

ESTIMATING ABOVEGROUND BIOMASS OF BAMBOO AND MIXED BAMBOO FOREST IN THUA THIEN-HUE PROVINCE, VIET NAM USING PALSAR-2 AND LANDSAT OLI DATA

Cat Tuong T.T.^{1,2}, Tani Hiroshi^{3,*}, Xiufeng Wang³, Van-Manh Pham⁴

¹ Mien Trung Institute for Scientific Research, Vietnam Academy of Science and Technology, 321 Huynh Thuc Khang Street, Thua Thien Hue Province 530000, Vietnam

² Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan - cattuong@env.agr.hokudai.ac.jp

³ Research Faculty of Agriculture, Hokkaido University, Sapporo 060-8589, Japan - tani@env.agr.hokudai.ac.jp;
wang@env.agr.hokudai.ac.jp

⁴ Faculty of Geography, VNU University of Science, 334 Nguyen Trai, Thanh Xuan, Ha Noi, Vietnam -
manh10101984@gmail.com

WG IV/4 & IC WG IV/III

KEY WORDS: Aboveground biomass, bamboo, mixed bamboo forest, remote sensing data, multivariate linear regression, Thua Thien Hue province, Vietnam

ABSTRACT:

In this study, above-ground biomass (AGB) performance was evaluated by PALSAR-2 L-band and Landsat data for bamboo and mixed bamboo forest. The linear regression model was chosen and validated for forest biomass estimation in A Luoi district, Thua Thien Hue province, Vietnam. A Landsat 8 OLI image and a dual-polarized ALOS/PALSAR-2 L-band (HH, HV polarizations) were used. In addition, 11 different vegetation indices were extracted to test the performance of Landsat data in estimating forest AGB. Total of 54 plots were collected in the bamboo and mixed bamboo forest in 2016. The linear regression is used to evaluate the sensitivity of biomass to the obtained parameters, including radar polarization, optical properties, and some vegetation indices which are extracted from Landsat data. The best-fit linear regression is selected by using the Bayesian Model Average for biomass estimation. Leave-one-out cross-validation (LOOCV) was employed to test the robustness of the model through the coefficient of determination (R^2) and Root Mean Squared Error (RMSE). The results show that Landsat 8 OLI data has a slightly better potential for biomass estimation than PALSAR-2 in the bamboo and mixed bamboo forest. Besides, the combination of PALSAR-2 and Landsat 8 OLI data also has a no significant improvement (R^2 of 0.60) over the performance of models using only SAR (R^2 of 0.49) and only Landsat data (R^2 of 0.58-0.59). The univariate model was selected to estimate AGB in the bamboo and mixed bamboo forest. The model showed good accuracy with an R^2 of 0.59 and an RMSE of 29.66 tons ha^{-1} . The comparison between two approaches using the entire dataset and LOOCV demonstrates no significant difference in R (0.59 and 0.56) and RMSE (29.66 and 30.06 tons ha^{-1}). This study performs the utilization of remote sensing data for biomass estimation in bamboo and mixed bamboo forest, which is a lack of up-to-date information in forest inventory. This study highlights the utilization of the linear regression model for estimating AGB of the bamboo forest with a limited number of field survey samples. However, future research should include a comparison with non-linear and non-parametric models.

1. INTRODUCTION

Recently, using multi-sources data has become increased to evaluate the forest biomass because of the potential improvement of the estimated accuracy. Various methods have been developed for combining different data, for example, the remotely sensed and ground data (Badreldin & Sanchez-Azofeifa, 2015; Zhang et al., 2019), or fusion technique between satellite images (Cutler, Boyd, Foody, & Vetrivel, 2012; Fayad et al., 2016; Tian et al., 2017). The combination of optical data and synthetic aperture radar (SAR) has been received much attention because it provides much better information and improves estimation accuracy. The optical images are rich in spectral and spatial information, while SAR has several advantages like sensitivity to dielectric properties surface roughness (Mahyoub, Fadil, Mansour, Rhinane, & Al-Nahmi, 2019), longer wavelength and can penetrate through the forest canopy (CEOS, 2018).

Among various optical images, Landsat has been commonly used because of the wide range of spectral bands, medium spatial resolution, and an open satellite imagery source. There are

several studies were related to the biomass estimation through the relationship between Landsat signals and forest structure parameters such as tree height or stand volume (Chrysafis, Mallinis, Gitas, & Tsakiri-Strati, 2017; Hall, Skakun, Arsenault, & Case, 2006). The sensitivity of spectral bands to biomass estimation was mentioned in (Powell et al. 2010; Lu et al. 2012). Besides, some vegetation indices (VIs), as an additional feature extracted from optical data, were found to be significant variables to calculate forest biomass (Foody et al., 2003; Propastin, 2012). The success of VIs application on forest biomass estimation depends on different forest ecosystems as thoroughly reviewed (Sarker & Nichol, 2011). Although VIs are limitedly considered for estimating biomass in tropical forests, some VIs are used as an approach of reducing saturation in simple spectral bands (Zhao, Lu, Wang, Liu, et al., 2016).

Synthetic aperture radar is one of the most promising remote sensors to map the global forest biomass (Mermoz et al., 2015). Various SAR data were popularly used, such as ERS-1 and -2, JERS-1, Envisat ASAR, RADARSAT, and ALOS PALSAR-1 and -2. PALSAR is an L-band frequency microwave sensor

* Corresponding author

which is a joint project between JAXA and the Japan Resources Observation System Organization. With the launch of PALSAR-2 in 2014, it becomes a unique and highly useful sensor now with high-resolution, wide swath width and image quality until now (EORC). At L-band, the penetration depth exceeds the crown layer and mechanism involving the lower part of the canopy presented a strong relationship with biomass from the major contribution of branches and trunk (Toan, Beaudoin, Riou, & Guyon, 1992). Much successful application of PALSAR-1/2 data for forest biomass mapping was recognized based on backscatter intensities (Suzuki, Kim, and Ishii 2013; Mermoz 2014) or different techniques such as polarimetry and interferometry (Chowdhury, Thiel, Schmullius, & Stelmaszczyk-Górska, 2013; Neumann, Saatchi, Ulander, & Fransson, 2012; Thiel & Schmullius, 2016). A recent approach of tomography showed a high potential for biomass estimation through its correlation to forest height (Ho Tong Minh et al., 2016; Tebaldini & Rocca, 2012).

