

Multiple Regression Models of the Volumetric Stem Biomass

NORAINI ABDULLAH¹, ZAINODIN HJ. JUBOK and NIGEL JONNEY J.B.

Mathematics with Economics Programme,
School of Science & Technology
Universiti Malaysia Sabah, Locked Bag No. 2073,
88999 Kota Kinabalu, Sabah,
MALAYSIA.

¹email: noraini@ums.edu.my
<http://www.ums.edu.my>

Abstract: - The development of a simple model was presented for obtaining the volumetric stem biomass of a tropical tree species. To model the volumetric stem biomass, *Cinnamomum* of family *Lauracea* was chosen. Mensuration data were collected based on two volumetric equations, namely, the Huber's and Newton's equations. During data collection, the variables considered were height of stem or trunk, height of tree, diameter at breast height, diameter at middle and diameter at top of the stem before the crown. Possible variables with their interactions were screened with spearman correlation tests and values greater than 0.95 were selected. The best model was determined using the process of eight selection criteria (8SC). However, the best model was found to be in the form of multiple regressions (MR) up to the fourth order interactions.

KeyWords: - stem volume, volumetric equations, best model, correlation tests, interactions, selection criteria, multiple regression.

1 Introduction

The last few decades had seen activities contributing to valuable forest depletion and global warming and since then, sustainable forest, its management and policies have become global issues. Deforestation has been blamed to be partially responsible for the global atmospheric CO₂ buildup, and alternatively, afforestation appears to be one of the feasible methods of reducing the concentration of CO₂ from the atmosphere. It uses solar energy and allows an economic fixation of CO₂ from the atmosphere which does not depend on concentrated CO₂ streams. Control of dispersed sources of CO₂ is also taken into account by photosynthetic extraction of CO₂ from the atmosphere. The photosynthetic process of using atmospheric carbon dioxide and water from the soil, will result in the releasing of oxygen into the air. However, these reaction steps of the oxygen cycle are evolved from water and not from CO₂ [1].

A preliminary study done by [2] had indicated that the saving in land area for a forest plantation adjacent to a fossil-burning power plant due to the increased stack gas CO₂ concentration was relatively small compared to planting in any area. It was estimated that a tract of 100 km² or 385 square mile (ml²) planted with moderately growing trees would be sufficient to fix all the CO₂ from a 1000-MW(e) coal-fired power plant over the lifetime of the plant.

Trees properly used in a landscape could increase property values by as much as 20 percent and provide food and shelter for birds and urban wildlife. Planted strategically, the right shade trees could further reduce building cooling costs by as much as 50 percent. [3] also discovered that trees were found to reduce the temperature of streets and parking lots by 8 to 10 degrees in the summer, making paved surfaces last longer without repairs. They would also improve air quality by trapping dust, absorbing air pollutants and converting carbon monoxide to oxygen which is essential towards mankind environment.

2 Biomass Estimation

The tree-crown biomass estimation was also studied by [4] on the forest species of the Ural and Kazakhstan. They had stated that there were two separate most economical and relatively precise regressions, one for broad-leaved while the other for coniferous species. Each of these species only use stem diameter at the lowest point of the crown, d_c . Approximation for coniferous foliage was found to have improved considerably by allowing parallel regressions, inclusive of mean diameter increment and diameter at breast height as predictors. However, tree age was found to be less influential than its mean increment.

An allometric equation relating biomass component to independent variables, such as, diameter at breast height (Dbh) and tree height (H)

was also developed by [5]. Using simple linear regression for 10 co-occurring tree species in China's temperate forests, [5] then executed the PROC GLM procedure in SAS for analysis. The foliage biomass was found to be more variable than other biomass components, both across and within tree species.

Estimation of crown characters and leaf biomass from leaf litter in a Malaysian canopy species, *Elateriospermum tapos* (*Euphorbiaceae*) was also studied by [6]. Estimated values were found to be similar to the values estimated from the allometric equation which used parameters such as, the diameter at breast height and the overall tree height. Forest productivity was evaluated and the characteristics in various forest were studied using litter trap method which in turn, estimated by the non-linear least square regression.

Tree biomass equations and estimation for urban tree planting in the tropics were also modelled using parameters measured from single trees of *Cinnamomum iners* grown in the Kota Kinabalu campus of Universiti Malaysia Sabah. Linear and allometric regression equations for biomass prediction were developed, and evaluations showed that the model can attribute to the tree stem biomass yield [7].

The increasing desire for total tree utilization and the need to express yield in terms of weight rather than volume had stimulated studies of biomass production by [8]. Even-aged stands of *Gmelina arborea* and *Nauclea diderrichii* in Nigeria were studied to obtain the biomass equations and estimation of both of the species. *Nauclea diderrichii*, an indigenous species was found to thrive well in plantations. [9] had assessed the above ground biomass production of even-aged stands of *Gmelina* by non-destructive method and recommended that models incorporating Dbh only to be used for estimating the biomass production.

3 Materials and Methods

3.1 Study Area

The scope of this research is focused on a commonly grown tropical tree, *Cinnamomum iners*, a species of *Cinnamomum*, which is found in Universiti Malaysia Sabah main campus in Sabah. The university is located on a 999-acres piece of land along the Sepanggar Bay in Kota Kinabalu and situated in a tropical rainforest zone of Sabah in East Malaysia, at latitude $6^{\circ} 00'$ and longitude $116^{\circ} 04'$. *Cinnamomum iners* thrives well in a tropical climate

with a mean annual rainfall of around 2000-2499 mm and the relative humidity of 81.2 ± 0.3 °C. The mean annual temperature is 27.2 ± 0.1 °C and its rainy season extends from October to January.

