

Aboveground biomass and carbon stocks modelling using non-linear regression model

Nurul Ain Mohd Zaki¹, Zulkiflee Abd Latif¹, Mohd Nazip Suratman² and Mohd Zainee Zainal³

¹Applied Remote Sensing & Geospatial Research Group, Centre for Studies Surveying Science & Geomatics, Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Email: nurulainzaki@gmail.com; zabdlatif@gmail.com

²Faculty of Applied Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

³Centre for Studies Surveying Science & Geomatics, Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA, Arau, 02600, Perlis, Malaysia

Abstract. Aboveground biomass (AGB) is an important source of uncertainty in the carbon estimation for the tropical forest due to the variation biodiversity of species and the complex structure of tropical rain forest. Nevertheless, the tropical rainforest holds the most extensive forest in the world with the vast diversity of tree with layered canopies. With the usage of optical sensor integrate with empirical models is a common way to assess the AGB. Using the regression, the linkage between remote sensing and a biophysical parameter of the forest may be made. Therefore, this paper exemplifies the accuracy of non-linear regression equation of quadratic function to estimate the AGB and carbon stocks for the tropical lowland *Dipterocarp* forest of Ayer Hitam forest reserve, Selangor. The main aim of this investigation is to obtain the relationship between biophysical parameter field plots with the remotely-sensed data using non-linear regression model. The result showed that there is a good relationship between crown projection area (CPA) and carbon stocks (CS) with Pearson Correlation ($p < 0.01$), the coefficient of correlation (r) is 0.671. The study concluded that the integration of Worldview-3 imagery with the canopy height model (CHM) raster based LiDAR were useful in order to quantify the AGB and carbon stocks for a larger sample area of the lowland *Dipterocarp* forest.

1. Introduction

Tropical forest plays a significant role in the climate change as a main carbon storehouse for the whole world. In the carbon sequestering photosynthesis process, tree and soils continuing to provide the oxygen for human being respiration cycle. Recent studies by [1], shows that there has been rapidly increasing in forest loss in most other parts of the tropics over the last decade, with South-East Asia displaying one of the highest rates of deforestation. It is confirmed in a recent forest cover research that shows the main cause of forest loss in South-East Asia is the conversion of the forest to cash crop plantation [2]. In the past few centuries, anthropogenic activities and natural consequences have increased the concentration of carbon dioxide including greenhouse gases (GHG) to the atmosphere [3]. The anthropogenic activities



such as degraded human-set fires, destructive timber logged, forest clear for agricultural and development purposes would affect the carbon cycle may lead to global warming affect [1]. There are several methods that can be used to measure aboveground biomass. Another method that usually used to estimate the aboveground biomass is using the method conversion of volume data to develop allometric equation. Tree stand allometric equations are developed by calculating the relationship between field measurement of the tree parameter such as the diameter of the trunk, tree height, and diameter at breast height (DBH), tree species, age, crown density, and also bioclimatic variable. Undeniable, field-based forest inventory remains vital in the carbon stocks estimation of forest structure [2] , [3]. Nevertheless, this can be improved by the use of remotely sensed data where it can reduce the massive field-based sampling if it comes to a huge area of forested land. Therefore, the objective of this research includes (i) investigating the relationship between carbon stocks and tree parameter (ii) developing a model of carbon stocks estimation using non-linear regression model develop from crown projection area (CPA) of the emergent and canopy layer and finally (iii) predicting the carbon estimation of the study area.

2. Material and methods

2.1. Study area

The research was conducted in a forest managed by University Putra Malaysia, Ayer Hitam Forest Reserved, Selangor State, Malaysia. The forest lies between Latitude 3°00'24.19"N, Longitude 101°38'25.24"E, the location of the study area is illustrated in Figure 1. The type of this forest is lowland *Dipterocarp* forest. It comprises of various species dominated by the family from *Dipterocarpaceae*, followed by family *Euphorbiaceae* as a second major family species. The average rainfall occurring on the average of 2178 mm annually while the humidity reaches 74% and the average temperature is 27.7 ° C minimum and 22.9 ° C maximum [5].

