Research Journal of

Forestry

ISSN 1819-3439

Academic
Journals Inc.

www.academicjournals.com

Research Journal of Forestry 6 (1): 1-15, 2012
ISSN 1819-3439 / DOI: 10.3923/rjf.2012.1.15

Sustainable Urban Forest using Multiple Regression Models

Noraini Abdullah, H.J. Zainodin and Amran Ahmed
School of Science and Technology, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia

Corresponding Author: Noraini Abdullah, School of Science and Technology, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia

ABSTRACT

Global Warming and carbon sequestration have expedited the urgency of proper forest management and its practices. This sustainability comes with stringent environmental policies, planning and management. With increased pressure for better and quality urban living, the existence of well-managed urban forest is of prime importance where the need for human recreational activities, balanced natural ecosystem and habitat, as well as the oxygen and carbon cycle, have to be sustained. Since, the green canopy is directly related to the bole and tree volume, urban forests sustainability can be mathematically modeled using multiple regressions. The Multiple Regression (MR) models are based on the tree-stem mensuration data. Three volumetric formulas are used to calculate the tree stem volume, namely, the Newton, Huber and Smalian equations. Data are collected and categorized according to sizes: small (S), medium (M) and large (L). Six independent variables based on measurable variables and five categorical variables based on location samples have been taken. Data transformations are done for normality and Spearman correlation coefficient matrix is used to identify bivariate relationships between them. Removal of high multicollinearity and insignificant variables and applying parameter tests are done on the models. Three selected best volumetric models from the equations are chosen based on the eight selection criteria (8SC). Comparisons are made to have the best model equation. The Newton's MR models on all the three sizes are found to be the best to represent the mensuration growth factors which affect the sustainability of the urban forests.

Key words: Sustainable urban forests, volumetric formulas, multiple regressions, growth factors, best model

INTRODUCTION

According to Bechtold (2004), the stem (bole) diameter was statistically significant and the strongest predictor in all models of tree estimations, and for some species a quadratic term was needed for model enhancement. John (2003) also wrote on models for estimation and simulation of crown and canopy cover, covering eight states in the United States, measuring the diameter at breast height, crown widths, crown ratio, number of stems, density measurements, elevation and various geographical or political divisions. Shawn *et al.* (2007) had stated that tree crowns are the sources of carbohydrates from photosynthesis, hence, serving as a vigor indicator of the tree, while Hasenauer (2006) used modelling to predict future forest stand development. These model approaches were hence used to support foresters' decision makings. An allometric equation using simple linear regression, relating biomass component to independent variables like diameter at breast height (DBH) and tree height (h) was developed by Wang (2006) for 10 co-occurring tree species in China's temperate forests, while Noraini *et al.* (2006) had also presented linear and allometric equations for biomass prediction based on the high coefficient of determination (R^2). The vitality for sustainable forest, its management and practices seeks the importance for an estimation tool. Hence, a modeling approach is developed, based on the volumetric equations used by the Forestry Department.

In this study, Multiple Regression (MR) technique is employed to model the sustainability of the urban forest whereby factors affecting tree growth are considered. By modelling tree stem volumes, the green canopy cover could be estimated and indirectly, the photosynthetic functions of tree crowns and gas cycles in the urban areas could be monitored and ascertained its environmental qualities.

STUDY SITE AND MATERIALS

The main campus of Universiti Malaysia Sabah (UMS) is on a 999-acres piece of land along Sepanggar Bay in Kota Kinabalu, Sabah. Sabah is located on the island of Borneo and had been part of Malaysia since 1963. This equatorial region has a tropical climate with a mean annual rainfall around 2000-2499 mm with a relative humidity of 81.2±0.3°C. The mean annual temperature is 27.2±0.1°C and its seasonal rain extends from October to January (http://www.met.gov.my). The study site is at latitude 6°00' and longitude 116°04' (http://maps of the world.com). Mensuration Data were collected from 130 trees, planted all around the city of Kota Kinabalu, as depicted by the pictures of the Google Maps in Fig. 1.

Fig. 1(a-e): Five sample locations of tree plots. (a) Entrance and exit gate of UMS, (b) UMS-CIMB parking lots, (c) Road toward the UMS chancellery, (d) Indah Permai housing estate and (e) Kota Kinabalu Padang Merdeka

The materials used in field data measurements would include a clinometer and a girth tape. The clinometer is used to measure the height of a tree while the fiberglass tape measures the diameter indirectly, by wrapping round the tree to measure the circumference in a perpendicular plane to the stem axis, and its value is divided by (π) to estimate the diameter. The diameter measurements are taken at the respective heights of the tree, namely, at the top, middle, at breast height and at the base of tree.

If the height of stem is denoted by H_s, then diameter at the top is measured at 0.9 H_s, middle at 0.5 H_s, base at 0.1 H_s while diameter at breast height is technically approximately 1.384 m above the ground level. Figure 2 depicts the schematic diagram of the tree stem measurements taken during data collection. These measurements are then used to calculate the stem volumes and also are regarded as independent variables in the multiple regression technique.

