PREDICTING ABOVE-GROUND BIOMASS AND CARBON STOCKS BY USING GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

Nurul Ain Mohd Zaki[1], Zulkiflee Abd Latif[1*], Mohd Zainee Zainal and Nurul Nadhirah Hemmy Shah.
[1]Applied Remote Sensing & Geospatial Research Group, Centre for Studies Surveying Science & Geomatics, Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[2]Centre for Surveying Science and Geomatics Studies, Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA, Arau, 02600, Perlis, Malaysia

Email: zabdlatif@gmail.com

KEY WORDS: Above-ground biomass, carbon stock, geographically weighted regression, ordinary least square, LiDAR.

ABSTRACT: Carbon dioxide (CO_2) believed to be one of the major greenhouse gases which impact the climate change. Tropical forest is well known as the world's most complex trees which embrace a large stock of carbon in the global carbon cycle and contributes an enormous amount of above-ground and below-ground biomass. Therefore, this study aims to estimate the carbon stocks prediction for tropical rainforest using geographically weighted regression (GWR). The predictor's variables used in this studies is the tree height derived from the canopy height model (CHM) of airborne light detection and ranging (LiDAR), HT_L and crown projection area (CPA) derived from a fusion of LiDAR and WorldView-3 (WV-3) images. This study contributes to the potential of linking the above-ground biomass and carbon stocks estimation and remotely-sensed data by using geographically weighted regression (GWR) and ordinary least square (OLS) approach. The GWR models presented the substantial improvement in the above-ground biomass and carbon stocks estimation and beneficial for future development and strategic planning of the forest resources.

1. INTRODUCTION

1.1 Above-ground biomass and carbon stock of tropical forest.

Tropical forest is heterogeneous in nature with tremendously rich in biodiversity of species of living plant and animal which provide an important basis for the ecosystem circulation. This lead to the developing numerous number of vegetation indices and the radiometric characteristic in assessing the vegetation using remotely sensed data. The use of remotely sensed data using optical sensors fusion with the empirical models is commonly used by the previous studies in order to assess the AGB (Basuki et al., 2012). Numerous studies have demonstrated the effectiveness of using the model based on regression can give an accurate result for the estimation of AGB and carbon stocks at many spatial scales (Hamdan et al., 2014; Muukkonen & Heiskanen, 2005).

However, recent studies show that there has been an increase rate of carbon over 100 times faster than anything ever observed in the past eight hundred thousand years (NOAA, 2016). Moreover, a huge scale study by Hansen et al., (2013) demonstrated an alarming rate of the increasing of forest clearing in most tropical areas over the last decade, with South-East Asia being ranked as one of the highest country of deforestation activity rate. Due to the economic influence, a lot of forests have been cleared for timber logging and broad conversion of forest to oil palm plantation. Subsequently, numerous habitats of living plants and animals are decreasing and this impacts the biodiversity of species at the forest floor (Mohd Zaki & Abd Latif, 2016).

Remote sensing imageries are useful in the assessment the above-ground biomass and also important for carbon monitoring (Baccini & Asner, 2013; Gibbs et al., 2007; UNFCC, 2015). Over the past three decades, the researcher has become increasingly interested in exploiting remote sensing methods for above-ground biomass (AGB) estimation for forest biome (Lu et al., 2014). A considerable amount of literature has been published for investigating the relationship between remote sensing vegetation indices and spectral analysis with the AGB estimation (Basuki et al., 2012; Eckert, 2012; Propastin, 2012). Nevertheless, accurate estimation of AGB, carbon stocks and biomass variation present a challenging aspect to derive accurately the remote sensing based for tropical forest. Therefore, the objective of this paper includes (i) applying the model above-ground biomass and carbon stock prediction using GWR and OLS and (ii) produce the carbon stocks mapping using GWR and OLS.

