VOLUMETRIC STEM BIOMASS MODELLING USING MULTIPLE REGRESSION

NORAINI ABDULLAH[1], 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.

Abstract: - This paper presented the development of a simple model for obtaining the stem volume of a tropical tree species, that is, *Cinnamomum iners* based on the two volumetric equations, namely, the Huber's and Newton's equations. Variables considered during data collection 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. Correlation tests were done to screen and select these possible variables with their interactions. The eight criteria model selection(8SC) process was also done to obtain the best model. However, the best model was found to be in the form of multiple regressions (MR) of interactions up to the fourth order.

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

1   Introduction

Sustainable forest, its management and policies have become global issues since the last few decades due to valuable forest depletion and global warming. While deforestation has been partially responsible for the global atmospheric CO_2 buildup, alternatively, aforestation appears to be one of the feasible methods of reducing the concentration of CO_2 from the atmosphere. It uses solar energy and allows an economic fixation of CO_2 from the atmosphere which does not depend on concentrated CO_2 streams. Control of dispersed sources of CO_2 is also taken into account by photosynthetic extraction of CO_2 from the atmosphere. The process of photosynthesis uses atmospheric carbon dioxide and water from the soil, hence, releasing oxygen into the air. However, these reaction steps of the oxygen cycle are evolved from water and not from CO_2 [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 CO_2 concentration was relatively small compared to planting in any area. It was estimated that a tract of 100 km^2 or 385 square mile (ml^2) planted with moderately growing trees would be sufficient to fix all the CO_2 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

[4] studied on the tree–crown biomass estimation in the forest species of the Ural and Kazakhstan, had stated that there were two separate most economical and relatively precise regressions, one for broad-leaved while the other for coniferous species, each 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 being less influential than its mean increment.

[5] also developed an allometric equation relating biomass component to independent variables, such as, diameter at breast height (Dbh) and tree height

(H) for 10 co-occurring tree species in China's temperate forests, by using simple linear regression and 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 strive 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* 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 6º 00' and longitude 116º 04'. *Cinnamomum iners* strives well in a tropical climate with a mean annual rainfall of around 2000-2499 mm and the relative humidity of 81.2 ± 0.3 ºC. The mean annual temperature is

27.2 ± 0.1 ºC and its rainy season extends from October to January.

Only 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 Data Collection
Clinometers and fiberglass diameter girth tape were used to do measurements towards data collection. Variables measured 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.

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

2

can be calculated from the angles measured. Figure 3 below shows how the height of tree is being calculated [10].
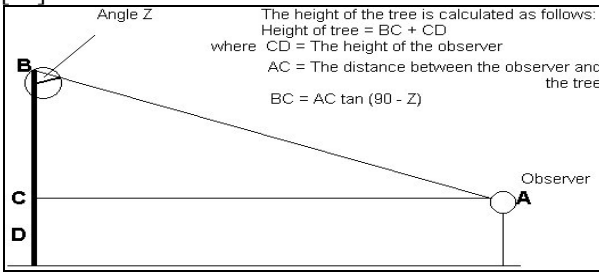

Figure 3 : Height of tree calculation [10].

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 л, to estimate the diameter. Often the tape will have normal units (mm and cm) on one side and л 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.
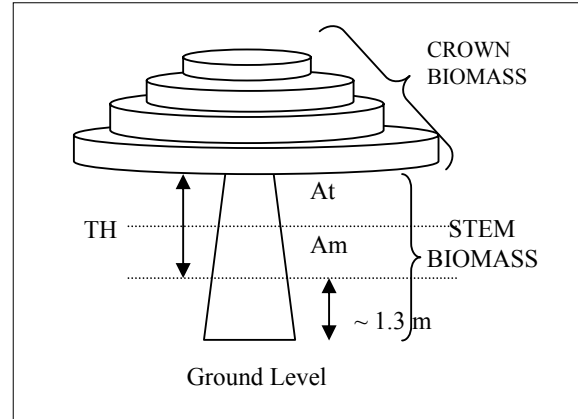

Figure 4 : Schematic diagram of *C.iners*.

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

In this study, there are two formulae that are being compared with. They are the Newton's and Huber's formulae. Both formulae are 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, $V_N = h(A_b + 4A_m + A_t)/6$, 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: $V_H = h$ x S_h. Taking the sectional area(cm^2) halfway along the log as S_h, then:-

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

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 more 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, ..., X_k, can be given by the model equation:
$$y = \beta_0 + \beta_1 X + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon,$$

with the assumption that the random deviation ε is normally distributed, with zero mean and variance σ^2 for any values of X_1, ..., 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 βi's are called parameters, i.e., regression coefficients, and $\beta_0 + \beta_1 X_1 + ... + \beta_k X_k$ 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 in 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.

A model with a lower value of a criterion statistics is judged to be preferable [26]. [27] had also shown the statistical procedures of getting the best model based on these eight selection criteria, namely, SGMASQ, AIC, FPE, GCV, HQ, RICE, SCHWARZ and SHIBATA.

5 MR Models Results

5.1 Descriptive Statistics
Since the data are not normally distributed, the spearman's correlation is used instead. With Lilliefors significance level more than 0.05 and n>50, the data are tested using Kolmogorov-Smirnov giving p-values of 0.139 for Huber's and 0.111 for Newton's equations respectively.

5.2 Analyses of Regression Functions
After a few initial regression procedures, a total of 22 possible models were obtained using the Huber's equation and 13 possible models using the Newton's equation. 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 is as follows:-

Based on the parameters regression coefficients of the models, say the p-value of 0.578 is related to variable X_3. Since 0.578 is the highest p-value, and greater than 0.05, therefore, variable X_3 is removed. Rerun the remaining variables and remove the variable with the highest p-value if the p-value is more than 0.05. The procedure is repeated until there is no more variable with a p-value > 0.05.

5.2.1 Newton's-Huber's MR Equations
Based on the parameters regression coefficients and since there are no more p-values that are more than 0.05, therefore the final selected regression model function, without considering the excluded variables for *C.iners* using Huber's equation is:-

$$\hat{y}_H = -0.0007377X_5 - 0.0007097X_1X_3 + 0.0009113X_1X_4 - 0.0000276X_3X_4 + 0.0000828X_4X_5 + 0.0000596X_1X_3X_4 + 0.0000055X_2X_3X_4 - 0.0000083X_2X_3X_5 + 0.0000015X_1X_2X_4X_5 - 0.0000002X_1X_3X_4X_5 \qquad (1)$$

By using the same elimination method as indicated previously, the final parameter regression coefficients and model function selected for *C.iners* using Newton's equation is therefore:-

$$\hat{y}_N = 0.160X_1X_3 + 0.644X_1X_4 + 0.196X_1X_5 \qquad (2)$$

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 also clear that Newton's volumetric equation was simpler and consistent as compared to that of Huber's. Under Newton's equation, it was also clear that the height of trunk would give a major contribution towards 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 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 [28], hence, modelling volumetric stem biomass can be a vital indicator for foresters and other relevant and related fields [29].

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 south-western 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] Bhatacharjee. *Hand Book of Indian Medicinal Plants*. Pointer Publishers, Jaipur, 1998.

[17] Hirasa, K.&Takemasa, M. *Spices Science and Technology*. MarcelDekker, 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'sJurong 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] 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.

[29] Romualdas, J.,&Augustaitis, A. Forest Health Monitoring Indicators and Their Interpretability: A Lithuanian case study. USDA Forest service, 1998.