Development and Validation of Height-Diameter Models for *Cupressus lusitanica* in Gergeda Forest, Ethiopia

Mengesha Tsega\(^a\), Awoke Guadie\(^b\), Zebene Lakew Tefera\(^a\), Yigez Belayneh\(^a\) and Dongjie Niu\(^c\)

\(^a\)College of Environmental Science and Engineering (CESE), Tongji University, Shanghai 200092, China; \(^b\)Department of Biology, College of Natural Sciences, Arba Minch University, 21, Arba Minch, Ethiopia; \(^c\)State Key Laboratory of Pollution Control and Resource Reuse, College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China

**ABSTRACT**

This study was conducted to fit and evaluate ten existing nonlinear height diameter functions for *Cupressus lusitanica* in Gergeda forest Ethiopia. A total of 260 trees were measured for their diameter at breast height (D) and height using destructive sampling methods. This data were randomly split into two datasets for model development (50%) and model validation (50%). Considering heteroscedasticity of variance, all functions were fitted using weighted nonlinear least square regression. To evaluate the performance of each model, five fit statistics—such as Coefficient of determination (R\(^2\)), root mean square error (RMSE), bias (E), absolute mean deviation (AMD), and coefficient of variation (CV\%)—were used. Among all the models tested, the Weibull-type function of the form H = 1.3 + a (1-exp (-bD\(^c\)))+c was observed to give the best fit based on the model’s goodness of fit and predictive ability. Therefore, this model with three parameters has been confirmed to provide reliable estimate of total tree height for *Cupressus lusitanica* in Gergeda forest, Ethiopia.

**Introduction**

The total height of a tree is the linear distance from ground level to the upper tip of the tree crown, through its main axis (Téo et al. 2017). It is a variable of fundamental importance for quantifiable description of a tree and stand. Models describing the relationship between total tree height and diameter at breast height are extremely valuable tools for forest management planning (Uzoh 2017). Because of their importance for a number of forest stand modelling applications, height-diameter (H-D) equations have received considerable attention (Burkhart and Tome 2012). Since H-D relationship has a tendency to differ among sites, stand densities, ages and other factors, allowing the relationship to vary by stand stand and growth period may yield more accurate results (Gonda 1998). In addition to these, H-D relationships of tree species vary in geographic region, depending on local climate, soil, altitude and ecological conditions (Peng et al. 2001; Zhang et al. 2002). Therefore, site specific H-D models are required to predict tree heights from field measurements of tree diameters.

Diameter at breast height of a tree can be measured easily, quickly and accurately with most cost effective; but measurements of total tree height is relatively complex, time consuming and expensive (Sharma and Parton 2007) especially in tropical forests where sloping land, tall trees, irregular shaped tree stems and crowns, several layers and dense canopy occur (Vanclay 1992; Téo et al. 2017). Therefore, H-D model which relates height to diameter of the tree is an alternative solution to such problems.

In practice of forest inventory, tree stand and site condition may prevent accurate height measurement. In such cases diameter at breast height for all samples are measured and height is measured for only subsamples of trees selected across the range of diameter classes (Huang et al. 1992; Calama and Montero 2004; Vibrans et al. 2015). H-D functions are then used to estimate the missing heights of trees for which only diameter measurements are taken (Larsen and Hann 1987; Sharma and Parton 2007). Generally, the use of mathematical modelling becomes interesting when the measurement of dependent variables is costly and the independent variables are easily obtainable. Although the use of such models are alternatives for studying the relationships between these variables and to assess vertical structure patterns of forest communities, they may not always provide exact predictions for situations that more accurate height values are needed (Vibrans et al. 2015). Considering these conditions, this study has been designed to fit and evaluate the relative performance of some commonly used H-D functions for predicting the height of *Cupressus lusitanica* tree in Gergeda forest, Ethiopia.
Materials and methods

Study area

Ethiopia is a country located near the equator between 3°24' to 14°53'N latitude and 32°42' to 48°12'E longitude in the North-eastern part of Africa (Balcha 1999). In Ethiopia, there are 58 forest priority areas (Mekonen et al. 2017), among which Gergeda forest is one. It is located in Oromia regional state of Ethiopia in Kellem Wollega zone about 670 km West of Addis Ababa (capital city of Ethiopia) between 34°31'01.600 to 34°54' E longitude and 8°37'12.700 to 8°58'40.300 N latitude (Figure 1). The altitude range of the area is between 1492 to 3273 meters above sea level.