STEM VOLUMETRIC EQUATIONS

The following equations viz., the Newton, Huber and Smalian formulae, are used to calculate the tree stem volumes based on the variables considered in Fig. 2. The areas at each respective log sections, that is, top, middle and at the base are calculated using the standard formula and the variables in Fig. 2. They are then substituted into the following equations for the stem volumes.

Newton's Formula: $$V_{Nw} = \frac{H_s}{6}(A_b + 4A_m + A_t)$$ (Fuwape *et al.*, 2001) (1)

Huber's Formula: $$V_{Hb} = \pi H_s \left(\frac{d_{bh}^2}{40000}\right)$$ (Brack, 2006) (2)

Smalian's Formula: $$V_{Sm} = \frac{H_s}{2}(A_b + A_t)$$ (Husch *et al.*, 2003) (3)

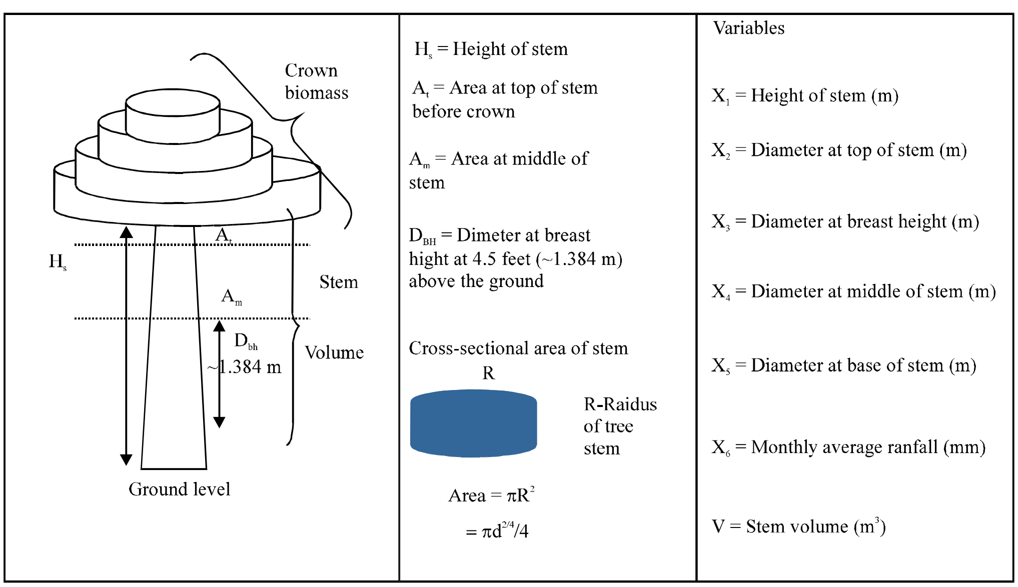


Fig. 2: Schematic diagram of tree stem measurements and variables

METHODOLOGY

Preliminary, modelling starts with the conceptual development of the importance of forest stands, its estimations and its contributions to the ecosystem, follows then by the application of mathematical theories for model building. The variables are initially tested for normality and transformations are done for any non-normal data. Characteristics of the non-linear data can be identified using the ladder-power transformations. Searching for the best transformations which are applicable to the data sets and then optimizes the range of transformation needed have also become a part of the models' simulations and optimization in the modelling procedures.

Figure 3 depicts the Modeling flowchart, in which this study will focus on the phases in Stage 3 of the model-building development; Phase 1 is the all possible models considered, Phase 2 is the multicollinearity removal and insignificant variables eliminated from the models selected, Phase 3 is the best model selection based on the eight selection criteria (8SC) and finally, Phase 4 is the model's goodness-of-fit based on the standardized residuals. Consequently, the best model can be employed for future predictions by using other forecasting measures, such as, the Mean Average Prediction Error (MAPE) or the Mean Average Error (MAE). Models developed for forecasting, using say the Mean Average Prediction Error (MAPE) of less than 10%, would give very good estimates.

The model-building is developed based on the method of multiple regressions, a statistical method of more than two independent variables that can be written in a general form as:

$$V = \Omega_0 + \Omega_1 W_1 + \Omega_2 W_2 + \ldots + \Omega_k W_k + \mathbf{u} \tag{4}$$

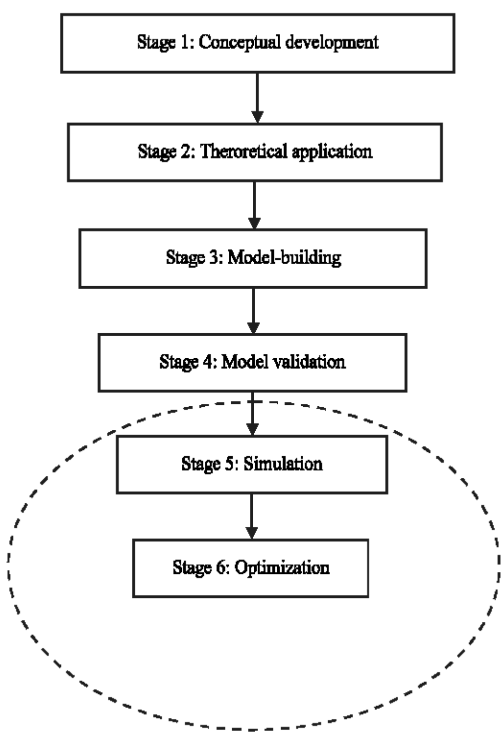


Fig. 3: Modelling flowchart

Table 1: Total number of models for each volumetric equation

No. of Variables	Single Independentvariables	Interactions 1st order	2nd order	3rd order	4th order	5th order	Total number of models
1	6						6
2	15	15					30
3	20	20	20				60
4	15	15	15	15			60
5	6	6	6	6	6		30
6	1	1	1	1	1	1	6
Total	63	57	42	22	7	1	192

