Transactions of the Nigerian Association of Mathematical Physics
Vol. 6 (Jan., 2018), pp. 247– 262

In honour of Prof. Ekhaguere at 70

Estimating tree growth parameters from existing height-diameter models

K.O. Obisesan[a*] and O.J Odewole[b]
[a,b]*Department of Statistics, University of Ibadan, Nigeria*

Abstract. Diameter at breast height is one of the easily measured variables with much lesser error of measurement when compared to tree height. This study presents five set of existing height-diameter equations for two species of tree. These species are the *gmelina*, also called white beech, and the *pine* specie. 189 tree stands of the latter and 130 tree stands of the former were considered. A non-linear least squares approach was adopted in estimating the growth parameters from the selected models. Curtis (1967) H-D model performed best among the five selected models in estimating tree growth parameters for *pine* specie. This was determined by the values of RSE and AIC. The Curtis(1967) model had RSE= 2.76 and AIC= 924.29 which are the least among others. Also, the Michaelis-Menten model performed very well next to the Curtis model for *gmelina* specie because both models had very close RSE and AIC. From the exploratory data analysis carried out, it was observed that height of *pine* tree species increases even when the diameter remains unchanged at a point. A series of tests was carried out to check if errors generated by the selected models were normally distributed. Shapiro-Wilk and Kolmogorov-Smirnov statistic established the normality of the residuals.

Keywords: gmelina, pine, diameter, least squares, RSE, AIC, model.

1. Introduction

The main aim of this study is to estimate the tree growth parameters from some existing height-diameter models in other to investigate the model that best estimates these parameters.

Forest, in a broader sense, can be classified into two: Animal forest and Plant forest. It could also be classified as either equatorial evergreen rainforest or moist forest. But looking at it beyond our narrow, human, not to mention urban, perspectives; forest provides habitats to diverse animal species and it also forms the source of livelihood for many different human settlements as well as for governments.

Forest resources are natural resources that are renewable, and many of these resources are richly and fairly distributed in various part of Nigeria. According to Badejo et al(2008), Nigeria is naturally endowed with vast forest land where a wide variety of wood producing tree species are found. The forest and woodlands in Nigeria play a major role in providing economic, ecological and social benefits and also supply numerous forest products and services for man's consumption and utilization.

In Nigeria, the tropical rainforest ecosystem is the major source of timber supply to various wood-based industries (Akindele and Akinsanmi, 2002). Regardless of the numerous advantages of forest, Nigeria had lost most of the forest cover; this may be due to mismanagment resulting from lack of proper care and planning, deforestation, drastic increase in the population, e.t.c.

Over the past fifty (50) years, about half of the world's original forest cover has been lost, the most significant cause for that is human beings' unsystematic use of resources (UNFAO, 2010). When we take the forest, it is not just the trees that go; the entire ecosystem begins to fall apart with dire consequences for all of us.

Tree height remains an intriguing component of a comprehensive growth and yield model owing to its significant place in volume estimation and its great contributions to reasonable valuations of standing tree species. This importance is much more substantial when its relationship is verged on interaction between diameter at breast height (d.b.h) which is 1.3m or 4.5ft above the ground

*Corresponding author. Email: Obidairo@gmail.com

(Leduc and Goelz, 2009).

Considering that the diameter at breast height (d.b.h) can be more accurately obtained, and at lower cost than total tree height, only a sub-sample of heights is usually measured in the field. Height-diameter equations are then used to predict the heights of the remaining trees, thus reducing the cost of data acquisition. For these reasons, developing suitable height-diameter equations (models) may be considered one of the most important elements in forest design and monitoring (Peng, 2001).

2. Review of literature

Forests, generally, play very important roles not only in timber, mining and recreation sectors, but also in global carbon cycles and climate change (K. C. Colbert, et al 2002). Tree height and diameter are the most commonly measured variables for estimating tree growth and volume (Leduc and Goelz 2009).

The Weibull model was first introduced by Ernst Hjalmar Waloddi Weibull in 1951. Initially it was described as a statistical distribution. It has many applications in population growth, agricultural growth, height growth and is also used to describe survival in cases of injury or disease or in population dynamic studies. In 1997, Lianjun Zhang used this model to describe tree height-diameter data of ten conifer species. In the paper by Fekedulegn et al, this model was used for the top height data of Norway spruce from the Bowmont Norway spruce Thinning Experiment. Colbert et al have tried to define some character developments such as forest trees height growth and diameter development by using the model.

In the paper by Karadavut et al, the Curtis model was used to evaluate the relative growth rate of silage corn. Ozel and ertekin in 2011 studied the chapman growth model and applied it to the oriental beach Juvenilities growth. Weibull model was also used to study the height growth of Pinus radiate by Colff and Kimberley in 2013. Lumbres et al used this model to describe the diameter at breast height of Pinus Kesiya.

In ARPN Journal of science and technology, the equations for predicting total tree height for eight rain forest species in Nigeria, including Mansards model, were presented. The species of interest in the study included Allanblackia floribunda,Anonidium mannii, Celtis Zenkeri, Diospyrous suavelens, Hylodendrum gambunense, Gossweilerodendron balsamiferon, Guarea cedrata and Strombosia pustulata. It was observed that only two species (Gossweilerodendron balsamiferum and Hylodendron gabunense) of the same family (Caesalpinioideae) significantly differ from each in their fit statistics; with Gossweilerodendron balsamiferum having R^2 value of 0.795 and RMSE value of 0.319 while Hylodendron gabunense had 0.873 and 0.267 forR^2 and RSME respectively.

This variation was said to be related to the wide difference in the abundance of the two species. Similar trends were observed in the works of Lootens et.al 2007; Larsen and Hann 1987 and Colbert et.al 2002 on heightdiameter relationships. This development also showed that fitted model coefficients are similar in sign and magnitude as well as conformed to the findings of various studies that had once used Monserud's equation is other geographical region (e.g Lootens et.al 2007).

