

Allometric Equation for Biomass Estimation of *Pouteria adolfi-friederici* (Engl.) Baehni in Gerba Dima Forest, South western Ethiopia

Abyot Dibaba Hundie¹, Teshome Soromessa², and Bikila Warkineh³

¹Debre Berhan University

²Addis Ababa University

³Addis Ababa University College of Natural Sciences

February 22, 2024

Abstract

Abstract: This study was conducted in Gerba Dima Forest, South Western Ethiopia to formulate allometric equations for *Pouteria adolfi-friederici*. Prior to collecting data for tree allometry, the study forest was stratified into 3 forest strata based on altitudinal variation. By employing semi-destructive technique, 30 individuals were systematically selected and sampled for measuring biomass along the three forest strata. Based on the data collected, several equations were developed. Before establishing the allometric equation, scatter plots were used to see whether the relationship between independent and dependent variables was linear. Furthermore, several allometric relationships between independent and dependent variables were tested. The best-fit model developed was validated by testing the regression assumptions. AGB was regressed against the various forms of predictors (i.e., DBH, H and WD) and three allometric models showed significant performance ($p < 0.05$) on their F-test. Among the three allometric equations which showed significant performance, the selection of the best-fit model was conducted based on their P-value, adjusted r^2 , AIC, RMSE. The two models are nested to the third model and hence the complete model, $\ln AGB = 1.806 + 1.419 \times \ln DBH + 0.628 \times \ln WD$, is selected as the best-fit model against the other two nested models since the p-values of coefficients of the complete model are significant ($p < 0.05$).

Allometric Equation for Biomass Estimation of *Pouteria adolfi-friederici* (Engl.) Baehni in Gerba Dima Forest, South western Ethiopia

Abyot Dibaba^{1*}, Teshome Soromessa², Bikila Warkineh³

¹College of Natural Sciences, Department of Biology, Debre Berhan University, P. O. Box 445, Debre Berhan, Ethiopia, e-mail: abyotdibaba77@yahoo.com

²Center for Environmental Science, Addis Ababa University, and P. O. Box No: 1176, Addis Ababa, Ethiopia, e-mail: soromessa@yahoo.com

³College of Natural Sciences, Department of Plant Biology and Biodiversity Management Addis Ababa University, P. O. Box 3434, Addis Ababa, Ethiopia, e-mail: kooket@gmail.com

Abstract

This study was conducted in Gerba Dima Forest, South Western Ethiopia to formulate allometric equations for *Pouteria adolfi-friederici*. Prior to collecting data for tree allometry, the study forest was stratified into 3 forest strata based on altitudinal variation. By employing semi-destructive technique, 30 individuals were systematically selected and sampled for measuring biomass along the three forest strata. Based on the data collected, several equations were developed. Before establishing the allometric equation, scatter plots were

used to see whether the relationship between independent and dependent variables was linear. Furthermore, several allometric relationships between independent and dependent variables were tested. The best-fit model developed was validated by testing the regression assumptions. AGB was regressed against the various forms of predictors (i.e., DBH, H and WD) and three allometric models showed significant performance ($p < 0.05$) on their F-test. Among the three allometric equations which showed significant performance, the selection of the best-fit model was conducted based on their P-value, adjusted r^2 , AIC, RMSE. The two models are nested to the third model and hence the complete model, $\ln \text{AGB} = 1.806 + 1.419 \times \ln \text{DBH} + 0.628 \times \ln \text{WD}$, is selected as the best-fit model against the other two nested models since the p-values of coefficients of the complete model are significant ($p < 0.05$).

Key words/Phrases: Allometric equation, Model fitting, Model validation, Regression

1. Background

Pouteria adolfi-friederici is a tall tree of high forest, up to 50 m tall, with the long, straight, more or less fluted trunk, and with buttresses at the base. It is an emergent or upper canopy species in montane moist forest and grows between 1350-2450m altitudinal ranges. The Phytogeographic distribution of the species includes Ethiopia, Sudan, Uganda, Kenya, the Democratic Republic of Congo, Zambia, Malawi, and Zimbabwe (Hedberg et al., 2003). This species is a high-value timber tree used for all types of woodwork. Although classified as 'not threatened' on a global basis, the species and the high forest associated with it is rapidly being cut down and very little is being done to either conserve representative forests or replant the species in Ethiopia (Tesemma, 1993; Hedberg et al., 2003).

