
Quercus aliena var.acuteserrata	1	0.2756200	1.0426643	0.033444	-0.025373	1.240574	0.052280
	2	0.2750195	1.0957233	0.008671	-0.000189	0.025215	0.043963
	3	0.2662959	1.1148816	0.008657	-0.000197	0.026106	0.043202
Pinus tabuliformis	1	0.2063066	0.9712463	0.027343	0.014792	0.733982	0.031515
	2	0.2081705	1.0625462	0.009669	-0.000567	0.071958	0.028979
	3	0.2027038	1.0780950	0.009904	-0.000635	0.077105	0.028552
Pinus armandii Franch	1	0.2576481	0.9104142	0.086994	0.087404	1.111761	0.040712
	2	0.2816293	0.9128282	0.015492	-0.001018	0.084816	0.033165
	3	0.2744900	0.9351187	0.014779	-0.000912	0.081089	0.032644
Betula albo-sinensis Burk	1	0.2790090	1.0099986	0.077015	-0.080495	1.150311	0.062596
	2	0.2922114	0.9661464	0.027217	-0.001870	0.075183	0.062686
	3	0.2779062	0.9956368	0.026544	-0.001876	0.077041	0.061354

Equations, "1", linear regression; "2", weighted linear regression; "3", nonlinear regression; $\hat{\sigma}_a$ – estimation of Variance of $\hat{a}$; $\hat{\sigma}_b$,estimation of Variance of $\hat{b}$; $\hat{\sigma}_{ab}$, estimation of Covariance of $\hat{a}, \hat{b}$; $\hat{\sigma}_0$, the estimated value of the standard error fitted by biomass

Model evaluation results

Table 3 shows the parameter estimates and the goodness-of-fit statistics of three equations. It is obviously seen that all parameter estimates were significant at $P<0.01$. All the equations performed well for all the species. According to the model evaluation methods, the higher R^2, while the lower *RSME* and *AICc*, the better fitting the equation has. As expected, only the equation for *Larix gmelinii's RSME* and *AICc* obtained by nonlinear regression is slightly larger than the weighted regression, but smaller than simple linear regressions. For the other 3 species, the order of coefficient R^2 is common linear regression< weighted linear regression < nonlinear regression, while the order of *RSME*, while the order of *AICc* is common linear regression> weighted linear regression>nonlinear regression. Therefore, for these three methods, the nonlinear method has best fit. Our results are depicted in the following tables.

Table 3. Results of the fit of three allometric equations for 4 species to predict aboveground biomass

Tree species	equations	R^2	RSME	AIC	P
Quercus aliena var.acute serrata	1	0.9890	0.5062	255.237	<0.001
	2	0.9923	0.4256	244.248	<0.001
	3	0.9925	0.4183	243.171	<0.001
Pinus tabuliformis	1	0.9827	0.2977	132.897	<0.001
	2	0.9855	0.2732	129.814	<0.001
	3	0.9859	0.2694	129.315	<0.001
Pinus armandii Franch	1	0.9831	0.3641	75.719	<0.001
	2	0.9888	0.2966	71.619	<0.001
	3	0.9892	0.2919	71.342	<0.001
Betula albo-sinensis Burk	1	0.9883	0.5598	84.522	<0.001
	2	0.9882	0.5606	84.422	<0.001
	3	0.9887	0.5487	84.081	<0.001

R^2 – coefficient of correlation; *SEE* – standard error of the estimate; *AIC* – information criterion; *P* – significant coefficient

Discussion

Most studies on tree biomass are based on allometric models due to the difficulty of direct measurements, which involve the cutting and weighing of trees. These allometric equations are widely used in estimating tree biomass, both coniferous species, and broadleaf species, in many countries, including china. It is beneficial to contrast results with the previous analysis of Shen et al. (2011) and Du et al. (2012), who designed the allometric equations utilized in this study. In their analysis, only one species (*Larix gmelinii* for shin and *Pinus tabuliformis* for du) was used as a dataset for designing the model. According to the final prediction model, the R^2, which depicts the individual tree biomass, is 0. 968 for *Larix gmelinii,* and 0.96 for *Pinus tabuliformis*. These calculations are less precise than the R^2 calculated in this study, which was more than 0.98 for four species. Even though our sample data was a little small, there are many existing studies utilizing this allometric equation, in which the researched had used a small sample. For example, Russell (1983) weighed 15 trees in Para, Brown et al. (1995) weighed 8 trees in the Rondonia, and Deans et al. (1996) weighed 14 trees in Cameroon. Furthermore, within this study, we concentrated on coefficient precision for the power allometric equation, which may have a certain implication for future research.