Data from 106 trees of *Cinnamomum iners*, located along the road towards the Chancellery, were collected. The total tree area of study covered was measured using Trimble's GeoXH handheld, for best accuracy GIS data collection.



Figure 1 : Study Area Location



Figure 2 : Satelite Image of Tree Location

3.2 *Cinnamomum iners*

Cinnamomum iners or in short, *C.iners* is in the family of Lauracea. Being an Asian native, it is also popular in the southern up across the western parts of USA [11,12]. Planted as shade trees in urban forests, studies on the *cinnamomum* species include toxic and antifungal properties of the essential oils [13], medicinal [14,15,16], antimicrobial agents [17,18], cosmetics, perfumes and beauty care [19,20]. Hence, cinnamomum plantation and propagation can be seen as environmentally significant with potential economic and commercial contributions [21,22].

3.3 Data Collection

Measurements towards data collection were done using clinometers and fiberglass diameter girth tape. Variables that were measured include height of stem or trunk, height of tree, diameter at breast height, diameter at middle and diameter at top of the stem before the crown.

3.2.1 Clinometer

Clinometer is a tool to measure the height of a tree. A clinometer is a fairly simple instrument which is used to measure the angle of a slope. By using the principles of trigonometry, the height of tall objects can be calculated from the angles measured. Figure 3 below shows how the height of tree is being calculated [10].

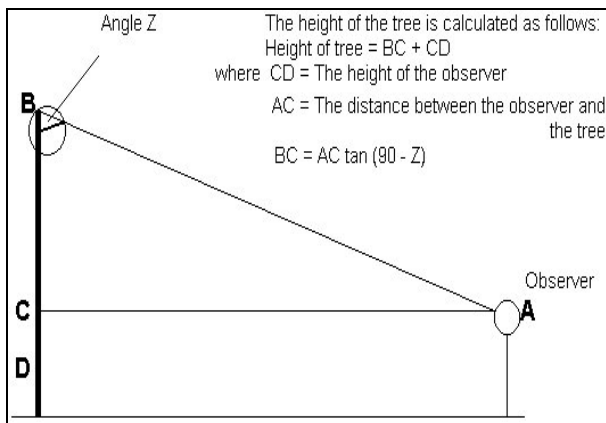


Figure 3 : Height of tree calculation [10].

3.2.2 Fiberglass tape

A girth tape measures diameter indirectly. The tape is wrapped round the tree to measure the circumference. This measurement was divided by π , to estimate the diameter. Often the tape will have normal units (mm and cm) on one side and π units on the other side. The tape should be held relatively firmly without stretching it and also be wrapped round the bole in a perpendicular plane to the stem axis. Keeping the tape numbers right side up reduces the chances of incorrect reading of the scale or parallax error; such as for instance, when upside down readings like 4 times being recorded as 6 times are common errors.

3.3 *Cinnamomum iners*

Cinnamomum iners or in short, *C.iners* is in the family of Lauracea. Being an Asian native, it is also popular in the southern up across the western parts of USA [11,12]. Planted as shade trees in urban forests, studies on the *cinnamomum* species include toxic and antifungal properties of the essential oils [13], medicinal [14,15,16], antimicrobial agents [17,18], cosmetics, perfumes and beauty care [19,20]. Hence, *cinnamomum* plantation and propagation can be seen as environmentally significant with potential economic and commercial contributions [21,22].

During the field study, the height of trunk of each tree was taken using two measurements. Firstly, it was measured from the land at ground level, to where twigs or branches started to grow and secondly, the height of the whole tree from the ground up to the tip or apex of the crown by using a clinometer. The diameter of each tree is also measured at breast height, middle and top of the trunk with a diameter girth tape. Diameter at breast height (Dbh) is often quoted by foresters to be technically approximately 1.3 meters from the ground level.

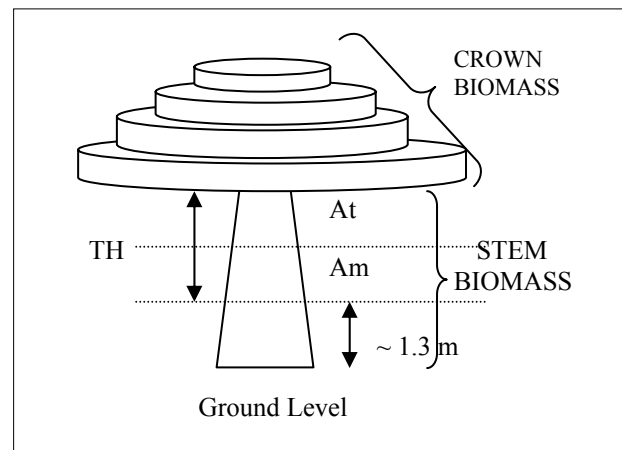


Figure 4 : Schematic diagram of *C.iners*.

- TH = height of trunk
- At = area at top before branches
- Am = area at middle of trunk
- Dbh = diameter at breast height
- Ba = basal cross-sectional area of trunk approx. 1.3m at Dbh

3.4 Volumetric Biomass Equations

There are two formulae that are being compared with in this study. They are the Newton's and

Huber's formulae. Both formulae are being used to calculate the stem volume of a tree but looking at different variables that are being measured.