However, a saturation phenomenon is a constraint in both optical and radar data which can lead to underestimating forest biomass. Saturation value varies in different satellite images and different forest ecosystems. Recent studies were conducted to indicate the saturation point which often found in dense canopy forests with high biomass values. Some instances identified the saturation point at 150 tons ha^{-1} , corresponding to HV backscatter at -11.52 dB for a semi-evergreen rain forests and savannas (Mermoz et al., 2015), and for subtropical forests (Zhang et al., 2019), while this point can be reached at 100 tons ha^{-1} for tropical forests (Häme, Rauste, Antropov, Ahola, & Kilpi, 2013). For Landsat data, a wide range of saturation levels was examined for different vegetation types in a subtropical region (Zhao, Lu, Wang, Wu, et al., 2016). Methods to eliminate the effects of saturation have been discussed in a limited number of studies for Landsat images (Avitabile, Baccini, Friedl, & Schmullius, 2012; Phua et al., 2017), and for SAR images (Carreiras, Vasconcelos, & Lucas, 2012; Mermoz et al., 2015). The combination of Landsat and SAR products is also possible to reduce this effect in some instances (Basuki, Skidmore, Hussin, & van Duren, 2013; Cutler et al., 2012; Zhao, Lu, Wang, Liu, et al., 2016).

Another issue in forest biomass calculation is the divergence in species composition and structure in different vegetation types. Therefore they have different data saturation values in Landsat or radar data (Zhao, Lu, Wang, Liu, et al. 2016), and require the diversity of remote sensing algorithms and datasets that have been used to retrieve forest biomass (Lucas et al., 2015). Despite the increasing researches on forest aboveground biomass, there are a limited number of studies developed biomass estimation for bamboo. Bamboo forests are distributed 0.8% of the world's total forested area with total ecosystem carbon in the range of 94–392 $Mg C \cdot ha^{-1}$. With the rapid growth rate, they contribute significantly to sequester substantial quantities of carbon with the estimated annual carbon accumulation rates of 8–14 $Mg C \cdot ha^{-1}$, thereby helping to mitigate the effects of climate change (Yuen, Fung, & Ziegler, 2017). Besides, because of its economic and environmental benefits, the bamboo forest is being considered as an alternative sustainable land-use strategy in the upland region, particular in Vietnam (Ly, Pillot, Lamballe, & de Neergaard, 2012). Therefore, there is a demand to evaluate the performance of satellite sources data for individual bamboo forest which provides useful information for our understanding to plan a sustainable development strategy.

In this study, we focused on evaluating the performance of ALOS/PALSAR-2 and Landsat OLI on forest biomass by using a single source and the combined data in order to develop the best model for the bamboo and mixed bamboo forest. The paper is structured as follows: Section 2 describes the Study area and field data; Section 3 explains the methods to process data and estimate forest biomass; Section 4 showed the performance of Landsat OLI and PALSAR-2 for aboveground biomass estimation and the best-fit model selection; Our findings are discussed in Section 5; Finally, a conclusion is showed in Section 6.

2. STUDY AREA AND FIELD DATA

2.1 Description of the study area

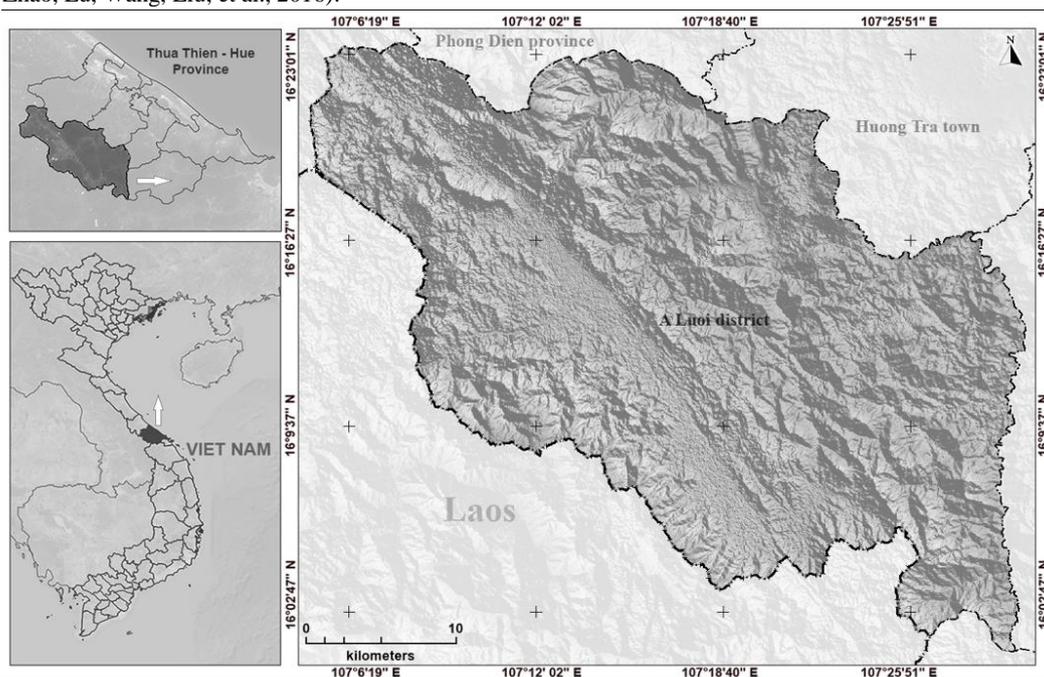


Figure 1. Location of the study areas (A Luoi district) in Thua Thien-Hue province, Vietnam.

Bamboo forest has a unique morphological structure that is easily identifiable from a distance, a secondary subtype formed on natural forest land after exploitation or shifting cultivation. Bamboo forests in Vietnam are widely distributed from an altitude nearly above sea level to 2,000 m. There are two groups of bamboos, including a group of herbaceous bamboos and another group of woody species of bamboo. Biomass and carbon accumulation are different among these bamboo groups. In the context of climate change, bamboo forests provide a number of ecosystem services that are beneficial for carbon sequestration. Bamboo can isolate significant amounts of carbon from above-ground biomass (AGB). Therefore, the bamboo forest is an important resource to minimize the greenhouse effect of climate change.

A Luoi is a mountainous district of Thua Thien Hue province in Central Vietnam (Figure 1). This region covers about 1,224.6 km² (accounting for 24.17% of the natural land area of Thua Thien - Hue province) with an average altitude of 600-800 m above sea level, the average slope of 2⁰-25⁰. The west part is steep mountains ranging from 500 to 1,700 m, while the east part is more flat with an average elevation of 600 m above sea level. A Luoi area has the most rainfall compared to other localities in Thua Thien-Hue province, the annual precipitation is an average of about 3,500 mm, and there is a dense hydrographic network with five major rivers flowing through. In general, A Luoi has a cool climate throughout the year with the characteristics of climate, local weather in the area of A Luoi has favorable for developing bamboo forest.

