Original Research

Combination of SAR Polarimetric Parameters for Estimating Tropical Forest Aboveground Biomass

**Truong Thi Cat Tuong[1, 2], Hiroshi Tani[3], Xiufeng Wang[3],
Nguyen Quang Thang[4], Ha Manh Bui[5, 6]***

[1]Mientrung Institute for Scientific Research, Vietnam Academy of Science and Technology,
Hue City, Thua Thien Hue Province, Vietnam
[2]Graduate School of Agriculture, Hokkaido University, Sapporo, Japan
[3]Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan
[4]Central Sub Forest Inventory and Planning Institute, Thua Thien Hue Province, Vietnam
[5]Institute of Research and Development, Duy Tan University, Da Nang City, Vietnam
[6]Department of Environmental Sciences, Saigon University, Ho Chi Minh City, Vietnam

Abstract

There is a demand for better information on forest biomass in tropical regions for use in carbon accounting. This needs robust above-ground biomass (AGB) estimation in different forest types. Our study sought to improve biomass estimation by selecting the best regression models based on observations of the contribution of radar signals to AGB in five forest types in Vietnam. Data from PALSAR and PALSAR-2, which covered the forest area, were used to extract 16 polarimetric radar (PolSAR) parameters in 2007 and 2016. This study was designed as a comparative experiment of four regression models: linear, polynomial, support vector machine (SVR) and random forest. First, the contribution of PolSAR data to AGB estimation was evaluated using two approaches: the sample data from all forest types, and the five individual forest types (rich, medium, poor, restoration and bamboo forest). Second, we examined the improvement of AGB prediction by selecting the important variables and assessing the best models for different forest types. The results showed an improvement in the value of R-squared and RMSE using the five individual forest types compared to the combined forest types. In particular, using a multivariate model, RMSE values were enhanced by 9-18% for the rich forest, and by 80-85% for the remaining forest types in all models. SVM provided the best performance for medium and poor forest (RMSE of 8.27 tons ha[-1] and 12.38 tons ha[-1], respectively), random forest for bamboo (RMSE of 23.18 tons ha[-1]), and the polynomial regression for the restoration forest (RMSE

---

*e-mail: manhhakg@sgu.edu.vn

of 10.11 tons ha⁻¹). Further research is required to derive a more robust AGB estimation model for the rich forest.

Introduction

Forest planning plays an important role in sustainable management for forest and land resources, and it provides protection for areas identified as significant for conservation [1]. In forest planning, it is important to observe and predict the change of forest land, as well as forest biophysical parameters. The observation of forest changes can be effectively obtained with the advantages of remote sensing techniques. For instance, Landsat multitemporal data was used to assess the land cover changes and CO_2 emissions in the tropical forest [2], or to extract construction land information [3]. In addition, the integration of remote sensing and a geographical information system (GIS) provides a useful tool to develop a forest plan. A recent study used GIS to evaluate the relationship between natural resources and human activities (social, economic, and cultural characteristics) on the forest land in order to develop a successful forest plan [4].

Biomass is one of the most important forest biophysical parameters that can be successfully estimated by various methods using remotely sensed data. Knowledge about aboveground forest biomass is of fundamental importance in quantifying the terrestrial carbon cycle [5]. In a tropical forest, it accounts for two-thirds of all terrestrial biomass [6] but it is frequently affected by human activities and climate change, making it essential to observe these regions. In recent years, research has focused on biomass research for tropical forest areas to supplement existing knowledge. Numerous studies have been conducted to improve the formula for estimating above-ground biomass (AGB) from the correlation with factors such as tree height, stem diameter and wood specific gravity (WSG) [7-9]. The development of methods based on remote sensing technology has also enhanced the ability of people to estimate biomass on a global scale. However, this method is only demonstrated as highly accurate for boreal and temperate forests, with low accuracy in the tropics, where there is decreased sensitivity of backscattering because of high AGB.