3.4.1 Newton's Formula

The basal area of every mean tree in the population was calculated with the formula, $Ba = \pi(Dbh)^2 / 4$, where Ba is the basal area, and Dbh, the diameter at breast height. Stem volumes of the mean tree were then estimated using the Newton's formula given by:-

$$V_N = h(A_b + 4A_m + A_t) / 6, \quad (1)$$

where V_N is the volume using Newton's formula (m^3), h is the height, and A_b , A_m and A_t are the areas at the base, middle, and top, respectively [8].

3.4.2 Huber's Formula

The main stem, up to merchantable height, is theoretically divided into a number of (mostly) standard length sections. The standard length is normally 3m (~10 feet). The exception to the standard section is the odd log - a section less than the standard length that fits between the last standard section and the merchantable height. These sections are assumed to be second degree paraboloids in shape. The bole from the merchantable height to the tip is assumed to be conoid in shape. Huber's formula is based on the assumption that the sections are second degree paraboloids. However, this may not be appropriate for the bottom or base log - which is often neiloid. Huber's formula will underestimate the volume of a neiloid. However this underestimate will be small if the difference in diameter between the bottom and the top of the section is small (i.e. small rate of taper or small sectional length). Thus, sections smaller than 3m may be necessary to avoid bias. Error in the standard sectional estimate of volume may also be introduced where the tip is not like a conoid. However the volume in the tip is relatively small, so this error is likely to be unimportant [23].

The Huber's formula was used to calculate the volume of the standard sections and the odd log. It is given by: $V_H = h \times S_h$. Taking the sectional area(cm^2) halfway along the log as S_h , then:-

$$V_H = h \times \pi \times (d_h^2) / 40000, \quad (2)$$

where V_H is the volume using Huber's formula (m^3), h is the height (m), d_h denotes diameter (cm) halfway along the log [23].

4 Statistical Analyses

4.1 Normality and Correlation Tests

The field data variables were initially typed in Microsoft Excel together with its all possible interactions. They were then tested for normality distribution of the dependent variable, volume (Y) and were based on Kolmogorov-Smirnov statistics with Lilliefors significance level less than 0.05, since the sample size is large ($n > 50$). This test was based on the null hypothesis that the data set was normally distributed. If data were not normally distributed, spearman's correlation matrix would be adopted instead and it is also used to check multicollinearity among the independent variables [24]. The correlation matrix for the main variables, without its interactions, would then include the followings:-

- Y = stem volume (m^3)
- X_1 = height of trunk (m)
- X_2 = height of tree (m)
- X_3 = diameter at breast height (m- Newton's; cm-Huber's)
- X_4 = diameter at middle of trunk (m- Newton's; cm-Huber's)
- X_5 = diameter at top of trunk (m-Newton's; cm-Huber's)

4.2 Multiple Regression (MR) Models

The multiple regression model of more than any two independent variables, that is, relating dependent variable Y to k predictor variables X_1, X_2, \dots, X_k , can be given by the model equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon, \quad (3)$$

with the assumption that the random deviation ε is normally distributed, with zero mean and variance σ^2 for any values of X_1, \dots, X_k .

The possible interactions of the predictor variables would include as such $X_1 X_2, X_1 X_3, X_2 X_3$, etc., up to the fourth order, i.e., $X_1 X_2 X_3 X_4 X_5$. The β_i 's are called parameters, i.e., regression coefficients, and $\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$ would be referred to as the regression function or model [25].

4.3 Backward Elimination Method

Backward elimination began with the full model of all affecting individual variables as well as its possible interactions were run using the Statistical Software for Social Sciences (SPSS), and sequentially eliminated from the model the least important variable. The importance of a variable was judged by the size of the t (or equivalent F) - statistic for dropping the variable from the model, i.e., the t-statistic was used for testing whether the corresponding regression coefficient is zero. After the variable with the smallest absolute t-statistic or highest p-value was dropped, the model was refitted and the t-statistics recalculated. Again, the variable with the smallest absolute t-statistic or highest p-value was dropped.

The process ended when all of the absolute values of the t-statistics were greater than some predetermined level. The predetermined level could be a fixed number for all steps or it could be changed depending on the step. When allowing it to change depending on the step, the process could be set up so that it stopped when all of the p-values were below a fixed or desired level [25].

4.4 Eight Criteria Model Selection (8SC)

In recent years, several criteria for choosing among models have been proposed. These entire selection criteria take the form of the residual sum of squares (ESS) multiplied by a penalty factor that depends on the complexity of the model. A more complex model will reduce ESS but raise the penalty. The criteria thus provide other types of trade-offs between goodness of fit and model complexity. Using the least square method, the estimator was obtained where the criteria was to minimize the sum of square of error (SSE),

$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$. A model with the least value of a criterion statistics was judged to be preferable [26]. The statistical procedures of getting the best model were based on the eight selection criteria, namely, SGMASQ, AIC, FPE, GCV, HQ, RICE, SCHWARZ and SHIBATA had been shown by [27]. These criteria were based on minimizing the SSE multiplied by a penalty factor where $m = k + 1$ is the number of parameters in the model and n is the number of the observations.