where, W is an independent variable which represents one of these types of variables, namely, single independent, interaction, generated, transformed or even dummy and categorical variables, and u as the error terms. The number of models is given by the formula:

$$\text{No. of models} = \sum_{j=1}^{q} j({}^{q}C_j)$$

where, q is the number of single independent variables. Based on six single independent variables, except for dummy or categorical variables the number of models then is 192 models (Table 1), with interactions up to the fifth order for each volumetric equation. Each of the 192 models created can be written as in the general model of Eq. 4. With three equations on the stem volume, hence, giving the total number of models for model comparisons as 576 models.

One of the possible models with different variables' attributes can be given by say model M121:

$$v = \beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_{12}X_{12}+\beta_{13}X_{13}+\beta_{23}X_{23}+\beta_{123}X_{123}+\beta_{D1}D1+\beta_{D2}D2+\beta_{D3}D3+\beta_{D4}D4+\beta+D5+u \quad (5)$$

where, X_1, X_2 and X_3 are the single independent variables, X_{12} (or X_1X_2) and X_{13}(or X_1X_3) are the 1st order interactions, X_{123} (or $X_1X_2X_3$) is the 2nd order interaction and D1,..., D5 are the categorical variables based on the five location plots.

Multicollinearity is a statistical phenomenon in which two or more independent variables are highly correlated, thus, affecting the SSE of the respective models. These effects of multicollinearity can be remedied by doing model transformations or by removing the highly correlated source variables (Gujarati, 1995). Absolute coefficient values that, are greater than 0.75, are normally considered to cause multicollinearity effects. However, in this study, high correlation coefficient of absolute values greater than 0.95 ($|r|\geq 0.95$) are chosen to be removed.

The Zainodin-Noraini multicollinearity remedial procedures had been illustrated in detail in Zainodin *et al.* (2011). These remedial techniques of multicollinearity removal are executed in Phase 2 of the model-building development. A case I type of multicollinearity problem occurs when there is a presence of a high frequency with no tie in the correlation coefficient matrix. This independent variable having the highest absolute $|r|\geq 0.95$ is removed and the model is thus, rerun to check any more occurrences of multicollinearity.

Contrarily, a case II type problem occurs when there is a presence of some independent variables of the highest frequencies with a tie found in the correlation matrix. The independent variables with the number of frequencies are listed in a frequency table for identification. The

Table 2: Multicollinearity removal procedures

Types	Description	Procedures and action taken
Case I	Highest frequency with no tie	An independent variable has high |r|≥0.95 among variables. Remove corresponding variable and model is rerun.
Case II	Highest frequency and tie found	Some independent variables have same frequency with high |r|≥0.95. Remove the weakest/least value with the dependent variable and rerun model.
Case III	Highest frequency with only one tie	Two pairs of multicollinearity variable have the same frequency of 1. Remove the weakest/least with the dependent variable and rerun model.

variable with the highest frequency and the least value or the weakest contribution to the dependent variable will be removed. The model is then rerun to check any more occurrence of multicollinearity.

A case III type will involve two pairs of multicollinearity source variables of the highest frequency with only a single tie. By comparing these two independent variables, the variable with least value or the weakest contribution to the dependent variable is removed. The model is then again rerun to check the presence of multicollinearity source variables.

All the models may exhibit any one of these cases, or a combination of the case-types in their correlation coefficient matrices (Zainodin *et al.*, 2011). Table 2 shows the Case Types for Multicollinearity removal procedures that are to be carried out on all the models.

There are four phases involved in obtaining the Multiple Regression (MR) models. Starting with Phase 1 of the all possible models of the single independent variables and their related interaction variables, the number of models will depend on the number of single independent variables of the data sets. This is then followed by Phase 2 of the selected model whereby two procedures are involved, viz., the Zainodin-Noraini remedial techniques of multicollinearity removal and the elimination procedures of the insignificant variables. Parameter tests are carried out during the model selection. This will include the Global (or F-) Test where the overall significance of the independent variables on the dependent variable is tested, the Coefficient Test where the backward elimination procedures are implemented and the Wald test which is to test the overall goodness of fit of the stem volume regression model, by checking the effect of the omitted independent variables on the volumetric stem.

Phase 3 incorporates all the eight criteria used for selection of the best model (8SC). The best model is the one which fulfils the least value of most of the criteria used (Ramanathan, 2002). All the criteria (8SC) is based on the (k+1) parameters where, k is the number of single independent variables and Sum of Square Error (SSE) with a penalty factor. The selection of the eight criteria (8SC) used in identifying the best MR model had been used.

