Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 2, 2021, page 203-212

Development of Local Tree Volume Tables for Eucalyptus Pellita and Acacia Mangium Using Regression Analysis

[1]Nizar Narudin and [2]Dr Muhammad Fauzee Hamdan

[1,2]Department of Mathematical Sciences
Faculty of Science, Universiti Teknologi Malaysia,
81310 Johor Bahru, Johor, Malaysia.

e-mail: [1]nizarnarudin@gmail.com 1, [2]mfauzee@utm.my

Abstract Sustainable forestry plays a vital role in our life. Thus, forestry management needs to leverage tree volume table to maximize efficiency in forest conservation. The purpose of this study is to investigate the application of regression analysis in formulating a model to best estimate the volume of local standing trees specifically *Eucalyptus pellita* and *Acacia mangium*. Statistical approaches such as residual plot and *R*-squared were used to measure the adequacy of the models. Linearity of data is tested. The accuracy of all models is tested and compared using RMSE and MAPE. The models applied to the datasets was studied and the effectiveness and shortcomings of independent variables were highlighted. Results were compared between MLRA, SLRA, and SNRA. Findings suggest that the best model for *Eucalyptus pellita* and *Acacia mangium* are SNRA and MLRA model respectively.

Keywords *Eucalyptus pellita; Acacia mangium;* regression analysis; RMSE; MAPE

1 Introduction

Local volume tables of specific species are very important for estimation of timber volume of standing trees [13]. However, site-specific volume tables lack two important tree species which are *Eucalyptus Pellita* and *Acacia Mangium*. Hence, this study was carried out to determine which model (linear or nonlinear) is the best in developing local volume tables for these 2 species via regression analysis.

According to Clark [2], Heinrich Cotta introduced the concept of volume table for the forest trees in around 1804. It took many years of data collection through extensive study to construct the first volume table. Norway spruce was the main subject of the study. Volume table is significant in calculating standing trees volume on a global scale [6].

The volume is calculated using the volume measurements of 40 to 50 fallen or standing trees of various diameter classes. It is based on the assumption that trees of the same species, having the same diameter, or both diameter and height should on average have the same volume provided that they have similar growing conditions.

1.1 Tree Volume Table

One of the important aspects of forest inventory for managing the forest ecosystem is establishing the tree volume table [10]. For a given species, the average contents of trees of various sizes can be shown using a volume table [3]. Similarly, according to Honer [5], standard volume tables give estimation of individual tree volume based on diameter at breast height, total or merchantable height, and sometimes a measure of form.

Reliable volume tables can be constructed with the aid of enumeration of a relatively large number of sample trees of various sizes classes as well as species or species group [4]. Additionally, accurate diameter measurements of several sections of the stem are required from each of the sample tree. In order to obtain such measurements, the sample trees have been measured either using felled or standing trees [4]. Hajar et al. [4] suggest that the best approach to estimate volume in most tropical forests is through volume tables developed via a volume equation.

Since volume tables were made using data from broad region, the current volume tables might underestimate or overestimate the tree volume of individual trees in a specific region [12]. Volume tables for *Eucalyptus Pellita* in Indonesia might be different than in Malaysia. Hence, local volume tables are strongly needed for the purpose of supporting the local forest administration in management of forest.

1.2 Regression Analysis

Regression analysis is a method commonly used in statistical modelling. According to Sykes [14], it is a statistical tool to investigate the relationships between variables. Similarly, it is a set of statistical processes where relationships between dependent variable (commonly called the ‘outcome variable’ and usually denoted as ‘*y*’) and independent variables (commonly called ‘predictors’ and usually denoted as ‘*x*’) are estimated.

Sykes [14] states that the investigator usually seeks to find the causal effect of one variable upon another. Sykes [14] also mentions that “statistical significance” of the estimated relationships, also known as, the degree of confidence that the true relationship is close to the estimated relationship is also being assessed by the investigator.

There are two types of regression. The first one would be univariate regression which can be linear or nonlinear and it has a dependent variable and one independent variable. Then there is multilinear regression which indicates that the regression models have one independent variable and more than one independent variables [15]. The efficacy of these different types of regression depends on the type of the data set that they are being implemented on since every data set has its own unique characteristics and pattern. Hence, multiple tests across different types of linear and nonlinear models are essential so as to achieve the best model. According to Khanna and Chaturvedi [7], a tree species' volume table is not used for another tree species.