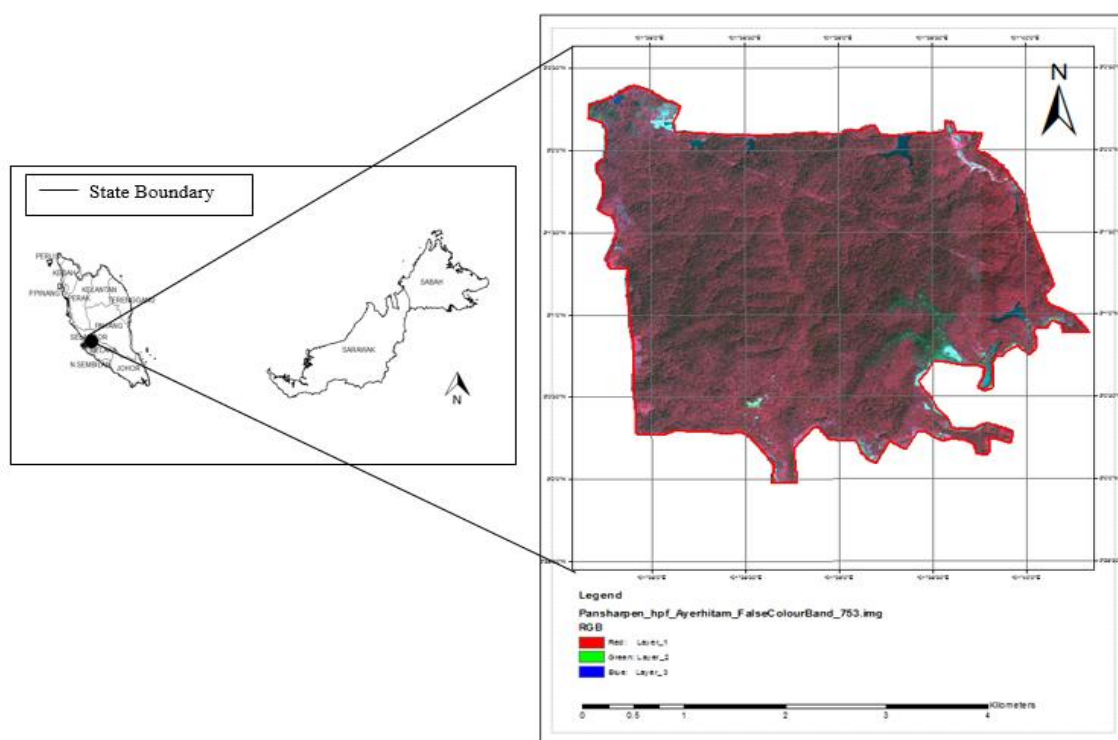


Figure 1. Location of Ayer Hitam Forest Reserve is shown in false colour band combination of 542 of WordView3 (WV-3).

2.2. Satellite image, LiDAR and supporting data

The light detection and ranging (LiDAR) data had been used in this dissertation were acquired in August 2013. The sensor used in this mission is by using LITEMAPPER-5600 that consist of RIEGL LMS-Q560 Laser Scanner. This equipment allows an airborne scanning laser to penetrates the distance between the aircraft and the ground and produce the precise digital surface models. The LiDAR system was mounted on board of aircraft, fixed with the real-time kinematic (RTK) survey that will provide the differential position of the aircraft and equip with the Inertial Measurement Unit (IMU). The altitude of the flying height was 1000 m above the ground level using the flying speed 90 knots. With the capacity of laser swath width of approximately, 1155m was a measure for every flight line, the pulse repetition frequency that had bet set during the LiDAR observation is 150 kHz (150, 000 pulses per second) in order to generate a number of return. The canopy height model (CHM) was created using LASTools plug-in software ArcGIS. Figure 2 show the CHM of the tree height in a different view.