Finally, the best model will undergo the goodness-of-fit tests of Phase 4. Phase 4 involves two goodness-of-fit tests which comprises of the randomness test and the normality test on the residuals of the best model chosen from Phase 3. The goodness-of-fit tests are done to validate the choice of the best model, where the assumptions of randomness and normality of the model's residuals are verified. The best model could then be used for forecasting and thus, provide a better prediction for future forest planning strategy and management.

RESULTS AND ANALYSES

Five tree mensuration variables, comprise of the height of stem or bole (H_s), diameter at the top of stem (D_t), diameter at breast height (D_{bh}), diameter at the middle of stem (D_m) and diameter at the base of stem (D_b), are measured using a girth tape and a clinometer. These mensuration data

Table 3: Definitions of data variables

Main variables	Definition of variables	New transformed variables
V_{nw}	Stem Volume.(m^3): Nw-Newton;	V1
V_{hb}	Hb-Huber	V2
V_{sm}	Sm-Smalian	V3
H_s	Height of Stem (bole)(m)	X1
D_t	Diameter at the top of stem	X2
D_{bh}	Diameter at breast height	X3
D_m	Diameter at the middle of stem	X4
D_b	Diameter at the base of stem	X5
R	Average rainfall over the months	X6
Di	Entrance Gate into UMS	D1
Dii	Exit Gate of UMS	D2
Diii	UMS-CIMB Parking Lots	D3
Div	Indah Permai Housing Estate	D4
Dv	Kota Kinabalu Padang Merdeka	D5

Table 4: Normality tests of Shapiro-Wilks statistics according to tree sizes

	Shapiro-wilks statistics											
Tree sizes	S				M				L			
volume equation	Stats.	df	Sig.p	Normality	Stats.	df	Sig. p	Normality	Stats.	df	Sig. p	Normality
Newton	0.950	41	0.069	/	0.964	46	0.163	/	0.989	43	0.946	/
Huber	0.948	41	0.059	/	0.951	46	0.052	/	0.989	43	0.941	/
Smalian	0.961	41	0.168	/	0.976	46	0.442	/	0.932	43	0.014	x

are then used to calculate the stem biomass volumes based on the three equations in section 2. Information on tree sample locations is treated as 'categorical' variables. In this study, five sample locations of tree plots had been taken during data collection, as shown in Table 3 as D1-D5 of the new transformed variables.

The independent variables for multiple regressions are comprised of these tree mensuration data and average rainfall (R) over the months and tree-locations. Overall, there are six single independent variables as quantitative variables, with five as qualitative variables. Transformations of non-normal data are carried out after normality tests are done on the data sets. The types of ladder transformations would depend on the concavity or the convexity of data set scatter plots. The data sets are categorized into three tree-sizes, namely, small (S; 41), medium (M; 46) and large (L; 43) and then substitute into the volume equations of the stem biomass used in section 2. Since the number of trees are less than 50 ($n<50$), the Shapiro-Wilks statistics with $\alpha>0.05$ (more than 5% significance level) is used to determine normality. The significant p-value of more than 0.05 indicates that the variables according to sizes are all normal except for the Smalian's large trees having a significant p-value of less than 0.05 (0.014), as shown in Table 4. However, the normality plot of the volume using Smalian equation of the L-sized trees in Fig. 4 indicates that it is relatively quite normal at 4%, hence, no transformation is needed.

Taking the variables using the Newton's equation as an example, the Spearman Correlation Coefficient matrix (since, $n<50$) was given in Table 5. Table 5 showed the correlation coefficients of the sizes (S, M and L) with their corresponding number of samples, that is (41, 46 and 43), respeclively using Newton's equation for model M63 with all the six main

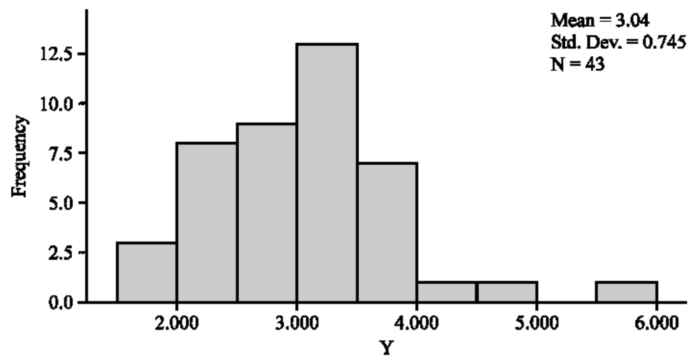


Fig. 4: Normality plots of smalian volume for large trees (L)

Table 5: Coefficient matrix using newton's equation of six main variables of model M63

M63	V1	X1	X2	X3	X4	X5	X6
V1							
41		S.699	S.958	S.971	S.875	S.092	S.927
46		M.33	M.870	M.951	M.810	M.826	M.793
43	1	L.519	L.868	L.884	L.746	L.635	L.667
X1							
41			S.603	S.600	S.600	S.094	S.625
46			M.142	M.107	M.115	M.147	M.073
43		1	L.224	L.162	L.447	L.111	L.201
X2							
41				S0.966	S.808	S.021	S.919
46				M0.868	M.585	M..66	M.536
43			1	L0.843B	L0.599	L.627	L.569
X3							
41					S.802	S.036	S.912
46					M.720	M.865	M.795
43				1	L.452	L.656	L.583
X4							
41						S.225	S.859
46						M.673	M.833
43					1	L.442	L.668
X5							
41							S.039
46							M.898
43						1	L.901
X6							
41							
46							
43							1

variables (Table 6). It could be seen that there were some highly positive correlation between the variables (highlighted). These problems of multicollinearity had to be remedied first so as to avoid biased estimations in the resultant models.