1.2 Spatial relationship for the tree variables using GWR and OLS.

Spatial statistics in ArcGIS was used to model the spatial pattern, distribution and relationship using technique space and area, length or spatial relationship (Scott & Janikas, 2010). Several steps need to be done before generate the GWR is the exploratory regression and ordinary least square (OLS). These steps can provide the trustworthy results of any estimation application but it also depends on the situation of data or the thing to be model. The first steps before OLS analysis generation are by using exploratory regression to identify the find the potential parameters if there have a ton of variables that are trying to be the model. Moreover, the analysis provides by this regression including the criteria such as Adjusted R^2, coefficient p-values, Variance Inflation Factor (VIF), Jarque-Bera p-values and spatial autocorrelation p-values (Esri, 2013). In order to predict or understand the incident or something is happening, the dependent variable (y) must be set up and known. Then, followed by the explanatory variable or independent variable (x) which is should be two or more variables so that the OLS can analyse the relationship between the selected variables (Esri, 2013). The system computed OLS regression tool for those parameters and formed a coefficient value that shows the strength of the relationship between dependent and independent variables.

$$Y = B_0 + B_1 X_1 + B_2 X_2 + ...B_n X_n + \in \qquad (1)$$

Where Y is the dependent variables, B_n is the coefficients, X_n is the explanatory variables and ϵ is the random error term or residuals. In contrary, GWR is a local model and the output result was generated to fit the model equation for each of the dataset (Esri, 2013). Therefore, every of the spatial feature has their own model equation. GWR was a tool that improved and explored the fit model equation by OLS analysis. It is a platform for visualization on something that we are predicted. Besides that, the model development is the key to the analysis which is can define the neighbourhood. The output including coefficient estimated, local R^2 and standard residual instead of predicted value. However, GWR is not consistent to differentiate whether the data generating is stationary or no stationary. If there has multicollinearity data in the estimated coefficient, it may have biased in GWR result.

2. STUDY AREA AND DATASET

2.1 Geographical Background

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 2. The type of this forest is lowland Dipterocarp forest. It comprises of various species dominated by the family from Dipterocarpaceae, followed by family Euphorbiaceaea as a second major family species. The average rainfall occurring on average of 2178 mm annually while the humidity reach 74% and the average temperature is 27.7 º C minimum and 22.9 º C maximum (Ibrahim, 1999). Topographically the terrain slope is characterized as undulation up to 34º and the altitude that comprise in this lowland forest varies from 15m to 233 height (Shida et al., 2014).