Climatic elements such as precipitation, temperature, wind, humidity and sunshine are strongly affected by altitude and geographic locations. Ethiopia which is found near the equator has a wide range of climatic features suitable for diverse natural vegetation and agricultural products. Therefore, climatic information from nearby meteorological station (Dembi Dollo station) with complete metrological data was collected and presented.

In addition to the last three years data (2015-2017) from the local station, precipitation and temperature data of the study area was obtained from Ethiopian National Metrology Agency (ENMA). These data (precipitation and temperature data from 1984-2017) were extracted, compiled and presented as climadiagram of Gergeda Forest (Figure 2).

According to the data from the ENMSA, the long term mean annual rainfall is calculated to be 1268.48 mm and showed unimodal pattern. It has one long rainy season which extends from March to November and one short dry season extending from December to February (Figure 2). Typically, the first two months receive small amount of rain during the onset of the rainy season and gradually reach to its peak in May, June, July and August. More than 71% of the mean annual rain fall is recorded from May to September. The mean monthly maximum temperature is about 25.91°C. The mean maximum temperature reaches to its peak during February with a temperature of 28.32°C. The lowest mean annual monthly temperature occurs during December with a temperature of 11.23°C (Figure 2).

The occurrence and distribution of plants are considerably influenced by the chemical composition and physical properties of the soil. According to the soil Atlas of Africa (Jones et al. 2013), the soils of western Ethiopia, where the Gergeda forest is found, falls into four major soil types. These are alisols, fluvisols, nitisols and leptosols. However, the soil type found in the sample area falls into leptosols and nitisols soil types as they are the most common soil types found on the western and southern parts of the forest.

Figure 1. Location of Gergeda forest.
Gergeda forest possesses 10 out of the 24 national priority tree species that are endangered (Albizia gumi-mifera, Celtis africana, Cordia africana, Croton macro-stachyus, Ekebergia capensis, Juniperus procera, Olea welwitschii, Pouteria adolf-friederici, Prunus africana, and Syzygium guineense subsp. afromontanum) (Dessalegn and Ali 2017). Among the Gergeda plantations, Eucalyptus and Cupressus species are the dominant ones. The major silvicultural systems used in Gergeda forest for the establishment of Cupressus lusitanica are weeding, thinning and pruning. Thinning and pruning schedules had no regular tending operations in the forest before the establishment of Oromia Forest and Wildlife Enterprise (OFWE) (2009). Nevertheless, special attention was given and regular tending operations were set for these activities after the establishment of the Enterprise. Accordingly, high pruning (pruning up to 8 m) and first and second thinning are being carried out in the stands of less than 20 years. First and second thinning are done at ages of 6 and 12 years by removal of 48 and 54 m³ stocks/ha respectively (Bekele 2011). However, for older trees (>20 years) thinning and pruning activities had not been done in many forest stands. Rather naturally self-pruned trees and some suppressed trees were observed while drying and being removed from the stands.