Taking model M63 of small(S)-sized tree and using Newton's equation as an example for illustration. A high multicollinearity existed between X2 and X3 and with the dependent

Table 6: Correlation matrix of small-sized using Newton's with Case I Type

M63		V1	X_1	X_2	X_3	X_4	X_5	X_6
V1	41	1	S.699	S.958	S.971	S.875	S.092	S.927
X1	41		1	S.603	S.600	S.600	S.094	S.625
X2	41			1	S.966	S.808	S.021	S.919
X3	41				1	S.802	S.036	S.912
X4	41					1	S.225	S.859
X5	41						1	S.039
X6	41							1

Table 7: Correlation matrix after 1st multicollinearity removal of source variable

M63.1		V1	X_1	X_2	X_4	X_5	X_6
V1	41	1	S.699	S.958	S.875	S.092	S.927
X1	41		1	S.603	S.600	S.094	S.625
X2	41			1	S.808	S.021	S.919
X4	41				1	S.225	S.859
X5	41					1	S.039
X6	41						1

Table 8: Correlation matrix after 2nd multicollinearity removal of source variable

M63.2		V1	X_1	X_4	X_5	X_6
V1	41	1	S.699	S.875	S.092	S.927
X_1	41		1	S.600	S.094	S.625
X_4	41			1	S.225	S.859
X_5	41				1	S.039
X_6	41					1

variable V_1, with a Case I type. X3 was then removed since it has the highest frequency of occurrences, that is, 2. After the first variable removal, the model hence became model M63.1, where the value of 1 indicated the 1st removal of multicollinearity source variable (Table 7).

By rerunning the model again, variable X2 was found to have a case III type and had to be removed due to multicollinearity. The resulting model thus became M63.2, where the number 2 signified the 2nd removal of multicollinearity source variable, as shown in Table 8.

The backward elimination procedures of the Coefficient test could then be carried out so as to eliminate any insignificant variables. With just four single independent variables remained, the model is rerun to test the significance of the remaining variables. Any insignificant variables (variables with p-values>0.05) would be omitted from the model. This is illustrated in Table 9 whereby variable X5 having a p-value of 0.81 is eliminated in the first iteration. Rerunning the model again for any insignificant variable, it is found that there is no variable with p-values >0.05 in the resulting model. This implies that all the variables are significant and would contribute to the dependent variable, the stem volume. The final model hence becomes M63.2.1, where the value of 1 indicates the 1st iteration of elimination of insignificant variables.

All the possible models would undergo the multicollinearity removal procedures, the backward elimination method of eliminating insignificant variables as well as the Wald test in order to ensure that the omitted variables were verified for elimination. Model comparisons were then made based on the eight selection criteria (8SC). The criteria used were then compared for all models according to the respective stem volumes. The best model for each respective volume would then be the one

Table 9: Eliminated insignificant variables of the coefficient test

Model M63.2		Unstandardized coefficients β	Standard error	Significant p-value
1	Constant	-1.638	0.305	0.000
	X1	0.163	0.058	0.008
	X4	1.254	0.468	0.011
	X5	-0.028	0.115	0.810
	X6	2.321	0.470	0.000
2	Constant	-1.684	0.234	0.000
	X1	0.162	0.057	0.007
	X4	1.211	0.429	0.008
	X6	2.357	0.441	0.000

Table 10: Comparisons of the best models using newton's, Huber's and Smalian's equations

Model	Vnw			Vhb			VSm		
tree size	S	M	L	S	M	L	S	M	L
k+1	14	12	12	9	9	13	11	8	10
SSE	3.0E-04	5.0E-04	9.0E-04	9.0E-04	1.0E-03	1.2E-03	3.8E-03	2.7E-03	2.4E-03
AIC	1.13E-05	1.88E-05	3.38E-05	3.50E-05	3.21E-05	4.93E-05	1.60E-05	8.45E-05	8.96E-05
FPE	1.14E-05	1.89E-05	3.42E-05	3.53E-05	3.23E-05	5.03E-05	1.62E-05	8.48E-05	9.04E-05
GCV	1.23E-05	2.01E-05	3.66E-05	3.71E-05	3.36E-05	5.53E-05	1.74E-05	8.75E-05	9.55E-05
HQ	1.33E-05	2.21E-05	3.98E-05	4.02E-05	3.67E-05	6.00E-05	1.89E-05	9.52E-05	1.04E-04
RICE	1.42E-05	2.23E-05	4.15E-05	4.03E-05	3.57E-05	6.81E-05	2.02E-05	9.15E-05	1.05E-04
SCHWARZ	1.78E-05	2.91E-05	5.30E-05	5.10E-05	4.59E-05	8.40E-05	2.53E-05	1.16E-04	1.35E-04
SGMASQ	9.00E -06	1.53E-05	2.72E-05	2.89E-05	2.70E-05	3.86E-05	1.28E-05	7.23E-05	7.33E-05
SHIBATA	1.01E-05	1.72E-05	3.06E-05	3.25E-05	3.02E-05	4.32E-05	1.44E-05	8.05E-05	8.24E-05