2 Materials and Methods

2.1 Method of Data Collection

For data collection method, the researcher decided to use traditional way which is called the destructive method. The most accurate method to estimate tree height and volume is through destructive measurement [11].

Destructive measurement is a conventional way of measuring the parameters of tree structure such as Diameter at Breast Height (DBH) and Total Length (Height) where the trees are cut into sections of logs where each diameter of both ends of every log and total length of every tree are measured. All the logs are cut into sections of 2m as shown in the Figure 1.

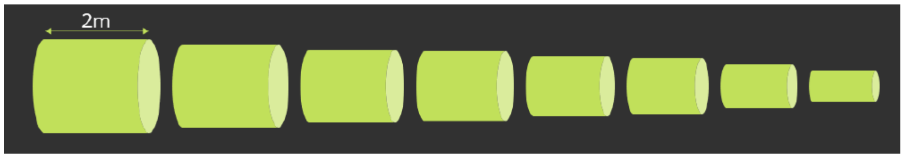

Figure 1. Sections of 2m logs

In the forest ecosystem, tree structure parameters like diameter at breast height (DBH) and height are important for evaluation of tree volume [11]. The Total Length (Height) is measured from the ground level to the top level of the tree. On the other hand, DBH of every tree is measured at approximately 1.3m (4.3ft) above ground.

2.2 Volume Calculation Method

Total tree volume of a tree will be calculated by summing-up the volume of every 2m logs of the tree. Each log volume will be calculated using Smalian's formula. According to Avery and Burkhart [1], Smalian's formula seems to be the most frequently used in the United States because it is easy to measure the diameter of both ends of logs. The Smalian's formula can be expressed as:

$$Volume\ of\ each\ section\ (m^3) = \left[\frac{B+b}{2}\right] L \tag{1}$$

Where B = the cross-sectional area at the large end of the log, b = the cross-sectional area at the small end of the log, and L= log length.
The volume at the tip section of tree will be calculated using cone formula.

$$Volume\ of\ a\ cone\ (m^3) = \frac{1}{3}\pi r^2 h \tag{2}$$

Larsen [9] claims that the diameter at the top of tree is assumed to be 0 and forming a cone.

2.3 Data Analysis

2.3.1 Regression Analysis

It is a set of statistical approach where the relationships between a dependent variable and one or more independent variable are estimated. The result of the analysis will then be utilized to measure the strength of the relationship between parameters and may also be used to model the future relationship between them. There are many variations to regression analysis, namely linear, multiple linear and nonlinear. The most commonly used model would be simple linear and multiple linear as it only shows the linear relationship between variables. More complicated data sets that exhibit nonlinear relationship between the independent and dependent variable will usually be analysed using the nonlinear regression analysis. Regression analysis are also widely used for forecasting, time series modelling and predict continuous values.

2.3.2 Simple Linear Regression

The most basic form of regression algorithms in machine learning is linear regression. The model consists of one dependent and independent variable where both of them share the same relationship which is linear. As the number of independent variables increases, it will then be called multiple regression models. Simple linear regression is denoted by the following equation below:

$$y = mx + c + e \tag{3}$$

where m is the slope of the line, c is an intercept, and e represents the error in the model.

Varying the values of m and c for different combinations will determine the best-fit decision boundary. Difference between the predicted values and observed values is called residuals or predictor error. Hence, the value of m and c will be selected in order to minimize the residuals or the predictor error.

2.3.3 Simple Nonlinear Regression

Nonlinear regression is a type of regression analysis that involves fitting data to a model and then expressing the result as a mathematical function. Nonlinear regression connects two variables (X and Y) in a nonlinear (curved) connection, whereas simple linear regression connects them with a straight line (y= mx + b).

The model's aim is to reduce the sum of the squares to the smallest feasible value. The sum of squares is a metric for determining how far the Y observations differ from the nonlinear (curved) function used to forecast Y.

It's calculated by calculating the difference between the fitted nonlinear function and each Y point in the data set. After that, each difference is squared. Finally, put all of the squared figures together. The better the function matches the data points in the collection, the lower the total of these squared numbers. Logarithmic, trigonometric, exponential, power, Lorenz curves, Gaussian functions, and other fitting methods are used in nonlinear regression.