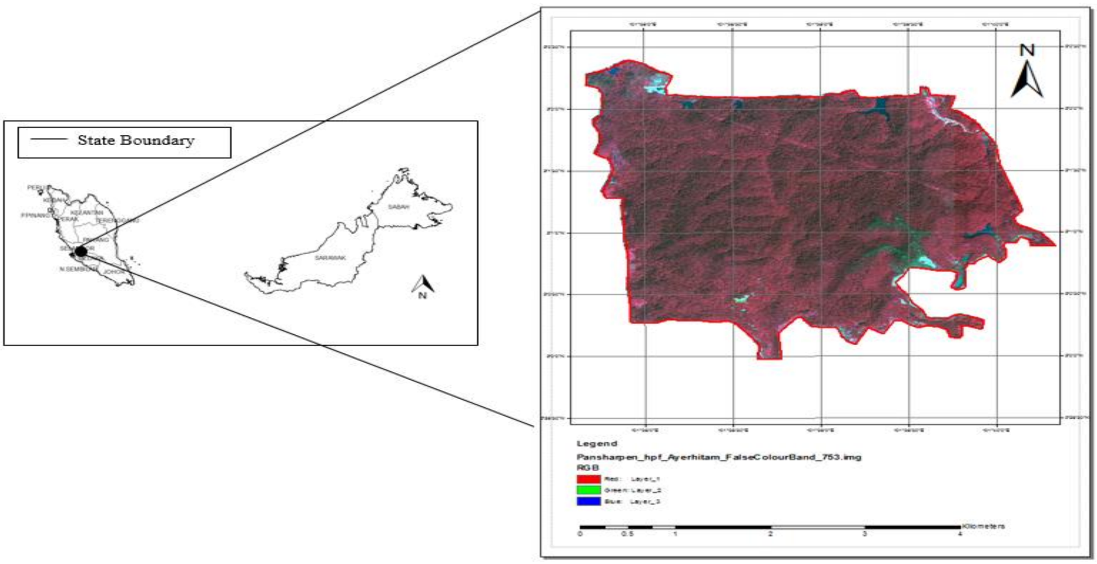

Figure 2. Location of the study area

2.2 LiDAR Data Acquisition

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 consists of RIEGL LMS-Q560 Laser Scanner. This equipment allows an airborne scanning laser to penetrate 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) for the orientation of the aircraft which is a roll, pitch, and bearing. The differential correction signal will give the exact location from the base station at known coordinate. 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 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 more number of return. The laser scans at 160 lines per seconds and the laser scan angle fixed minimum at 46º and maximum at 60º while the image scan angle at 45º. The LiDAR data had been classified into different classes according to the ASPRS and ICSM standard specification in order to filter the ground and vegetation (ASPRS, 2013; ICSM, 2010). Using the Terra Scan, in Bentley Microstation software, the macro tools were used to generate the classification. The point clouds contains the (X, Y, Z) of the coordinates of the laser point, a number of return, intensity, scan direction, the edge of flight line, user data, point sources ID, GPS time, classification, scale factor, offset to point data and the scan flight direction.

A very high resolution satellite imagery used in this study WorldView-3 is provided by commercial company main in Colorado, United States which is known as Digital Globe. WorldView-3 is the first multi-payload, super spectral and high resolution commercial satellite which had been launch in 2014 (DigitalGlobe, 2014). It has an average revisit time of less than one day with capabilities of collecting about 680,000 km^2 per day (DigitalGlobe, 2014). This imagery comprises of 8-bands, which is 4 standard VNIR colours: blue, green, red, near-IR2, 4 added VNIR colours: coastal, yellow, red edge and near-IR2, 8 SWIR bands: penetrates haze, fog, smog dust and smoke and 12 CAVIS bands, but in this research only use 8 bands multispectral imagery for the purposes of studies.

2.3 Ground Data Collection

The above-ground biomass and carbon stocks was estimated using model allometric equation developed, Equation 2 model. This equation was developed using a regression analysis of 245 of trees with a diameter ranging from 10 to 113 cm and consisting of 112 species. The equation to estimates the above-ground C stocks was:

$$ln\ \hat{C} = -4.092 + 0.898\ ln\ HT_L + 2.073\ ln\ DBH - 0.058 ln\ CPA \quad (2)$$

Where C = carbon stocks (kg / tree), HT_L = Height derived from LiDAR, CPA = Crown projection area and DBH = diameter at breast height. The adjusted R^2 of the model is 0.951. The carbon stocks conversion that had been used is 0.47 which had been suggest by the International Panel on Climate Change (IPCC) (IPCC, 2006). The above-ground C stocks was obtained by transforming *ln* C stocks values. The descriptive statistic of the above-ground C stocks of the individuals trees for training and validation data are presented in Table 1.

Table 1. Descriptive statistics of the above-ground C stocks (kg tree^{-1}) for the training (166 of trees) and the validation (79 of trees) data.

C stocks (CS) in kg	Training samples (166 of tree)	Validation (79 of trees)
Minimum	15	26
Mean	359.090	354.734
Maximum	8068	2503
Standard deviation	702.011	458.74