Super-spectral high resolution of WV-3 (8 bands) 25 km² with the projection of UTM47 N and WGS84 datum, have been used in this study. The data had been obtained on 9th December 2014 that consists of a 31cm panchromatic band, 4 bands Very Near Infrared (VNIR) colours (blue, green, red and near infrared NIR-1) and 4 added VNIR colours which are coastal, yellow, red edge and NIR2. The spatial resolution was 0.30 m (panchromatic) and 1.2 m (multispectral). The classification was performed on a pan-sharpen WV-3 image using hyperspherical colour sharpening (HCS) techniques for better visualisation.

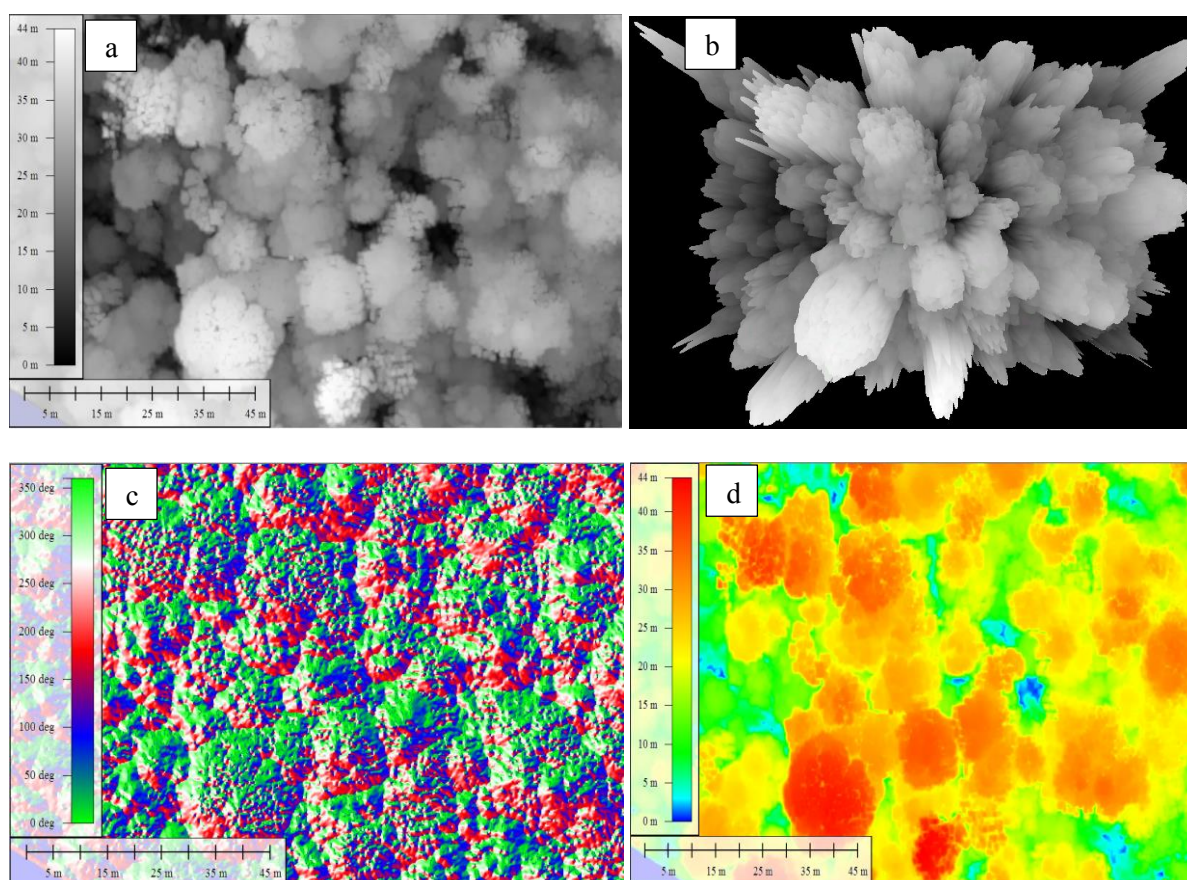


Figure 2. CHM derived from the digital surface model and the digital terrain model. **a).** The top view of the CHM of the study area. **b).** The 3D view from the top of the CHM. **c).** The slope direction shade using CHM and finally **d).** The colour ramp shade of the CHM.