**Data collection and analysis methods**

The data used in this study were collected by destructive sampling. Stratified random sampling and proportional allocation procedure was used to allocate sample plots based on the area of stratum. The stratification of the forest was made based on the age class and topographic features. According to this division, the forest stand of Cupressus lusitanica (602 ha) was divided into three strata: Stratum I (age between 1-10 years with an area of 216 ha), Stratum II (age between 11-20 years with an area of 195 ha) and Stratum III (age between 21-30 with an area of 191 ha). In a total of 52 circular sample plots, 260 sample trees (93, 84 and 83 trees from age classes of 1-10, 11-20 and 21-30, respectively) were taken for measurement. The sample allocation according to the size of stratum was estimated by equation 1 (Laar and Akça 2007):

\[
    n_h = \frac{N_h}{N} \cdot n
\]

Where: \(n_h\) = number of sample units to be drawn from \(h\)th stratum, \(n\) = total number of sample units in the strata, \(N_h\) = total size of stratum \((h = 1, 2, 3)\), \(N\) = total size of all strata

Since circular plots are easier to establish (West 2015), circular plot shape was used. A total of 52 circular plots with the radius of 11.28 m were laid out and data collection was made on five trees nearest to the center of the sample plots (Teshome 2005). Diameters at the breast height of 260 sample trees were measured by calliper while the trees were standing. All the sample trees were then felled and measured for their heights by measuring tape from 30 cm stamp to the tip (Sharma and Breidenbach 2015). The data collected were then randomly divided into fitting (50%) and validation (50%) datasets with sample size of each 130 trees. The summary of the datasets were presented in Table 1.

**Table 1. Summary of fitting and validation datasets of Cupressus lusitanica.**

| Statistics | Fitting Data (50%) | Validation Data (50%) |
|------------|--------------------|-----------------------|
| Number of observation | 130 | 130 |
| Mean | 20.04 | 20.11 |
| Minimum | 7.80 | 8.80 |
| Maximum | 35.40 | 35.20 |
| Std. deviation | 7.4823 | 7.4454 |

Where: \(D\) = Diameter at breast height, \(H\) = height

Height is usually defined by specific models, where the tree height \(H\) is the function of its diameter \(D\) (Petřáš et al. 2014):

\[
    H = f(D)
\]

The selection of H-D functions was based on the examination of height diameter relationship as revealed by plotting height against diameter (Huang et al. 1992). This is because the main focus in these studies...
lies in finding the best functional form of the model. Based on this examination, concave shaped scatter plot was observed which can best fit with typical sigmoid functions and second degree polynomials (Figure 3).

Considering the plot of height versus diameter, nonlinear functions having similar graph were selected from different literatures for testing and validation. Lists of these candidate model forms are presented in Table 2.

These model forms have been used in several studies for different tree species in various countries (Huang et al. 1992; Lumbres et al. 2013; Stanikova and Diéguez-Aranda 2013). Parameter estimation for all the H-D models was done by nonlinear curve fit procedure of Origin pro 9.0 software (OriginLab 2012, 2013). Weighted least square method was also used for correcting the heteroscedasticity of the variance. Except model 7 which was weighted by a weight factor $\omega = 1/D^2$, all the equations were weighted by instrumental weight of Origin which uses a weight factor $\omega = 1/\sigma_i^2$ where $\sigma_i$ are values from the rows of the error column.

In order to evaluate the performance of these models, five common fit statistics were used; including the coefficient of determination ($R^2$), root mean square error (RMSE), bias (E), absolute mean deviation (AMD), and coefficient of variation (CV%) (Equations 3–7). To identify the best model, rank analysis was used. The model with the lowest values for AMD, RMSE and CV% had the best rank. For $R^2$, the model with a value closest to 1 had the best rank. Whereas for E, values closest to zero was considered best and negative values of E indicates that the model over predicts and vice versa. These statistical criteria were computed as follows:

\[
R^2 = 1 - \frac{\sum_{i=1}^{n}(h_i - \hat{h}_i)^2}{\sum_{i=1}^{n}(h_i - \bar{h})^2} \tag{3}
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n}(h_i - \hat{h}_i)^2} \quad \tag{4}
\]

\[
E = \frac{1}{n} \sum_{i=1}^{n}(\hat{h}_i - \bar{h}) \tag{5}
\]

\[
AMD = \frac{1}{n} \sum_{i=1}^{n}|\hat{h}_i - \bar{h}| \tag{6}
\]

\[
CV\% = \frac{1}{\bar{h}} \times 100 \sqrt{\sum_{i=1}^{n}(h_i - \bar{h}_i)^2} \tag{7}
\]