The selection criteria to select the best model were as follows [27]:

- i) SGMASQ : $\left(\frac{SSE}{n}\right) \left[1 - \left(\frac{m}{n}\right)\right]^{-1}$
- ii) AIC : $\left(\frac{SSE}{n}\right) e^{(2m/n)}$
- iii) FPE : $\left(\frac{SSE}{n}\right) \frac{n+m}{n-m}$
- iv) GCV : $\left(\frac{SSE}{n}\right) \left[1 - \left(\frac{m}{n}\right)\right]^{-2}$
- v) HQ : $\left(\frac{SSE}{n}\right) (\ln n)^{2m/n}$
- vi) RICE : $\left(\frac{SSE}{n}\right) \left[1 - \left(\frac{2m}{n}\right)\right]^{-1}$
- vii) SCHWARZ : $\left(\frac{SSE}{n}\right) n^{m/n}$
- viii) SHIBATA : $\left(\frac{SSE}{n}\right) \frac{n+2m}{n}$

After having chosen the best model which had satisfied most of the eight criteria with the least value, the Wald Test was carried out. This enhancement test was used to test whether the best model from the selected model (reduced model) was acceptable than the initial selected model (complete model) [27]. The best model and the initial possible models were expressed as in equations (4) and (5) below:-

The complete model (best model);

$$Y = \gamma + \beta_1 x_1 + \dots + \beta_g x_g + \beta_{g+1} x_{g+1} + \dots + \beta_k x_k + \varepsilon \quad (4)$$

The reduced model (initial possible model);

$$Y = \gamma + \beta_1 x_1 + \dots + \beta_g x_g + \varepsilon \quad (5)$$

The hypothesis used to carry out the Wald Test was given by:-

$$H_0: \beta_{g+1} = \beta_{g+2} = \dots = \beta_k = 0$$

$$H_1: \text{At least one } \beta \text{ is nonzero}$$

The null hypothesis would be rejected when the F-statistic as shown in equation (6), F_C was greater than $F_{k-g, n-(k+1), \alpha}$;

$$F_C = \frac{(SSE_{\text{Reduced model}} - SSE_{\text{Complete model}}) / (k-g)}{(SSE_{\text{Complete model}}) / (n-[k+1])} \quad (6)$$

The randomness test was used to check the distribution of the residuals [28]. If $z_i (i=1,2,\dots,n)$ were independent on i , then the random variable given by (7):

$$T_n = R \sqrt{\frac{(n-k)}{(1-R^2)}} \quad (7)$$

would follow a t-distribution with $\nu = n - k$ degrees of freedom,

where $R = \frac{\frac{1}{n} \sum_{i=1}^n iz_i - \bar{z}\bar{K}}{S_z S_1}$ and $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$,

$$S_z^2 = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2, \bar{K} = \frac{n+1}{2} \text{ and } S_1 = \frac{n^2 - 1}{12}.$$

The assumption of the residual randomness would be met when $|T_n| < T_{\alpha, n}$.

5 MR Models Results

5.1 Descriptive Statistics

Since the data were not normally distributed, the spearman's correlation was used instead. With Lilliefors significance level less than 0.05 and since $n > 50$, the null hypothesis was accepted and therefore normality was assumed. Based on the correlation matrix outputs, the correlation coefficients for variables were taken at equal to and greater than 0.95, indicating a strong correlation between the dependant variable Y to the independant variables X 's and its interactions. The data were then tested using Kolmogorov-Smirnov giving p-values of 0.139 for Huber's and 0.111 for Newton's equations respectively.

Table 1: Correlation Coefficient Matrix of *C.iners*-Huber's .

	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Y	1	0.017	0.752	0.980	0.985	0.969
X ₁	0.017	1	-0.031	-0.087	-0.125	-0.159
X ₂	0.752	-0.031	1	0.744	0.759	0.765
X ₃	0.980	-0.087	0.744	1	0.982	0.976
X ₄	0.985	-0.125	0.759	0.982	1	0.990
X ₅	0.969	-0.159	0.765	0.976	0.990	1

Table 2: Correlation Coefficient Matrix of *C.iners*-Newton's .

	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Y	1	0.182	0.734	0.949	0.940	0.922
X ₁	0.182	1	-0.031	-0.087	-0.125	-0.159
X ₂	0.734	-0.031	1	0.744	0.759	0.765
X ₃	0.949	-0.087	0.744	1	0.982	0.976
X ₄	0.940	-0.125	0.759	0.982	1	0.990
X ₅	0.922	-0.159	0.765	0.976	0.990	1

5.2 Analyses of Regression Functions

Using the same variables (as in section 4.1) together with its combinations of interactions, a total of 80 possible models were then obtained, before any regression procedures were done (Table 3).

Table 3: The number of all possible models

No of Variables	Single	Interactions				Total
		1st order	2 nd order	3 rd order	4 th order	
1	5	-	-	-	-	5
2	10	10	-	-	-	20
3	10	10	10	-	-	30
4	5	5	5	5	-	20
5	1	1	1	1	1	5
Total	31	26	16	6	1	80

However, after a few initial regression procedures, a total of 22 possible models were obtained using the Huber's equation, with three main variables (X_3, X_4, X_5), six first order interactions ($X_1X_3, X_1X_4, X_1X_5, X_3X_4, X_3X_5, X_4X_5$), seven second order interactions ($X_1X_3X_4, X_1X_3X_5, X_1X_4X_5, X_2X_3X_4, X_2X_3X_5, X_2X_4X_5, X_3X_4X_5$), five third order interactions ($X_1X_2X_3X_4, X_1X_2X_3X_5, X_1X_2X_4X_5, X_1X_3X_4X_5, X_2X_3X_4X_5$) and a single fourth order interaction ($X_1X_2X_3X_4X_5$).