Allometric equations relate simply recorded variable (e.g. tree diameter, height) to other structural and useful features (Niklas, 1994), which is the most common and reliable method for estimating biomass, net primary production, and biogeochemical budgets in forest ecosystems (Gower et al., 1999). They have been developed to satisfy various purposes in forest ecology and management. Diameter at breast height (DBH) is commonly used as the only independent variable and most allometry develops an allometric relationship between DBH and component biomass (Gower et al., 1999). Some investigations proposed to incorporate tree height (H) as the second predictor and produce the DBH–H combined equation to increase the accuracy of biomass estimates (Ketterings et al., 2001). The allometric technique estimates the entire or partial (by compartments) mass of a tree from measurable tree dimensions, including trunk diameter and height, using allometric equations (Kangas and Maltamo, 2006). Thus, dendrometric parameters of all the trees are measured and using the allometric equation the biomass of the stand is estimated by summing the biomass of individual trees.

Generalized models have immense potential for large-scale carbon budgets computed from inventory data, (Pastor et al., 1984). When estimating the above-ground biomass of a forest, the uses of species-specific equations are preferred because trees of different species may differ greatly in tree architecture and wood density. However, due to the great number of different tree species in humid tropical rainforests and enormous efforts needed to develop these equations. Species-specific allometric equations for the humid tropics are almost absent while relatively few mixed-species equations have been developed (Ketterings, et al., 2001). This study was aimed to develop a species-specific allometric equation for *Pouteria adolfi-friederici*, which can help the precise estimation of forest carbon stock in south western Ethiopia.

2. Material and methods

2.1 The study area

This study was conducted in Gerba Dima forest found in south western Ethiopia between $7^{\circ} 45'$ to $8^{\circ} 10'$ North latitude and $35^{\circ} 29'$ to $35^{\circ} 50'$ East longitude. The studied forest is part of the mountainous highlands west of the Great Rift Valley and is situated on undulating and dissected mountain ranges between 1582m to 2285m altitudinal ranges.

Figure 1. Map of the study area

The rainfall data collected from Gore meteorological station indicated that the study area receives very high annual rainfall. The climate diagram constructed using 20 years of climate data of the nearest Gore meteorological station shows unimodal rainfall pattern with the monthly mean maximum and mean minimum temperature of the area is 27.2⁰C and 13.3 ⁰C, respectively. The mean annual temperature is 19.2⁰C and with slight variation from year to year (Figure 2). The mean annual rainfall of the study area is 1854mm. The rainfall pattern exhibits low rainfall in December, January and February, steadily rising to the peak period in August. The vegetation type at the studied site is part of the moist evergreen Afromontane forest, which is characterized by one, or more closed strata of evergreen trees that may attain a height of 30 to 40 m. The characteristic emergent species that form the upper canopy include *Pouteria adolfi-friederici* , *Albizia gummifera*, *A. schimperiana*, *A. grandi bracteata*, *Sapium ellipticum*, *Ekebergia capensis*, *Ficus sur*, *Hallea rubrostipulata*, *Olea welwitschii*, *Polyscias fulva* and *Schefflera abyssinica* .

Figure 2. Climate diagram of Gore

2.2 Biomass Data collection

The studied forest was stratified into three forest strata based on the altitudinal variation. Strata one was found at an altitudinal range of between 1500-1800m while strata two and three were located between the altitudinal range of 1801-2000m and 2001-2300m respectively. Following the methodologies recommended in Picard *et al* ., (2012), area and species-specific allometric equations were formed involving a semi-destructive approach. Thirty individuals were systematically selected and sampled for measuring biomass along the three forest strata. In order to minimize the error of sampling and to obtain the representative tree and samples, diameter distributions were taken into account during tree selection. Accordingly, the trees were classified into five DBH classes and each class having six individuals per DBH class ranging from 10-20 cm, 20.1-30 cm, 30.1-40 cm, 40.1-50 cm and greater than 50 cm. Experienced persons climbing on live trees using ropes did the measurement of diameter at 2m interval in each section of a tree. Using data obtained by climbing live trees, aboveground biomass of trunk, branches and leaves were estimated. Generally, three random small branches with less than 10cm basal diameter per individual plant were destructed and trimmed. The measurements of fresh biomass (in kg) were divided into two parts: measuring trimmed fresh biomass and measuring untrimmed fresh biomass.