2.2 Field data collection

The forest ground data was provided by the Central Sub Forest Inventory and Planning Institute, Thua Thien Hue Province, Vietnam (Sub-FIPI). The in-situ measurements were conducted in 273 field plots from 16 January to 3 July 2016 over the whole Thua Thien Hue Province. The satellite images covered 54 plots including the bamboo and mixed bamboo forest. A sample plot size had a rectangular shape of 30 m × 33 m with a longer aspect in an east-west direction and the shorter aspect in the north-south direction. In each plot, there were four sub-plots 5 m × 5 m in size. Diameter at breast height (DBH) was measured for all trees with a diameter over 6 cm, while total tree height (H) was measured for five normal growth trees near the center of the plot. Allometric equations were used to estimate the height of the remaining trees in the plots. The main ecosystem was the broad-leaved tropical forest and bamboo forest. The bamboo forest includes 21 plots, and the forest mixed bamboo with broad-leaved species includes 33 plots.

For the bamboo forest, the number of trees, diameter and the average height of bamboo in each sub-plot were measured. If the bamboo grew like a clump, it was necessary to count the number of clumps in a plot and the number of stems per clump. For estimating the aboveground biomass (AGB) of bamboo forest, we used the formula given by Ly and partner in 2012:

$$AGB = 0.3002 DBH^2 + 0.115 DBH + 1.7632 \quad (1)$$

For estimating the biomass of broadleaved species tree, the one-factor formula to estimate AGB (in tons ha⁻¹) given by Bao Huy and colleagues is:

$$AGB = 0.104189x DBH^{2.491453} \quad (2)$$

3. METHODOLOGY

3.1 Satellite data and pre-processing

A dual-polarized radar data (HH, HV polarizations) in single look complex (SLC) format were comprised in May 2016. The preprocessing data was operated to convert the digital number value into sigma naught (σ^0) values. A refined Lee filter was used with a window size of 7x7 to reduce the speckle noise. The topography effect was eliminated by using Range-Doppler Terrain Correction with a digital elevation model (DEM) from Shuttle Radar Topography Mission and resampled all of the product images to reach 6.5 meters in pixel spacing.

Landsat 8 OLI are used for optical data, which is provided by the United States Geological Survey (USGS) with moderate resolution and wide spectral coverage. A scene Landsat OLI was acquired in April 2016 cover entire the study area with a spatial resolution of 15 m in panchromatic and 30 m in the multi-spectral band. Landsat digital numbers (DNs) were converted to reflectance and atmospheric correction by using the FLAASH tool (Fast line-of-sight atmospheric analysis of hypercubes). Then, different vegetation indices were extracted from the pre-processed image (Table 1).

Vegetation indices can be calculated by ratioing, differencing, ratioing differences and sums, and by forming linear combinations of spectral band data. Vegetation indices are intended to enhance the vegetation signal (Jackson & Huete, 1991) and successfully used for estimating biophysical properties (Anderson & Hanson, 1993; Sarker & Nichol, 2011).

In this study, we used vegetation indices calculated using band Near-Infrared (0.7-1.1 μ m), red (0.6-0.7 μ m), blue (0.45-0.52 μ m) and green band (0.52-0.60 μ m) in Landsat OLI data. For the Enhanced vegetation index (EVI) formula, L is a soil adjustment factor, and C₁ and C₂ are coefficients used to correct aerosol scattering in the red band by the use of the blue band. In general, G=2.5, C₁=6.0, C₂=7.5, and L=1 (Huete, Liu, Batchily, & J., 1997).

For TSAVI, the equation was developed by (Baret & Guyot, 1991), where *s* and *a* are the soil line parameters with the default value of 0.5, and *X* is an adjustment factor that is set to minimize soil noise with value of 0.08.

3.2 Performance of parameters extracted from remotely sensed data in forest biomass estimation

Features derived from remotely sensed data using a 5x5 window size were identified as variables for AGB estimation. Different features for sample areas, such as sigma value of HH and HV polarization from PALSAR-2, multispectral data from Landsat OLI includes Red and Near-infrared (NIR), and vegetation indices (EVI, NRVI, NDVI, TNDVI, GNDVI, RVI, TTVI, TVI, RatioNR, SAVI, TSAVI, MSAVI). The linear relationship of in-situ AGB and these features were determined through analyzing the pairwise Pearson's correlation coefficient. Next, multiple linear regression was conducted to identify the best variables for accurate AGB estimation. Different sets of variables were used.

A total of 22 models were created from these variables. The best multiple linear models were selected through Bayesian Model Averaging (BMA). BMA accounts for the model uncertainty inherent in the variable selection problem by averaging over the best model in the model class according to approximate posterior model probability (Raftery et al., 2018). The best-fit model is tested for AGB estimation in the study site.

Table 1. Predictor variables from PALSAR-2 and LANDSAT 8 OLI used in this study.

Indepent variables	Explanations/Name	Describe/Wavelength	Formula/Resolution
LANDSAT 8 OLI	Band 2: Blue	0.45 – 0.51 μm	30 m
	Band 3: Green	0.53 – 0.59 μm	30 m
	Band 4: Red	0.64 – 0.67 μm	30 m
	Band 5: Near Infrared	0.85 – 0.88 μm	30 m
Spectral indices from LANDSAT 8 OLI	EVI	Enhanced Vegetation Index	$G \times \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + (C_1 \times \rho_{red} - C_2 \times \rho_{blue}) + L}$
	RVI	The simple Ratio Vegetation Index	$\frac{\rho_{red}}{\rho_{nir}}$
	NRVI	The Normalized Ratio Vegetation index	$\frac{RVI - 1}{RVI + 1}$
	NDVI	Normalized Difference Vegetation Index	$\frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$
	TNDVI	Transformed Normalized Difference Vegetation Index	$\sqrt{NDVI + 0.5}$
	GNDVI	Green Normalized Difference Vegetation Index	$\frac{\rho_{nir} - \rho_{green}}{\rho_{nir} + \rho_{green}}$
	TTVI	Thiam's Transformed Vegetation Index	$\sqrt{ABS\left(\frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} + 0.5\right)}$
	RatioNR	Ratio of Near-infrared and Red	$\frac{\rho_{nir}}{\rho_{red}}$
	SAVI	Soil Adjusted Vegetation Index	$\frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red} + L} \times (1 + L)$
	MSAVI	Modified Soil Adjusted Vegetation Index	$\frac{1}{2}(2(\rho_{nir} + 1) - \sqrt{2(\rho_{nir} + 1)^2 - 8(\rho_{nir} - \rho_{red})})$
	TSAVI	Transformed Soil Adjusted Vegetation Index	$\frac{s(\rho_{nir} - s \times \rho_{red} - a)}{a \times \rho_{nir} + \rho_{red} - a \times s + X \times (1 + s^2)}$
PALSAR-2	HH	HH polarization (σ, dB)	6.5 m
	HV	HV polarization (σ, dB)	6.5 m

* ρ_{nir} , ρ_{red} , ρ_{blue} , and ρ_{green} : reflectance values of Near-Infrared, Red, Blue, and Green band.