2.3.4 Multiple Linear Regression

Multiple linear regression is similar to simple linear regression except that the independent variables are more than one. It is commonly used to draw mathematical relationship among number of random variables. In other word, MLR investigate how multiple independent variables are connected to a single dependent variable. The information on multiple variables will then be utilized for an accurate prediction on how significant are they in terms of affecting the outcome of the dependent variable. We denote multiple linear regression as follows:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip} + e \tag{4}$$

where, for $i = n$ observations:
$y_i =$ dependent variable
$x_i =$ independent variable
$b_0 =$ y-intercept (constant term)
$b_p =$ slope coefficients for each of the independent variable
$e =$ the model's error term

3 Results and Discussion

The raw data of sample felled trees in Kuala Segan, Bintulu, Sarawak was collected back in 2019. Originally, there is a total of 238 and 126 of raw sample felled trees data of *Eucalyptus Pellita* and *Acacia Mangium* respectively. Due to human errors occurred in the data sets of both species which leads to irrelevant measurement of parameters, some of the raw data needs to be excluded. Therefore, only a total of 126 and 48 raw data of *Eucalyptus Pellita* and *Acacia Mangium* were used for this study.

The total length of each tree was measured and recorded. The length and diameter of the log for every 2m of the tree was also recorded for actual tree volume calculation. Volume of every section will then be calculated using Smalian's and area of circle formulae. The section where the tip of the felled sample tree however will have additional formulae which is the cone formulae for the volume of the top portion (tip) of the section. The volume for every section will then be sum up to get the total volume of the felled sample tree.

3.1 Demographic Information (Descriptive Statistics)

On this section, data on Total Length (TL) and Diameter at Breast Height (DBH) were divided into different size of classes. Then, descriptive statistics such as histograms and summary statistics were generated using the previous data. Figures below show the result of the analysis of DBH and Total Length for both species:

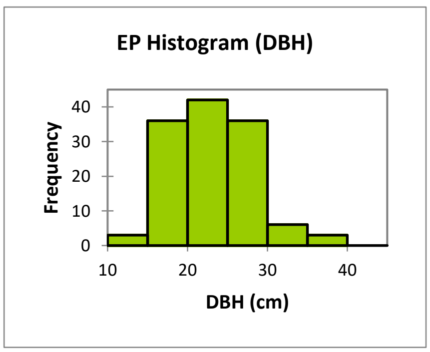

Figure 2. Histogram of DBH of *Eucalyptus Pellita*

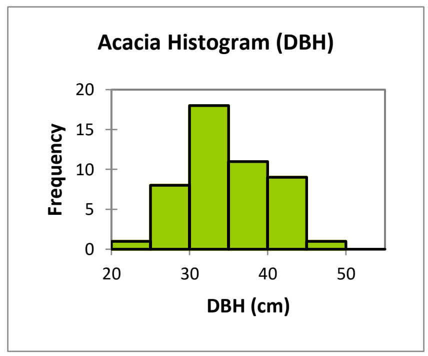

Figure 3. Histogram of DBH for *Acacia Mangium*

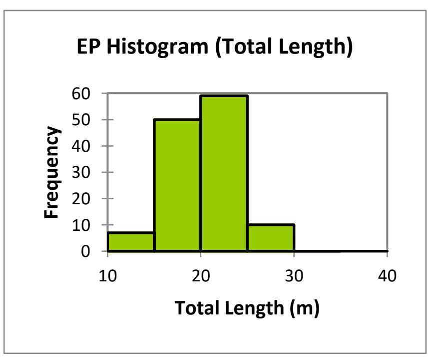

Figure 4. Histogram of TL of *Eucalyptus Pellita*

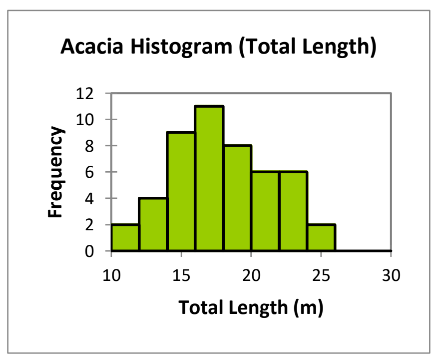

Figure 5. Histogram of TL for *Acacia Mangium*

The histogram of DBH in Figure 2 shows that the distribution for *Eucalyptus Pellita* is multimodal because there exist 3 distinct peaks in its histogram. On the other hand, Figure 3 shows that the distribution for *Acacia Mangium*'s DBH is unimodal where there is only one peak in the histogram. The 3 dominant classes for DBH of *Eucalyptus Pellita* is between 15-20cm, 20-25cm, and 25-30cm. Meanwhile, for *Acacia Mangium* the only dominant class is between 30-35cm.