Using the Newton's equation, 13 possible models were obtained with two main variables ($X_3,$

X_4), four first order interactions ($X_1X_3, X_1X_4, X_1X_5, X_3X_4$), three second order interactions ($X_1X_3X_4, X_1X_3X_5, X_1X_4X_5$), three third order interactions ($X_1X_2X_3X_4, X_1X_2X_3X_5, X_1X_3X_4X_5$) and also a single fourth order interaction variable ($X_1X_2X_3X_4X_5$).

The best model for each volumetric equation of *C.iners* was then selected by applying the backward elimination method [25]. This was further done using SPSS. An example of the elimination procedure would be as follows:-

Table 4(i): Model M22 with Entered Variables for Backward Elimination Process.

Model	Unstandardized Coefficients	Standardized Coefficients	t	p-value
	B	Std. Error	Beta	
M22 (Constant)	-.003	.002		
X_3	.000	.000	.057	1.278
X_5	-.002	.000	-.235	-4.283
X_1X_3	.000	.000	-.077	-1.739
X_1X_4	.001	.000	.270	9.545
X_1X_5	.000	.000	.059	1.183
X_4X_5	.000	.000	.861	10.630
$X_2 X_3 X_4$.000	.000	.800	15.542
$X_2 X_3 X_5$.000	.000	-.783	-11.382
$X_3 X_4 X_5$.000	.000	-.728	-11.283
$X_1X_2X_3X_5$.000	.000	-.107	-1.229
$X_1 X_2 X_4 X_5$.000	.000	.084	1.265
$X_1 X_3 X_4 X_5$.000	.000	.799	14.571

a Dependent Variable: y

Table 4(ii): Excluded Variables Involving Multicollinearity.

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
					Tolerance
M22 x_3x_4	.643(a)	3.984	.000	.440	.000
x_3x_5	-.088(a)	-.623	.535	-.076	.000
$x_1x_3x_4$.748(a)	8.289	.000	.714	.000
$x_1x_3x_5$.213(a)	1.606	.113	.194	.000
$x_1x_4x_5$	1.411(a)	10.007	.000	.776	.000
$x_2x_4x_5$	-.371(a)	-1.494	.140	-.181	.000

a Predictors in the Model: (Constant), $x_1x_3x_4x_5, x_1x_3, x_1x_2x_4x_5, x_5, x_1x_4, x_3, x_2x_3x_4, x_3x_4x_5, x_1x_5, x_2x_3x_5, x_4x_5, x_1x_2x_3x_5$

b Dependent Variable: y

Let say, based on the parameters regression coefficients of the above model as in Table 4(i), the p-value of 0.241 was related to variable X_1X_5 . Since

0.241 being the smallest absolute t-statistics or the highest p-value and was greater than 0.05, variable X_1X_5 was therefore removed and the model was then refitted. The excluded variables such as in Table 4(ii) were ignored but they were not to be removed from the model since these variables involved multicollinearity.

Table 5(i): Model M22 after 5 Backward Elimination Processes

Model	Unstandardized Coefficients	Standardized Coefficients	t	p-value
	B	Std. Error	Beta	
M22.5 (Constant)	-0.003	.002		
x_3	0.000	.000	0.010	0.324
x_5	-0.001	.000	-0.187	-6.81
x_1x_3	0.000	.000	-0.041	-1.441
x_1x_4	0.001	.000	0.290	11.345
x_4x_5	0.000	.000	0.869	10.16
$x_2x_3x_4$	0.000	.000	0.784	15.27
$x_2x_3x_5$	0.000	.000	-0.839	-14.27
$x_2x_4x_5$	0.000	.000	0.027	0.313
$x_3x_4x_5$	0.000	.000	-0.689	-11.30
$x_1x_2x_4x_5$	0.000	.000	0.023	0.460
$x_1x_3x_4x_5$	0.000	.000	0.755	14.36

a Dependent Variable: y

Table 5(ii): Excluded Variables of selected model M22.5

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
					Tolerance
M22.5 x_3x_4	.523(a)	3.335	.001	.377	.000
x_3x_5	-.105(a)	-.675	.502	-.082	.000
$x_1x_3x_4$.630(a)	6.325	.000	.611	.000
$x_1x_3x_5$.205(a)	1.655	.103	.198	.000
$x_1x_4x_5$.677(a)	6.121	.000	.599	.000
$x_1x_2x_3x_5$	-.464(a)	-2.262	.027	-.266	.000

a Predictors in the Model: (Constant), $x_1x_3x_4x_5, x_1x_3, x_1x_2x_4x_5, x_5, x_1x_4, x_3, x_2x_3x_4, x_3x_4x_5, x_2x_3x_5, x_4x_5, x_2x_4x_5$

b Dependent Variable: y

Rerun the model with the remaining variables and the variable with the highest p-value, if the p-value was more than 0.05, would be sequentially removed. From Table 5(i), the p-value of 0.755 was related to variable $X_2X_4X_5$. Since it was more than 0.05, variable $X_2X_4X_5$ was thus removed and the model was refitted again. The procedures were subsequently repeated until there would be no more variable with a p-value greater than 0.05, as shown

in Table 6.