2.3. Field sampling plot

Field sampling plot were conducted on May 2015 at 2-hectare rectangular plots that consist of 50 number of subplots (20 x 20 m) using stratification number of sampling. In each plot, every tree that has more and equal to 10 cm DBH were inventoried and recorded. The parameter that had been collected includes tree id, tree location, species, stem diameter at breast height, crown base height, crown diameter, tree height and leaf area index of the subplot area. The field plot and tree coordinates were tied by the traversing using total station from a known Global Positioning System (GPS) reference stations. The reference stations were observed using GPS using Topcon GTS with static L1/L2 and L5 bands.

$$AGB_{est} = \exp [[-1.803 - 0.976 E + 0.976 \ln \rho + 2.673 \ln (D) - 0.0299 [\ln(D)^2]]] \quad (1)$$

Where AGB = Aboveground biomass (kg / tree), ρ = wood density, E = Bioclimatic variables and D = DBH. The AGB and carbon stocks were calculated using allometric equation develop by Equation (1), [7] which incorporated the usage of wood density, height, DBH and bioclimatic variables as predictors. Despite height and DBH, wood density is one of the important predictors in the AGB calculation [8]. In addition, the AGB (kg) were converted to Mg ha⁻¹ and converted to carbon stocks by applying the conversion factor 0.47 which represents 47% of the dry biomass assume to be carbon for all part of the tree that had been recommending by Intergovernmental Panel on Climate Change (IPCC) [9].

2.4. Non-linear regression modelling

The previous researcher had been used regression modelling [4], [5], and machine learning approaches [6], [7] in order to quantify the complexity of the AGB and carbon stocks for the tropical rain forest. The quadratic model of non-linear regression model were used to develop the model to estimate AGB using this mathematical function (2) and (3):

$$y = \alpha x^2 + \beta x + c \quad (2)$$

$$Carbon\ Stocks = \alpha x^2 + \beta x + c \quad (3)$$

2.5. Validation of the statistical analysis

The statistical analysis was performed by calculating the statistical parameter such as residual R^2 , Root Mean Square Error (RMSE) of the absolute accuracy had been done. (Equation 4). The data was analysed in the scatter plots and box plots to identify the outliers and had been removed in order to get the robust model for regression. Out of 911 of tress, 183 were used to develop the model, while other 30% of the field sampling were used as a validation.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (4)$$

Where n is the number of samples, y_i is the calculated value of carbon, \hat{y}_i is the predicted carbon by the model of the response variables.

3. Results and Discussion

3.1. Descriptive statistics

The selection of the independent variables of the carbon stocks for tropical lowland was made through broad review of the literature on the subject matter[8]. Overall, the dominant tree species occurred in the area was *Dipterocarpaceae* spp, *Euphorbiaceaea* spp, *Burseraceae* spp, and *Sapotaceae* spp. The

most dominant species was *Endospermum diadenum* with 24.69% occurrence, *Hopea sulcata*, *Dipterocarpus verrucosus* and *Shorea macroptera* (Figure 3).