3.3 Assessment of model accuracy

Leave-one-out cross-validation (LOOCV) was obtained to test the robustness of the model through the coefficient of determination (R^2) and Root Mean Squared Error (RMSE). LOOCV is a special case of k -fold cross-validation, in which the number of folds equals the number of observations (Wong, 2015). This type of estimate is obtained by carrying out N repetitions of a learn+test cycle, where n is the size of the given data set. On each repetition one of the n observations is left out to serve as a test set, while the remaining $n-1$ cases are used to obtain the model. The process is repeated n times by leaving aside each of the n given observations (Torgo, 2015).

To assess the accuracy of models, a correlation coefficient (R) and RMSE were used. The ratio of R^2 is so called the coefficient of determination, and varies between a range of 0 and 1. With the predicted values of the estimator (\hat{y}) and the observed values of y , the ratio R^2 and RMSE are explained as:

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

4. RESULT

4.1 Correlation results between parameters extract from remotely sensed data and aboveground biomass

Figure 2 shows the linear correlation between forest AGB and various parameters derived from satellite images. For Landsat OLI data, the weak correlation was observed in NIR band (R of 0.6) while this correlation showed stronger in other bands (R of 0.73-0.77). The reflectance in Red band showed the best performance for AGB estimation (R of 0.77). Some of vegetation indices (NDVI, RatioNR) were not defined in the model, basically there were other variables in whose linear combination can fulfill their contribution to models. Therefore, nine out of eleven VIs were retained to evaluate the linear correlation with forest biomass. Generally, VIs provided similar results to single spectral bands and had no improvement in correlation with AGB. The RVI poorly correlated to AGB with R of 0.48 although it represents the simple ratio of Red and NIR band which have good relationship to AGB. SAVI correlated well to AGB (R of 0.73), however, its transformed index TSAVI was found no sensitivity to AGB (R of 0.04). Apart from TSAVI and RVI, the remaining of VIs are inversely proportion to AGB in the range of 0.7-0.76. For SAR data, the biomass sensitivity for L-band backscattering was weaker than for optical data due to the presence of saturation,

as shown in Figure 3c. The sigma value of HV (R of 0.7) polarization has a better correlation than HH (R of 0.63).

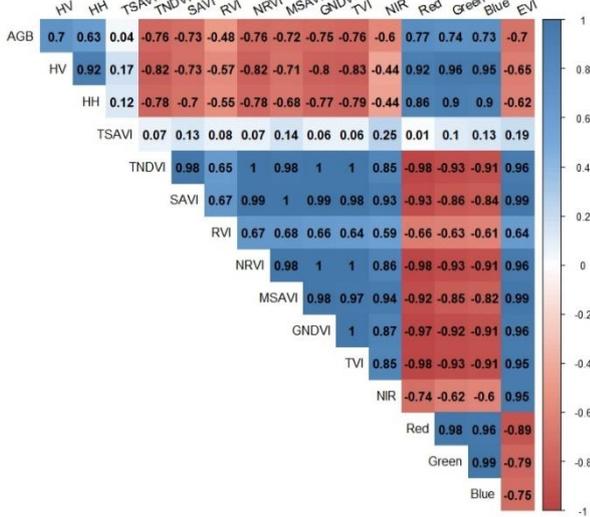


Figure 2. Pearson's correlation coefficient (R) between aboveground biomass and different remotely sensed parameters.

Figure 3 shows the parameters extracted from Landsat data and PALSAR-2 response to AGB in the entire samples. VIs did not show a clear saturation to biomass, but the saturation was considered in the NIR band. For PALSAR-2, the backscattering saturates at around 200 tons ha⁻¹ which is corresponding at -13 dB for HV and -8 dB for HH.

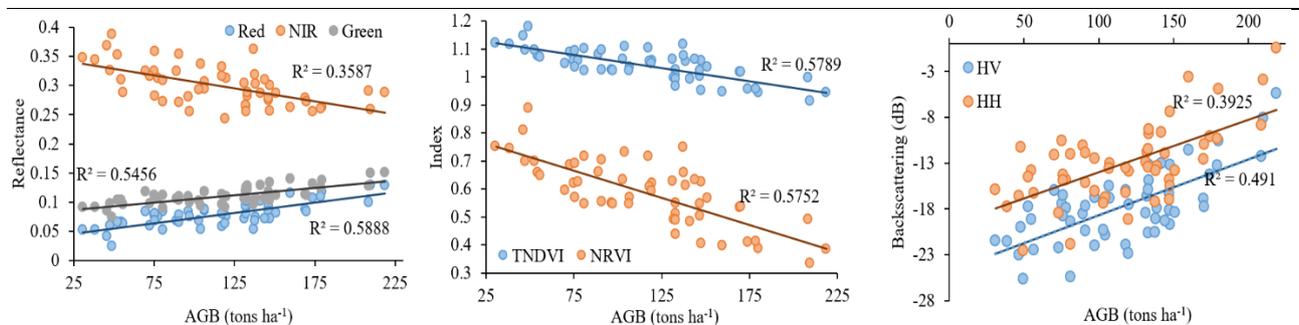


Figure 3. Linear regression between aboveground biomass and Landsat reflectance in Green, Red, and NIR (left panel); Vegetation indices in TNDVI and NRVI (middle panel); and PALSAR-2 backscattering in HH and HV (right panel).