Table 6(i): Variables with p-values less than 0.05 .

Model	Unstandardized Coefficients		Standardized Coefficients	t	p-value
	B	Std. Error	Beta		
M22.9(Constant)	.0022794	.0012186		1.871	.066
x5	-.0007377	.0001215	-.110	-6.072	.000
x1x3	-.0007097	.0000599	-.236	-11.841	.000
x1x4	.0009113	.0000627	.284	14.544	.000
x3x4	-.0000276	.0000111	-.172	-2.495	.015
x4x5	.0000828	.0000097	.515	8.524	.000
x1x3x4	.0000596	.0000044	.792	13.573	.000
x2x3x4	.0000055	.0000009	.372	6.453	.000
x2x3x5	-.0000083	.0000007	-.580	-11.497	.000
x1x2x4x5	.0000015	.0000002	.210	6.593	.000
x1x3x4x5	-.0000002	.0000001	-.076	-2.456	.017

a Dependent Variable: y

Table 6(ii): Excluded Variables of selected model M22.9 .

Model	Beta In	t	Sig.	Partial	Collinearity
				Correlation	Statistics
				Tolerance	
M22.9	x3x5	-.396(a)	-3.740	.000	-.413
	x1x3x5	-.251(a)	-2.498	.015	-.290
	x1x4x5	.456(a)	5.374	.000	.546
	x3x4x5	-.116(a)	-1.123	.265	-.135
	x1x2x3x5	-.276(a)	-3.823	.000	-.421

a Predictors in the Model: (Constant), x1x3x4x5, x1x3, x1x2x4x5, x5, x1x4, x2x3x4, x1x3x4, x4x5, x2x3x5, x3x4

b Dependent Variable: y

Based on the parameters regression coefficients and since there are no more p-values that are more than 0.05, therefore the selected regression model functions, without considering the excluded variables for *C.iners* using Huber's equation were found to be given by:-

Table 7: Significant Selected Models – Huber's.

Models	No. of Eliminations	SSE	m	n
M5	3	0.005083	3	80
M12	7	0.004452	6	80
M15	4	0.00039	12	80
M17	5	0.0000963	8	80
M20	5	0.0003975	14	80
M22	10	0.000005	11	80
M25	4	0.0013558	9	80
M31	12	0.000233	13	80

By using the same elimination method as indicated previously, the final parameter regression

coefficients and model functions selected for *C.iners* using Newton's equation were therefore given by:-

Table 8: Significant Selected Models – Newton's.

Models	No. of Eliminations	SSE	m	n
M3	5	1.70644	4	80
M5	5	0.97579	5	80
M11	10	0.00267	6	80
M13	9	0.00262	9	80
M14	12	0.0858	12	80
M22	14	0.00276	4	80
M30	22	0.0026	9	80
M35	2	0.0032909	6	80

5.3 Best Model Selection Criteria

The best model was then chosen from the selected models by using the eight selection criteria (8SC). Table 9 below showed some significant selected models with the values of 8SC using both Huber's and Newton's equations.

Table 9: Values of 8SC for significant selected models – Huber's.

Model	SGMASQ	AIC	FPE	GCV	HQ
M5.3	6.6E-05	6.35E-05	6.85E-05	6.85E-05	6.35E-05
M12.7	6.01E-05	5.5E-05	6.47E-05	6.5E-05	5.56E-05
MM17.5	1.34E-06	1.2E-06	1.47E-06	1.49E-06	1.2E-06
M20.5	1.91E-05	1.695E-05	2.12E-05	2.15E-05	1.69E-05
M22.10	7.25E-08	8.23E-08	8.24E-08	8.40E-08	9.38E-08
M25.4	6.02E-06	4.97E-06	7.08E-06	7.3E-06	4.97E-06
M31.12	3.48E-06	2.91E-06	4.04E-06	4.15E-06	2.91E-06

Continue Table 9:

RICE	SCHWARZ	SHIBATA
6.6E-05	1.1786	6.35E-05
6.01E-05	1.38915	5.56E-05
1.34E-06	1.54992	1.204E-06
1.91E-05	1.6372	1.695E-05
8.62E-08	1.1433	7.97E-08
6.02E-06	2.1529	4.97E-06
3.48E-06	2.0382	2.91E-06

Table 10: Values of 8SC for significant selected models – Newton’s.

Model	SGMASQ	AIC	FPE	GCV	HQ
M3.5	0.2245	0.0213.	0.0257	0.0236	0.0213
M5.5	0.01301	0.0122	0.0138	0.0139	0.012197
M11.10	3.61E-05	3.31E-05	3.88E-05	3.9E-05	3.34E-05
M13.9	3.54E-05	3.3E-05	3.8E-05	3.83E-05	3.33E-05
M14.12	0.0013	0.0011	0.0014	0.0015	0.00107
M22.5	0.0010	0.00087	0.0012	0.0012	0.00087
M23.10	3.68E-05	3.17E-05	4.19E-05	4.3E-05	3.17E-05
M30.22	3.66E-05	3.27E-05	4.07E-05	4.1E-05	3.25E-05
M35.2	4.45E-05	4.11E-05	4.78E-05	4.81E-05	4.11E-05

Continue Table 10:

RICE	SCHWARZ	SHIBATA
0.0224	0.0266	0.0213
0.0130	0.01604	0.0122
3.6E-05	4.64E-05	3.34E-05
3.5E-05	4.5E-05	3.33E-05
0.0013	0.00206	0.00107
0.0010	0.00178	0.00087
3.7E-05	5.8E-05	3.175E-05
3.6E-05	5.32E-05	3.25E-05
4.45E-05	1.3891	4.11E-05

The best model was chosen based on the model having a majority of the least values of the 8SC, as indicated in bold (Table 9 and Table 10).