Trimmed fresh biomass

The three small branches destructed per tree were trimmed using a machete. A tree caliper was used to determine the diameter at the base of each branch. The basal diameters were determined at the first point at which the branch became cylindrical, typically within 5 cm of the trunk then the leaves were separated from the trimmed branches. The fresh biomass of the leaves from the trimmed branches (Btrimmed fresh leaf) and the fresh biomass of the wood from the trimmed branches (Btrimmed fresh wood) were determined (by weighing separately). Three random samples of the leaves from three different trimmed branches were taken to constitute the aliquot. Its fresh weight (B aliquot fresh leaf in g) was measured. Without debarking, an aliquot of the wood was also taken at random from the trimmed branches. Its fresh mass was measured (B aliquot fresh wood in g) in the field, immediately after cutting. Samples about 10 cm long were taken from twigs and 6 disks were taken from the destructed three branches. These aliquots were placed in numbered plastic bags and sent to the laboratory. The subsamples were then dried at 65⁰C for 48 h or until a constant weight was obtained to determine the moisture content. The fresh volume of the wood aliquot was measured in the lab, and the value was used to determine mean wood density. These data (dry biomass of trimmed component) were used to develop regressions between basal diameter and trimmed branch component weight (trimmed branches, twigs and leaves) (Vann *et al* ., 1998; Picard *et al* ., 2012).

Untrimmed fresh biomass

Untrimmed biomass is measured indirectly as non-destructive. The small untrimmed branches with less than 10cm basal diameter were processed differently from the large branches and the trunk. For the small branches, only basal diameter was measured. The biomass of these small untrimmed branches with their accompanying twigs and leaves was estimated from the relationship between their basal diameter and the

regression developed from the trimmed branches (Vann *et al.* , 1998; Picard *et al.* , 2012).

The biomass of the trunk and the large branches was estimated from measurements of volumes (V_i in cm^3) and mean wood density ($\bar{\rho}$ in g cm^{-3}). The large branches and trunk were divided virtually into sections that are then materialized by marking the tree. The volume V_i of each section i was obtained by measuring its diameter and its length. Diameter measurements were carried out in sections of about two meters in length along the length of the trunk. For larger branches greater than 10 cm in basal diameter, diameter measurements were conducted in sections of 50 cm in length using tree caliper.

The volume (V aliquot fresh wood) of the wood aliquot taken from the trimmed compartments was measured by measuring the volume of water displaced when the sample is immersed in water using a graduated tube of suitable dimensions for the sample (Picard *et al.* , 2012). The dry biomass of the tree was obtained by the sum of the trimmed dry biomass and the untrimmed dry biomass:

$$B_{\text{dry}} = B_{\text{trimmed dry}} + B_{\text{untrimmed dry}}$$

2.3 Data analysis

Calculating trimmed biomass

From the fresh biomass $B_{\text{fresh wood}}^{\text{aliquot}}$ of a wood aliquot and its dry biomass $B_{\text{dry wood}}^{\text{aliquot}}$, the moisture content of the wood (including bark) was calculated as follow:

$$X_{\text{wood},i} = \frac{B_{\text{aliquot dry wood},i}}{B_{\text{aliquot fresh wood},i}}$$

Likewise, the moisture content of the leaves was calculated from the fresh biomass $B_{\text{fresh leaf},i}^{\text{aliquot}}$ of the leaf aliquot and its dry biomass.

$$X_{\text{leaf},i} = \frac{B_{\text{dry leaf},i}^{\text{aliquot}}}{B_{\text{fresh leaf},i}^{\text{aliquot}}}$$

Trimmed dry biomass can then be calculated:

$$B_{\text{trimmed dry}} = B_{\text{trimmed fresh wood}} \times X_{\text{wood}} + B_{\text{trimmed fresh leaf}} \times X_{\text{leaf}}$$

Where $B_{\text{trimmed fresh leaf}}$ is the fresh biomass of the leaves stripped from the trimmed branches and $B_{\text{trimmed fresh wood}}$ is the fresh biomass of the wood in the trimmed branches (Picard *et al.* , 2012).

Calculating untrimmed biomass

Two calculations were required to calculate the dry biomass of the untrimmed part (i.e. that still standing): one for the small branches accompanied with leaves and twigs and the other for the large branches. The untrimmed biomass is the sum of the two results:

$$B_{\text{untrimmed dry}} = B_{\text{untrimmed dry branch}} + B_{\text{dry section}}$$

Each section i of the large branches was considered as a cylinder of volume (Smalian's formula):

$$V_i = \pi/8 L_i (D_{1i}^2 + D_{2i}^2)$$

Where V_i is the volume of section i , L_i its length, and D_{1i} and D_{2i} are the diameters of the two extremities of section i . The dry biomass of the large branches is the product of mean wood density and the total volume of the large branches:

$$B_{\text{dry section}} = \bar{R} \times \sum_i V_i$$

Where the sum corresponds to all the sections in the large branches, and where mean wood density is calculated by:

$$R = \frac{B_{\text{dry wood}}^{\text{aliquot}}}{V_{\text{fresh wood}}^{\text{aliquot}}}$$

The dry biomass of small untrimmed branches, twigs and leaves was calculated from the regression model developed between the basal diameter of the trimmed branches and biomass of trimmed branches, twigs and leaves. This model is established by following the same procedure as for the development of an allometric model. Power type equations were used:

$$B_{\text{dry branch}} = a + bD^c$$

Where a, b and c are model parameters and D branch basal diameter. Using a model of this type, the dry biomass of the untrimmed branches and their components (twigs and leaves) is:

$$B_{\text{untrimmed dry branch}} = \sum_j (a + bD_j^c)$$

Where the sum is all the untrimmed small branches and their components and D_j is the basal diameter of the branch j (Picard *et al.*, 2012).