2.4 Object based image segmentation (OBIA) tree crown delineation

Starting with the convolution filter, the processed is to assure the image has a smoothly view for interpreting the shape of the tree crown. After that, the WV-3 image had been segmented using the scale 50, shape 0.8 and compactness 0.5. Accordingly, to the result, these three factors had a relationship toward another that will affect how the tree crown could be delineated. But actually, it was undertaken several times at different level in order to get the suitable crown shape for this research. Next, the shadow should be masked out from the image to remain only tree crown for carbon stock estimation. It was a processed defining the suitable range of image brightness which is the shadow has been masked out at the range between 46 until 182-pixel value while the value more than 182 considered as the crown shape. Then, the watershed transformation is a processed to split or separate the image object from an overlapping tree crown into individual tree crown (Karna et al., 2013; Mohd Zaki et al., 2015). It was applied in this study to improve the image segmentation. According to (Belaid & Mourou, 2009), the watershed transformation is suitable and very useful to use for image segmentation because this processed always provides closed contours and requires low computation times to compare with another segmentation method. Last but not least, the tree crown can be defined much better after morphology processed which is refining and reshaping the boundaries of the segmented result. The tree crown look clearer and beautiful as nearly similar as in the image. Then, the tree was classified either it was a short or tall tree by using CHM data and then both of crown layer will be export to the shapefile (.shp) format for carbon stock modelling and mapping in ArcMap 10.3.

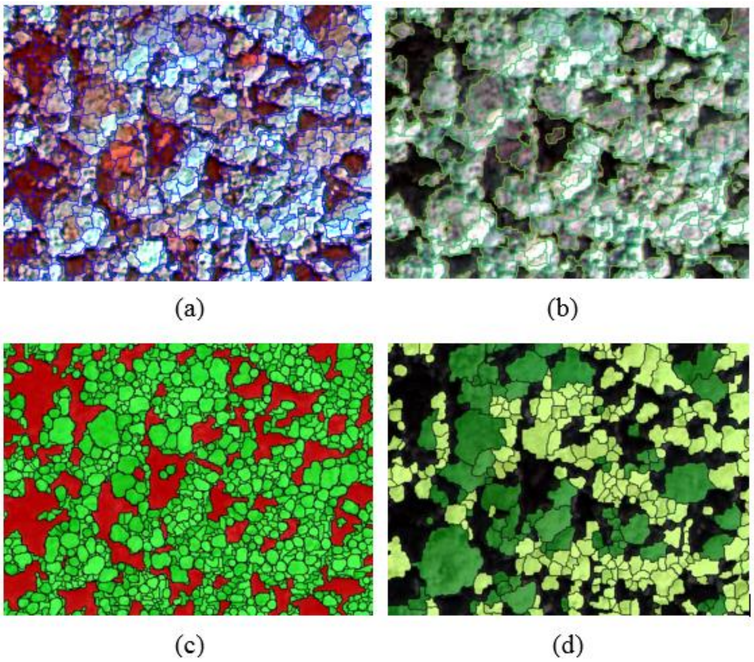

Figure 3. Result of tree crown delineation (a) image after multi-resolution segmentation process (b) shadow masked out from the image (c) image after morphology and (d) image classification of short and tall trees

3.0 RESULTS AND DISCUSSION

There were 240 individuals trees had been examined at this stage by using the Model 1 equation, carbon stocks estimation as dependent variables. Using the data generated from the study area, NDVI, IPVI, WV-VI,WV-BI, EVI, OSAVI, DVI, RVI, NIR/GREEN, GRVI, PCI, CPA, HT_L and reflectance 1 to 8 bands were tested as explanatory variables to the carbon stocks estimation. Using spatial statistics tools in ArcGIS (Arc Toolbox), the vegetation indices, reflectance value was calculated in related to the above-ground biomass and carbon stocks estimation for the study area.

There are 21 explanatory variables from remote sensing variables to execute the outputs. For the initial stage, the large VIF with (VIF > 7.5) values was removed and the OLS were run again. There are several assumptions of using the GWR and OLS in ArcGIS:

1. There is an asterisk (*) sign next to the number shows that there is a statistical significance ($p < 0.05$).

2. The rule of thumbs of the Variance Inflation Factors (VIF) is the explanatory variables related to the VIF values larger than 7.5 will be removed.

3. R^2, adjusted R^2 and Akaike's Information Criterion (AIC) used to measure the model performance. The value of R^2 ranges from zero to hundreds in percentage (%).

4. The model significance of Joint-F Statistics and Wald-Statistics denotes that if the asterisk (*) exists, indicates that the overall model was significant. In the OLS, if Koenker (BP) statistic (f) is statistical significance, the Joint Wald Statistic will be used to determine the overall model significance.