Where: $h_i =$ observed height for the $i^{th}$ tree, $\hat{h}_i =$ predicted height for the $i^{th}$ tree, $\bar{h} =$ observed mean tree height, $n =$ number of observations and $p =$ the numbers of model parameters

**Results**

**Initial model fitting and model validation**

In this study, among the total sample ($n = 260$), 50% of the datasets were used to fit ten candidate H-D models. As shown in Table 3, the candidate H-D models have two to three parameters. For all the models, the parameters were estimated and since the ten H-D equations were appropriately selected based on the shape of the scatter plot distribution of height versus diameter (Figure 3), all the models were found to be

---

**Figure 3.** Height versus Volume diameter of the combined data.

**Table 2.** A list of candidate H-D models.

| Model code | Model form | References |
|------------|------------|------------|
| 1          | $H = 1.3 + a/(1 + b^{-D^p})$ | Ratkowsky and Reedy 1986, Huang et al 1992 |
| 2          | $H = 1.3 + aD/(b + D)$ | Huang et al 1992 |
| 3          | $H = 1.3 + 10^pD^b$ | Huang et al 1992 |
| 4          | $H = 1.3 + D^p/(a + bD)$ | Huang et al 1992 |
| 5          | $H = 1.3 + aD/(1 + Di)^p$ | Huang et al 1992 |
| 6          | $H = a + bD + cD^2$ | Trorey 1932, Vrbans et al 2015 |
| 7          | $H = 1.3 + \exp(a + b/(D + 1))$ | Stanikova and Diéguez-Aranda 2013 |
| 8          | $H = 1.3 + \exp(a + b/D)$ | Adapted form Model 7 |
| 9          | $H = 1.3 + a(1 - \exp(-bD^2))$ | Weibull 1951; Huang et al 1992 |
| 10         | $H = a\exp(b/D)$ | Calama and Muntero 2004 |

---

**Table 3.** Estimated parameters of H-D models using 130 samples.

| Model | Estimated parameters |
|-------|----------------------|
| a     | b                     | c                     |
| 1     | 24.62906894           | 0.00544312            | 2.01782213 |
| 2     | 43.21458413           | 32.30073028           |            |
| 3     | 0.41232842            | 0.61391796            |            |
| 4     | –1.51619353           | –0.16873295           |            |
| 5     | 31.65878551           | 12.85003169           |            |
| 6     | –2.83729790           | 1.55169702            | –0.02400437|
| 7     | 3.49734339            | –14.05079392          |            |
| 8     | 3.43545343            | –12.10895084          |            |
| 9     | 21.36011909           | 0.01072310            | 1.68870984 |
| 10    | 31.95236474           | –11.18570521          |            |
best fit to the dataset significantly at $p = 0.05$. The coefficient of determination ($R^2$) values for all the fitted functions on weighted observations were also ranged from 0.93 to 0.99 (data not shown).

After the model fittings were done, the remaining 50% ($n = 130$) datasets were used for initial H-D model validation. The five fit statistics such as $R^2$, RMSE, E, AMD and CV% were used to evaluate the performance of each model. The values for fit and validation statistics of these models were presented in Table 4.

The H-D validation results showed that models 7 and 2 are the most biased models. The bias value of both is negative which means the models over predict the estimation of the tree height. Six of the ten candidate models were also observed with negative bias meaning over prediction of the estimation and four are observed with positive bias meaning under prediction. The coefficients of determinations of the models vary from 0.83 to 0.89. It has been reported that the $R^2$ value closer to 1 is better and otherwise poor (Lumbres and Lee 2013; Lumbres et al. 2013). Therefore, the results of developed models are well fitted since the $R^2$ values are closer to 1. All the models have small RMSE. The larger RMSE is 1.59018 for model 3 and the smaller is 1.25548 for model 9. Model 9 is the candidate model developed from the Weibull function (Weibull 1951). The result of the AMD also varied from 0.00767 to 0.01267.