In Figure 4 and 5, Total Length's histogram for *Eucalyptus Pellita* however showed more of a bimodal distribution since there exist two distinct peaks while *Acacia Mangium*'s Total Length histogram showed a unimodal distribution with a single peak. The 2 dominant classes for *Eucalyptus Pellita*'s Total Length are 15-20m and 20-25m while *Acacia Mangium* has only one dominant class for Total Length which is 16-18m.

3.2 Linearity Test

Before we start doing any regression analysis, linearity test needs to be done for both of *Eucalyptus Pellita* and *Acacia Mangium* data. This is to determine that the relationship between independent variables and the dependent variable is linear or not. Table below shows the result of the linearity test for *Eucalyptus Pellita* and *Acacia Mangium:*

Species	Relationship	sig. Deviation from Linearity
Eucalyptus Pellita	DBH*TreeVol	0.001
	TotalLength*TreeVol	0.555
Acacia Mangium	DBH*TreeVol	0.094
	TotalLength*TreeVol	0.224

Table 1: Linearity test for *Eucalyptus Pellita* and *Acacia Mangium*

For *Eucalyptus Pellita*, DBH showed nonlinear relationship to Tree Volume because it has significance value of Deviation from Linearity < 0.05 while the Total Length showed linear relationship to Tree Volume because the significance value of Deviation from Linearity > 0.05. Since both of them are different, we decide to just use DBH as independent variable as it shows higher R-squared value compared to Total Length. Therefore, we will only conduct simple nonlinear regression analysis on *Eucalyptus Pellita's* data.

On the other hand, for *Acacia Mangium*, DBH and Total Length showed linear relationship to Tree Volume because both of the independent variables have significance value of Deviation from Linearity > 0.05. Hence, we will conduct simple and multiple linear regression analysis for *Acacia Mangium's* data.

3.3 *Eucalyptus Pellita* Regression Analysis

The first nonlinear model that will be used is a widely used model especially for estimation total stem volume.

$$\log V = a_0 + a_1 logD + a_2 logHt \tag{5}$$

According to Krisnawati [8], the logarithmic model was the best in determining estimated total stem volume both over and under bark. He also stated that, this type of total stem volume model has been widely used before by many studies in estimation of stem volume. Since logarithmic models are good in determining estimated total stem volume both over and under bark, we will be using it and compare it with other nonlinear models such as exponential, polynomial, and power.

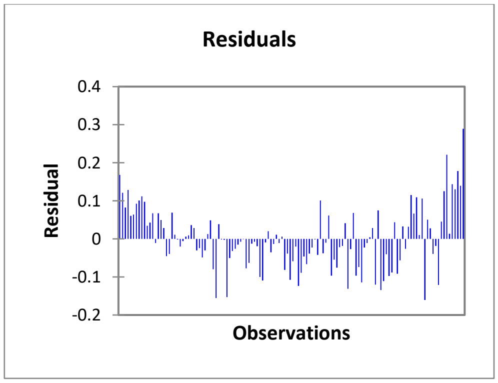

Figure 6. Residuals using SNRA (log)

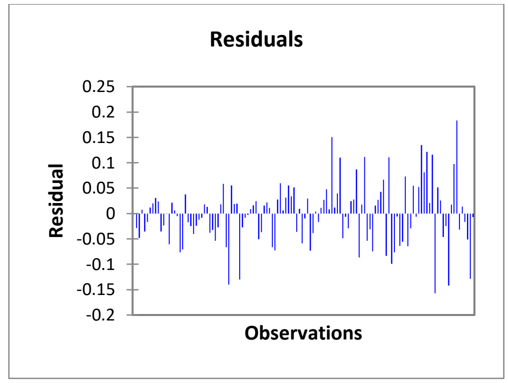

Figure 7. Residuals using SNRA (poly)

