ABOVE GROUND BIOMASS ESTIMATION USING MULTISPECTRAL SENTINEL-2 MSI DATA: A PRELIMINARY EXPERIMENT IN WANGCHAN FOREST LEARNING CENTER, THAILAND.

Sorasak Khoomboon[1,a], Tranid Prasertsri[1], Pudtraporn Napang[1], Natnicha Yooyen[2], Mattana Pongsopon[2], Suwit Navakam[2], Napak Karnasuta[2], and Sasawat Soontaros[2,b]

[1]AI and Robotics Ventures Co., Ltd. Bhiraj Tower at Sathorn Building C 33, 31 31/1 S Sathorn Road, Yannawa, Sathorn, Bangkok 10120, Thailand
[2]Reforestation and Ecosystem Institute, PTT Public Co., Ltd.
555 Vibhavadi Rangsit Road, Chatuchak Bangkok 10900 Thailand
Email: [a]sorasakk@arv.co.th, [b]sasawat.s@pttplc.com

ABSTRACT

Dry Evergreen Forest above ground biomass (AGB) directly indicates conditions of the ecosystem, carbon cycle, and biodiversity conservation. Accurate AGB estimation is essential for the monitoring and supervision of the ecosystem and important parameters to calculate carbon sequestration, which is a method to determine carbon adsorption capacity of forest areas. This is an indicator to monitor the integrity of the forest within the area of interest that has been restored. The calculation of carbon sequestration is the most accurate and accepted method by Thailand Greenhouse Gas Management Organization (Public Organization) (TGO) through field survey to measure the parameters of several trees and calculation using allometric equations, which require more human resource, time, and expenses in large forest areas. However, using Sentinel-2, we can collect reflection data in various wavelengths within associated properties of plants and trees with large spatial measurements. It also has a high image resolution and can be downloaded for free to use in the interpretation and analysis of AGB estimation to handle resource problems. This study aims to carry out an experiment employing Sentinel-2 Multispectral Instrument (MSI) data to test the correlation between reflectance in the spectral channels and vegetation indices derived from imagery to model Dry Evergreen Forest AGB using Machine learning regressors (Random Forest, AdaBoost, Support vector machine, etc.) in the Dry Evergreen Forest at Wangchan Forest Learning Center in Rayong, Thailand, where the feasibility of model application was demonstrated.

1. INTRODUCTION

The above ground biomass (AGB) of forest accounts for 70-90% of total forest biomass, which directly indicates the abundance of ecosystems, carbon cycle, and biodiversity within the area (Prado-Junior et al., 2016). Moreover, forest AGB estimation is an important parameter to calculate Carbon Sequestration which leads to an assessment of the carbon absorption capacity in the area including forest resources management as well as monitoring and evaluating forest restoration, managing forest ecosystems to cope with climate change, and carbon credit trading (Muralikrishna, 2014).

The current method for forest AGB estimation in Thailand has been established by the Thailand Greenhouse Gas Management Organization (TGO), which involves field data collection by counting trees and plot sampling (*Thailand Voluntary Emission Reduction Program Reference Manual: Forestry and Agriculture Sector*, 2016). Tree counting is used for areas with plantation subplots sizes not exceeding 30 rai (4.80 hectare) and multiplied by a constant determined by the TGO. The sample plots for data survey should cover more than

1.0% of the total area to measure tree dimensions: diameter at breast height (DBH) and height. Biomass and capacity of carbon sequestration are determined by calculating from the allometric equations (Senpaseuth, 2009). The obtained result is multiplied with the Carbon Fraction according to TGO's guideline. Then, the carbon content in the plantation plot is determined and converted to the total amount of carbon dioxide. However, the area for sample plots will be larger according to the forest size, as well as a greater number of human resources, time, and expenses differently required in different areas of sample plots.