Calculating the biomass of the trunk

The truncated cone volume formula was used instead of the cylinder formula, for larger trees their diameter measured at two meters interval. For each log i, the circumferences at the two extremities were measured: circumference C_{1i} is the circumference of the log taken at the bottom end and circumference C_{2i} is the circumference of the log taken at the upper end. This was conducted to calculate the volume of the fresh log using the truncated cone volume formula (or Newton's formula):

$$V_{\text{frais},i} = L_i \times \frac{\pi}{3} \times (R_{1i}^2 + R_{1i}R_{2i} + R_{2i}^2)$$

Where L_i is the length of log i, and $R_{1i} = \frac{C_{1i}}{2\pi}$; $R_{2i} = \frac{C_{2i}}{2\pi}$ are the radii of log i at its two extremities. The dry biomass of the trunk is the product of mean wood density and total volume of the trunk

$$B_{\text{dry section}} = \bar{\rho} \times \sum_i V_i, \text{ where the sum corresponds to all the sections in the trunk (Picard } et al., 2012).$$

Formulation of allometric equation

Based on the data collected, several equations were developed. Before establishing the allometric equation, scatter plots were used to see whether the relationship between independent and dependent variables was linear. Furthermore, several allometric relationships between independent and dependent variables were tested. The independent variables included DBH, Height, and wood density, whereas the dependent variable was the dry weight of the AGB. Because the data exhibited heteroscedasticity, a power function was an inappropriate model in this study, so the data was transformed for linear regression using logarithmic transformation. The transformation equalized the variance over the entire range of biomass values, which satisfies the prerequisite of linear regression (Sokal and Rohlf, 1995; Sprugel, 1983).

To develop an allometric model for the above-ground biomass, Diameter at breast height ((DBH), Height (H) and Wood density (WD) were taken as explanatory variables. Given that there are p effect variables X_1, X_2, \dots, X_p , there are $2^p - 1$ models that include all or some of these effect variables (Picard *et al.*, 2012). Thus, in this study, the 3 effect variables can result in $2^3 - 1$ which becomes 7 candidate regression model (Table 1). Table 1. The seven candidate regression model which serve to select the best models

Model	Allometric equations
1	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \epsilon$
2	$\ln(\text{AGB}) = c + \beta \ln(H) + \epsilon$

Model	Allometric equations
3	$\ln(\text{AGB}) = c + \chi \ln(\text{WD}) + \varepsilon$
4	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \beta \ln(\text{H}) + \varepsilon$
5	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \chi \ln(\text{WD}) + \varepsilon$
6	$\ln(\text{AGB}) = c + \beta \ln(\text{H}) + \chi \ln(\text{WD}) + \varepsilon$
7	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \beta \ln(\text{H}) + \chi \ln(\text{WD}) + \varepsilon$

where: c is the intercept, and α, β and χ are the slope coefficient of the regression.

Model comparison and selection were based on average deviation (Brand and Smith, 1985; Chave *et al.* , 2005), slope coefficient of the regression (Nelson *et al.* , 1999), Akaike Information Criterion (AIC) (Burnham and Anderson, 2002; Chave *et al.* , 2005), confidence interval (CI) of the predictions, paired t-test and Coefficients of determination (r^2).

Model validation

The best-fit models developed for trees and shrubs have been validated by testing the regression assumptions. Model validation confirms that the model complies with underlying assumptions. All regression-type techniques are based on a series of assumptions, those of prime importance is the residuals are independent, residuals follow a normal distribution and residuals variance is constant (homoscedasticity).

Testing the assumption of Normality and Homoscedasticity

A quantile-quantile graph was plotted to check the residuals follow a normal distribution and whether the cluster of points form a straight line was visually verified. To check that the variance of the residuals is constant, we can plot the residuals against the fitted values and verify visually that the cluster of points does not show any particular trend (Picard *et al.* , 2012). Furthermore, the scale-location plot and each points leverage plot were also constructed to check the homoscedasticity assumption and the existence of possible outliers respectively.