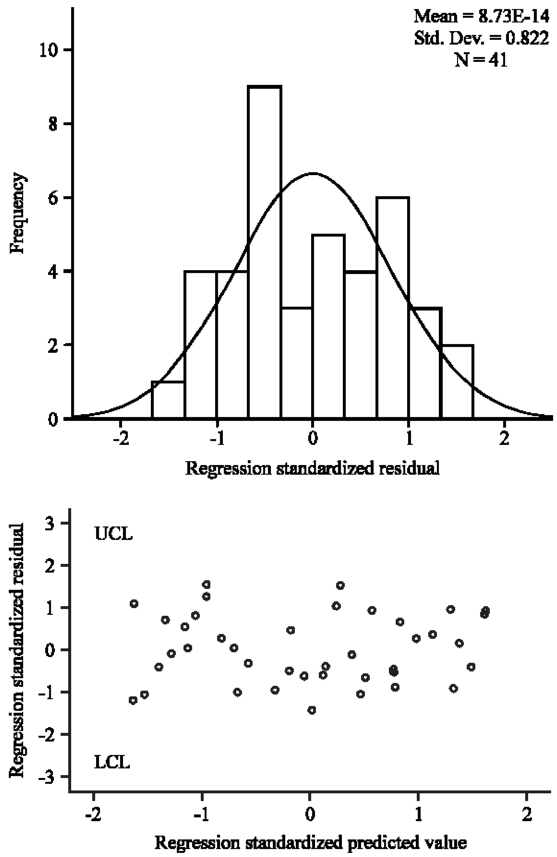
Table 11: Normality tests on residuals of newton's MR models According to tree sizes

Tree sizes newton's	Shapiro-Wilks Statistics S			M			L		
MR models	Stats.	df	Sig. p	Stats.	df	Sig. p	Stats.	df	Sig. p
Residuals	0.947	41	0.054	0.961	46	0.128	0.989	43	0.944

that had the least value of majority of the criteria. They were then compared and the one having the least value among all the best models would be the best MR model, as shown by the following Table 10.

Table 10 signified the comparisons of the MR models based on the stem biomass equations. It could be seen that the Newton's MR models could represent the best model which gave a better estimation on the stem volume. With a significance level of more than 0.05 to indicate normality (α>0.05), the numerical tests of normality on the residuals gave the Shapiro-Wilks statistics of 0.947(S), 0.961(M) and 0.989(L), respectively, with significant p-values of 0.054(S), 0.128(M) and 0.944(L) on the residuals (Table 11).

The residuals histogram and residuals scatter plots of S-sized tree of the best Newton's MR Model had all indicated that they were normally and randomly distributed, as shown in Fig. 5. The goodness-of fit tests on the residuals of the best model (M192.52.2) had thus indicated that the assumptions of randomness and normality of the residuals had been satisfied.

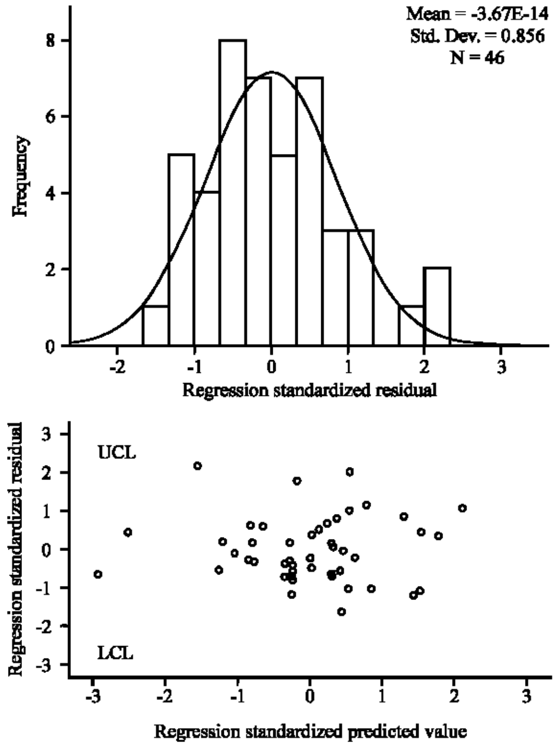


Model M192.52.2	Unstandardized coefficients β	Unstandardized coefficients Standard error	Sig.
Constant	0.394	0.061	0.000
X1	-0.028	0.009	0.004
X4	-1.226	0.140	0.000
X5	-0.103	0.036	0.008
X12	0.111	0.027	0.000
X23	-1.981	0.080	0.000
X25	-0.392	0.043	0.000
X45	0.478	0.071	0.000
X56	0.063	0.017	0.001
X134	0.477	0.021	0.000
X246	-0.748	0.303	0.020
X1235	0.047	0.016	0.007
X12345	0.306	0.060	0.000
X23456	-0.739	0.228	0.003
D4	0.019	0.002	0.000