3. Methodology

3.1 *Height-diameter models*

Several model forms are frequently used for estimating height-diameter relationships(Huang and others 2000, Trincado and Leal 2006). For the purpose of this study, five basic H-D models were considered. The table below shows the description of these selected models:

Name	Model	Reference
Chapman	$H = 1.3 + D^2/(a + b * D + c * D^2)$	lmfor package for H-D equations
Wykoff	$H = 1.3 + (a * D/(D + 1)) + b * D$	Wykoff et al(1982)
Curtis	$H = 1.3 + a * (D/1 + D)^b$	Curtis(1967)
Power	$H = 1.3 + a * D^b$	lmfor package for H-D equations
Michaelis-menten	$H = 1.3 + (D/a + b * D)$	lmfor package for H-D equations

Where H is the total height, D is the diameter at breast height and a,b and c represent the parameters to be estimated.

However, the models were chosen due to the following reasons

- H= 1.3 when D= 0
- The functions are simple, yet quite flexible in form.
- The parameters have reasonable biological interpretation.
  a = asymptotic maximum height (maximum height obtainable)
  b= exponential decay
  c= shape parameter (Huang and others, 2000)

Accurate height-diameter equations are a valuable tool for forest managers. While diamter at breast height(DBH) is easily and accurately measured, heightis time consuming and prone to error (Arabatzis and Burkhart 1992, Colbert and others, 2002, Trincado and Leal 2006).

While several equations are frequently used, many of these equations are applicable only to a limited geographical area or limited range of stand ages(Trincado and Leal, 2006). As a result, Height-Diameter equations are often classified as local or general(Arabatzis and Burkhart, 1992, Trincado and Leal, 2006).

General height-diameter equations typically modify a local equation by incorporating additional variables such as stand age, site index, dominant height, max-height or other stand variables(Huang and others, 2000).

3.2 *Non-linear Least Squares*

It is not new that many problems encountered by the experimental scientists are formulated in terms of determining the values of the parameters in a regression function of this form:

$$y = h\left(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_m\right) \tag{3.1}$$

$$Q = \sum_{i=1}^{n} \left[f_i(y_i, \theta)\right]^2 \tag{3.2}$$

$$f = y_i - h_i \tag{3.3}$$

where y is a response variable,$(x_1, x_2, \ldots, x_n)$ are the explanatory variables and $(\theta_1, \theta_2, \ldots, \theta_m)$ are the parameters to be estimated. When the relationship between y and x is not linear in parameter and in variables, we need to apply iterative techniques in estimating the parameters involved. In this case, we define a gradient matrix A as

$$A = \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & \frac{\partial f_1}{\partial \theta_2} & \cdots & \frac{\partial f_1}{\partial \theta_m} \\ \frac{\partial f_2}{\partial \theta_1} & \frac{\partial f_2}{\partial \theta_2} & \cdots & \frac{\partial f_2}{\partial \theta_m} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial \theta_1} & \frac{\partial f_n}{\partial \theta_2} & \cdots & \frac{\partial f_n}{\partial \theta_m} \end{bmatrix}$$

By differentiating the error sum of squares, we find the elements of the gradient vector necessary for the application of steepest descent which are given by:

$$\frac{\partial Q}{\partial \theta_k} = \sum_{i=1}^{n} 2 f_i \frac{\partial f_1}{\partial \theta_k} \tag{3.4}$$

So that

$$g = \begin{bmatrix} \frac{\partial Q}{\partial \theta_1} \\ \frac{\partial Q}{\partial \theta_2} \\ \vdots \\ \frac{\partial Q}{\partial \theta_m} \end{bmatrix} = 2 \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & \frac{\partial f_2}{\partial \theta_1} & \cdots & \frac{\partial f_n}{\partial \theta_1} \\ \frac{\partial f_1}{\partial \theta_2} & \frac{\partial f_2}{\partial \theta_2} & \cdots & \frac{\partial f_n}{\partial \theta_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_1}{\partial \theta_m} & \frac{\partial f_2}{\partial \theta_m} & \cdots & \frac{\partial f_n}{\partial \theta_m} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$$

$$\Rightarrow \quad g = 2A^{'}f \tag{3.5}$$

where $f^{'} = [f_1, f_2, \ldots, f_n]$. Differentiating (2.4) with respect to θ_j , we find that:

$$\frac{\partial^2 Q}{\partial \theta_k \partial \theta_j} = 2\sum_{i=1}^{n} \frac{\partial f_i}{\partial \theta_j} \cdot \frac{\partial f_1}{\partial \theta_k} + 2\sum_{i=1}^{n} \frac{\partial^2 f_i}{\partial \theta_j \partial \theta_k}$$

By assumption, let the second term in the derivative be neglected such that

$$\frac{\partial^2 Q}{\partial \theta_k \partial \theta_j} \approx 2\sum_{i=1}^{n} \frac{\partial f_i}{\partial \theta_j} \cdot \frac{\partial f_1}{\partial \theta_k}$$

These are the elements of the Hessian matrix H, which may therefore be written in the form:

$$H \approx 2A'A$$

Recall that the Newton Raphson Iteration is given as

$$\begin{aligned} x_{n+1} &= x_n + \frac{f' x_n}{f''(x_n)} \\ \Longrightarrow \quad \theta_{i+1} &= \theta_i - (2A'A)^{-1}(2A'f) \\ \Longrightarrow \quad \theta_{i+1} &= \theta_i - (A'A)^{-1}(A'f) \end{aligned}$$

where $f'(x_n)$ is the first derivative of the function at point x_n , $f^{''}(x_n)$ = the second derivative of the function at point x_n, x_n = current iterative value whilex_{n+1} is the iterative future value.

This is the iterative procedure which we need to undergo before the estimates of the parameters are generated. However, a starting value is assigned to each parameter and the final values of the parameters are gotten as the iteration converges. For the case of the selected models, taking Wykoff model as an example.

$$y = 1.3 + \frac{\theta_1 x}{x+1} + \theta_2 x + \epsilon \tag{3.6}$$

where ϵ is the random error component

$$\epsilon = y_i - h_i$$

where

$$h_i = 1.3 + \frac{\theta_1 x}{x+1} + \theta_2 x$$

$$Q = \sum_{i=1}^{n} \left[y_i - 1.3 - \frac{\theta_1 x}{x+1} - \theta_2 x \right]^2$$

$$\frac{\partial Q}{\partial \theta_k} = \sum_{i=1}^{n} 2 f_i \frac{\partial f_1}{\partial \theta_k}$$

where

$$f_i = y_i - 1.3 - \frac{\theta_1 x}{x+1} + \theta_2 x$$

Therefore, the partial derivatives are given as:

$$\frac{\partial f_i}{\partial \theta_1} = \frac{-x}{x+1} \tag{3.7}$$

$$\frac{\partial f_i}{\partial \theta_2} = -x \tag{3.8}$$

$$\frac{\partial Q}{\partial \theta_1} = \sum_{i=1}^{n} 2 f_i \left(\frac{-x}{x+1}\right) \tag{3.9}$$

$$\frac{\partial Q}{\partial \theta_2} = \sum_{i=1}^{n} 2 f_i (-x) \tag{3.10}$$

For the case of the data to be considered, n= 189 for pine specie; therefore, from $g = 2A'f$ matrix A becomes

$$A = \begin{bmatrix} -0.9091 & -0.9375 & \cdots & -0.9047 \\ & & & \\ -10 & -15 & \cdots & -9.5 \end{bmatrix}$$