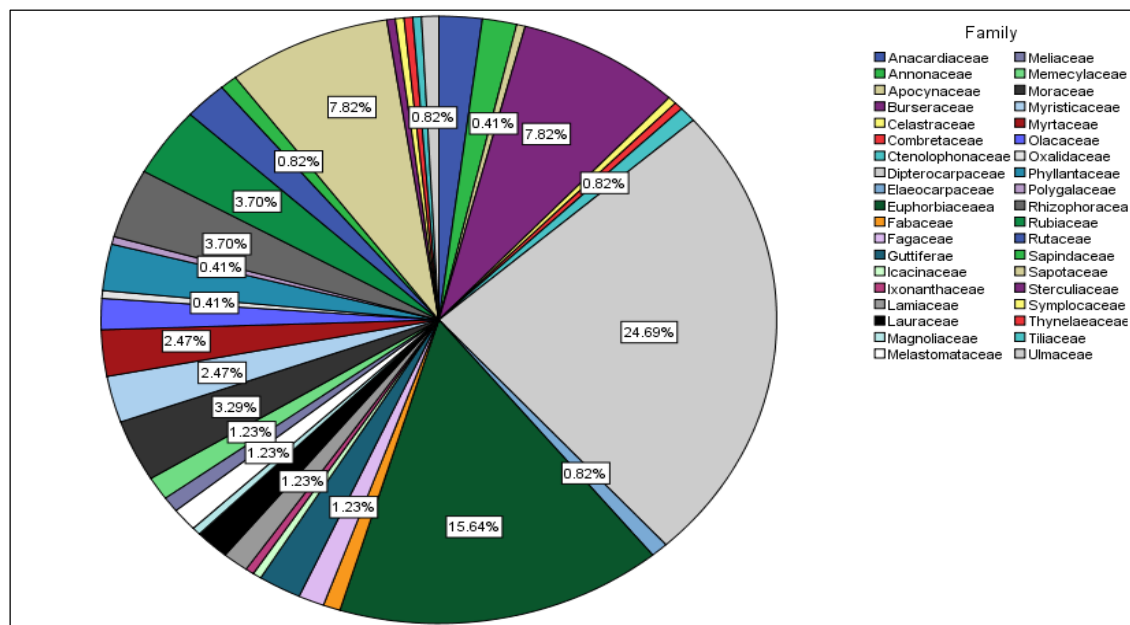


Figure 3. The occurrence of the different species of the study area.

Table 1. Summary statistics of variables collected from field sampling, LiDAR and satellite segmentation output and used for model development ($n = 183$) and validation dataset ($n = 62$), respectively.

| Variable and unit | Model-building data set | | | | | Validation data set | | | | |
|-------------------------------------|-------------------------|------------------|------------------|--------|-----------------|---------------------|------------------|------------------|---------|-----------------|
| | No. of tree | Min ^a | Max ^b | Mean | SD ^c | No. of tree | Min ^a | Max ^b | Mean | SD ^c |
| Height from LiDAR | 183 | 10.850 | 33.299 | 20.574 | 5.108 | 62 | 12.331 | 37.822 | 21.054 | 5.297 |
| Crown Projection Area | 183 | 7.514 | 117.26 | 31.094 | 19.517 | 62 | 10.034 | 214.28 | 29.456 | 28.789 |
| Height from field | 183 | 40.421 | 72.060 | 55.506 | 8.640 | 62 | 12.00 | 37 | 20.497 | 5.354 |
| DBH | 183 | 10 | 72.50 | 28.301 | 13.566 | 62 | 10.200 | 113.00 | 29.071 | 16.258 |
| Aboveground Biomass | 183 | 32 | 5325 | 697.54 | 901.42 | 62 | 46.00 | 17167 | 948.677 | 2276.4 |
| Carbon Stocks (kg) | 183 | 15 | 2503 | 327.81 | 423.71 | 62 | 22.00 | 8068.0 | 21.054 | 5.296 |
| Carbon Stocks(Mg ha ⁻¹) | 183 | 0.0075 | 1.2514 | 0.1639 | 0.2118 | 62 | 0.0108 | 4.03 | 0.2229 | 0.5349 |

(Min^a = Minimum, Max^b = Maximum and SD^c = Standard Deviation)

3.2. Image segmentation for crown projection area (CPA) delineation

The image multispectral WV-3 and CHM derived from LiDAR were segmented using Multiresolution segmentation algorithm using eCognition software was used in order to obtain the appropriate segments. Then, the watershed transformation was done and split the tree cluster[3]. Morphology was applied to the segmented objects and give an appropriate figure to the trees. Finally, the undesired objects were removed and output the crown projection area of the study plots. The accuracy assessment used were calculated and tabulated using equation develop by [9].

3.3. Correlation analysis – Pearson correlations

In order to see the relationship between the variables which is diameter at the breast height (DBH), height derived from LiDAR (H_LDR), crown projection area (CPA), height derived from field (H_Field) and carbon stocks (CS), Pearson correlation had been done to measure the strength of correlation between independent and dependent variables of the carbon estimations.