Fig. 5: Goodness-of-Fit plots of the best Newton's MR model of S-sized tree

From Fig. 5, the Newton's MR model of S-sized tree is given by:

$$
\begin{aligned}
V_{Nw}^{S} &= 0.394 - 0.028X_1 - 1.226X_4 - 0.103X_5 + 0.111X_{12} + \\
&1.981X_{23} - 0.392X_{25} + 0.478X_{45} + 0.063X_{56} + 0.477X_{134} - \\
&0.748X_{246} + 0.047X_{1235} + 0.306X_{12346} - 0.739X_{23456} + 0.019D_4 + u
\end{aligned}
\tag{6}
$$



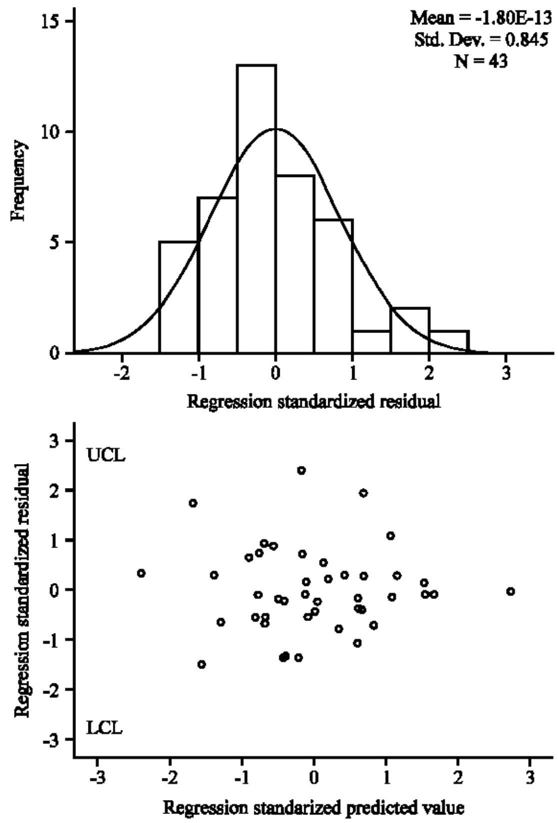
Model M192.52.2	Unstandardized coefficients		
	β	Standard error	Sig.
Constant	-2.353	0.169	0.000
X1	0.133	0.020	0.000
X3	1.926	0.095	0.000
X4	1.930	0.154	0.000
X5	0.408	0.037	0.000
X123	0.526	0.034	0.000
X125	-0.046	0.009	0.000
X145	-0.038	0.012	0.003
X245	-0.435	0.079	0.000
X456	-0.177	0.049	0.001
X2346	-1.524	0.148	0.000
X2356	-0.444	0.068	0.000
X12345	0.272	0.562	0.000

Fig. 6: Goodness-of-Fit plot of the best Newton's MR model of M-size tree

The residuals histogram and residuals scatter plots of M-sized tree of the best Newton's MR Model had all indicated that they too were normally and randomly distributed, as shown in Fig. 6. The goodness-of fit tests on the residuals of the best model (M192.52.5) had also indicated that the assumptions of randomness and normality of the residuals had also been satisfied.

From Fig. 6, the Newton's MR model of M-sized tree is given by:

$$V_{Nw}^{M} = -2.353 - 0.133X_1 - 1.926X_3 - 1.930X_4 + 0.408X_5 + 0.526X_{123} - 0.046X_{125} - 0.038X_{145} - 0.435X_{245} - 0.177X_{456} - 1.524X_{2346} - 0.444X_{2356} + 0.272X_{12345} + u \quad (7)$$



Model M192.52.4	Unstandardized coefficients β	Standard error	Sig.
Constant	-2.093	0.233	0.000
X1	0.147	0.027	0.000
X3	0.953	0.219	0.000
X4	0.802	0.123	0.000
X5	0.195	0.057	0.000
X125	0.067	0.010	0.000
X135	-0.081	0.019	0.000
X235	0.916	0.132	0.000
X256	-1.860	0.078	0.000
X1236	0.575	0.030	0.000
X1246	0.178	0.042	0.000
X2346	-2.661	0.338	0.000
X3456	0.917	0.108	0.000

Fig. 7: Goodness-of-Fit plots of the best Newton's MR model of L-sized tree

The residuals histogram and residuals scatter plots of L-sized tree of the best Newton's MR Model had both indicated that they also were normally and randomly distributed, as shown in Fig. 7. The goodness-of fit tests on the residuals of the best model (M192.53.4) had also indicated that the assumptions of randomness and normality of the residuals had also been satisfied.

From Fig. 7, the Newton's MR model of L-sized tree is thus given by:

$$V^{L}_{Nw} = -2.093 + 0.147X_1 + 2.953X_3 + 0.802X_4 + 0.195X_5 + 0.067X_{125} - 0.081X_{135} + 0.916X_{235} - 1.860X_{256} + 0.575 X_{1236} + 0.178X_{1246} - 2.661X_{2346} + 0.917X_{3456} + u \quad (8)$$

DISCUSSION

The stem volumetric best MR model using the Newton's equation in Eq. 6 showed that for small trees, tree plots and locations (D4) as a qualitative (categorical) variable, had become one of the determining factors in the best model equation. This is expected that since small plants and trees need care and attention during its early stages of growth, hence noting that the locations of tree plots of plant nurseries and small trees will help arborists to plan and strategize their inventories. It should also be noted that the S-sized MR model has more variables (that is, 14 variables) compared to the other models which had contributed to the growth of the tree stem volume. From the original possible model of 68 variables, 52 variables are removed due to multicollinearity with 2 insignificant variables eliminated.