The elements of this matrix are gotten by inserting the values of x (DBH) into the partial derivatives, taking the values 10, 15 and 9.5 as the first, second and the last values of the data on DBH respectively. Since the sample size is large, the matrix cannot be easily inverted which implies that the iterative process may not be solved in a closed form easily. However, the use of "nls" package in R to estimate the parameters in a closed form.

Considering the Chapman Model also;

$y_i = 1.3 + \frac{x^2}{\theta_1 + \theta_2 x + \theta_3 x^2} + f_i$; $\quad f_i = y_i - 1.3 - \frac{x^2}{\theta_1 + \theta_2 x + \theta_3 x^2}$

$$Q = \sum_{i=1}^{n} [y_i - h_i]^2 \tag{3.11}$$

$$= \sum_{i=1}^{n} \left[y_i - (1.3 + \frac{x^2}{\theta_1 + \theta_2 x + \theta_3 x^2}) \right]^2 \tag{3.12}$$

$$\frac{\partial Q}{\partial \theta_k} = \sum_{i=1}^{n} 2 f_i \frac{\partial f_1}{\partial \theta_k} \tag{3.13}$$

$$\frac{\partial f_i}{\partial \theta_1} = \frac{x^2}{(\theta_1 + \theta_2 x + \theta_3 x^2)^2} = \left[\frac{x}{(\theta_1 + \theta_2 x + \theta_3 x^2)^2} \right]^2$$

$$\frac{\partial f_i}{\partial \theta_2} = \frac{x^2 x}{(\theta_1 + \theta_2 x + \theta_3 x^2)^2} = \left[\frac{x^{3/2}}{(\theta_1 + \theta_2 x + \theta_3 x^2)^2} \right]^2$$

$$\frac{\partial f_i}{\partial \theta_3} = \frac{(x^4)}{(\theta_1 + \theta_2 x + \theta_3 x^2)^2} = \left[\frac{x^2}{(\theta_1 + \theta_2 x + \theta_3 x^2)^2} \right]^2$$

So that;

$$\frac{\partial Q}{\partial \theta_1} = \sum_{i=1}^{n} 2 f_i \left[\frac{x}{(\theta_1 + \theta_2 x + \theta_3 x^2)^2} \right]^2$$

$$\frac{\partial Q}{\partial \theta_2} = \sum_{i=1}^{n} 2 f_i \left[\frac{x^{3/2}}{(\theta_1 + \theta_2 x + \theta_3 x^2)^2} \right]^2$$

$$\frac{\partial Q}{\partial \theta_3} = \sum_{i=1}^{n} 2 f_i \left[\frac{x^2}{(\theta_1 + \theta_2 x + \theta_3 x^2)^2} \right]^2$$

It becomes a very complicated mathematical problem to generate the hessian matrix since all the partial derivatives contain all parameters of the model; therefore, we seek for a better way to estimate the least squares estimators by using the "nls" approach in R.

Considering the Curtis (1967) model also;

$$y_i = 1.3 + \theta_1 \left[\frac{x}{1+x} \right]^{\theta_2} + f_i;$$

$$f_i = y_i - 1.3 - \theta_1 \left[\frac{x}{1+x} \right]^{\theta_2}$$

$$Q = \sum_{i=1}^{n} [y_i - h_i]^2 \tag{3.14}$$

$$= \sum_{i=1}^{n} \left[y_i - (1.3 + \theta_1 (\frac{x}{1+x})^{\theta_2} \right]^2 \tag{3.15}$$

$$\frac{\partial Q}{\partial \theta_k} = \sum_{i=1}^{n} 2 f_i \frac{\partial f_1}{\partial \theta_k} \tag{3.16}$$

$$\frac{\partial f_i}{\partial \theta_1} = - \left[\frac{x}{1+x} \right]^{\theta_2} ;; \qquad \frac{\partial f_i}{\partial \theta_2} = -\theta_1 \left[\frac{x}{1+x} \right]^{\theta_2} \log_e (\frac{x}{1+x})$$

So that;

$$\frac{\partial Q}{\partial \theta_1} = - \sum_{i=1}^{n} 2 f_i \left[\frac{x}{1+x} \right]^{\theta_2} \quad \text{and} \quad \frac{\partial Q}{\partial \theta_2} = - \sum_{i=1}^{n} 2 f_i \left[\frac{x}{1+x} \right]^{\theta_2} \log_e (\frac{x}{1+x})$$

We also need to make use of a computer package to solve this iteration in a closed form.

For Michaelis-Menten H-D model, the partial derivatives needed to generate the hessian matrix are as follows;

$$\frac{\partial f_i}{\partial \theta_1} = \left[\frac{x^{1/2}}{\theta_1 + \theta_2 x} \right]^2, \qquad \frac{\partial f_i}{\partial \theta_2} = \left[\frac{x^2}{\theta_1 + \theta_2 x} \right]^2$$

3.3 *Models selection criterion*

For the purpose of this study, the following selection criteria will be used to determine the best H-D model, among others, for fitting the relationship between Height and diameter of both Pine and gmelina tree species:

- Residual standard error, RSE.
- Akaike Information Criterion, AIC.
- Bayesian Information Criterion, BIC.