In 2005, the European Space Agency proposed the BIOMASS mission to apply PolInSAR technology to estimate biomass through forest height using P-band synthetic aperture radar (SAR). The European Space Agency's initial studies showed that P-band was well suited for global biomass estimation with a larger range of AGB values. However, pending the launch of the BIOMASS satellite in 2020, SAR L-band remains the most effective tool [10-11]. In some instance (e.g., low density of forest canopy), L band has proved more sensitive to forest AGB than P band. It is well-known that the penetration of the L band is shorter than the P band. Therefore, the L band signal is largely dominated by volume scattering reflected by tree trunks, branches, and canopy; it is less affected by the double bounce and surface mechanism contributed from the ground. A study assessing the performance of L band and P band PolInSAR data in estimating boreal forest AGB [12] proved that L band provided the best correlation to the forest height, which is used for estimating AGB. Although PolInSAR is superior, PolSAR was chosen for this study because the quad-pol L band image sources needed to generate the full functions of PolInSAR products were not available in the study site.

Before 1990, Vietnam was one of the countries with the least degraded forest from a potential carbon density, with degradation ratios of 0.6-0.8 [13]. From 1990 to 2010, recorded deforestation was 3,900 ha per year. According to the FAO, in Vietnam there were large variations in natural forest change in the period of 1990-2015. For example, the annual change rate in the primary forest was −5.9% with a peak in the period 2000-2010 of −7.8%, but there was no change in the period 2010-2015 [14]. During 1993-2003, statistical data showed that natural forests increased in areas managed for protection/regeneration. However, the cover of other natural forests under the management of parastatal forestry organizations tended to decrease or remain static, especially more recently when the organizations increasingly turned to multi-purpose plantation forestry [15]. Such variations resulted in a reduction in the carbon uptake by forests. In recent years, several studies have examined natural forest biomass in Vietnam. These were carried out at national and regional scales, mainly based on inventory measurement, and exploited the correlation between AGB and stand structure parameters to improve the allometric equation [16-20]. Remote sensing is also commonly used to estimate biomass carbon stocks and disturbance assessment. The combination of NDVI in optical bandwidths with SAR textures and polarization can be processed by regression analysis in order to map tropical forest biomass [21]. In addition, forest disturbances and regrowth can be extracted through HH and HV backscattering from the L-band ALOS PALSAR mosaic [22].

SAR is one of the most promising remote sensors for mapping forest carbon [10]. In the past decade, the increase in SAR techniques has resulted in improved remote observation of vegetation such as classification, forestry structure, and biomass. The SAR techniques are known as polarimetry SAR (PolSAR), and polarimetric interferometric SAR (PolInSAR).

PolInSAR has superior performance because it combines the advantages of PolSAR and InSAR and it is an important technique for determining the height of reflectors [23, 24]. In recent years, a new technique – SAR tomography – has been proposed for forest height detection, which has enhanced the accuracy of biomass prediction, especially in tropical forest areas [11, 25].

PolSAR is a well-known technique for biomass estimation based on polarimetric intensities and phases. Combinations of polarimetric parameters such as the coherency matrix and decomposition products have resulted in a significant improvement in estimations. The scattering components obtained from polarimetric decomposition are used to estimate AGB [26-27]. Some indices such as the biomass index, canopy structure, and volume scattering index (which can be retrieved from SAR data), are also expected to show a linear regression with AGB [28]. The PolSAR approach only performs well for AGB, which is less than 200-300 tons ha^{-1} because of saturation in signal intensity [11, 27, 29]. However, the benefit of PolSAR is that it does not require a pair of images with a high coherence in the acquired time, geometry and other radar parameters. Although PolInSAR performs better than PolSAR in vegetation measurement, the lack of image sources in some instances causes difficulties in meeting the requirement for coherence in a pair of images. Therefore, the PolSAR approach is still being using in parallel with PolInSAR for monitoring and measuring vegetation.

Many studies have indicated the power of using PolSAR for predicting biomass. In this paper, we focused on investigating the behavior of SAR backscattering over multiple forest types using the PolSAR data with a case study in Thua Thien Hue Province, Vietnam. We proposed an approach for estimating natural forest AGB by observing the level of woody volume and the species composition. In this study area, the natural forest can be divided into two categories based on the species composition: ropical evergreen broad-leaved forestsropical evergreen broad-leaved forests and bamboo forests. Tropical evergreen broad-leaved forests can be divided into four levels of woody volume, including rich, medium, poor and restoration forest. The aim of this study was to show the contribution of polarimetry to biomass estimation following the individual forest types, which in previous assessments were not recorded or were incomplete. The study also selected a suitable function for AGB estimation in different forest types based on the comparison of various regression models. Furthermore, a preliminary assessment of the effectiveness of using PolSAR data for bamboo biomass, which was not documented in previous studies, has been carried out and forms a foundation for further study in various bamboo species.