Test for multicollinearity

Multicollinearity is the strong correlation between the independent variables. Some correlation between them is highly expected as all variables are ‘causing factors’ to tree biomass. However, ‘strong correlation’ causes ‘strong multicollinearity’ by which the true effect of estimated regression coefficients would be lost. Thus, we can no longer depend on standard statistical tests. In such circumstances, a variable that has less explanatory power with the dependent variable must be removed from the regression model. Nevertheless, while doing so, there should not have strong collinearity between the remaining variable with other potential explanatory variables. A variance inflation factor (VIF) quantifies how much the variance is inflated. It is a measure of how much the variance of the estimated regression coefficient b_k is “inflated” by the existence of correlation among the predictor variables in the model. A VIF of 1 means that there is no correlation between the k^{th} predictor and the remaining predictor variables, and therefore the variance of b_k is not inflated at all. The general rule of thumb is that VIFs exceeding 4 warrant further investigations, while VIFs exceeding 10 are signs of serious multicollinearity requiring correction (Belsley *et al.* , 1980).

3. Results

Trimmed and untrimmed Biomass

The average wood and leaf aliquot moisture contents were 0.61% and 0.53%. Mean trimmed fresh leaf biomass, mean trimmed fresh wood biomass, mean trimmed dry leaf biomass, mean trimmed wood dry biomass and the overall mean trimmed dry biomass were 6.02kg, 7.23kg, 3.19kg, 4.41kg and 7.60kg respectively. The dry biomass of small untrimmed branches, twigs and leaves was calculated from the linear regression model developed between the biomass of trimmed branches, twigs, leaves and the basal diameter of trimmed

branches (equation 1) and accordingly the mean basal diameter was 7.25cm and the mean untrimmed biomass for small branches was 9.01kg.

$$B = 547.5 \text{ BD} + 15382, r^2 = 0.92 \dots\dots\dots \text{ (Equation 1)}$$

The mean wood density was 0.45 gcm⁻³ while the mean untrimmed biomass for large branches calculated from the volume of each section of a branch obtained by Smalian's formula of a tree was 105.78kg. The mean biomass of trunks was 1543±312kg while the mean DBH, Height and the mean overall aboveground biomass (AGB) were 69.29±7.97cm, 22.87±0.9m and 1785.12±344kg respectively.

The person's correlation analysis between above-ground biomass and with biomass predictor variables (DBH, Height and wood density) showed that the above-ground biomass was strongly correlated with DBH while Height is the second important factor correlating with biomass showing moderate correlation. However, the AGB was not significantly correlated with wood density (Table 2).

Table 2. Pearson's product moment correlations coefficient and P value between the response variable (AGB) and explanatory variables(WD, DBH and H)

	AGB	WD	DBH	H
AGB	1	0.287	0.961**	0.539**
		0.124	0.000	0.002
WD	0.287	1	0.257	0.205
	0.124		0.171	0.276
DBH	0.961**	0.257	1	0.623**
	0.000	0.171		0.000
H	0.539**	0.205	0.623**	1
	0.002	0.276	0.000	

**. Correlation is significant at the 0.01 level.

Model fitting

AGB was regressed against the various form of predictors (i.e., DBH, H and WD) and three allometric models showed significant performance ($p < 0.05$) on their F-test (Model 1, 3 and 6)(Table 3). Among the three allometric equations which showed significant performance, the selection of the best-fit model was conducted based on their P-value, adjusted r^2 , AIC, RMSE. The first two models (models 1 and 3) are nested to model 6 and hence the complete model 6 is selected against the nested models 1 and 3 since the p-values of coefficients of the complete model are significant ($p < 0.05$). Model 6 selected as the best-fit model is indicated below.

$$\ln AGB = c + \alpha \ln DBH + x \ln WD + \varepsilon, \text{ with values of coefficient inserted the model form is: } \ln AGB = 1.806 + 1.419 \times \ln DBH + 0.628 \times \ln WD \dots\dots\dots \text{ (Equation 2)}$$

$$(r^2 = 0.951, \varepsilon = 0.219, AIC = -1.026, RMSE = 0.208)$$

Where AGB is in kg, DBH in cm and WD in cm³. The ANOVA also showed that the best-fit model was found to be statistically significant ($F=283.042, p < 0.000$).