Residual standard error

This is a measure of determining the accuracy of a particular model. It is statistically known that a model with a lesser RSE (residual standard error) is said to be a better and precise model over others. However, the residual standard error is mathematically given as:

$$\hat{\sigma} = \frac{\sum_{i=1}^{n} (Y - \hat{Y})^2}{n-2} \tag{3.17}$$

Akaike Information Criterion

Suppose that we have a statistical model of some data sets. let L be the maximized value of the likelihood function for the model, let k be the number of estimated parameters in the model. Then the AIC value of the model is given as:

$$AIC = 2k - 2ln(L) \tag{3.18}$$

Bayesian Information Criterion

This is another statistical tool of assessing and selecting a model among a set of finite models. BIC is also based on the likelihood function of the model as in the case of AIC.

$$BIC = -2ln(\hat{L}) + kln(n) \tag{3.19}$$

3.4 *Test for normality*

Kolmogorov-Smirnov test

The Kolmogorov-Smirnov statistic is based on the differences between the hypothesized cummulative distribution function F_o (x) and the empirical distribution function of the sample observation S_n (x) for all x. The empirical distribution S_n (x) is defined as the proportion of sample observation that are less than or equal of x for all real numbers x.

The K-S statistic is then defined mathematically as:

$$D_n = Sup_x|S_n(x) - F_X(x)| \tag{3.20}$$

This is for any n, a reasonable measure of accuracy of our estimate. D_n statistic is particularly useful in non-parametric statistical inference because the probability distribution of D_n does not depend on F_X(x) as long as F_X(x) is continuous. Therefore, D_n is called a distribution-free statistic. The directional deviations defined as

$$D_n^+ = Sup_x[S_n(x) - F_X(x)] \tag{3.21}$$

$$D_n^- = Sup_x[F_X(x) - S_n(x)] \tag{3.22}$$

These deviations are called the Kolmogorov-Smirnov statistics which are as well distribution-free. The hypothesis to be tested is given by:
H_o : The data follow a specified distribution.
against the alternative hypothesis
H_1 : The data do not follow the specified distribution.
The hypothesis regarding the distributional form is rejected if the test statistic D is greater than the critical value obtained from the table or if the p-value generated is lesser than the chosen alpha.

Shapiro-Wilk test

The Shapiro-Wilk test is a test for normality in frequentist statistics. It was published in 1965 by Samuel Sanford Shapiro and Martin Wilk. The shapiro-wilk test utilizes the null hypothesis principle to check whether a sample $x_1, x_2, ..., x_n$ come from a normally distributed population . The test statistic is given by:

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3.23}$$

where $x_{(i)}$ is the ith order statistic, $\bar{x}$ is the sample mean and

$$(a_1, a_2, ..., a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^T}$$

where $m = (m_1, ..., m_n)^T$ and $m_1, ..., m_n$ are the expected value of the order statistics of independent and identical distributed random variables sampled from the standard normal distribution and V is the variance-covariance matrix of those order statistics. If the p-value is lesser than the chosen alpha level, the null hypothesis is rejected and there is evidence that the data tested ar not from normal population. However, if the p-value exceeds the chosen alpha level, then the null hypothesis cannot be rejected which means that the data come from a normal population. Since the test is biased by sample size, the test may be statistically significant from a normal distribution in any large samples. Thus a Q-Q plot is required for verification in addition to the test.

4. Results and discussion

Descriptive statistics

Table 1: Table of summary of the data

Species Name	No of Sample tree	DBH(cm)				Total tree height (m)			
		Mean	Min.	Max.	S.D.	Mean	Min.	Max.	S.D.
Pine	189	13.88	5.0	25.0	3.848	13.33	5.0	22.00	3.38
Gmelina	130	27.06	3.50	61.50	12.81	15.67	2.00	25.00	4.91

The table above shows the statistics summary of the data for both species of tree used in this study. It is observed that minimum and the maximum heights of both pine and gmelina are respectively given as 5m and 22m, 2m and 25m. This may imply that gmelina are generally taller than pine tree specie. the case remains the same with their diameter at breast height (DBH).

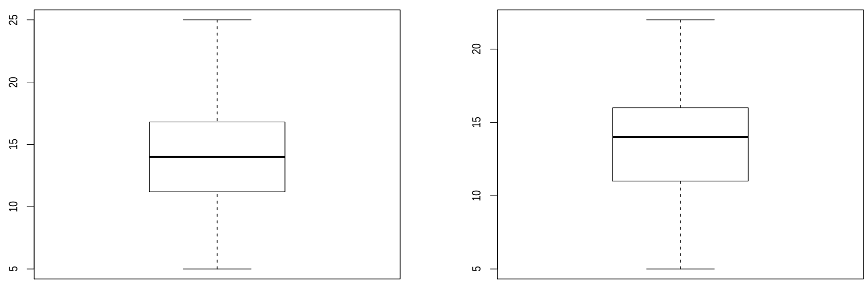


Figure 1. A boxplot for the diameter and height of pine specie

The box plots above show the minimum, maximum and the mean distribution of the data for the height and diameter of pine specie. Min= 5(cm) and max= 25(cm) and mean= 13.88 (cm) for the diameter of pine specie while min= 5m, max= 22m and mean= 13.33m for the total height. Figure 2 shows the histograms, scatter plots and box plots of the data used for the study. The scatter plots established the non-linear relationships between the diameters and Heights of the pine specie.

Figure 3 shows the histograms, scatter plots and box plots of the data used for the study. The scatter plots established the non-linear relationships between the diameters and Heights of the gmelina specie.

Parameter estimation

It is observed that the Curtis(1967) H-D model has the smallest Residual Standard Error and AIC (2.76, 924.29 respectively), followed by that of Michaelis-Menten H-D model with RSE=2.77 and AIC= 924.74. This means that the Curtis (1967) model best estimate the tree growth parameters for Pine specie and this is followed by the Michaelis-Menten H-D model. These are the curves generated for each model after the estimation of parameters.