Based on this performance evaluation and rank analysis, model 9 which has been designed based on the Weibull type function (model 9), the best to describe the H-D relations. The sigmoidal function that is the Weibull type function (model 9), the quadratic function (model 6) and the modified logistic function (model 1), gave the most satisfactory results. This is similar to the work reported by other scholars (Huang et al. 1992; Lumbres et al. 2013; Zhang et al. 2014). In this study, the examination and comparison of the weighted values of these parameters showed that the concave and sigmoidal function are the best to describe the H-D relations. The sigmoidal function that is the Weibull type function (model 9), the quadratic function (model 6) and the modified logistic function (model 1) gave the most satisfactory results. This is similar to the work reported by other scholars (Huang et al. 1992; Lumbres et al. 2013; Zhang et al. 2014; Lumbres et al. 2015). For instance, Huang et al. (1992) compared nonlinear H-D functions for major Alberta tree species (16 species in 9 groups) and found that the Weibull and the generalized logistic

Table 4. Values for fit and validation statistics of the fitted H-D models for Cupressus lusitanica.

| Model code | $R^2$ | RMSE   | E    | AMD  | CV% | $\sum E$ Rank |
|------------|-------|--------|------|------|-----|----------------|
| 1          | 0.89332 | 1.28554 | 0.00771 | 1.02637 | 7.49816 | 16 3 |
| 2          | 0.86265 | 1.45869 | 0.01895 | 1.14577 | 8.47476 | 45 10 |
| 3          | 0.83677 | 1.59018 | 0.00576 | 1.25548 | 9.23968 | 42 9 |
| 4          | 0.87732 | 1.37859 | 0.00763 | 1.07559 | 8.00939 | 34 7 |
| 5          | 0.88474 | 1.32200 | 0.01196 | 1.05534 | 7.73867 | 30 6 |
| 6          | 0.89714 | 1.62631 | 0.00962 | 1.01391 | 7.36263 | 13 2 |
| 7          | 0.88060 | 1.36004 | 0.03638 | 1.08704 | 7.90161 | 39 8 |
| 8          | 0.88679 | 1.23434 | 0.01267 | 1.04970 | 7.69421 | 27 5 |
| 9          | 0.89774 | 1.25862 | 0.01365 | 1.00877 | 7.34108 | 12 1 |
| 10         | 0.88741 | 1.32071 | 0.02019 | 1.04765 | 7.67310 | 17 4 |

The two (fit and validation) datasets were combined for final model fitting. The parameter estimates and their standard errors were determined for model 9. The results showed that a, b, and c had parameter values 21.37802984, 0.01215534, and 1.64351889, respectively (Table 5). Residual analysis and predictive ability of the best model (model 9) is also presented in Figure 5.

**Discussion**

The relationship between tree height and diameter is one of the most important elements of forest structure (Zhang et al. 2014). In this study, we developed and validated H-D models for Cupressus lusitanica using Gergeda Forest (Ethiopia). For model development and validation purpose, we used five fit statistics such as $R^2$, RMSE, E, AMD and CV%. Previous studies stated that for any appropriate H-D model, the asymptotic $t$-statistic for each coefficient has to be significant, the model RMSE has to be small and the standardized residual plot should show approximately homogeneous variance over the full range of predicted values (Huang et al. 1992). In the current study, the examination and comparison of the weighted values of these parameters showed that the concave and sigmoidal function are the best to describe the H-D relations. The sigmoidal function that is the Weibull type function (model 9), the quadratic function (model 6) and the modified logistic function (model 1) gave the most satisfactory results. This is similar to the work reported by other scholars (Huang et al. 1992; Lumbres et al. 2013; Zhang et al. 2014; Lumbres et al. 2015). For instance, Huang et al. (1992) compared nonlinear H-D functions for major Alberta tree species (16 species in 9 groups) and found that the Weibull and the generalized logistic

Figure 4. Graph of predicted height (m) versus diameter (cm) of the best three H-D models.