4.2 Selection of the best-fit model for AGB estimation

In an attempt to improve accuracy for AGB estimation, multiple linear regression was developed by various sets of variables. The selection of variable sets and the best-fit model was conducted based on the Bayesian model averaging algorithm. Consequently, the best-fit models were chosen for AGB estimation based on the high value of the coefficient of determination, Bayesian information criterion (BIC), and posterior model probability (Table 2). Comparatively, multiple linear models are slightly better than univariate models with improving R² values from 0.58 to 0.60. This table shows that the increasing number of variables in models performed higher R², however it led to increase value of BIC and therefore, increase the risk of overfitting.

Table 3 provides a comparative analysis of models using only SAR data, only Landsat data and the combination of SAR and Landsat based on R² and RMSE. In the table, each cell has two values, the first is the evaluation using the entire data, and the second is the Leave-one-out cross-validation value. It indicates that the combination of multisource data considerably improved compared to the models using only PALSAR-2 data. The R² value increased from 0.49/0.44 for the models using HH and HV backscattering to 0.60/0.56 for the combined data. The error showed a decreasing in the multisource with the difference of 4 tons ha⁻¹ compared to the models using PALSAR-2. However, the combination of SAR with Landsat (NIR band) or vegetation index (EVI) has no significant improvement in comparison with the model using the single Landsat (Red band). Besides, the comparison between two approaches using the entire data set and LOOCV showed no significant difference with R² of 0.59 and 0.56 (p<0.05), respectively. The biomass RMSE values also showed nearly similar in two approaches with 29.66 tons ha⁻¹ for the entire data and 30.06 tons ha⁻¹ for LOOCV.

Table 2. The result of selecting the best fit model using Bayesian model averaging.

	p!=0	EV	SD	model 1	model 2	model 3	model 4	model 5	model 6	model 7
Intercept	100	3.84E+02	438.49	-11.73	680.23	757.24	289.46	292.50	372.61	288.09
HV	16.5	6.10E-01	1.70	4.88	6.27	4.10
TNDVI	19.7	-1.37E+02	478.18	.	.	-616.53
SAVI	9.7	-2.55E+01	260.39	-263.55
NRVI	16.7	-4.36E+01	190.26	.	.	.	-296.00	.	.	.
TVI	21.1	-1.07E+02	296.70	.	-447.71
NIR	9.9	1.55E+01	172.70	-486.61	.
Red	28.2	4.54E+02	804.65	1635.97
EVI	8.2	-2.74E+00	31.54	-123.15	.	.
nVar				1.00	1.00	1.00	1.00	2.00	2.00	2.00
r2				0.59	0.58	0.58	0.58	0.60	0.60	0.60
BIC				-44.00	-42.98	-42.72	-42.24	-41.21	-41.09	-41.04
post prob				0.09	0.05	0.05	0.04	0.02	0.02	0.02

Table 3. The correlation of determination (R^2) and Root mean squared error (RMSE) in biomass estimation models using the entire data and Leave-one-out cross-validation (LOOCV).

Models	R_squared	RMSE
Red	0.59/0.56	29.66/30.06
HV+EVI	0.60/0.56	29.62/30.27
HV+NIR	0.60/0.55	29.65/30.32
HV+HH	0.49/0.44	33.27/34.01

4.3 Aboveground biomass prediction and mapping

Figure 4 illustrates the relationship between actual AGB and predicted AGB using LOOCV in some linear regression models. The distribution of AGB values in different models are quite similar. Analyzing the density distribution of AGB showed that overestimation occurs in the AGB range from 75 to 130 tons ha^{-1} , however, the remaining showed an underestimation. For the predicted AGB using LOOCV, a high density is considered in values from 85-150 tons ha^{-1} , and the highest density concentrates in the range of 100-125 tons ha^{-1} . For the actual AGB, its density distributes more equally and mainly concentrates between 50-170 tons ha^{-1} , with the highest density in 125-150 tons ha^{-1} .

Based on the different sets of variables in regression models, the AGB was predicted for the Bamboo and mixed bamboo forest in the study site. Figure 5 showed the predicted AGB maps which focus on a small area in this site using the Red band, and the combination of HV with NIR band and EVI. For all three models, RMSE of biomass showed similar values with around 29.6 tons ha^{-1} for the entire data and 30 tons ha^{-1} for LOOCV. However, the AGB maps showed the differences in the range of AGB values. The model using Red band has the highest value with 262 tons ha^{-1} . The appearance of negative biomass values was in all three models with the lowest in the map using HV and EVI with -87 tons ha^{-1} . The proportion of negative AGB varied in different models, particularly it accounts for 3.92% for the model using Red band, 0.37% for the model of HV and EVI, and 2.67% for the model of HV and NIR.

In summary, the linear regression produces negative AGB because of the extrapolation of very low values of parameters extracted from remotely sensed data. Using the combination of HV and EVI or NIR enables to decrease the portion of negative values, however, as analyzed above, the increasing number of variables can increase the risk of overfitting. The model using the Red band has the best performance for AGB estimation in the

Bamboo and mixed bamboo forest with the good regression evaluation in R^2 and RMSE, but it still has the presence of negative AGB with the proportion of 3.92%. Therefore, other regression models, such as non-linear or non-parametric models should be studied in the future with the expectation of finding a sufficient model for AGB estimation in the Bamboo and mixed bamboo forest.

5. DISCUSSION

This study focused on evaluating the performance of Landsat OLI and PALSAR-2 on above-ground biomass estimation. The linear regression was selected to train the ground data and was developed in the bamboo and mixed bamboo forest.

The results indicated that Landsat OLI has a slightly better relationship to AGB than PALSAR-2 for the bamboo and mixed bamboo forest. The reason is that the sensitivity of L-band backscattering to AGB had a trend of decreasing after the saturation point for the dense canopy forest (Mermoz et al., 2015; Suzuki et al., 2013), while the reflectance of Red band did not show a clear saturation to biomass and was sensitive to higher AGB. Visually, the saturation threshold in bamboo was found higher than other ecosystems with 100 tons ha^{-1} for the subtropical area (Häme et al., 2013), or 150 tons ha^{-1} for dense rainforest and savanna (Mermoz et al., 2015). This discrepancy confirms that saturation threshold widely fluctuates in different types of forest, therefore, it has different impacts on biomass estimation accuracy. The sensitivity of satellite imagery data to biomass should be examined separately to develop regression models for improving the prediction accuracy.