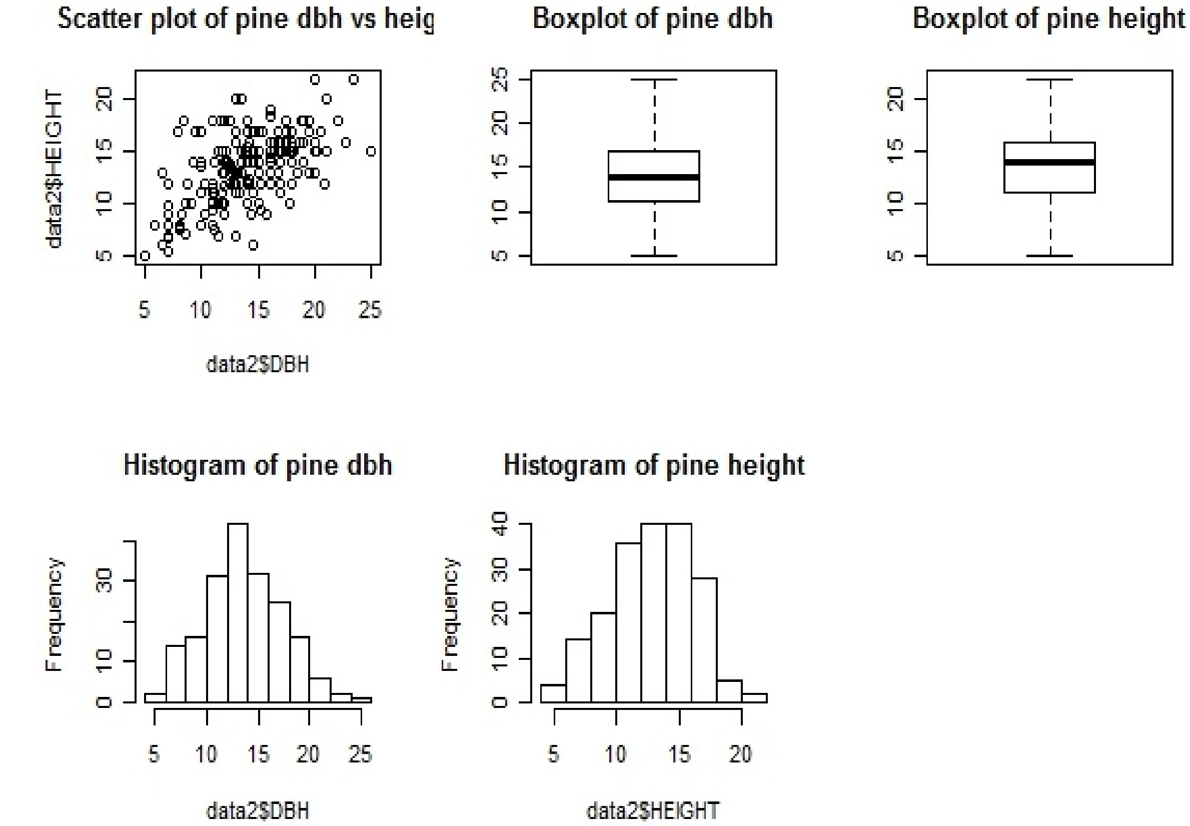


Figure 2. Relationship between diameter and height of pine specie

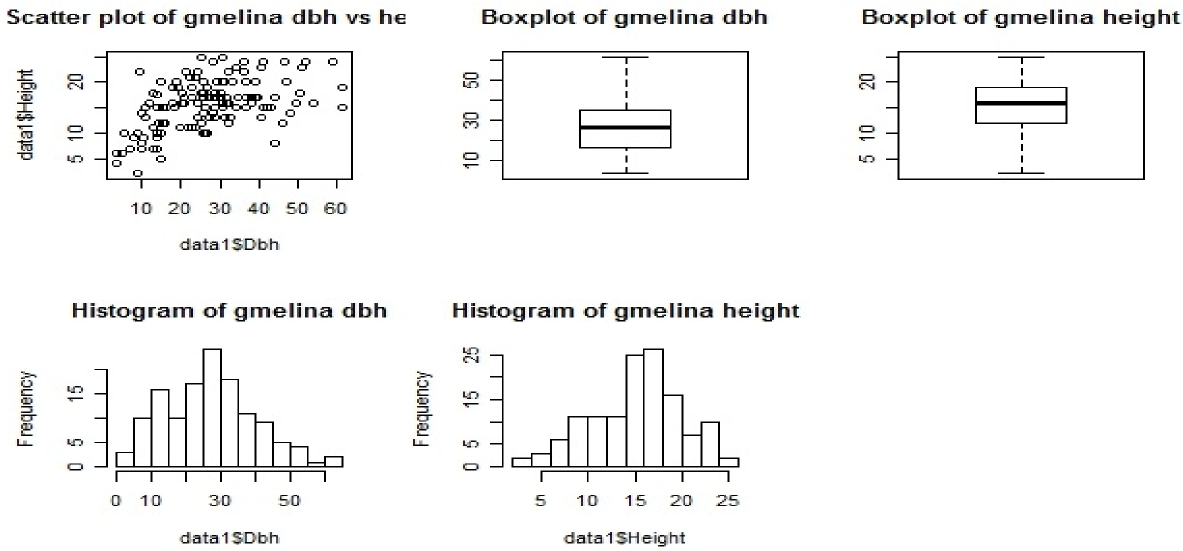


Figure 3. Relationship between diameter and height of pine specie 2

It is observed that the Curtis(1967) H-D model has the smallest Residual Standard Error and AIC (3.96, 730.34 respectively), followed by that of Chapman-Richard H-D model with RSE=3.97 and AIC= 736.3. This means that the Curtis (1967) model best estimate the tree growth parameters for Pine specie and this is followed by the Chapman-Richard H-D model.

Table 2: A table showing the estimates of the parameters for the selected models

Estimation of Parameter for Pine Species

Model	Parameter	Estimate	Std	t value	P-value	RSE	AIC
Chapman	a	1.3039	2.2215	0.587	0.5579		
	b	0.4011	0.3628	1.106	0.2702	2.769	930.33
	c	0.0452	0.0141	3.198	0.0016		
Power	a	2.808	0.4670	6.013	$9.38 * 10^{-9}$		
	b	0.5572	0.0618	9.016	$2.30 * 10^{-16}$	2.774	926.0288
Curtis	a	21.0432	1.3115	16.045	$2 * 10^{-16}$		
	b	7.5982	0.8755	8.679	$1.94 * 10^{-15}$	2.761	924.292
Michaelis-menten	a	0.6137	0.0751	8.170	$4.5 * 10^{-14}$		
	b	0.0372	0.0051	7.243	$1.11 * 10^{-11}$	2.765	924.7368
Wykoff	a	6.0549	0.8910	6.796	$1.4 * 10^{-10}$		
	b	0.4628	0.0574	8.060	$8.8 * 10^{-14}$	2.79	927.7198