Table 2. Pearson correlation of the variables

| Variable | | DBH | H_LDR | CPA | CS |
|----------|------|---------|---------|---------|---------|
| DBH | PC | 1 | 0.763** | 0.718** | 0.909** |
| | Sig. | | 0.000 | 0.000 | 0.000 |
| | N | 183 | 183 | 183 | 183 |
| H_LDR | PC | 0.763** | 1 | 0.549** | 0.709** |
| | Sig. | 0.000 | | 0.000 | 0.000 |
| | N | 183 | 183 | 183 | 183 |
| CPA | PC | 0.718** | 0.549** | 1 | 0.671 |
| | Sig. | 0.000 | 0.000 | | 0.000 |
| | N | 183 | 183 | 183 | 183 |
| H_Field | PC | 0.763** | 0.995** | 0.556** | 1 |
| | Sig. | 0.000 | 0.000 | 0.000 | 0.000 |
| | N | 183 | 183 | 183 | 183 |

**Correlation is significant at the 0.01 level (2-tailed)

(PC= Pearson Correlation, N = number of samples, H_LDR = tree height from LiDAR, H_Field = tree height from field, CPA = crown projection area, .Sig = significant level)

Based on Table 2, tabulated that the highest correlation strength exists between height derived from field sampling (H_Field) and height derived from LiDAR (H_LDR), with 0.995 with ($P < 0.01$). There is also a high positive correlation between carbon stocks (CS) and crown projection area (CPA) where the correlation between this two variables (0.671) with ($P < 0.01$) significant level. Next, the non-linear regression model was develop using CPA as independent variables and carbon stocks as dependent variables as the data yield in this result shown the strong evidence to develop a model.

3.4. Modelling AGB using non-linear regression models

The non-linear regression model was carried out using CPA and output the coefficient of determination of 0.67 which is ($R^2 = 0.45$). The correlation coefficient of the CPA and carbon was high but not strongly correlated which is (0.671) indicated there is a significant correlation between these two variables. According to ANOVA, the model is significant at 99% significance level. The details summarization of the regression of the developed model is tabulated in Table 3, 4 and 5.

Table 3. Model Summary^b

| R | R ² | Adj. R ² | Std. Error of the Estimate | Change Statistics | | | | | Durbin- Watson |
|--------------------|----------------|------------------------|-------------------------------------|--------------------------|-------------|---------|-----|-----------------|-------------------|
| | | | | R ² Change | F Change | df 1 | df2 | Sig.F Change | |
| 0.673 ^a | 0.453 | 0.447 | 0.14751 | 0.453 | 74.589 | 2 | 180 | 0.000 | 1.866 |

a. Predictors: (Constant), CPA, CPA²

b. Dependent Variable: Carbon Stocks (Mg ha⁻¹)

Table 4. ANOVA^a

| Model | Sum of Squares | df | Mean Square | F | Sig |
|-------|-------------------|-------|-------------|-------|--------------------|
| 1 | Regression | 3.701 | 2 | 1.851 | 0.000 ^b |
| | Residual | 4.466 | 180 | 0.025 | |
| | Total | 8.167 | 182 | | |

a. Dependent Variable: Carbon Stocks (Mg ha⁻¹)

b. Predictors: (Constant), CPA, CPA²

Table 5. Coefficients^a

| | Unstandardized Coefficients | | Standardized Coefficients | t | Sig |
|------------------|-----------------------------|------------|------------------------------|--------|-------|
| | B | Std. Error | Beta | | |
| Intercept | 1.333 | 0.354 | | 3.763 | 0.000 |
| CPA | -2.166 | 0.487 | -2.416 | -4.443 | 0.000 |
| CPA ² | 0.919 | 0.165 | 3.030 | 5.574 | 0.000 |

a. Dependent Variable: Carbon Stocks (Mg ha⁻¹)

Based on the summary obtained from the regression analysis (Table 4), a model was developed to estimate the carbon stocks of the study area (Equation 5 and 6) derived from the non-linear regression model using quadratic function (Equation 3).

$$\ln \text{Carbon Stocks} = \beta_1 \ln x^2 + \beta_2 \ln x + \beta_0 \quad (5)$$

$$\ln CS = 0.919 \ln CPA^2 - 2.166 \ln CPA + 1.333 \quad (6)$$

Where *CS* is the carbon stocks in Mg Ha-1, *CPA* is the crown projection area, and β_1 , β_2 and β_0 is the constant derived from the regression model. The variables *CPA* and carbon had been transforming and the natural logarithmic had been applying to the formula.