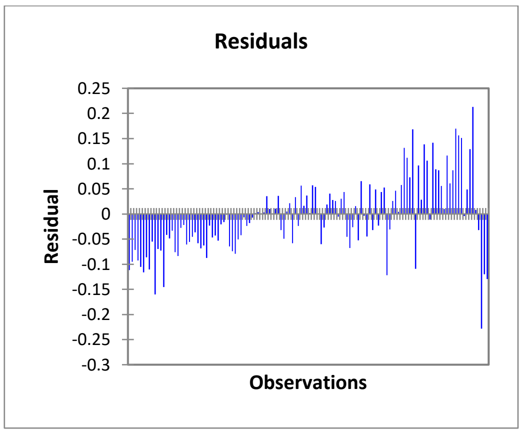

Figure 8. Residuals using SNRA (exponential)

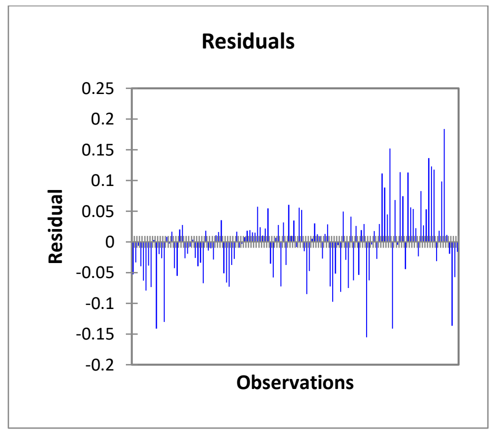

Figure 9. Residuals using SNRA (power)

Most of the residuals the models are scattered randomly. All the residuals in all nonlinear models are also centered around the *x*-axis which also indicate that the models are suitable in explaining the variation in the dependent variable.

Table 2: R-squared, MAPE, and RMSE values for all models for *Eucalyptus Pellita*

Model	R-squared	MAPE	RMSE
Polynomial	0.9428	10.447	0.059
Logarithmic	0.8957	17.872	0.092
Exponential	0.9105	15.091	0.075
Power	0.9419	10.611	0.060

Table above shows the R-squared, MAPE, and RMSE values for all the models for *Eucalyptus Pellita*. The best model is Polynomial model via SNRA because it has the highest *R*-squared value and lowest MAPE and RMSE values compared to other models.

3.4 *Acacia Mangium* Regression Analysis

The same type of regressions as in *Eucalyptus Pellita* are applied to *Acacia Mangium.* Figures below shows the residual plots extracted from every analysis:

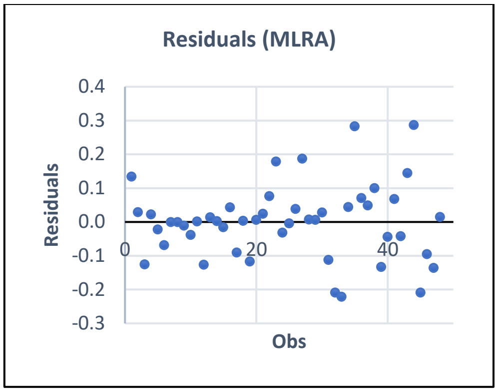

Figure 6. Residuals using MLRA

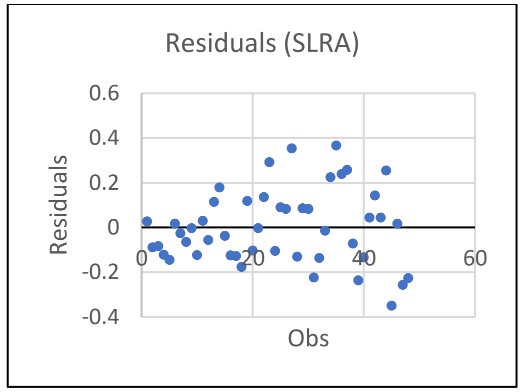

Figure 7. Residuals using SLRA

For *Acacia Mangium* the residuals for both of the models are scattered randomly and there exist no pattern in any of the models. This suggest that the models are a good fit for the data set. All the residuals in all analysis are also centered around the *x*-axis which also indicate that the models are also suitable in explaining the variation in the dependent variable.

Table 3: MAPE and RMSE values for all models for *Acacia Mangium*

Model	**MAPE**	**RMSE**
MLRA	7.7228	0.1084
SLRA	12.9767	0.1633

Table above shows the R-squared, MAPE, and RMSE values for all the models for *Acacia Mangium*. The best model is MLRA because it has the lowest MAPE and RMSE values compared to SLRA model.