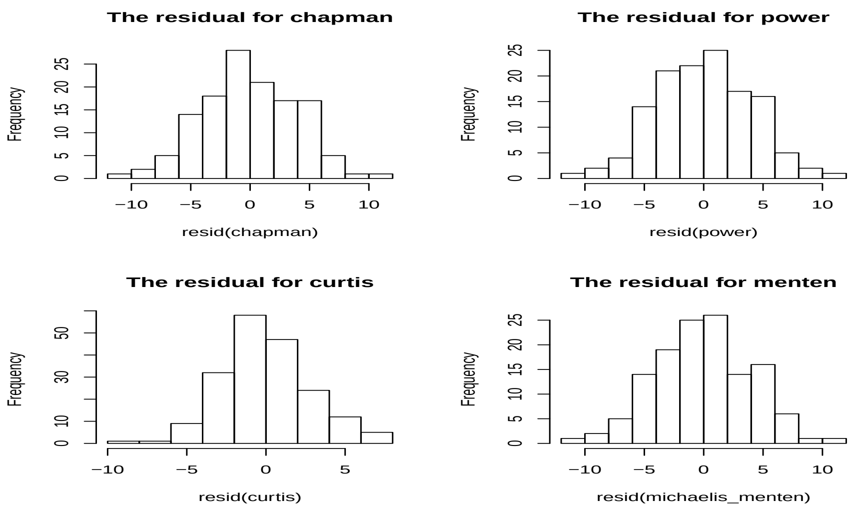
Table 3: A table showing the estimates of the parameters for the selected models

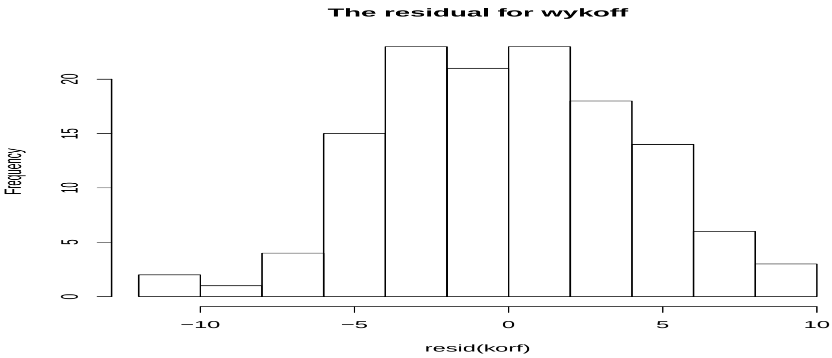
Estimation of Parameter for gmelina Specie

Model	Parameter	Estimate	Std	t value	P-value	RSE	AIC
Chapman	a	1.9264	2.253	0.855	0.394		
	b	0.3110	0.263	1.183	0.2392	3.97	736.3395
	c	0.0508	0.0064	7.932	$9.67 * 10^{-13}$		
Power	a	4.7184	0.8168	5.777	$5.47 * 10^{-8}$		
	b	0.3472	0.0512	6.786	$3.84 * 10^{-10}$	4.061	737.2604
Curtis	a	20.0518	0.9998	20.056	$2 * 10^{-16}$		
	b	7.3183	1.1514	6.356	$3.34 * 10^{-9}$	3.956	730.34
Michaelis-menten	a	0.5316	0.0941	5.65	$9.9 * 10^{-8}$		
	b	0.046	0.0036	12.93	$2.0 * 10^{-16}$	3.968	741.2685
Wykoff	a	10.3317	0.9552	10.817	$2.0 * 10^{-16}$		
	b	0.1687	0.0304	5.549	$1.58 * 10^{-7}$	4.161	743.607

5. Residual analysis

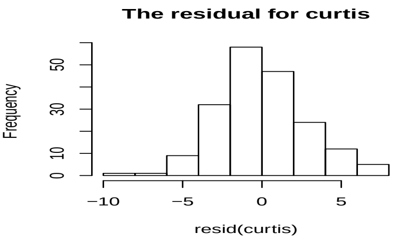
5.1 *Residual plots for the models used on gmelina specie*

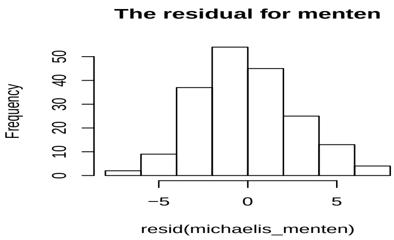


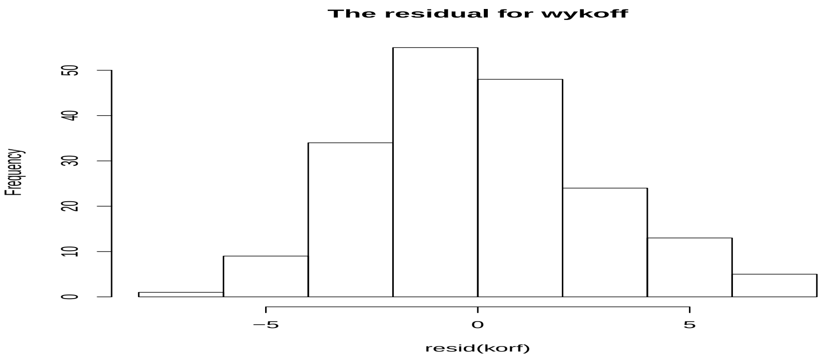


The histograms suggest that almost all the models generated residuals that are normally distributed due to the shape of the histograms.

5.2 *Residual plots for the models used on pine specie*







The histograms suggest that almost all the models generated residuals that are normally distributed due to the shape of the histograms.

Predicting the height of a tree when Dbh is known

Since it has been established from the previous results that the Curtis(1967) H-D best estimates the growth parameters for both pine and gmelina species, the following forecasted values of the height of trees were derived from the model.