Another result was found that VIs also is potential for AGB estimation, except for the simple ratio of Red and NIR. However, the combination of VIs in the multivariable model was not effective because they have highly correlation (Foody et al., 2001) and the contribution of a VI can be substituted by another one (for example, NDVI can be replaced by EVI in this case study). The application of VIs for biomass calculations needs to be careful, although they are considered a solution in reducing saturation in some studies (Gasparri, Parmuchi, Bono, Karszenbaum, & Montenegro, 2010; Zhu & Liu, 2015), others have demonstrated the inefficiency in applying VIs for biomass calculations (Sarker & Nichol, 2011). An example indicated that the sensitivity of NDVI to variation in land surface properties varies in space and time (Foody et al., 2003).

Previous studies pointed out the improvement of the combination of optical and SAR data for biomass estimation compared to single data sources (Häme et al., 2013; Hoan, Tateishi, Bayan, Ngigi, & Lan, 2011; Shao & Zhang, 2016; Zhao, Lu, Wang, Liu, et al., 2016).

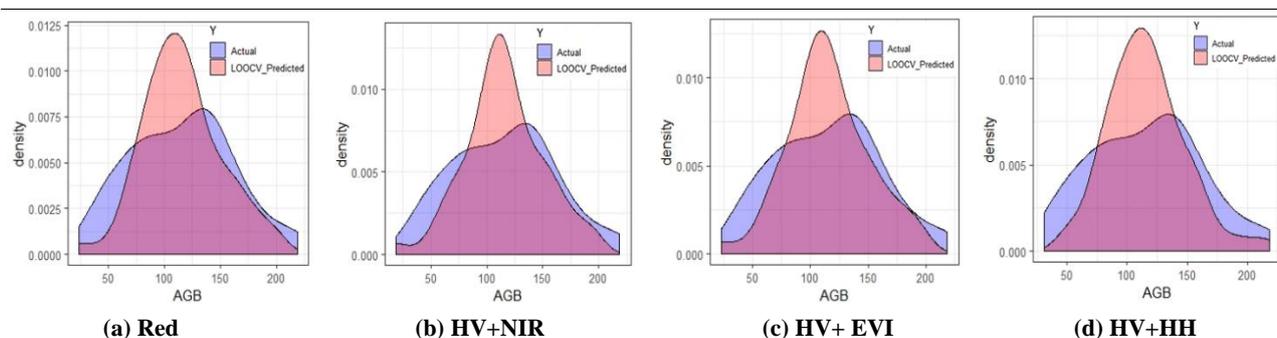


Figure 4. Comparison between the actual aboveground biomass and the predicted aboveground biomass using Leave-one-out cross-validation in different models (a) Red, (b) HV+NIR, (c) HV+ EVI, and (d) HV+HH.

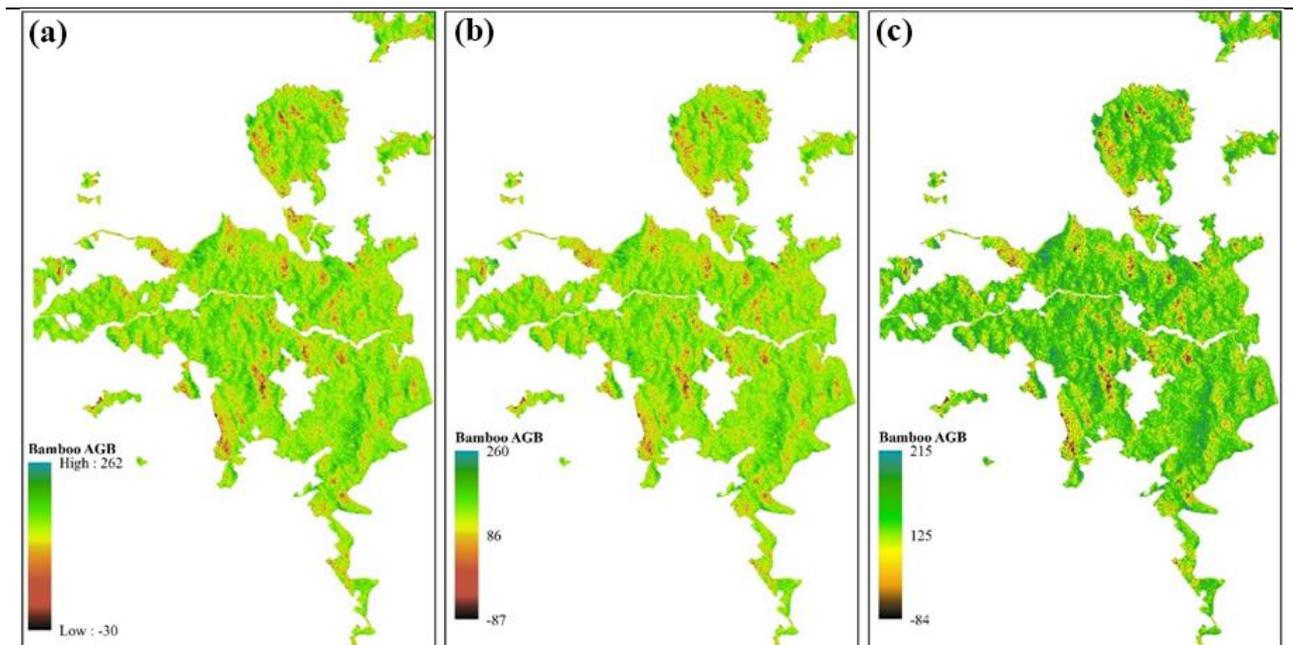


Figure 5. The predicted biomass using different models with sets of variables: (a) Red, (b) HV+EV1, and (c) HV+NIR

On the contrary, this study indicated the combination of multisource remotely sensed data has a better performance compared to SAR data, but this is no significant improvement compared to using the single optical band. The reason is maybe the fusion technique of Landsat and SAR data cannot effectively incorporate radar information into the newly fused data (Zhao, Lu, Wang, Liu, et al., 2016).

Another finding is the effort of using multivariate regression did not provide an improvement in model accuracy as expected. The increasing number of variables showed slightly higher R^2 values than univariate models, but they also increased the risk of overfitting with the gain of BIC. The reason might be increasing number of variables is likely to lead to overfitting with the limited number of training samples (54 samples). More study is needed to assess the possibility of multivariate regression with larger samples to enhance the estimation accuracy of forest biomass.

Finally, the Landsat Red band provided the best correlation to AGB in linear regression with R^2 of 0.59 and RMSE of 29.66 tons ha^{-1} . However, the problem of linear regression is that the extrapolation of very low values (Neumann et al., 2012) in spectral signaled to the presence of negative values. The integration of PALSAR-2 and Landsat or VIs reduced the portion of negative values, but it widens the range of negative values with the minimum value from -30 tons ha^{-1} to -87 tons ha^{-1} . Furthermore, the overestimation and underestimation in the predicted AGB indicated that the linear regression may be an insufficient approach to perform the AGB estimation. However, the proportion of negative values is small and the Red band has an important potential for AGB estimation in the bamboo and mixed bamboo forest.