Allometric model of $W = a(DBH^2H)^b$ is generally fitted using the log transformation approach, followed by the linear regression. It is important to note that the initial linear regression would not be used if heteroscedasticity was not detected. However, the data for trees are strongly heteroscedastic, exhibiting increasing variation in biomass, with an increase in diameter (Chave et al. 2005, Mascaro et al. 2011). These considerations are extremely important for evaluating heteroscedasticity, which was previously mentioned by Gujarti and Porter[40]. There are two ways in which the treatment for the heteroscedastic error, commented by Carlos R. Sanquetta et al. (2015), can be avoided. One treatment is utilizing the method of weighted least squares when heteroscedasticity is known. The other treatment involves utilizing nonlinear fitting without log transformation. These two treatments were frequently applied to the allometric equation utilized within this research.

Using DBH^2H (unit: m^3) as the X-axis, and observed biomass W (unit: kg) as the Y axis, the scatter diagram and fitting curve of AGB are shown in Figure 1. As you can see from Figure 1, the curve fitted by the weighted linear regression, was also fitted by the nonlinear least square regression in this approach. These calculations are far from the curve fitted by the common linear regression.

Within Figure 1, the red dashed line represents the curve fitted by the common linear regression, the black dashed line represents the curve fitted by weighted linear regression, the solid green line represents the curve fitted by the nonlinear least square regression, and the purple square dots represent the observed biomass value.

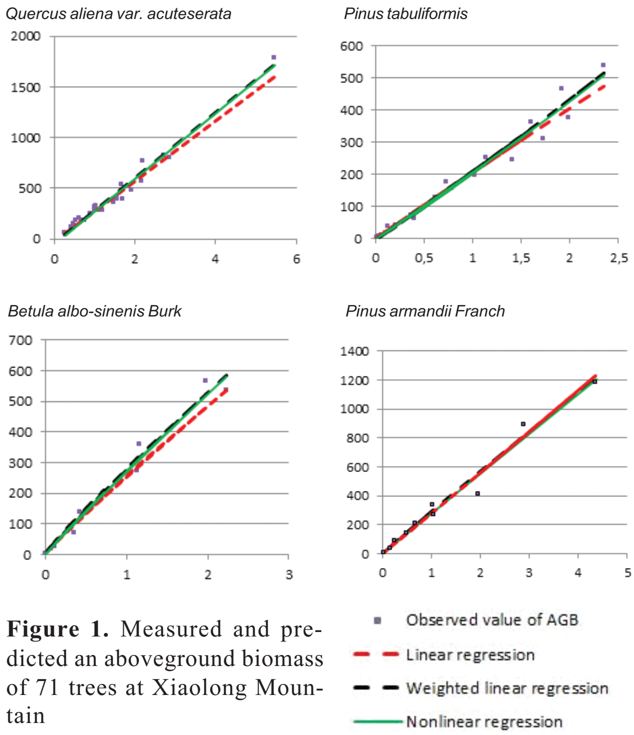

Figure 1. Measured and predicted an aboveground biomass of 71 trees at Xiaolong Mountain

The coefficients for the calculation utilizing the power model, the nonlinear least square regression method had a higher precision, but needed iterative calculations that are relatively difficult to conduct. These calculations weight the linear regression method with slightly lower precision than nonlinear regression method; however, the linear method that calculates one degree term. This is a relatively simple method because the precision for the linear regression method is commonly low.

In addition, the observational error for y should follow the same normal distribution without gross error. When studying the forest biomass utilizing the allometric equation, and constructing models with the two methods above, an emphasis needs to be placed on the samples with large diameters. If, for example, gross error of the organ's biomass exists, then great influence will result on the regression outcome.

Conclusion

This paper researches on the methods of estimating biomass of a single tree. In the region of Xiaolong Mountain, we used three methods of common linear regression, weighted linear regression, and nonlinear regression, in order to establish the allometric equations for 4 species. Our results indicate that nonlinear regression has the best precision but hardest to calculate. Therefore, when utilizing allometric equation, we highly recommend weighted linear regression out of all the possible options, which has similar accuracy but more simple compared with nonlinear regression method.

Acknowledgements

Financial support for this study was provided through National Science Fund for Distinguished Young Scholars (No.41401650), the Fundamental Research Funds for the Central University (2015ZCQ-LX-01), and Young Excellent Talent Program for the Beijing University (YETP0738). We are grateful to the undergraduate students and staff of the Laboratory of Forest Management and "3S" technology, Beijing Forestry University.