Table 4: A table showing the predicted heights of pine and gmelina trees with assumed diameters

Pine specie		Gmelina specie	
D.b.h(cm)	Predicted Height(m)	D.b.h(cm)	Predicted Height(m)
11	12.16057134	11	11.92326175
18.4	15.37239584	18.4	14.92485398
23	16.52551012	23	15.99553886
14.5	13.97423182	14.5	13.62194871
16	14.57228813	16	14.17991878
18.9	15.51909862	18.9	15.06125649
13.5	13.52321805	13.5	13.20047978
19	15.54771226	19	15.08785462
17.7	15.15627041	17.7	14.72380064
17	14.92656836	17	14.50998169
15.1	14.22390857	15.1	13.85501388
16.2	14.64577042	16.2	14.24840637
14.5	13.97423182	14.5	13.62194871
34.8	18.26434074	34.8	17.60409748
14	13.75446963	14	13.4166579
11	12.16057134	11	11.92326175
10	11.49673959	10	11.29879755
9.04	10.77871771	9.04	10.62154318
16.3	14.68200369	16.3	14.28217133
25	16.91681109	25	16.35812854

Tests for normality

Kolmogorov-Smirnov test

The hypothesis to be tested
H_o: The residuals are normally distributed.
H_1: The residuals are not normally distributed.
$\alpha = 0.05$.

Table 5: Table of summary for K-S test for normality

Model	D_n	P-value
Chapman	0.3226	$2.183 * 10^{-07}$
Power	0.337	$5.051 * 10^{-08}$
Curtis	0.2381	$4.445 * 10^{-05}$
Michaelis-menten	0.3255	$1.636 * 10^{-07}$
Korf	0.3399	$3.737 * 10^{-08}$

If the p-value is greater than the significant level (0.05), the null hypothesis will be rejected and if otherwise, we accept the null hypothesis. Since the p-values for all the models are lesser than 0.05 which is the chosen alpha, there is no evidence to reject the null hypothesis; therefore, the residuals generated from the models are normally distributed.

Shapiro-Wilk test for normality

The hypothesis to be tested
H_o: The residuals are normally distributed.
H_1: The residuals are not normally distributed.
$\alpha = 0.05$.

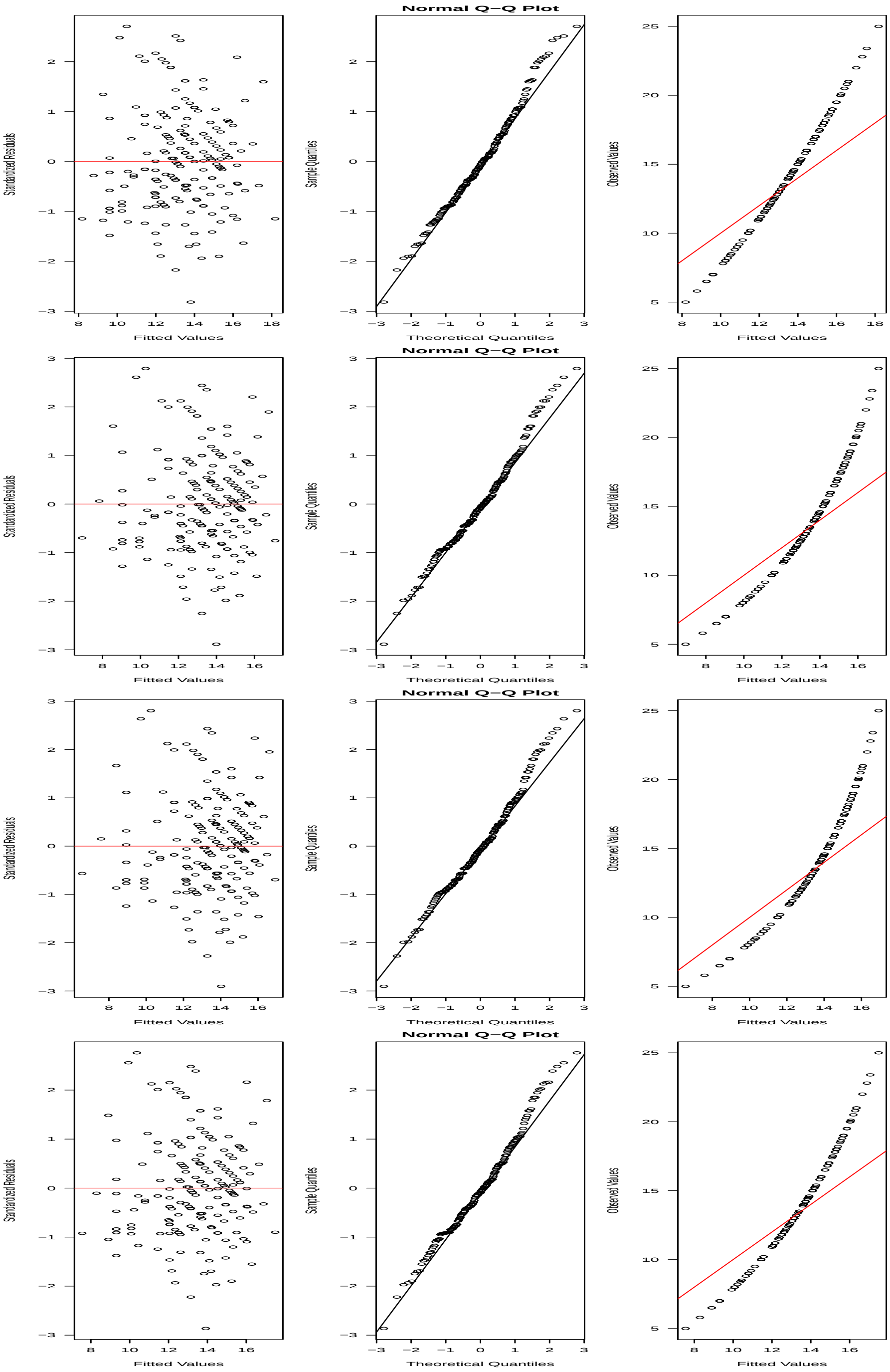
If the p-value is greater than α, we accept the null hypothesis; otherwise, we reject it. Since all the

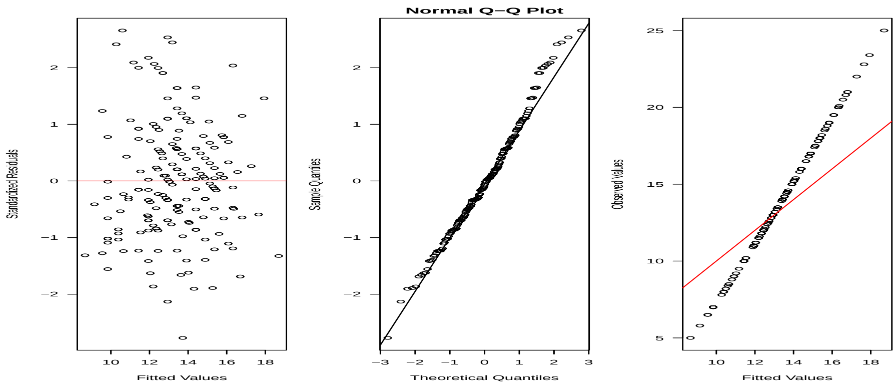
Table 6: Table of summary for Shapiro-Wilk test for normality

Model	W	P-value
Chapman	0.9861	0.0601
Power	0.9876	0.096
Curtis	0.9864	0.0648
Michaelis-menten	0.9863	0.064
Korf	0.9882	0.1192

p-values for the models are greater than 0.05 which is the chosen alpha, there is no evidence to reject the null hypothesis; therefore, the residuals generated from the models are normally distributed.