3.5 Tree Volume Table

For *Eucalyptus Pellita*, the Simple Nonlinear Regression Model (SNRA) is the best model to forecast the value of tree volume, hence the model will be used to generate the volume table. Figure below shows the tree volume table for *Eucalyptus Pellita* generated using the SNRA model:

Table 4: One-way volume table (Overbark) *Eucalyptus Pellita*

DBH (cm)	Tree Volume (m3)	DBH (cm)	Tree Volume (m3)	DBH (cm)	Tree Volume (m3)	DBH (cm)	Tree Volume (m3)
10.00	0.074	18.00	0.284	26.00	0.628	34.00	1.105
10.50	0.084	18.50	0.302	26.50	0.654	34.50	1.139
11.00	0.093	19.00	0.320	27.00	0.680	35.00	1.174
11.50	0.104	19.50	0.338	27.50	0.707	35.50	1.209
12.00	0.114	20.00	0.358	28.00	0.734	36.00	1.245
12.50	0.126	20.50	0.377	28.50	0.762	36.50	1.281
13.00	0.137	21.00	0.397	29.00	0.791	37.00	1.318
13.50	0.150	21.50	0.418	29.50	0.820	37.50	1.355
14.00	0.163	22.00	0.439	30.00	0.849	38.00	1.393
14.50	0.176	22.50	0.461	30.50	0.880	38.50	1.432
15.00	0.190	23.00	0.483	31.00	0.910	39.00	1.471
15.50	0.204	23.50	0.506	31.50	0.941	39.50	1.510
16.00	0.219	24.00	0.529	32.00	0.973	40.00	1.550
16.50	0.235	24.50	0.553	32.50	1.005		
17.00	0.251	25.00	0.577	33.00	1.038		
17.50	0.267	25.50	0.602	33.50	1.071		

Conversely, the Multiple Linear Regression Model (MLRA) is the best model in estimating the Tree Volume for *Acacia Mangium* species specifically in Kuala Segan, Bintulu, Sarawak. Therefore, table below shows the tree volume for *Acacia Mangium* generated using MLRA model:

Table 5: Two-way volume table (Overbark) *Acacia Mangium*

DBH (cm)	Height (m)										
	10	12	14	16	18	20	22	24	26	28	30
20.00	0.0447	0.1200	0.1953	0.2706	0.3459	0.4212	0.4965	0.5718	0.6471	0.7224	0.7977
22.00	0.1438	0.2191	0.2944	0.3697	0.4450	0.5203	0.5956	0.6709	0.7462	0.8215	0.8968
24.00	0.2428	0.3181	0.3934	0.4687	0.5440	0.6193	0.6946	0.7699	0.8452	0.9205	0.9958
26.00	0.3419	0.4172	0.4925	0.5678	0.6431	0.7184	0.7937	0.8690	0.9443	1.0196	1.0949
28.00	0.4410	0.5163	0.5916	0.6669	0.7422	0.8175	0.8928	0.9681	1.0434	1.1187	1.1940
30.00	0.5401	0.6154	0.6907	0.7660	0.8413	0.9166	0.9919	1.0672	1.1425	1.2178	1.2931
32.00	0.6392	0.7145	0.7898	0.8651	0.9404	1.0157	1.0910	1.1663	1.2416	1.3169	1.3922
34.00	0.7382	0.8135	0.8888	0.9641	1.0394	1.1147	1.1900	1.2653	1.3406	1.4159	1.4912
36.00	0.8373	0.9126	0.9879	1.0632	1.1385	1.2138	1.2891	1.3644	1.4397	1.5150	1.5903
38.00	0.9364	1.0117	1.0870	1.1623	1.2376	1.3129	1.3882	1.4635	1.5388	1.6141	1.6894
40.00	1.0355	1.1108	1.1861	1.2614	1.3367	1.4120	1.4873	1.5626	1.6379	1.7132	1.7885
42.00	1.1346	1.2099	1.2852	1.3605	1.4358	1.5111	1.5864	1.6617	1.7370	1.8123	1.8876
44.00	1.2336	1.3089	1.3842	1.4595	1.5348	1.6101	1.6854	1.7607	1.8360	1.9113	1.9866
46.00	1.3327	1.4080	1.4833	1.5586	1.6339	1.7092	1.7845	1.8598	1.9351	2.0104	2.0857
48.00	1.4318	1.5071	1.5824	1.6577	1.7330	1.8083	1.8836	1.9589	2.0342	2.1095	2.1848
50.00	1.5309	1.6062	1.6815	1.7568	1.8321	1.9074	1.9827	2.0580	2.1333	2.2086	2.2839