5. The Koenker (BP) Statistic or also known as Koenker's studentized Bruesch-Pagan statistic is a test to evaluate whether the explanatory variables in the model have a consistent relationship to the dependent variables in term of geographic and data space. In order to determine the coefficient significance, one ought to rely on the Robust Pr reading column whether significant or not.

6. The Jarque-Bera statistic indicates the residuals are normally distributed or not. When the test is statistically significant, model distribution is not normally distributed.

7. Spatial autocorrelation of the residuals using Moran's I to ensure that the data equally random.

The first test examined the explanatory variables (Height from LiDAR) HT_L and CPA by using dependent variables above-ground carbon stocks from Equation 2. The results show that all the explanatory variables were statistically significant with adjusted R^2 value was 0.690 with Koenker (BP) test significant. The overall model was significance with Wald Statistic and Robust_Pr was used to evaluate the coefficient significant.

Nevertheless, the Jarque-Bera statistic shows that the residuals were not statistic normally distributed ($p < 0.05$). Therefore, the spatial auto-correlation (Morans I) need to run to ensure that the regression residuals are spatially random distributed. In the auto correlation report in Morans I summarized that the p-value 0.272 and z-score is 1.0995 which is not statistically significant, so the null hypothesis of complete spatial randomness had been accepted. This indicates that the regression residuals are randomly distributed. The CPA and HT_L of OLS results reacted confidently to the above-ground carbon stocks estimation by using the dependent variables from Equation 2. The maximum cut-off value of the VIF is must be less than 7.5, and the value showed that there are no multi-collinearity exist as the value of VIF shown as 1.436 which is acceptable.

Table 2. Remotely-sensed above-ground carbon stocks using ArcGIS OLS

Variables	Coefficient	Standard error	T-Statistic	Probability	Robust SE	Robust t	Robust Pr	VIF
Intercept	-882.795	79.822	-11.060	0.000000*	113.496	-7.778	0.000000*	-
HT_L	39.468	4.346	9.082	0.000000*	7.429	5.313	0.000000*	1.436
CPA	12.956	1.019	12.713	0.000000*	4.287	3.022	0.000000*	1.436

Table 3. Remotely-sensed above-ground carbon stocks using ArcGIS OLS diagnostic

Number of Observations	240	Dependent variable	Equation1 (Carbon stock)
Degree of Freedom	237	Akaike's Information Criterion (AICc)	3414.698
Multiple R^2 [3]	0.692	Adjusted R^2	0.690
Joint F Statistic	266.629	Prob(>F), (2,237) degrees of freedom:	0.000000*
Joint Wald Statistic	94.382	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic	79.333	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Jarque-Bera Statistic	4489.694	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

The second test examined the explanatory variables of NDVI, IPVI, WV-VI,WV-BI, EVI, OSAVI, DVI, RVI, NIR/GREEN, GRVI, PCI, REF 1-8 and HT_L by using dependent variables above-ground carbon stocks from Equation 2. The results show that out of the 20 explanatory variables, only two variables are significant ($p < 0.05$) against dependent variables carbon stocks from Equation 2, which is HT_L and Reflectance band 5 which indicates by an asterisk (*) sign at the Joint-F statistic and Wald- statistics. However, the explanatory variables of vegetation indices and reflectance imagery did not improve the carbon stocks estimation and display the moderate value of R^2 which is 0.49. The VIF value also high for the variables Ref_B2 and Ref_B5 and more than 10 which is not acceptable as the multi-collinearity exist.