3.5. Validation of Non-Linear regression model

The Pearson product model of the non-linear regression using quadratic function were validated using a random selection of 30% of the dataset which is 62 tress from various species and tree height from out of the total of 245 total stands. The resulting coefficients of determination between carbon predicted from a non-linear model and the carbon observed from field dataset was ($R^2=0.7463$) (Figure 4).

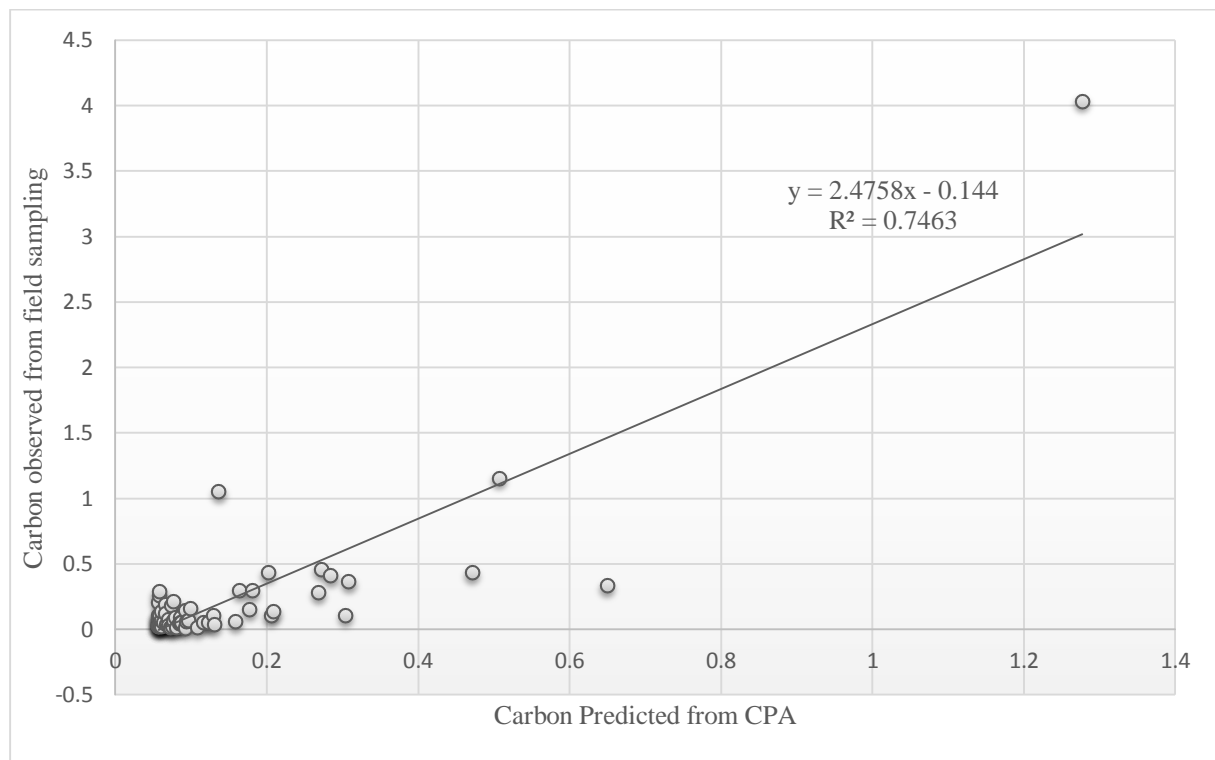


Figure 4. Scatter plots of non-linear regression model validation

The validation of the carbon predicted and carbon observed from field sampling were tabulated using scatter plots, and the coefficient of determination shows that the fit explain 74.63% of the total variation in the data about the average.