Although field data measurement is a method that has been recognized for its accuracy, it is not a practical approach for broad-scale assessments that require more resources and time. Therefore, to address these problems, remote sensing technology is introduced. This method has various advantages, especially in terms of resource reduction along with non-sampling whole plantation area assessment and accessibility in all areas. It has been reported the possibility of AGB estimation in sub-tropical buffer zone community forests from data collection on remote sensing technologies consisting of satellite data and optical remote sensing data using Sentinel-2 data (Pandit, Tsuyuki, & Dube, 2018).

This study aims to carry out an experiment by using Sentinel-2 Multispectral Instrument (MSI) data, which is a free optical imagery with 10 m resolution remote sensing data, to test the correlation between reflectance in the spectral channels and vegetation indices derived from imagery to estimate Dry Evergreen Forest AGB using Machine learning regressors at Wangchan Forest Learning Center in Rayong province, Thailand and to compare this data with the field measurement in order to find out the method and closest equation for this estimation, so that it can be applied to other forest areas.

2. MATERIALS AND METHODS

2.1 Study area

The study area is within the premises of Wangchan Forest Learning Center in Rayong province, Thailand (12.9948°N, 101.4421°E). The dry evergreen forest has been grown for about 6 years (since 2014) in the Forest Carbon Model Zone (FCMZ) covering a total area of 18.68 hectare.

2.2 Sampling strategy and field data collection

Field data collection and AGB calculation were performed following the TGO's guideline (*Thailand Voluntary Emission Reduction Program Reference Manual: Forestry and Agriculture Sector*, 2016). Three sample plots were selected from different locations of FCMZ, marked red in Figure 1. Each sample plot was further divided into 16 subplots, where 1 subplot has the size of 10 x 10 m to relate to the resolution of Sentinel-2. AGB of one data point is the summation of AGB of every tree within a single subplot. Hence, there are a total of 48 data plots in this study.

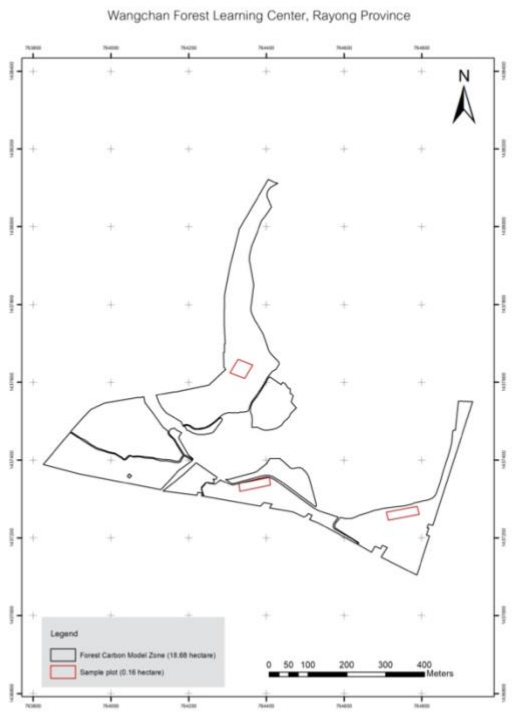

Figure 1. Study area for AGB estimation using sentinel-2 imagery.

2.3 Satellite image acquisition and variables for forest AGB estimation

Data of Single Tile Standard Sentinel-2 Level 2A on 26 July 2018 was downloaded from AWS earth ("Open Data on AWS," 2018), selected based on the cloudless condition in the area of interest and the time taken is similar to that of field data collection. The data we used in this study is acquired in 10 spectral bands or 'raw bands' which extend through (i) visible wavelength; B2, B3, and B4, (ii) near-infrared (NIR); B8, (iii) red-edge; B5, B6, B7, and B8A, and (iv) shortwave infra-red (SWIR); B11 and B12 wavelengths at 10 and 20 m spatial resolution. Subsequently, we re-sampled the resolution with B5, B6, B7, B8A, B11, and B12 to 10 m spatial resolution. To study the applicability of the Sentinel2 MSI sensor to estimate dry evergreen forest restoration AGB, raw bands and vegetation indices (VIs) were used. The selection of VIs was based on their performance in forest biophysical parameters introduced by Pandit and team as shown in Table 1 (Pandit et al., 2018).