References

Brown, I.H., Martinelli, L.A., Thomas, W.W., Moreira, M.Z., Victoria, R.A. and Ferreira, C. 1995. Uncertainty in the biomass of Amazonian forests: an example from Rondônia, Brazil. *Forest Ecology and Management* 75: 175–189.

Brown, J.H., Gillooly, J.H., Allen, A.P., Savage, V.M. and West, G.B. 2004. Toward a metabolic theory of ecology. *Ecology* 85: 1771–1789.

Carmela, B.M., Timothy, A.V., Eddie, B. and Lawrence, A. 2007. Development and validation of aboveground biomass estimations for four *Salix* clones in central New York. *Biomass and Bioenergy* 31: 11–12.

Cháidez, J. J. N., Barrientos, N. G., Luna, J. J. G., Dale, V. and Parresol, B. 2004. Additive biomass equations for pine species of forest plantations of Durango, Mexico. *Madera y Bosques* 10(2): 17–28.

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Folster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.P., Nelson, B.W., Ogawa, H., Puig, H., Riera, B. and Yamakura, T. 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* 145: 87–99.

Chave, J., Rejou-Mechain, M., Burquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B. and Vielledent, G. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology* 20: 3177–3190.

Claesson, S., Sahlen, K. and Lundmark, T. 2001. Functions for biomass estimation of young *Pinus sylvestris*, *Picea abies* and *Betula* spp. From stands in Northern Sweden with high stand densities. *Scandinavian Journal of Forest Research* 16: 138–146.

Cunia, T.1987. Error of forest inventory estimates: its main components. *Proc. of the Workshop on Tree biomass regression functions and their contribution to the error of forest inventory estimates*. USDA Forest Service General Technical Report NE, 1 17: 303.

Deans, J.D., Moran, J. and Grace, J. 1996. Biomass relationships for tree species in regenerating semi-deciduous tropical moist forest in Cameroon. *Forest Ecology and Management* 88:215–225.

Delitti, W.B., Meguro, M. and Pausas, J.G. 2006. Biomass and mineralmass estimates in a "cerrado" ecosystem. *Revista Brasil Bot* 29(4):531–540.

Du, W.Z. 2012. Study on the Individual Tree Biomass of *Pinus tabuliformis* in Xiaolong Mountain, Gansu Province. *Gansu Science and Technology* 28(23):153-155

Dong, L.H., Li, F.R. and Jia, W.W. 2013. Linear mixed modeling of branch biomass for Korean pine plantation. *Chinese Journal of Applied Ecology* 24(12): 3391-3398

Feng, Z.H. 1999. The biomass and productivity of forest ecosystems in China. *Science Press, Beijing*: 2-9.

Healy, J. and Zoback, M.D. 1988. Hydraulic fracturing stress measurements in the Cajon Pass research well to 2 km depth. *Geophysical Research Letters* 15:1005-1008.

Kale, M., Sing, S., Roy, P.S., Desothali, V. and Ghole, V.S. 2004. Biomass equations of dominant species of dry deciduous forests in Shivupuri district, Madhya Pradesh. *Current Science* 87(5):683–687.

Li, H.K. and Zhao P.X. 2013. Improving the accuracy of tree-level aboveground biomass equations with height classification at a large regional scale. *Forest Ecology and Management* 289:153–163

Lott, J.E., Howard, S.B., Black, C.R. and Ong, C.K. 2000. Allometric estimation of above-ground biomass and leaf area in managed Grevillea robusta agroforestry systems. *Agroforest System* 49:1–15.

Maddala, G.S. 2001. Introduction to Econometrics. Englewood Cliffs: Prentice Hall.

Mascaro, J., Litton, C. M., Hughes, R. F., Uowolo, A. and Schnitzer, S. A. 2011. Minimizing bias in biomass allometry: model selection and log-transformation of data. *Biotropica* 43(6): 649–653.

Mary, A.A., Steven, P.H. and Thomas, G.S. 2001. Validating allometric estimates of aboveground living biomass and nutrient contents of a northern hardwood forest. *Canadian Journal of Forest Research* 31(1): 11-17.

Nelson, B. W., Mesquita, R., Pereira, S J. L. G., Souza, G. A., Batista, G.T. and Couto, L. B. 1999. Allometric regressions for improved estimate of secondary forest biomass in the central Amazon. *Forest Ecology and Management* 117: 149–167.

Norris, D.N., Blair, J.M., Johnson, L.C. and McKane, R.B. 2001.Assessing changes in biomass, productivity, and carbon stores following *Juniperus virgiana* forest expansion into tallgrass prairie. *Canadian Journal of Forest Research* 31: 1940–1946.