4 Conclusion and Recommendation

4.1 Conclusion

The purpose of this study is to develop the local tree volume tables for *Eucalyptus Pellita* and *Acacia Mangium* by fitting various types of linear and nonlinear model via regression analysis. Based on the analysis conveyed, it can be concluded that multiple regression analysis is better at describing the variation of the tree volume for both of these species. The findings also suggest that the forecast accuracy in tree volume estimation is improved when both independent variables, DBH and Total Length, being paired instead of being used as the only parameter. Further exploration into development of tree volume for other species could be useful to find better and more robust model that can be used for tree volume estimation in a global scale. The amount this could improve the efficacy of forestry management in achieving sustainable forest for the use of future generations is worth exploring.

4.2 Recommendation

The merchantable volume of Eucalyptus Pellita and Acacia Mangium stands have been adequately described by the equations generated in this study, and the equation can be applied with relative ease.

References

[1] Avery, T. E., & Burkhart, H. E. (1983). Forest measurements (No. Ed. 3). McGraw-Hill Book Company.

[2] Clark, J. F. 1902. Volume tables and the bases on which they may be built. Forestry 1: 6–11.

[3] Demeritt, D. B., & McIntyre, A. C. (1932). A SIMPLE METHOD OF CONSTRUCTING TREE VOLUME TABLES'. Journal of Agricultural Research, 44(6), 529-539.

[4] HAJAR, Z., SHUKRI, W. W. M., Samsudin, M., RAZALI, W., & Ismail, H. (2010). Development of local volume table for second growth forests using standing tree measurements. THE MALAYSIAN FORESTER, 73(2), 163-170.

[5] Honer, T. G. (1965). A new total cubic foot volume function. The Forestry Chronicle, 41(4), 476-493.

[6] Husch, Bertram, Beers, Thomas W. and Kershaw, John A. 2003. Forest Mensuration. John Willey and Sons Inc. Hoboken New Jersey, Canada

[7] Khanna, L. S. and Chaturvedi, A. N. 1982. Forest Mensuration. International Book Distributer, Deharadun, India.

[8] Krisnawati, H. (2016). A Compatible estimation model of stem volume and taper for Acacia Mangium Willd. plantations. Indonesian Journal of Forestry Research, 3(1), 49-64.

[9] Larsen, D. R. (2017). Simple taper: Taper equations for the field forester. In In: Kabrick, John M.; Dey, Daniel C.; Knapp, Benjamin O.; Larsen, David R.; Shifley, Stephen R.; Stelzer, Henry E., eds. Proceedings of the 20th Central Hardwood Forest Conference; 2016 March 28-April 1; Columbia, MO. General Technical Report NRS-P-167. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station: 265-278. (pp. 265-278).

[10] Liu, J., Feng, Z., Mannan, A., Khan, T. U., & Cheng, Z. (2019). Comparing Non-Destructive Methods to Estimate Volume of Three Tree Taxa in Beijing, China. Forests, 10(2), 92.

[11] Mayamanikandan, T., Reddy, R. S., & Jha, C. S. (2019, October). Non-destructive tree volume estimation using terrestrial LiDAR data in teak dominated central indian forests. In 2019 IEEE Recent Advances in Geoscience and Remote Sensing: Technologies, Standards and Applications (TENGARSS) (pp. 100-103). IEEE.

[12] Shin, M. Y., Yun, J. W., & Cha, D. S. (1996). Local correction of tree volume equation for Larix leptolepis by ratio-of-means estimator. Journal of Korean Society of Forest Science, 85(1), 56-65.

[13] Shrestha, H. L., kafle, M. R., Khanal, K., & Mandal, R. A. (2018). Developing local volume tables for three important tree species in Nawalparasi and Kapilvastu districts. Banko janakari, 84-91.

[14] Sykes, A. O. (1993). An introduction to regression analysis. Coase-Sandor Institute for Law & Economics Working Paper No. 20. University of Chicago Law School

[15] Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. Procedia-Social and Behavioral Sciences, 106, 234-240.