Model Number	Allometric equation	Coefficients	Coefficients	Standard Error of coefficients	Adjusted r^2	Residuals standard Error	AIC	RMSE
1	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \epsilon$	Symbol	Value	0.268	0.945	0.233	1.594	0.225
2	$\ln(\text{AGB}) = c + \beta \ln(\text{H}) + \epsilon$	c	1.123***	0.065	0.536	0.676	65.549	0.653
		α	1.462***					
3	$\ln(\text{AGB}) = c + \chi \ln(\text{WD}) + \epsilon$	c	-3.270	1.759	0.122	0.930	84.706	0.898
		β	3.318***	0.565				
4	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \beta \ln(\text{H}) + \epsilon$	c	9.193***	0.975	0.944	0.235	0.223	0.223
		χ	2.668 *	1.190				
5	$\ln(\text{AGB}) = c + \beta \ln(\text{H}) + \chi \ln(\text{WD}) + \epsilon$	c	0.694	0.671	0.565	0.655	64.550	0.621
		α	1.411 ***	0.099				
6	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \chi \ln(\text{WD}) + \epsilon$	β	0.204	0.293	0.951	0.219	-1.026	0.206
		c	-1.332	2.056				
7	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \beta \ln(\text{H}) + \chi \ln(\text{WD}) + \epsilon$	β	3.074 ***	0.566	0.950	0.222	0.463	0.206
		c	***	0.566				
8	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \beta \ln(\text{H}) + \chi \ln(\text{WD}) + \epsilon$	χ	1.460	0.867	0.951	0.219	-1.026	0.206
		c	1.806***	0.410				
9	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \beta \ln(\text{H}) + \chi \ln(\text{WD}) + \epsilon$	α	1.419 ***	0.065	0.950	0.222	0.463	0.206
		c	***	0.065				
10	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \beta \ln(\text{H}) + \chi \ln(\text{WD}) + \epsilon$	χ	0.628 *	0.296	0.950	0.222	0.463	0.206
		α	1.411	0.722				
11	$\ln(\text{AGB}) = c + \alpha \ln(\text{DBH}) + \beta \ln(\text{H}) + \chi \ln(\text{WD}) + \epsilon$	β	0.621 *	0.299	0.950	0.222	0.463	0.206
		χ	0.621 *	0.299				

Model Number	Allometric equation	Coefficients	Coefficients	Standard Error of coefficients	Adjusted r ²	Residuals standard Error	AIC	RMSE
Significant codes: 0 (***)	Significant codes: 0 (***)	Significant codes: 0 (***)	Significant codes: 0 (***)	Significant codes: 0 (***)	Significant codes: 0 (***)	Significant codes: 0 (***)	Significant codes: 0 (***)	Significant codes: 0 (***)
0.001 (***) 0.01 (*) 0.05 (.) 0.1 (.)	0.001 (***) 0.01 (*) 0.05 (.) 0.1 (.)	0.001 (***) 0.01 (*) 0.05 (.) 0.1 (.)	0.001 (***) 0.01 (*) 0.05 (.) 0.1 (.)	0.001 (***) 0.01 (*) 0.05 (.) 0.1 (.)	0.001 (***) 0.01 (*) 0.05 (.) 0.1 (.)	0.001 (***) 0.01 (*) 0.05 (.) 0.1 (.)	0.001 (***) 0.01 (*) 0.05 (.) 0.1 (.)	0.001 (***) 0.01 (*) 0.05 (.) 0.1 (.)
1	1	1	1	1	1	1	1	1

Table 3. Model description for the estimation of the above-ground biomass of *Pouteriaadolphi-friederici* in Gerba Dima forest

Model Validation

Testing the Independence Assumption

The Durbin-Watson statistical test was conducted for the occurrence of serial correlation between residuals which can be used to check the independence of residuals for the best-fitted model. The Durbin-Watson statistics obtained for the model selected was 1.942.

Testing the assumption of Normality and Homoscedasticity

A quantile-quantile graph was plotted in R- software for the best-fitted model and visually verified that almost the cluster of points forms a straight line. Homoscedasticity characterized by a constant variance of the residuals was also checked by plotting the residuals against the fitted values in R software and visually verified that more or less the cluster of points does not show any particular trend (Figure 3).

Figure 3. Residuals plotted against fitted values (upper-left), quantile-quantile plot (lower left), scale-location plot (upper - right) and plots for each point leverage (lower-right)for best fitted model of *Pouteriaadolphi-friederici*

4. Discussion

Fitting Allometric Regression Models

The above-ground biomass was strongly correlated with DBH so that DBH is the most influential factor affecting the biomass of the trees. Height is the second important factor correlating with biomass showing a strong correlation coefficient while the biomass was poorly correlated with wood density. Thus, if there is no multicollinearity between DBH and H, the combination of these two variables gives the best regression model. Conducting a correlation test between response and explanatory variables is an important step to select an appropriate explanatory variable and to develop the best-fit regression model (Maraseni *et al.* , 2005).