Parresol, B.R. 1999. Assessing tree and stand biomass: A review with examples and critical comparisons. *Forest Science* 45: 573–593.

Parresol, B.R. 2001. Additivity of nonlinear biomass equations. Canadian. *Journal of Forest Research* 31: 865–878.

Reed, D.D. and Green, E.J. 1985. A method of forcing additivity of biomass tables when using nonlinear models. Canadian. *Journal of Forest Research* 1: 1184–1187.

Russell, C. 1983.Nutrient cycling and productivity of native and plantation forests at Jari Florestal, Para, Brazil. PhD thesis, University of Georgia, Athens.

Saatchi, S.S., Houghton, A, Dos Santos, Alvala, R.C., Soare, J.V. and Yu, Y. 2007. Distribution of aboveground biomass in the Amazon. *Global Change Biology* 13: 816–837.

Saint-Andre' L., M'bou Mabiala, A., Mouvondy, W., Jourdan, C., Roupsard, A., Deleporte, P., Hamel, O. and Nouvellon, Y. 2005. Age-related equations for above- and below-ground biomass of Eucalyptus hybrid in Congo. *Forest Ecology Management* 205: 199–214.

Saglan, B., Kucuki, O., Bilgili, E., Durmaz, D. and Basal I. 2008. Estimating fuel biomass of some shrub species (Maquis) in Turkey. *Turkish Journal of Agriculture and Forestry* 32: 349–356.

Sanquetta, C. R., Wojciechowski, J., Dalla Corte, A. P., Behling, A., Péllico Netto, S., Rodrigues, A. L. and Sanquetta, M. N. I. 2015. Comparison of data mining and allometric model in estimation of tree biomass. *BMC Bioinformatics* 16: 247-56.

Schlaegel, B.E. 1982. Testing, reporting, and using biomass estimation models. In: Gresham, C.A. (ed.): *Proc. of the 3rd Annual Southern Forest Biomass Workshop*. Belle W. Baruch Forest Sci. Institute, Clemson University, Clemson, SC. 13: 95-l 12.

Shen, Y.Z., Sun, X.M. and Zhang, J.T. 2011. Study on the Individual Tree Biomass of *Larix kaempferi* Plantation in Xiaolong Mountain. *Gansu Province. Forest Research,* 24 (4): 517-522.

Tang, S.Z. and Wang, Y.H. 2002. A parameter estimation program for the error-in-variable model. *Ecological Modelling* .156:225-223.

Siddique, M. R. H., Hossain, M. and Chowdhury, R. K. 2012.Allometric relationship for estimating above-ground biomass of *Aegialitis rotundifolia* Roxb. of Sundarbans mangrove forest, in Bangladesh. *Journal of Forestry Research* 23: 23–28.

Soares, M. L. G. and Schaeffer-Novelli, Y. 2005 Aboveground biomass of mangrove species. I. Analysis of models. *Estuarine, Coastal and Shelf Science* 65: 1–18.

Tang, S.Z., Li, Y. and Wang, Y.H 2001. Simultaneous equations, error-in-variable models, and model integration in systems ecology. *Ecology Modelling* 142: 285–294.

Thomas, G. C. and John, J. E. 2006. Allometric equations for four valuable tropical tree species .*Forest Ecology and Management* 229: 351–360.

Van, T.K., Rayachhetry, M.B. and Centre, D. 2000. Estimating aboveground biomass of *Melaleuca quinquenenervia* in Florida, USA. *Journal of Aquatic Plant Management* 38: 62–67.

Wadham-Gagnon, B. and Sharpe, D. 2006. Estimating carbon stocks in tropical hardwood plantations: using species-specific and non-destructive parameters to estimate aboveground biomass for six native species in Panama. Internship.

Wang, Z.F and Feng, Z.K. 2006. On the problem and improved measure in some quasi-linearization regression with one argument. *Journal of Northeast Normal University (Natural Science Edition)*.38(4):45-52.

Xu, H. 2003. A Comparison between CAR and VAR Biomass Models. *Journal of Southwest Forestry College* 23(2): 36-39.

Williams, M.S. and Gregoire, T.G. 1993. Estimating weights when fitting linear regression models for tree volume. *Canadian Journal of Forest Research* 23: 1725-1731.

White, H.A. 1980. Heteroskedasticity consistence covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica*.48: 817–838.

Williams, C.J., Lepage, B.A., Vann, D.R., Tange, T., Ikeba, H. and Ando, M. 2003. Structure, allometry, of plantation *Metasoquoia glyptostroboides* in Japan. *Forest Ecology Management* 180 (1–3): 287–301.

Received 16 December 2016
Accepted 02 October 2017