The normal Q-Q plots for the residuals





Based on the theory behind the use of Q-Q plot, if the data indeed follow the distribution (Normal distribution), then the points on the q-q plot will fall approximately on a straight line. Therefore, the q-q plots show that the residuals are normally distributed since almost all points fall approximately on a straight line.

6. Conclusion

This study was aimed at estimating tree growth parameters from some existing H-D models, and two species of tree were used for the study. It was observed that:

- The curtis (1967) model performed best among the selected models used for modelling height-diameter relationship of *gmelina specie* based on the values of the RSE and the AIC which are 2.76 and 924.23 respectively.
- Curtis(1967) H-D model has the smallest Residual Standard Error and AIC (3.96, 730.34 respectively), followed by that of Chapman-Richard H-D model with RSE=3.97 and AIC= 736.3. This means that the Curtis (1967) model best estimate the tree growth parameters for Pine specie and this is followed by the Chapman-Richard H-D model.
- The residuals generated from the models used followed normal distribution. This was established by the tests for normality adopted on the residuals. Both Shairo-Wilk and Kolmogorov-Smirnov tests showed that the residuals are normally distributed.

As a result of the summary compiled above, the following are my conclusions with respect to the proposed objectives of the study :

- It has been determined that the Curtis (1967) model for Height-Diameter relationship best estimates the tree growth parameters among the selected models.
- It was observed from the exploratory data analysis conducted that the two species of tree selected for this study are different in their heights and diameters, this means that they grow in different ways. Also, from the parameter estimation, the parameters of a particular model using both *pine* and *gmelina* data are different. Therefore, the tree species selected for this study do not grow the same way.
- It has been observed that the error term added to the models (Using additive method) are normally distributed based on the residual analyses conducted.

Having known that it is much easier to predict the height of a tree when its diameter at beast height is known (being the most accurate measurement), it is therefore recommended that Forest researcher should adopt the use of Curtis (1967) model for prediction. The non-linear least squares approach in estimating the parameters of the models used was quite encouraging in R, this because most of the iterations converged faster. Therefore, it is recommended that researchers should make use of the "nls" package in R for parameter estimation for non-linear models. The Government should implement a policy to ensure that people plant two or three trees in place of any single tree cut down for either commercial purpose or other purposes, since it has been earlier listed above that trees are greatly important to humanity biologically.

References

[1] Arabatzis A.A.; Burkhart, H.E. 1992. An evaluation of sampling methods and model forms for estimating height-diameter relationships in loblolly pine plantations. Forest Science. 38: 192-198.
[2] Bi, H.Q.; Jurskis, V.; O'Gara, J. 2000. Improving height prediction of regrowth eucalypts by incorporating the mean size of site trees in a modified Chapman-Richards equation. Australian Forestry, 63: 257-266.
[3] Clutter, J.L.; Fortson, J.C.; Pienaar, L.V.; Brister, G.H.; Bailey, R.L. 1983. Timber management: a quantitative approach. Wiley. New York. 333 p.
[4] Colbert, K.C.; Larsen, D.R.; Lootens, J.R. 2002. Height-diameter equations for thirteen midwestern bottomland hardwood species. Northern Journal of Applied Forestry. 19: 171176.
[5] Curtis, R.O. 1967. Height-diameter and height-diameter-age equations for second growth Douglas-fir. For. Sci. 13:365375.
[6] DenUyl, D. 1954. Indiana's old growth forest. Proceeding, Indiana Academy of Sciences 63: 73-79.
[7] Diguez-Aranda, U.; Barrio Anta, M. ; Castedo Dorado, F.; lvarez Gonzlez, J.G. 2005.
[8] FAO, "Scientific facts on forests", 2006, http://www.greenfacts .org/en/forests/forests-greenfacts-level2.pdf.
[9] Furnival, G.M. 1961. An index for comparing equations used in constructing volume tables. Forest Science. 7: 337-341.
[10] Generalized height-diameter relationship for Pinus sylvestris plantations in North-western Spain. Investigacin Agraria, Sistemas y Recursos Forestales. 14:229-241.
[11] George, D.W.; Fischer, B.C. 1989. The effect of site and age on tree regeneration in young upland hardwood clearcuts. In: Rink, G.; Budelsky, C.A., eds. Proceedings, 7th Central Hardwood forest conference. 1989 March 5-8; Carbondale, IL. Gen. Tech. Rep. NC-132. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 40-47.
[12] Mehtatalo, L. 2008. Forest Biometrics with examples in R. Lecture notes for the forest biometrics course. 186 pages. Available at http://cs.uef.fi/ lamehtat/.
[13] R. O. Curtis, Height-diameter and height-diameter-age equations for second-growth Douglas-fir, Forest Science, vol. 13, pp.365375, 1967.
[14] Trincado, G., C.L. Vanderschaaf, and H.E. Burkhart. 2007. Regional mixed-effects heightdiameter models for loblolly pine (Pinus taeda L.) plantations. Eur.J. For. Res. 126(2): 253-262.
[15] VOL. 2, NO. 5, June 2012 ISSN 2225-7217 ARPN Journal of Science and Technology.
[16] Wykoff, W.R., C.L. Crookston, and A.R Stage. 1982. User's guide to the Stand Prognosis Model. USDA For. Serv. Gen. Tech. Rep. INT-133. 122 p.
[17] Z. Fang and R. L. Bailey, Height-diameter models for tropical forests on Hainan Island in southern China, Forest Ecology and Management, vol. 110, no. 13, pp. 315327, 1998.