It should be noted too that the contribution of tree plots and locations to the model, however, diminishes as the tree size increases. This can be seen by the absence of 'categorical' variables in the model Eq. 7 and 8, respectively. It should also be noted that both the M-sized MR model and the L-sized MR model have less variables (that is, 12 variables) compared to the S-sized model, which had contributed to the growth of the tree stem volume. For the M-sized MR model equation, 52 variables are removed due to multicollinearity with 5 insignificant variables being eliminated, while for the L-sized MR model equation, 53 variables are removed due to multicollinearity and 4 insignificant variables eliminated.

It was obvious that from Eq. 6 that the tree stem (H_s), diameter at the middle (D_m) and diameter at the base (D_b) are the main common variables or major contributors. It was also obvious from Eq. 7 and 8 that the tree stem (H_s), diameters at breast height (D_{bh}), middle (D_m) and base (D_b) are the main common variables or major contributors to the final models of the medium (M) and large (L) trees. Other contributions are from the second and the third-order interaction variables, except for S- and M-sized which had included the fourth-order interaction variables in their models.

CONCLUSIONS

The eight selection criteria is effective in identifying the best model, where formally the criteria used is only based on the R^2 or the adjusted-R^2 for model selection. Tree sizes, location of samples, and inclusion of other environmental factors, such as rainfall, will not affect much on the resulting outcome of the choice of the best model. In fact, as the tree size increases, sample locations as 'categorical' variables in the models have no contributions at all to the final best model. However, as for small trees, locations may play a vital role in tree-planting strategy and management of nurseries. Comparisons between the best models of the three stem biomass volumetric equations, the Newton's MR model again has appeared to be able to represent a better model for volumetric prediction (Abdullah *et al*., 2008). The best model equations were also found to have reduced to a more simplified form after multicollinearity removal procedures. Zabek and Prescott (2006) had stated that accurate assignments of carbon credits required accurate estimations of the amount of carbon stored in the biomass. Since, tree shapes do influence the choice of the equations and ultimately, the best model, it is therefore suggested that further works be done on identifying these models. Hence, derivation of biomass equations and models are vital, as a tool in estimating aboveground woody carbon sequestration and ultimately, can estimate the ranges and rates of forest stands over time.

REFERENCES

Abdullah, N., Z.H.J. Jubok and J.B.N. Jonney, 2008. Multiple regression models of the volumetric stem biomass. WSEAS Trans. Math., 7: 492-502.

Bechtold, W.A., 2004. Largest-crown width prediction models for 53 species in the Western United States. West J. Applied For., 19: 245-251.

Brack, C., 2006. Tree crown: Forest measurement and modelling. http://sresassociated.anu.edu.au/mensuration/volume.htm.

Fuwape, J.A., J.C. Onyekwelu and V.A.J. Adekunle, 2001. Biomass equations and estimation for gmelina arborea and nauclea diderrichii stands in Akure forest reserve. Biomass Bioenergy, 21: 401-405.

Gujarati, D.N., 1995. Basic Econometrics. 3rd Edn., McGraw-Hill, Inc., New York, pp: 540-583.

Hasenauer, H., 2006. Concepts within Tree Growth Modelling. Sustainable Forest Management. Springer, New York, USA.

Husch, B., T.W. Beers and J.A. Jr. Kershaw, 2003. Forest Mensuration. 4th Edn., John Wiley and Sons, Inc., New Jersey, USA., pp: 949.

John, D.S., 2003. Models for estimation and simulation of crown canopy cover. Proceedings of the 5th Annual Forest Inventory and Analysis Symposium. November 18-20, 2003, New Orleans, LA. Gen. Tech. Rep. WO-69. Washington, D.C., U.S., pp: 222-222.

Noraini, A., S.R.M. Hashim and K. 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 Managemant Conference, November 21-23, 2006, University of Malaysia Sabah, Science and Technology, Malaysia. pp: 250-255.

Ramanathan, R., 2002. Introductory Econometrics with Applications. 5th Edn., South-Western, Thomson Learning, Ohio, USA.

Shawn, X.M., V.J. Lieffers and H. Shongming, 2007. Modeling crown volume of lodgepole pine based upon the uniform stress theory. For. Ecol. Manage., 251: 174-181.

Wang, C., 2006. Biomass allometric equations for 10 co-occurring tree species in Chinese temperate forests. For. Ecol. Manage., 222: 9-16.

Zabek, L.M. and C.E. Prescott, 2006. Biomass equations and carbon content of aboveground leafless bioass of hybrid poplar in coastal British Columbia. For. Ecol. Manage., 223: 291-302.

Zainodin H.J., A. Noraini and S.J. Yap, 2011. An alternative multicollinearity approach in solving multiple regression problem. Trends Applied Sci. Res., 6: 1241-1255.