Materials and Methods

Study Area

The study area is situated in Thua Thien Hue Province in central Vietnam, extending from 16°44'30" N to 15°59'30" S. The climate is a humid subtropical climate. The monsoon climate also influences changes in tropical humidity with an average rainfall of 1500-2000 mm and humidity of around 80%.
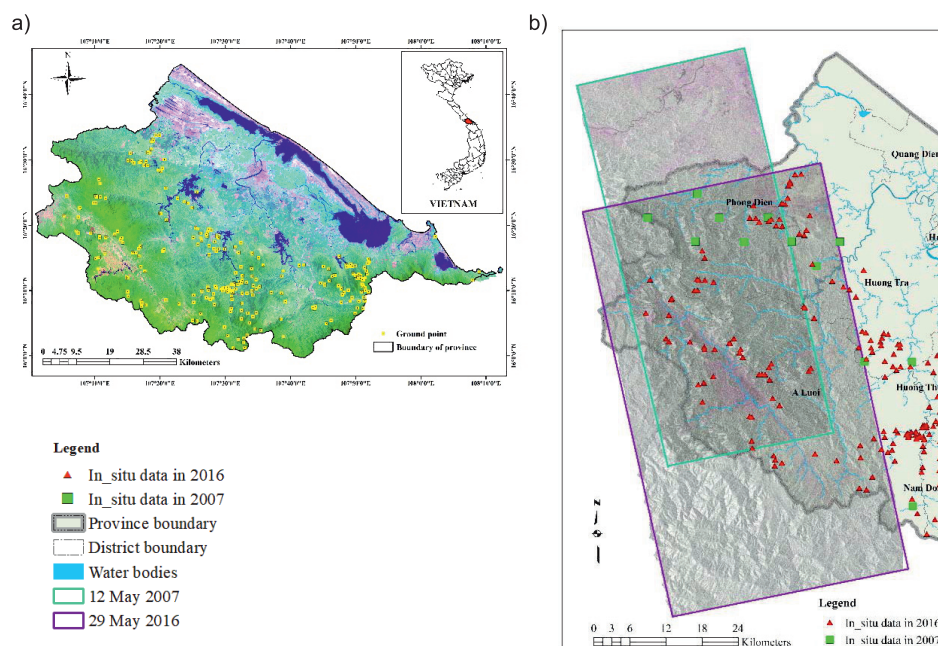

Fig. 1. Location map of the study area in Landsat TM in pseudo-colors band 7,5,2 a), and the cover of SAR images and *in-situ* data in 2007 and 2016 b).

In an effort to reduce climate change and improve livelihoods for people living near forests, the Vietnamese government has implemented programs on emission reduction and carbon service payment. Thua Thien Hue is one of the important pilots of these programs. However, the serious shortage of information on forest carbon and forest emission reductions is a challenge and is essential to be additional.

The total area of the province is 5033 km^2, including 2026 km^2 of natural forest. The natural vegetation mainly consists of lowland tropical forest with characteristic species including the dominant distribution of *Rhodomyrtus tomentisa, Melastoma candidum,* and *Imperata cylindrica.* The natural forest links to the Annamite range provide high forest continuity and integrity. The protected forest in Thua Thien Hue Province includes Bach Ma National Park and Phong Dien Nature Reserve.

According to Circular Number 34/2009/TT-BNNPTNT of June 10, 2009 [30] published by the Vietnam Ministry of Agriculture and Rural Development on criteria for forest identification and classification in Vietnam, as well as based on the specific conditions of the study site, we classified the natural forest here into five types, including:

– Rich forests (RF), which are forests with a timber reserve of standing trees of between 201 and 300 m^3/ha.
– Medium forests (MF) (or average forest), which are forests which have a timber reserve of standing trees of between 101 and 200 m^3/ha.
– Poor forests (PF), which are forests with a reserve of standing trees of between 10 and 100 m^3/ha.
– Restoration forest (ReF) (forests with no reserve in the case of our study site), which are forests having timber trees with an average diameter of less than 8 cm and a timber reserve of standing trees of less than 10 m^3/ha.
– Bamboo forests (BAM): in our study area, these are a mixture of bamboo and some woody species.