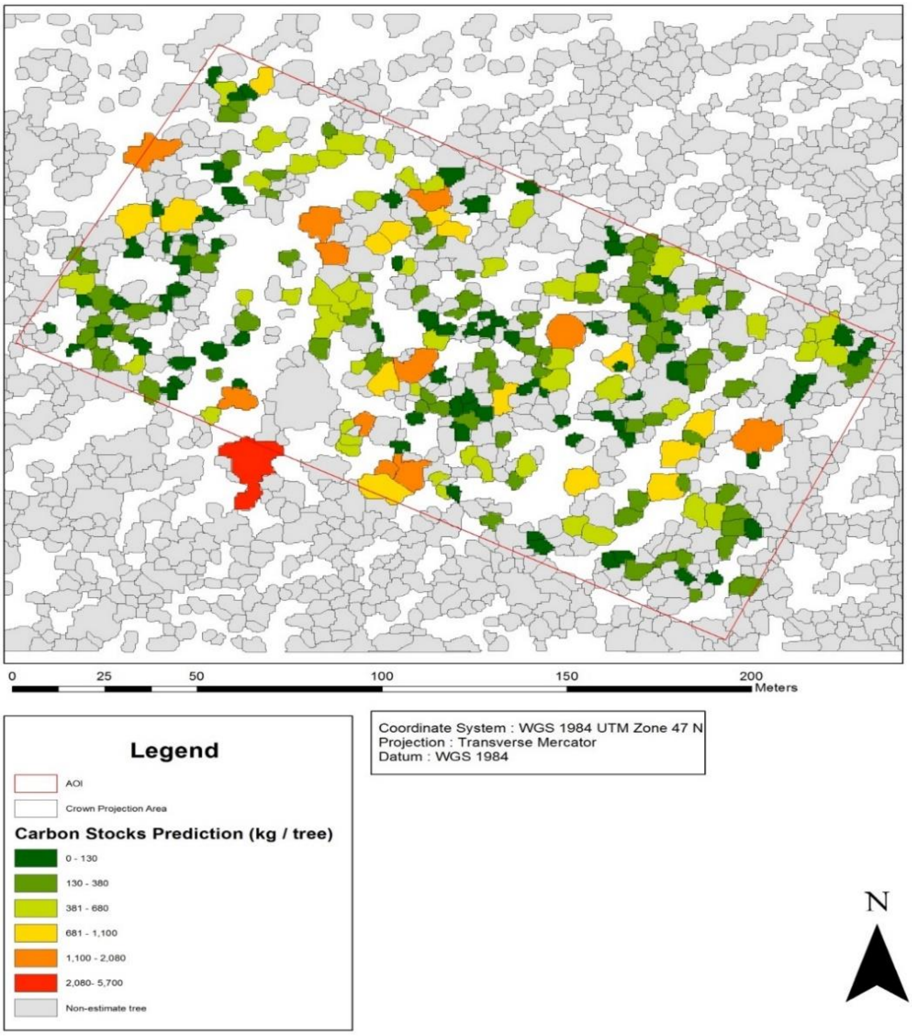

Figure 4. Above-ground carbon stocks prediction by using OLS and GWR in ArcGIS.

The Geographically Weighted Regression (GWR) is a local regression analysis that also contains dependent and explanatory variables. GWR can generate a hundred features of data set for the better result but this tools not working with multipoint or redundant data. In this research, GWR is using after generate the OLS analysis to see such as the intercept, probability value, residual and VIF value for the selected model. The final output of GWR according to the shape and extension of bandwidth corresponding to the user input the independent variable. The result of GWR gives the predicted value of carbon stock estimation as well as local R^2,a coefficient intercept of each variable, standard error, and standard error intercept.