Table 1. Vegetation indices calculated from Sentinel-2 previously introduced.

Vegetation Indices	Equations
NDVI	(B8-B4)/(B8+B4)
RGR	B4/B3
EVI	2.5*((B8-B4)/(1+B8+6*B4-7.5*B2))
SR	B8/B4
PSRI	((B4-B3)/B6)
NDII	(B8-B11)/(B8+B11)
SAVI	((B8-B4)/(B8+B11+1.5)) *1.5
IRECI	(B8-B4)/(B5/B6)
S2REP	705+(35*((0.5*(B7 +B4)/2)-B5)/(B6-B5))
RE1NDVI	(B8 -B5)/(B8+B5)
RE2NDVI	(B8 -B6)/(B8+B6)
RE3NDVI	(B8 -B7)/(B8+B7)
RE4NDVI	(B8 -B8A)/(B8+B8A)

2.4 Machine learning regression models

We used Sentinel-2 Multispectral Instrument (MSI) data to estimate AGB between reflectance in the spectral channels, vegetation indices derived from imagery, and field measurement data of 48 data points each data point represents a sample subplot. We divided these data points into 19 data points for training model and 13 data points for testing model using Machine learning regression models in Scikit-learn library (Pedregosa et al., 2011) which are random forest (RF) (Breiman, 2001), Adaptive Boosting/AdaBoost (ADA) (Freund & Schapire, 1997), Bagging (BAG) (Breiman, 1996), Gradient Boosting (GB) (Friedman, 2001), Support Vector Machine (SVM) (Platt, 1999), and least squares Linear Regression (LIN). Performance was tested to estimate AGB with the remaining unseen 16 data points. The overview method used in this study is presented in Figure 2.

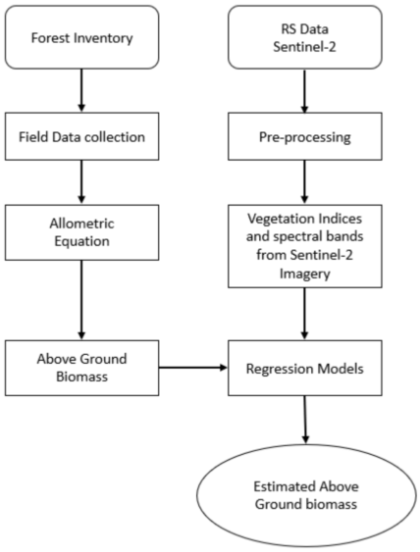

Figure 2. Method overview for AGB estimation using Sentinel-2 imagery.

3. RESULTS AND DISCUSSION

The experimental results of training with random RF, ADA, BAG, GB, SVM, and LIN from 19 data points of both data sets, data set A consisting of only raw bands (B2, B3, B4, B5, B6, B7, B8, B8A, B11 and B12) and data set B consisting of raw bands and VIs (B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, NDVI, RGR, EVI, SR, PSRI, NDII, SAVI, IRECI, S2REP, RE1NDVI, RE2NDVI, RE3NDVI, and RE4NDVI), revealed the feasibility in biomass estimation with the other 13 data points used in the testing model. Accuracy test of the experiment was obtained with R^2 showed greater than 0.80 as shown in Table 2. Using SVM with data set B, the most accurate value obtained was 0.88, which demonstrates the possibility of applying Machine learning regression models to estimate AGB in dry evergreen forest. However, when the resulting model is applied to 16 unseen data points, none of the models were able to estimate AGB as shown in Table 2. This could be caused by two reasons, one is the small sample size in which the amount of data points used for training is very low, since the number of sample subplots in this experiment is 48 in 3 sample plots, whereas the previously reported model study had a total of 113 plots (Pandit et al., 2018). The other reason is that the subplot location does not align with the direction of Sentinel-2 imagery (North-South direction), or the data measurement of the subplot location does not fit the pixel position of the satellite image.