Model fitting for *Pouteria adolphi-friederici* involves choosing the best model among the 3 statistically significant candidate models. The third complete model (Model six) is selected against the first two nested models since the p-values of coefficients of the complete model are significant ($p < 0.05$). In the case of nested models, a statistical test can be used to test one of the models against the other. The null hypothesis of this test is that $\emptyset = 0$, i.e. the additional terms are not significant, which can also be expressed as the nested model is better than the complete model. If the p-value of this test proves to be below the significance level (typically 5 %), then the null hypothesis is rejected, i.e. the complete model is best. Conversely, if the p-value is above the significance threshold, the nested model is considered to be the best (Picard *et al.*,

2012). Since the best-fit model selected for *P. adolfi-friederici* involves two predictor variables, the absence of Multicollinearity was also checked using the Variance Inflation Factor. The VIF value obtained was 1.8 for each predictor variable and this value assures that multicollinearity is not a problem in this model. A strong correlation between independent variables causes ‘strong multicollinearity’ by which the true effect of estimated regression coefficients would be lost. The general rule of thumb is that VIFs exceeding 4 warrant further investigations, VIFs exceeding 10 are signs of serious multicollinearity requiring correction while VIFs less than 4 indicates an absence of multicollinearity problem (Belsley *et al.*, 1980). The overall accuracy of the model is 95.11% with an overall p-value < 0.001, in this model the Adjusted R-squared (0.9511) is closer to 1, showing that 95.11% of the variability of the above-ground biomass is explained by this model. The ANOVA test conducted also shows that the overall allometric equation was found to be statistically significant ($F=283.042$, $P < 0.001$). **Model validation** The best-fit regression models developed were validated based on a series of assumptions, those of major importance are the residuals are independent, residuals follow a normal distribution and residuals variance is constant (homoscedasticity). Violation of these assumptions may result in biased parameter estimates and type I errors (Quinn & Keough 2002; Picard *et al.*, 2012). The Independence Assumption was tested by Durbin-Watson statistics which has values ranging between 0 and 4. However, the residuals are considered not correlated (independent) if the Durbin-Watson statistic is between 1.5 and 2.5 (Field, 2009). The Durbin-Watson statistics of models is less than 2.5, indicating that the residuals for the model selected are uncorrelated; therefore, the independence assumption is met in this study. The independence assumption is a significant assumption that should be investigated prior to any interpretation of multiple regression analysis, as a violation of this assumption could hold critical implications (Stevens, 2009). Even a slight violation of the independence assumption should be taken seriously, as it can greatly increase the risk of Type 1 error, resulting in the risk of falsely rejecting the null hypothesis several times greater than the level of error assumed for the test (Stevens, 2009). The assumption of normality and homoscedasticity was tested by thoroughly investigating a quantile-quantile graph (Q-Q plot) and the residuals against the fitted values plot developed. The residual errors plotted versus their fitted values for the best-fit model indicated that the residuals are randomly distributed around the horizontal line representing a residual error of zero; that is, no distinct trend in the distribution of points is observed and confirms the homoscedasticity assumption of the model. The standard Q-Q plot of the best-fit models also suggests that the residual errors are normally distributed.

Screening for normality is an important early step when conducting multiple regressions, as residuals are normally distributed is assumed (Stevens, 2009; Tabachnick & Fidell, 2006). Non-normal distributions that are positively or negatively skewed, contain large kurtosis or have extreme outliers which can distort the obtained significance levels of the analysis, resulting in the standard errors becoming biased (Osborne & Waters, 2002).

The assumption of homoscedastic indicates that the variance of errors is equal and constant across all levels of the variables (Osborne & Waters, 2002; Stevens, 2009). Homoscedasticity is related to the assumption of normality because when the assumption of normality is met, the relationship between the variables is homoscedastic (Tabachnick & Fidell, 2006). Heteroscedasticity occurs when the variance of errors differs at different values of the independent variables (Osborne & Waters, 2002). Slight heteroscedasticity has little effect on significance tests; however, when heteroscedasticity is marked it can lead to serious distortions of findings and seriously weaken the analysis thus increasing the possibility of a Type 1 error for a small sample size (Osborne & Waters, 2002).

The scale-location plot for all models shows the square root of the standardized residuals (sort of a square root of relative error) as a function of the fitted values. Again, there is no obvious trend in this plot which further confirms the absence of heteroscedasticity. Finally, the plot in the lower right (Figure 3) shows each point’s leverage, which is a measure of its importance in determining the regression result. Superimposed on the plot are contour lines for the Cook’s distance, which is another measure of the importance of each observation to the regression. Smaller distances mean that removing the observation has little effect on the regression results. Distances larger than 1 are suspicious and suggest the presence of a possible outlier or a poor model. In this study, the model selected exhibits Cook’s distance of less than 1 which confirms the

absence of possible outliers.