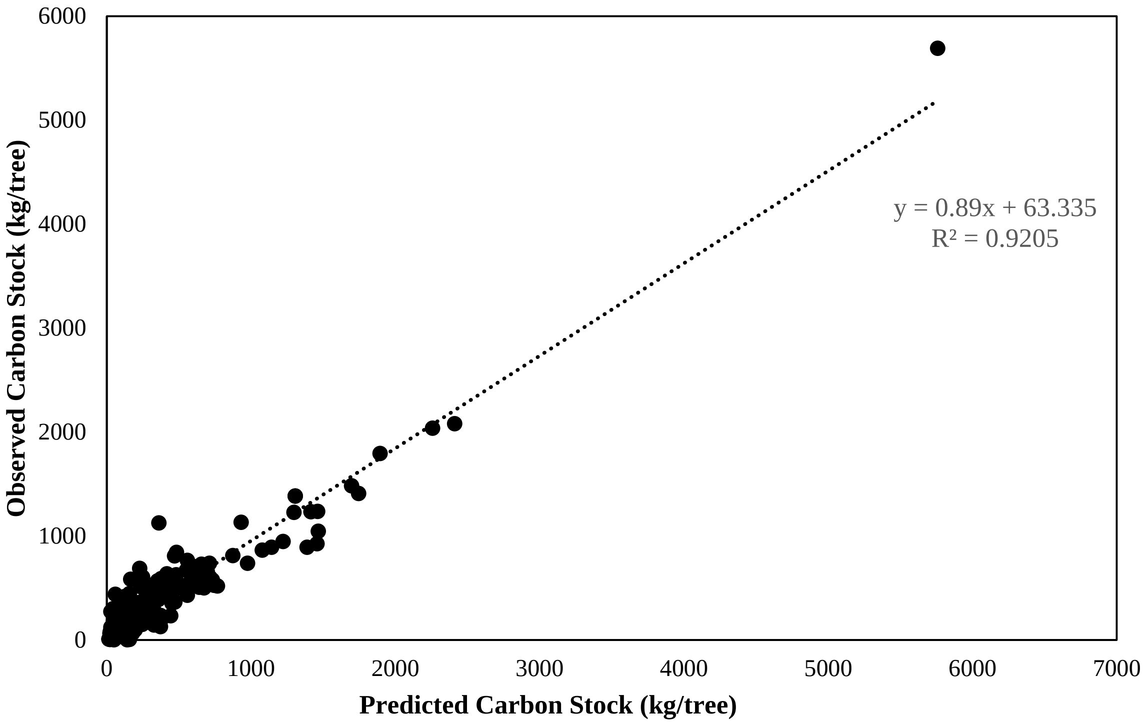

Figure 5. Scatter plot of predicted and observed carbon stock

Based on Figure 5, demonstrated the validity of the predicted (GWR results) against observed dataset (Equation 2) by using analytic methods. The scatter plot shows that $R^2 = 0.921$ which is the strong relationship between observed and predicted carbon stock. Therefore, these two variables, CPA and HT_L are confident can be used to estimate carbon stock for the wider area of tropical forest.

4. CONCLUSION

The study demonstrated that the advancement of remote sensing technologies nowadays helps to reduce the time and cut the cost of hiring the manpower for the forest inventory data collection. The combination of LiDAR data and Worldview-3 imagery was suitable for estimate carbon stock in the tropical rainforest. This research also proved the use of remotely sensed data can estimate and mapping the carbon stock for tropical rainforest by using GWR at AHFR. Particularly, the approach of OLS analysis and GWR tool give the best model equation for any incident or something that we need to predict. Height variables of tropical forest have potential to be used as one of the most important predictors for studying above-ground biomass and carbon stocks. More detailed work needs to be carried out in the future studies focus on monitoring changes of the height for the analysis of the emergent and canopy layers of the tropical forest. With the absence of field data inventories of the tree details, these model can fill in the gaps by deriving spatial input to the model, applying the model for tropical forest estimation by using remote sensing which will benefit the forest management and monitoring purposes. This research has identified the potential of linking the above-ground biomass and carbon stocks estimation and remotely-sensed data and beneficial for future development and strategic planning of the forest resources.

5. ACKNOWLEDGMENT

The authors would like to express their gratitude to the Ministry of Higher Education (MOHE), Malaysia, for financing the research under Fundamental Research Grant Scheme (FRGS/1/2015/WAB07/UITM/02/1) Universiti Teknologi MARA (UiTM), and also to Faculty of Forestry, Universiti Putra Malaysia (UPM) for granting access to the study area.

REFERENCES

Baccini, A., & Asner, G. P. (2013). Improving pantropical forest carbon maps with airborne LiDAR sampling. *Carbon Management*, *4*(6), 591–600.