Moreover, the study area is dry evergreen forest with a complex and multi-tiered canopy structure, hence it is necessary to take forest type and topography into account as well.

Table 2. R^2 with 13 test and 16 unseen data points.

Regression model	13 Test data points		16 Unseen data points	
	A	B	A	B
RF	0.83	0.82	-6.82	-4.65
ADA	0.84	0.85	-3.86	-6.14
BAG	0.81	0.80	-5.09	-5.1
GB	0.83	0.83	-4.98	-4.76
SVM	0.83	0.88	-5.95	-5.97
LIN	0.81	0.81	-4.46	-5.4

4. CONCLUSION

We could carry out an experiment using Sentinel-2 Multispectral Instrument (MSI) data to demonstrate the feasibility of application of Machine learning regression models to estimate AGB in dry evergreen forest. However, this depends on the amount of data points used for training and strategic field data collection. Field data collection following the TGO's guideline resulted in misalignment to that of Sentinel-2 imagery which might have caused failure in AGB estimation using the unseen data. In future work, we plan to design a new strategy for field measurement to align with the direction of Sentinel-2 image sensing and divide contour level for location selection in the forest along with a higher number of sample size. Unfortunately, field experiments have been limited due to the situation of COVID-19 outbreak in Thailand which occurred during the study period. Hence, we have not been able to test the idea with larger samples yet. Nonetheless, once the situation is better to an extent that we can implement our field work, we are certain to develop and test the AGB assessment using Machine learning regression models to compute carbon storage estimations at the commercial level.

5. ACKNOWLEDGEMENTS

This research was co-supported by the Reforestation and Ecosystem Institute of PTT Public Co., Ltd. and the AI and Robotics Ventures Co., Ltd. We thank Thailand Greenhouse Gas Management Organization (TGO) and Faculty of Forestry, Kasetsart University for utilizing their expertise and providing us suggestions that greatly assisted the research.

6. REFERENCES

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123-140. doi:10.1007/BF00058655

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi:10.1023/A:1010933404324

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences, 55*(1), 119-139. doi:https://doi.org/10.1006/jcss.1997.1504

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189-1232, 1144. Retrieved from https://doi.org/10.1214/aos/1013203451

Muralikrishna, I. (2014). Biomass Calculations for Carbon Sequestration in Forest Ecosystem. *Journal of Energy and Chemical Engineering*.

Open Data on AWS. (2018). Retrieved from https://aws.amazon.com/opendata/

Pandit, S., Tsuyuki, S., & Dube, T. (2018). Estimating Above-Ground Biomass in Sub-Tropical Buffer Zone Community Forests, Nepal, Using Sentinel 2 Data. *Remote Sensing, 10*(4), 601. Retrieved from https://www.mdpi.com/2072-4292/10/4/601

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res., 12*(null), 2825–2830.

Platt, J. (1999). *Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods*: MIT Press.

Prado-Junior, J. A., Schiavini, I., Vale, V. S., Arantes, C. S., van der Sande, M. T., Lohbeck, M., & Poorter, L. (2016). Conservative species drive biomass productivity in tropical dry forests. *Journal of Ecology, 104*(3), 817-827. doi:https://doi.org/10.1111/1365-2745.12543

Senpaseuth, P., Navanugraha, C., & Pattanakiat, S. . (2009). The Estimation of Carbon Storage in Dry Evergreen and Dry Dipterocarp Forests in Sang Khom District, Nong Khai Province, Thailand. *Environment and Natural Resources Journal, 7*(2), 1-11. Retrieved from https://ph02.tci-thaijo.org/index.php/ennrj/article/view/82564

Thailand Voluntary Emission Reduction Program Reference Manual: Forestry and Agriculture Sector. (2016). (3rd ed.). Bangkok, Thailand: Thailand Greenhouse Gas Management Organization (Public Organization).