Using the Wald Test on the case of the best model selection using the Newton’s equation, the completed model (M13) was thus taken as the initial possible model and M13.9 as the reduced model. This was done to test the significance of the omitted variables in the best model.

The complete model (M13):

$$Y = \gamma + \beta_3 X_3 + \beta_4 X_4 + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{15} X_{15} + \beta_{34} X_{34} + \beta_{35} X_{35} + \beta_{45} X_{45} + \beta_{234} X_{234} + \beta_{235} X_{235} + \beta_{345} X_{345} + \beta_{2345} X_{2345} + \epsilon$$

The reduced model (M13.9):

$$Y = \gamma + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{15} X_{15} + \epsilon$$

The hypothesis would be:-

$$H_0: \beta_3 = \beta_4 = \beta_{13} = \beta_{14} = \beta_{15} = \beta_{34} = \beta_{35} = \beta_{45} = \beta_{234} = \beta_{235} = \beta_{345} = \beta_{2345} = 0$$

$$H_1: \text{Atleast one } \beta \text{ is nonzero}$$

The F_C value, as in equation (6), was calculated to be equal to 0.912384 and F critical value, from the F Distribution Curve at 0.05, was 1.860. Hence, the null hypothesis (H_0) was accepted since $F_C < F_{Critical, 0.05}$. Consequently, the best model was justified. Based on equation (7), T_n (value) equals 0.008775 and $T_{\alpha, n} = 2.132$. Since $|T_n| < T_{\alpha, n}$, the assumption of randomness residual from the best model using Newton’s was hence met.

From the reduced model using Huber’s equation, the best regression model would therefore be represented by: -

$$\hat{y}_H = -0.0007377X_5 - 0.0007097X_1X_3 + 0.0009113X_1X_4 - 0.0000276X_3X_4 + 0.0000828X_4X_5 + 0.0000596X_1X_3X_4 + 0.0000055X_2X_3X_4 - 0.0000083X_2X_3X_5 + 0.0000015X_1X_2X_4X_5 - 0.0000002X_1X_3X_4X_5 \tag{8}$$

On the other hand, using Newton’s equation, the best regression model was thus be represented by: -

$$\hat{y}_N = 0.160X_1X_3 + 0.644X_1X_4 + 0.196X_1X_5 \tag{9}$$

6 Discussions

Stem volume is a good estimation in determining the biomass of a certain tree species. [9] concluded that most of the biomass accumulation of *Gmelina arborea* (ROXB) plantations in south-western Nigeria, much of which is stored in the stem, indicating that a very high percentage of a tree wood is merchantable either for timber or other uses. The work’s of [8] also showed that more than 75% of total biomass yield for both species of *Gmelina arborea* and *Nauclea diderrichii* stands in the Akure forest reserve that they studied were from the stem.

In this study, however, the results showed that both volumetric equations using Huber’s and Newton’s formula for *C.iners* were best fitted with multiple regression equations. Using Huber’s volumetric equation, the best model was found to be as in equation (1), meanwhile, under Newton’s volumetric equation instead, the best model was found to be as in equation (2).

By observing these models, however, it was obvious that the Newton’s volumetric equation was simpler and consistent in its variables as compared

to that of Huber's. It was also clear that under the Newton's equation, the height of trunk would give a major contribution towards the stem volume and a significant contribution from diameters at breast height (Dbh), middle and top [12].

7 Conclusions

In this study, the process of eight selection criteria (8SC) was found to be a convenient way to determine or select the best model for stem volume estimation. From this study too, it can be concluded that the *C.iners* trees can best modelled the volumetric stem biomass by adopting the Newton's volumetric equation since it was simpler, and its variables were consistent in all its possible models, whereas for the Huber's volumetric equation, variables of different tree species would contribute differently to the stem volume. As in the case of *C.iners*, the Newton's volumetric equation was significantly preferable. Since maximum volumetric biomass is also related to the circumferential area of tree trunk, optimisation of merchantable tree log and its economic values (essential oils, timber, chemicals, etc.) can be further explored as there is no indifference within this tree species (*Cinnamomum*) except for the distribution and observation [12]. The benefits of greenery are not just environmental but also recreational, aesthetic and emotional. Model approaches to support decision-making would include crucial elements and interactions of the systems [29], hence, modelling volumetric stem biomass can be a vital indicator for foresters and to other relevant and related fields [30].

Acknowledgments

This work was partially funded by Universiti Malaysia Sabah under the grant UMS: 40/2000 of the School of Science and Technology.