Basuki, T. M., Skidmore, A. K., van Laake, P. E., van Duren, I., & Hussin, Y. A. (2012). The potential of spectral mixture analysis to improve the estimation accuracy of tropical forest biomass. *Geocarto International*, *27*(4), 329–345.

Belaid, L. J., & Mourou, W. (2009). Image segmentation: A watershed transformation algorithm. *Image Analysis and Stereology*, *28*(2), 93–102.

DigitalGlobe. (2014). *WorldView-3*.

Eckert, S. (2012). Improved forest biomass and carbon estimations using texture measures from worldView-2 satellite data. *Remote Sensing*, *4*(4), 810–829.

Esri. (2013). *Regression analysis basics*. *Esri (GIS) Mapping Software, Solutions, Services, Map Apps, and Data*. Retrieved December 27, 2016.

Gibbs, H. K., Brown, S., Niles, J. O., & Foley, J. a. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, *2*(4), 45023.

Hamdan, O., Khali Aziz, H., & Mohd Hasmadi, I. (2014). L-band ALOS PALSAR for biomass estimation of Matang Mangroves, Malaysia. *Remote Sensing of Environment*, *155*(October), 69–78.

Hansen, M. C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S. a, Tyukavina, a, … Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science (New York, N.Y.)*, *342*(6160), 850–3. 3

Ibrahim, F. (1999). Plant Diversity and Conservation Value of Ayer Hitam Forest, Selangor, Peninsular Malaysia. *Pertanika Journal of Tropical Agricultural Science*, *22*(2), 73–83.

IPCC. (2006). Guidelines for National Greenhouse Gas Inventories Volume 4 Agriculture,Forestry and other Land Use, *4*. Retrieved from http://www.ipcc-nggip.iges.or.jp/public/2006gl/

Karna, Y. K., Hussin, Y. A., Gilani, H., Bronsveld, M. C., Murthy, M., Qamer, F. M., … Bhattarai, T. (2013). Integration of WorldView-2 and airborne LiDAR data for tree species level carbon stock mapping in Khayar Khola watershed, Nepal. *ISPRS Journal of Photogrammetry and Remote Sensing*, *In Review*, 280–291.

Lu, D., Chen, Q., Wang, G., Liu, L., & Moran, E. (2014). International Journal of Digital Earth A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems, (January 2015), 37–41.

Mohd Zaki, N. A., & Abd Latif, Z. (2016). Carbon sinks and tropical forest biomass estimation: a review on role of remote sensing in aboveground-biomass modelling. *Geocarto International*, *6049*(May), 1–16.

Mohd Zaki, N. A., Abd Latif, Z., Zainal, M. Z., & Zainuddin, K. (2015). Individual Tree Crown (ITC) Delineation Using Watershed Transformation Algorithm For Tropical Lowland Dipterocarp. *Proceedings of the IEEE 2015*, *2*, 237–242.

Muukkonen, P., & Heiskanen, J. (2005). Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. *Remote Sensing of Environment*, *99*(4), 434–447.

NOAA. (2016). Carbon dioxide levels race past troubling milestone. Retrieved November 10, 2016, from http://www.noaa.gov/stories/carbon-dioxide-levels-race-past-troubling-milestone

Propastin, P. (2012). Modifying geographically weighted regression for estimating aboveground biomass in tropical rainforests by multispectral remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, *18*(1), 82–90.

Scott, L. M., & Janikas, M. V. (2010). Spatial Statistics in ArcGIS. In *Handbook of Applied Spatial Analysis Software Tools, Methods and Applications* (pp. 27–42). Heidelberg, Germany: Springer.

Shida, N., Hanum, F., W.M, W. R., & K, K. (2014). Community Structure of Trees in Ayer Hitam Forest Reserve , Puchong,. *The Malaysian Forester*, *77*(1), 73–86.

UNFCC. (2015). UNFCC. Retrieved May 20, 2015, from http://unfccc.int/kyoto_protocol/items/2830.php