4. Conclusion

The relationship between carbon stocks with CPA obtained in this study showed a good relationship and impressive validation results. However the coefficient of determination R^2 of the regression model was not high ($R^2 = 0.453$), but there is an evidence that the correlation between CPA and carbon stocks was high which is 0.671. Due to the irregularities of the canopy of the tropical rainforest, the result can be considered as accepted based on comparison with the output of other research that had been done at other tropical rainforests for an example studies by [10] and [11].

References

- [1] FAO, *Global Forest Resources Assessment 2015*, ISBN 978-9. 2015.
- [2] Z. Latif and H. Zaqwan, "Deforestation and carbon loss estimation at tropical forest using multispectral remote sensing: Case study of Besul Tambahan Permanent Forest Reserve," *IEEE proceeding*, 2015.
- [3] N. A. Mohd Zaki, Z. Abdul Latif, M. Z. Zainal, and K. Zainuddin, "Individual Tree Crown (ITC) Delineation Using Watershed Transformation Algorithm For Tropical Lowland Dipterocarp," *Proc. IEEE 2015*, vol. 2, pp. 237–242, 2015.
- [4] S. Mebrat Tikabu, "Estimating Carbon Stock Using Very High Resolution Imagery and Terrestrial Laser Scanning In Tropical Rain Forest of Royal Belum, Malaysia," University of Twente, 2015.
- [5] H. Omar, M. H. Ismail, H. Shafri, and K. Hamzah, "Estimating Biomass in Logged Tropical Forest Using L-Band SAR (PALSAR) Data and GIS Estimating Biomass in Logged Tropical Forest Using L-Band," *Sains Malaysiana*, no. October, 2015.

- [6] Y. Zhu, K. Liu, L. Liu, S. Wang, and H. Liu, "Retrieval of Mangrove Aboveground Biomass at the Individual Species Level with WorldView-2 Images," *Remote Sens.*, vol. 7, pp. 12192–12214, 2015.
- [7] M. E. J. Cutler, D. S. Boyd, G. M. Foody, and a. Vetrivel, "Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions," *ISPRS J. Photogramm. Remote Sens.*, vol. 70, pp. 66–77, 2012.
- [8] Z. A. Latif, I. Zamri, and H. Omar, "Determination of tree species using Worldview-2 data," *2012 IEEE 8th Int. Colloq. Signal Process. its Appl.*, pp. 383–387, 2012.
- [9] N. Clinton, A. Holt, J. Scarborough, L. Yan, and P. Gong, "Accuracy Assessment Measures for Object-based Image Segmentation Goodness," *Photogramm. Eng. Remote Sising*, vol. 76, no. 3, pp. 289–299, 2010.
- [10] P. R. Mbaabu, Y. A. Hussin, M. Weir, and H. Gilani, "Quantification of carbon stock to understand two different forest management regimes in Kayar Khola watershed, Chitwan, Nepal," *J. Indian Soc. Remote Sens.*, no. DoF 2012, 2014.
- [11] Y. A. Hussin, H. Gilani, L. Van Leeuwen, M. S. R. Murthy, R. Shah, S. Baral, N. E. Tsendbazar, S. Shrestha, S. K. Shah, and F. M. Qamer, "Evaluation of object-based image analysis techniques on very high-resolution satellite image for biomass estimation in a watershed of hilly forest of Nepal," *Appl. Geomatics*, vol. 6, no. 1, pp. 59–68, 2014.

Acknowledgments

The authors would like to express their deepest gratitude to the Ministry of Higher Education (MOHE), Malaysia for financing the under Research Management Centre (RMC), (Ref.: RAGS/1/2014/STWN10/UITM//1) Universiti Teknologi MARA (UiTM), and also Faculty of Forestry, Universiti Putra Malaysia (UPM) for granting access to the study area. Thanks to Alhaffiz bin Abd Radzak (RS GIS Consultancy Sdn Bhd) for helping us in term of LiDAR pre-processing. Special thanks to a member of Ayer Hitam Rainforest Research Expedition 2015 (AHRRE 2015) who involve in data collection of forest inventories.