References

- [1] Mukherji, S & Ghosh, A.K. *Plant Physiology*. New Delhi: Tata McGraw-Hill, 1996.
- [2] Hallmann, M.M., & Steinberg, M. *Greenhouse gas carbon dioxide mitigation*. Florida: Lewis Publishers, 1999.
- [3] Burns, R. *City Trees: Forestry plays a significant economic and environmental role in urban areas*.
<http://agcomwww.tamu.edu/lifescapes/fall01/trees.html>, 2006.
- [4] Hoffmann, C.W. & Usoltsev, V.A. Tree-crown biomass estimation in forest species of the Ural and Kazakhstan. *Forest Ecology and Management* 158, 2002, pp.59-69.
- [5] Wang, C. Biomass allometric equations for 10 co-occurring tree species in Chinese temperate forest. *Forest Ecology and Management* 222, 2006, pp.9-16.
- [6] Osada, N., Takeda, H., Kawaguchi, A.F., & Awang, M. Estimation of crown characters and leaf biomass from leaf litter in a Malaysian canopy species, *Elateriospermum tapos* (Euphorbiaceae). *Forest and Ecology Management*, 177, 2003, pp.379-386.
- [7] Noraini Abdullah, Siti Rahayu Mohd. Hashim & Kamsia Budin, 2006. Modelling Of Tree Biomass Equations and Estimation For Urban Tree Planting In The Tropics. *Proceedings of the 2nd Southeast Asian Natural Resources and Environmental Management Conference (SANREM), Malaysia*, pp.250-255.
- [8] Fuwape, J.A., Onyekwelu, J.C., & Adekunle, V.A.J., 2001. Biomass equations and estimation for *Gmelina arborea* and *Nauclea diderrichii* stands in Akure forest reserve. *Biomass & bioenergy* 21, 2001, pp. 401-405.
- [9] Onyekwelu, J.C. Above-ground biomass production and biomass equations for even-aged *Gmelina arborea* (ROXB) plantations in southwestern Nigeria. *Biomass & bioenergy* 26, 2004, pp39-46.
- [10] <http://www.offwell.free.online.co.uk/newpage2.htm>
- [11] Lee, Y.F. *Preferred check-list of Sabah trees*. 3rd Edition. Natural History Publications (Borneo) Sdn.Bhd.Sabah, 2003, pp.34.
- [12] Lemmens, R.H.M.J., Soerianegara, I., & Wong, W.C.(eds.), 1995. *Plant Resource of South-East Asia*. Timber Trees: Minor commercial timbers 5(2). Bogor: Prosea Foundation.
- [13] Ibrahim, J. & Goh, S.H. The essential oils of cinamomum species from Peninsular Malaysia. *Journal of Essential Oil Research* 4, 1992, pp.161-172.
- [14] Khan, A., Safdar, M., Khan, M.M.A., Khattak, K.N., & Anderson, R.A. Cinnamomum Improves Glucose and Lipids of People with Type 2 Diabetes. *American Diabetes Association* 26, 2003, pp. 3215-3218.
- [15] Agarwall, V.S. *Drug Plants of India*. Vol.I, Kalyan Publishers, New Delhi, 1997.
- [16] Bhattacharjee. *Hand Book of Indian Medicinal Plants*. Pointer Publishers, Jaipur, 1998.
- [17] Hirasa, K. & Takemasa, M. *Spices Science and Technology*. Marcel Dekker, New York, 1998.
- [18] Ibrahim, J., Rasadah, M.A. & Goh, S.H. Toxic and Antifungal properties of the essential oils

- of Cinnamomum species from Peninsular Malaysia. *Journal of Tropical Forest Science* 6(3), 1994, pp.286-292.
- [19] Stoltz, C. Sepicontrol A5. *SEPIC Newsletter*, 1998, pp. 152-156.
- [20] Coppen, J.J.W. Flavours and Fragrances of Plant Origin. FAO, Rome, 1995.
- [21] Perry, P.Y. & Ong, B.I., Applying ecosystems concept to the Planning of Industrial areas: a case study of Singapore's Jurong Island. *Journal of Cleaner Production* 12, 2004, pp.1011-1023.
- [22] Sandström, U.G. Anglestam, P. & Khakee, A., Urban Comprehensive Planning-identifying Habitat Networks. *Landscape and urban Planning* 75, 2006, pp.43-57.
- [23] Brack, C. Tree crown: Forest measurement and modelling, 2006. <http://sres-associated.anu.edu.au/mensuration/crown.htm>.
- [24] Lind, D.A., Marchal, W.G. & Wathen, S.A. *Statistical Techniques in Business & Economics*. Boston: McGraw-Hill, 2005, pp.491.
- [25] Devore, J.L., & Peck R.. *Statistics The Exploration and Analysis of Data..* (2nd edition). California: Wadsworth Publishing Company, 1993.
- [26] Christensen, R. *Analysis of Variance, Design and Regression: Applied Statistical Methods*. London: Chapman & Hall, 1996.
- [27] Ramanathan, R. *Introductory Econometrics with application*. (5th ed.)Texas: The Dryden Press, 2002.
- [28] Ismail, B.M. Unimodality tests for Global Optimization of single variable function using statistical method. *Malaysian Journal of Mathematical Sciences*, 2007, 1(2): 1-11.
- [29] Sterba, H., Vospernik, S., Soderbergh, I., Ledermann, T., & Hasenauer, H., *Harvesting Rules and Modules for Predicting Commercial Timber Assortments*. Sustainable Forest management. Springer, New York, 2006.
- [30] Romualdas, J., & Augustaitis, A. Forest Health Monitoring Indicators and Their Interpretability: A Lithuanian case study. USDA Forest service, 1998.