6. CONCLUSION

This study examined AGB estimation in bamboo and mixed bamboo forest in the mountainous district of Vietnam using PALSAR L-band and Landsat 8 OLI. The linear regression was used to train the entire data set, and then compare the trained model with the Leave-one-out cross-validation model.

In general, the result showed that the simple linear regression using the Red band gave the best performance for AGB estimation with R^2 of 0.59 and RMSE of 29.66 tons ha^{-1} . The biomass sensitivity for optical bands is better than L-band SAR data due to the presence of the saturation phenomenon. Furthermore, the combination of optical band and SAR data did not indicate better performance than a single spectral of Red band. Besides, the attempt using multivariate regression was not successful because the increasing number of variables can lead to overfitting caused by the constraints of ground samples. Despite this limitation, this study provided the evaluation of behavior of remotely sensed data on biomass sensitivity and performed the potential of using PALSAR-2, Landsat OLI, as well as vegetation indices on AGB estimation. This study contributed to fill the gap in our understanding of biomass inventory for the bamboo and mixed bamboo forest. More studies should be further analyzed by adding more samples with other approaches, such as nonlinear and nonparametric models.

REFERENCES

- Anderson, G. L., & Hanson, J. D. (1993). 1-S2.0-0034425793900405-Main.Pdf, 175(165), 165–175.
- Antropov, O., Rauste, Y., Häme, T., & Praks, J. (2017). Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. *Remote Sensing*, 9(10), 1–24. <https://doi.org/10.3390/rs9100999>
- Avitabile, V., Baccini, A., Friedl, M. A., & Schmullius, C. (2012). Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sensing of Environment*, 117, 366–380. <https://doi.org/10.1016/j.rse.2011.10.012>
- Badreldin, N., & Sanchez-Azofeifa, A. (2015). Estimating forest biomass dynamics by integrating multi-temporal Landsat satellite images with ground and airborne LiDAR data in the Coal Valley Mine, Alberta, Canada. *Remote Sensing*, 7(3), 2832–2849. <https://doi.org/10.3390/rs70302832>
- Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of*

- Environment*, 35(2–3), 161–173.
- Basuki, T. M., Skidmore, A. K., Hussin, Y. A., & van Duren, I. (2013). Estimating tropical forest biomass more accurately by integrating ALOS PALSAR and Landsat-7 ETM+ data. *International Journal of Remote Sensing*, 34(13), 4871–4888. <https://doi.org/10.1080/01431161.2013.777486>
- Carreiras, J. M. B., Vasconcelos, M. J., & Lucas, R. M. (2012). Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sensing of Environment*, 121, 426–442. <https://doi.org/10.1016/j.rse.2012.02.012>
- CEOS. (2018). *Table of Contents. A Layman's Interpretation Guide to L-band and C-band Synthetic Aperture Radar data* (Vol. 28).
- Chowdhury, T. A., Thiel, C., Schmullius, C., & Stelmaszczyk-Górska, M. (2013). Polarimetric parameters for growing stock volume estimation using ALOS PALSAR L-Band data over siberian forests. *Remote Sensing*, 5(11), 5725–5756. <https://doi.org/10.3390/rs5115725>
- Chrysafis, I., Mallinis, G., Gitas, I., & Tsakiri-Strati, M. (2017). Estimating Mediterranean forest parameters using multi seasonal Landsat 8 OLI imagery and an ensemble learning method. *Remote Sensing of Environment*, 199, 154–166. <https://doi.org/10.1016/j.rse.2017.07.018>
- Cutler, M. E. J., Boyd, D. S., Foody, G. M., & Vetrivel, A. (2012). Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 70, 66–77. <https://doi.org/10.1016/j.isprsjprs.2012.03.011>
- EORC, J. (n.d.). ALOS-2 Project- PALSAR-2. Retrieved October 1, 2019, from <https://www.eorc.jaxa.jp/ALOS-2/en/about/palsar2.htm>
- Fayad, I., Baghdadi, N., Guitet, S., Bailly, J. S., Hérault, B., Gond, V., ... Tong Minh, D. H. (2016). Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data. *International Journal of Applied Earth Observation and Geoinformation*, 52, 502–514. <https://doi.org/10.1016/j.jag.2016.07.015>
- Foody, G. M., Boyd, D. S., & Cutler, M. E. J. (2003). Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment*, 85(4), 463–474. [https://doi.org/10.1016/S0034-4257\(03\)00039-7](https://doi.org/10.1016/S0034-4257(03)00039-7)
- Foody, G. M., Cutler, M. E., Mcmorrow, J., Pelz, D., Tangki, H., Boyd, D. S., & Douglas, I. (2001). Mapping the Biomass of Bornean Tropical Rain Forest from Remotely Sensed Data Published by: Blackwell Publishing Stable URL: <http://www.jstor.org/stable/2665383>. *Global Ecology & Biogeography*, 10(4), 379–387.
- Gasparri, N. I., Parmuchi, M. G., Bono, J., Karszenbaum, H., & Montenegro, C. L. (2010). Assessing multi-temporal Landsat 7 ETM+ images for estimating above-ground biomass in subtropical dry forests of Argentina. *Journal of Arid Environments*, 74(10), 1262–1270. <https://doi.org/10.1016/j.jaridenv.2010.04.007>
- Hall, R. J., Skakun, R. S., Arsenault, E. J., & Case, B. S. (2006). Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management*, 225(1–3), 378–390. <https://doi.org/10.1016/j.foreco.2006.01.014>
- Häme, T., Rauste, Y., Antropov, O., Ahola, H. A., & Kilpi, J. (2013). Improved mapping of tropical forests with optical and sar imagery, part II: Above ground biomass estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(1), 92–101. <https://doi.org/10.1109/JSTARS.2013.2241020>
- Ho Tong Minh, D., Le Toan, T., Rocca, F., Tebaldini, S., Villard, L., Réjou-Méchain, M., ... Chave, J. (2016). SAR tomography for the retrieval of forest biomass and height: Cross-validation at two tropical forest sites in French Guiana. *Remote Sensing of Environment*, 175, 138–147. <https://doi.org/10.1016/j.rse.2015.12.037>
- Hoan, N. T., Tateishi, R., Bayan, A., Ngigi, T. G., & Lan, M. (2011). Improving tropical forest mapping using combination of optical and microwave data of ALOS. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 34(1), 736–739. <https://doi.org/10.1109/IGARSS.2011.6049235>
- Huete, A. R., Liu, H. Q., Batchily, K., & J. L. van W. (1997). A comparison of vegetation indices over a Global set of TM images for EO -MODIS. *Remote Sensing of Environment*, 59(Table 1), 440–451. [https://doi.org/10.1016/S0034-4257\(96\)00112-5](https://doi.org/10.1016/S0034-4257(96)00112-5)
- Jackson, R. D., & Huete, A. R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine*, 11(3–4), 185–200. [https://doi.org/10.1016/S0167-5877\(05\)80004-2](https://doi.org/10.1016/S0167-5877(05)80004-2)
- Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., ... Saah, D. (2012). Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *International Journal of Forestry Research*, 2012(1), 1–16. <https://doi.org/10.1155/2012/436537>
- Lucas, R. M., Mitchell, A. L., & Armston, J. (2015). Measurement of forest above-ground biomass using active and passive remote sensing at large (Subnational to global) scales. *Current Forestry Reports*, 1(3), 162–177. <https://doi.org/10.1007/s40725-015-0021-9>
- Ly, P., Pillot, D., Lamballe, P., & de Neergaard, A. (2012). Evaluation of bamboo as an alternative cropping strategy in the northern central upland of Vietnam: Above-ground carbon fixing capacity, accumulation of soil organic carbon, and socio-economic aspects. *Agriculture, Ecosystems and Environment*, 149, 80–90. <https://doi.org/10.1016/j.agee.2011.12.013>
- Mahyoub, S., Fadil, A., Mansour, E. M., Rhinane, H., & Al-Nahmi, F. (2019). Fusing of optical and synthetic aperture radar (SAR) remote sensing data: A systematic literature review (SLR). *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4/W12), 127–138. <https://doi.org/10.5194/isprs-archives-XLII-4-W12-127-2019>
- Mermoz. (2014). Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sensing of Environment*, 155(May), 109–119. <https://doi.org/10.1016/j.rse.2014.01.029>
- Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., & Gourlet-Fleury, S. (2015). Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sensing of Environment*, 159(January), 307–317. <https://doi.org/10.1016/j.rse.2014.12.019>
- Neumann, M., Saatchi, S. S., Ulander, L. M. H., & Fransson, J. E. S. (2012). Assessing Performance of L- and P-Band Polarimetric Interferometric SAR Data in Estimating Boreal Forest Above-Ground Biomass. *IEEE Transactions on*