5. Conclusion

The species-specific and area specific allometric model developed for *P. adolfi-friederici* has good quality as evaluated by testing the regression assumptions. In the regression models developed, DBH, H and wood density were predictors in the estimation of the Above ground biomass.

Acknowledgements

We would like to thank Addis Ababa University for providing logistics for fieldwork and laboratory analysis. IDEA WILD is acknowledged for donating equipments used for field data collection. The International Foundation for Science financially supported the study.

Data Accessibility Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request and will be deposited in Dryad repository.

References

1. Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: John Wiley.
2. Brand, G.J., Smith, W.B. (1985). Evaluating allometric shrub biomass equations fit to generated data. Canadian Journal of Botany 63: 64–67.
3. Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D. & Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145: 87–99.
4. Field, A.P. (2009). Discovering statistics using SPSS: and sex and drugs and rock ‘n’ roll (3rd edition). London: Sage.
5. Gower, S.T., Kucharik, C.J., Norman, J.M., (1999). Direct and indirect estimation of leaf area index, F(APAR), and net primary production of terrestrial ecosystems. Remote Sens. Environ. 70: 29–51.
6. Hedberg, I., Edwards, S., Nemomissa, S., editors (2003). Flora of Ethiopia and Eritrea. Apiaceae to Dipsaceae, vol. 4(1). Addis Ababa: The National Herbarium, Addis Ababa University.
7. Kangas A. and Maltamo M. (2006). Forest inventory: Methodology and Applications, Managing Forest Ecosystems, Vol. 10. Springer, Dordrecht, the Netherlands.
8. Ketterings, Q.M., Coe, R., Van Noordwijk, M., Ambagau, Y., Palm, C.A. (2001). Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. For. Ecol. Manage 146: 199–209.
9. Maraseni, T.N., Cockfield, G., Apan, A. and Mathers, N. (2005). Estimation of Shrub Biomass: Development and Evaluation of Allometric Models Leading to Innovative Teaching Methods. International Journal of Business & Management Education Special Issue: Postgraduate Research in Innovative Methods of Teaching & Learning ISSN 1832- 0236.
10. Nelson, B.W., Mesquita, R., Pereira, J.L.G., de Souza, S.G.A., Batista, G.T., Couta, L.B., (1999). Allometric regressions for improved estimate of secondary forest biomass in the Central Amazon. Forest Ecology and Management 117: 149–167.
11. Niklas, K.J., (1994). Plant Allometry: The Scaling of Form and Process. University of Chicago Press, Chicago, Illinois.
12. Osborne and Waters (2002). Four assumptions of multiple regression that researcher should always test. Practical Assessment, Research & Evaluation, 8(2).

13. Pastor, J., Aber, J.D., Melillo, J.M., (1984). Biomass prediction using generalized allometric regressions for some northeast tree species. *For. Ecol. Manage* 7: 265–274.
14. Picard N., Saint-Andre L., Henry M. (2012). Manual for building tree volume and biomass allometric equations: from field measurement to prediction. Food and Agricultural Organization of the United Nations, Rome, and Centre de cooperation Internationale en Recherche Agronomique pour le Development, Montpellier, 213 pp.
15. Quinn, G.P. & Keough, M.J. (2002). *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge.
16. Sokal, R.R., Rohlf, F.J., (1995). *Biometry: The Principles and Practice of Statistic in Biological Research*, 3rd ed. W.H. Freeman and Co., New York, 859 pp.
17. Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations. *Ecology* 64 (1): 209–210.
18. Stevens, J.P. (2009). *Applied multivariate statistics for the social sciences*, 5th ed. New York: Routledge.
19. Tabachnick, B. & Fidell, L. (2006). *Cleaning up your act screening data prior to analysis. Using multivariate analysis*, 5th ed. Needham Heights, MA: Allyn & Bacon.
20. Tessema, A.B.(1993). *Useful Trees and Shrubs for Ethiopia: Identification, Propagation and Management for Agricultural and Pastoral Communities*. Regional Soil Conservation Unit (RSCU) Swedish International Development Authority (SIDA) Embassy of Sweden, Nairobi, Kenya.
21. Vann, D. R., Palmiotto, P. A., Strimbeck, G. R. (1998). Allometric equations for two South American conifers: Test of a non-destructive method. *Forest Ecology and Management* 106: 55–71.