Table 5. Parameter estimates and their standard errors of the final model (model 9) using combined data ($n = 260$ samples).

| Parameter | Parameter estimates | Standard error | 95% LCL | 95% UCL |
|-----------|---------------------|---------------|--------|--------|
| a         | 21.37802984         | 0.04177545    | 21.29576405 | 21.46092563 |
| b         | 0.01215534          | 0.00022311    | 0.01171598 | 0.01259469 |
| c         | 1.64351889          | 0.00761806    | 1.62857172 | 1.65852067 |
functions were the models which gave most satisfactory results. Zhang et al. (2014) also fitted H-D functions using classical and Bayesian method for Chinese fir (Cunninghamia lanceolata) of age 24 years and 26 years using samples of 445 and 424 trees respectively and found that the Weibull function was the best for describing H-D relationship of Chinese fir for both classical and Bayesian method. Moreover, Lumbres et al. (2013) fitted different H-D models using six nonlinear growth functions for Pinus kesiya in Benguet Province, Philippines using data collected by non-destructive method from pine forest of Benguet provinces in CAR, Philippines. It was observed that the modified logistic nonlinear growth function was the best model in predicting the total height of Pinus kesiya, specifically in Benguet province, Philippines.

On the other hand, the parabolic H-D function (model 6), first presented by Trorey (1932) and latter supported by Ker and Smith (1955) was found to be one of the best fit models (second in overall rank) to the dataset. However, this function has its own drawback in using because the equation would give decreasing values of height for values of diameter greater than \( b/2c \). This ratio is the theoretical maximum diameter and beyond this height is assumed constant and the function gives a close representation of the relation between height and diameter (Trorey 1932). In the beginning, Trorey found that parabolic equation of form \( H = a + bD - cD^2 \) can be used to describe the H-D relationships of many forest stands. He assumed that the constant “a” is the breast height and only the ascending portion of the calculated curve was applicable and once the maximum value was reached the predicted height remained constant. Re assessment and extensive tests of this work by other scholars showed that the parabolic curve was reasonable expression of the vast majority of H-D relationships(Ker and Smith 1955). The same holds true in our case and when the scatter plot of height versus diameter of the fitting dataset was plotted for H-D function selection, the shape of the plot revealed polynomial and consequently become the best fit in regression with significant values of parameters at \( p = 0.05 \) level. This function constantly works up to \( D = b/2c \) and starts decreasing in the form of quadratic equation \( H = a + bD - cD^2 \). In our case, the function can be used as follows: The value of \( b \) and \( c \) are estimated to be 1.55169702 and 0.02400437 respectively and maximum diameter \( (D = b/2c) \) is equal to 1.55169702/2*0.02400437 which gives 32.32cm. Therefore this function can be used to estimate height for trees less than 32.32cm and beyond this diameter size the predicted height has to be assumed constant.

Conclusion

This study presents an assessment of H-D relationships with ten existing H-D models using Cupressus lusitania stands in Gergeda forest. Based on the different performance evaluations, the Weibull function of this study provides more satisfactory results as compared to the polynomial, the modified logistic and other exponential functions. The local H-D functions which are only dependent on tree diameter can normally be applied to the stand where the data were gathered. Therefore the models developed and validated could provide more reliable estimate of height for Cupressus lusitania in Gergeda forest by optimizing cost and time spent on forest inventory.

Acknowledgement

The research was financially supported by the National Natural Science Foundation of China (Grant No.21677112). The authors thank Oromia Forest and Wildlife Enterprise (OFWE) Kellem Wollega District for permitting the data collection from Gergeda forest following destructive methods. The College of Environmental Science and Engineering, Tongji University also deserve thanks for providing necessary facilities during the study period.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Balcha G. 1999. The status of forest genetic resources in Ethiopia. A paper presented at the regional training workshop