- Geoscience and Remote Sensing*, 50(3), 714–726.
<https://doi.org/10.1109/TGRS.2011.2176133>
- Phua, M. H., Johari, S. A., Wong, O. C., Ioki, K., Mahali, M., Nilus, R., ... Hashim, M. (2017). Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests. *Forest Ecology and Management*, 406(October), 163–171.
<https://doi.org/10.1016/j.foreco.2017.10.007>
- Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., & Ohmann, J. L. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, 114(5), 1053–1068. <https://doi.org/10.1016/j.rse.2009.12.018>
- 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. <https://doi.org/10.1016/j.jag.2011.12.013>
- Raftery, A. A., Hoeting, J., Volinsky, C., Painter, I., Yeung, K. Y., & Sevcikova, M. H. (2018). Package ‘BMA.’
- Sarker, L. R., & Nichol, J. E. (2011). Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sensing of Environment*, 115(4), 968–977.
<https://doi.org/10.1016/j.rse.2010.11.010>
- Shao, Z., & Zhang, L. (2016). Estimating forest aboveground biomass by combining optical and SAR data: A case study in genhe, inner Mongolia, China. *Sensors (Switzerland)*, 16(6).
<https://doi.org/10.3390/s16060834>
- Suzuki, R., Kim, Y., & Ishii, R. (2013). Sensitivity of the backscatter intensity of ALOS/PALSAR to the above-ground biomass and other biophysical parameters of boreal forest in Alaska. *Polar Science*, 7(2), 100–112.
<https://doi.org/10.1016/j.polar.2013.03.001>
- Tebaldini, S., & Rocca, F. (2012). Multibaseline polarimetric SAR tomography of a boreal forest at P- and L-Bands. *IEEE Transactions on Geoscience and Remote Sensing*, 50(1), 232–246. <https://doi.org/10.1109/TGRS.2011.2159614>
- Thiel, C., & Schmullius, C. (2016). The potential of ALOS PALSAR backscatter and InSAR coherence for forest growing stock volume estimation in Central Siberia. *Remote Sensing of Environment*, 173, 258–273.
<https://doi.org/10.1016/j.rse.2015.10.030>
- Tian, X., Li, Z., Chen, E., Yan, M., Han, Z., & Liu, Q. (2017). Modeling of forest above-ground biomass dynamics using multi-source data and incorporated models: A case study over the Qilian mountains. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017-July(June), 5770–5773.
<https://doi.org/10.1109/IGARSS.2017.8128319>
- Toan, T. Le, Beaudoin, A., Riom, J., & Guyon, D. (1992). Relating Forest Biomass to SAR Data.pdf. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 403–411.
<https://doi.org/10.1109/36.134089>
- Torgo, A. L. (2015). DMwR package in R.
- Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, 48(9), 2839–2846.
<https://doi.org/10.1016/j.patcog.2015.03.009>
- Yuen, J. Q., Fung, T., & Ziegler, A. D. (2017). Carbon stocks in bamboo ecosystems worldwide: Estimates and uncertainties. *Forest Ecology and Management*, 393, 113–138.
<https://doi.org/10.1016/j.foreco.2017.01.017>
- Zhang, R., Zhou, X., Ouyang, Z., Avitabile, V., Qi, J., Chen, J., & Giannico, V. (2019). Estimating aboveground biomass in subtropical forests of China by integrating multisource remote sensing and ground data. *Remote Sensing of Environment*, 232(May 2018), 111341.
<https://doi.org/10.1016/j.rse.2019.111341>
- Zhao, P., Lu, D., Wang, G., Liu, L., Li, D., Zhu, J., & Yu, S. (2016). Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. *International Journal of Applied Earth Observation and Geoinformation*, 53, 1–15.
<https://doi.org/10.1016/j.jag.2016.08.007>
- Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., & Yu, S. (2016). Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sensing*, 8(6).
<https://doi.org/10.3390/rs8060469>
- Zhu, X., & Liu, D. (2015). Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102, 222–231. <https://doi.org/10.1016/j.isprs.2014.08.014>