The collection and classification of samples were conducted by staff in the Central Sub Forest Inventory and Planning Institute, Thua Thien Hue Province, Vietnam. These data showed that poor forests often had timber reserves of between 50 and 100 m^3/ha. All forests with a reserve lower than 50 m^3/ha were then assigned as restoration forests. We called this forest type "restoration" instead of "no reserve forest" because it still contained some wood reserves.

SAR and Data Processing

A scene of the PALSAR on 12 May 2007 and one PALSAR-2 image on 29 May 2016 were used. They covered a part in a forest area of Thua Thien Hue Province, Vietnam. The phased array type L-band synthetic aperture radar (PALSAR) is an active microwave sensor using L-band frequency. The benefit compared with optical images is that it achieves cloud-free and day-and-night land observation. PALSAR was launched in 2006 by a joint project between JAXA and the Japan Resources Observation System Organization. In 2014, they launched ALOS-2/PALSAR-2 which was based on ALOS/PALSAR, but with a higher resolution, wide swath width, and better image quality. The data comprised full-polarized images (HH, HV, VH and VV polarizations) in a single look complex (SLC) format. For the year 2007, the data has the incidence angle at the image center of 23.89° with a size of 3.55 × 9.3 m in azimuth and range, respectively. The 2016 scene has the incidence angle at the image center of 38.99° and the azimuth and range spacing was 3.12 × 4.57 m, respectively.

The data were then converted into a 3 × 3 coherency matrix to describe the polarimetric behavior of a target. A refined Lee filter was used with a window size of 7 × 7 to reduce the speckle noise. The topography effect was eliminated by using Range-Doppler Terrain Correction with a digital elevation model (DEM) from the Shuttle Radar Topography Mission, and all the product images were resampled to reach 8 m in pixel spacing (1 and 2 looks in azimuth and range). Polarization techniques were used to produce parameters such as span, pedestal height, radar forest degradation index (RFDI), canopy structure index (CSI), volume scattering index (VSI), and biomass index (BMI). The ratio among HH, HV, and VV was also calculated. The four-component Yamaguchi parameter was applied to decompose the backscattered power. The pixel values were derived using the mean value of 9 pixels (3 × 3 pixels).

A coherency matrix is a representation of the product of a Pauli basis with its transposition of the complex conjugate as $\mathbf{T} = K_p \times K_p^{*\mathrm{T}}$, where * represents the conjugate and denotes the Hermitian transpose [31]. The simple form of the coherency matrix is:

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \tag{2}$$

… where T11, T22, and T33 are $S_{HH}+S_{VV}$, $S_{HH}-S_{HV}$, and S_{HV}, respectively, called by the three real diagonal elements; others are the three real and three imaginary parts of the three complex off-diagonal elements. These matrix elements are sensitive to the size, dielectric constant, and orientation of the main scatter in the medium [12]. $S_{HH}+S_{VV}$ is related to single bounce scattering on a rough surface and $S_{HH}-S_{VV}$ is an indicator of double bounce, while S_{HV} is an indicator of volume scattering.

In this study, the general Yamaguchi four components decomposition with unitary transformation model (G4U) was used [32, 33]. The Yamaguchi method is a target decomposition technique to decompose a scattering matrix into the volume, double bounce, surface and helix scattering. The approach included helical scattering as a fourth component based on

the coherency matrix to deal with the problem of reflection asymmetry. In 2013, Singh et al. [34] developed the generalized four-component decomposition by a unitary transformation of the coherency matrix (G4U). The decomposition employs an extended volume scattering model, which discriminates volume scattering between dipole and dihedral scattering structures caused by the cross-polarized HV component.

Various intensity ratios and indices were computed from the coherency matrix element to predict AGB. The ratios of HH, HV, and VV such as co-pol HH/VV ratio, cross-pol HH/HV ratio, and VV/VH ratio were extracted based on the polarization measurement. In this study, some polarimetric parameters were computed from the coherency matrix element to predict AGB such as span, pedestal height, and various biophysical indices such as RFDI, CSI, VSI, and BMI. These indices are popularly applied for vegetation classification but have scarcely been examined in forest AGB estimation. The formulas to calculate these indices are shown below.

Span is the percentage of the total power and can be calculated based on the span of the coherence and covariance matrix, given as Span = trace ([**T3**]) [35].

The pedestal height of a polarization signature plot is the lowest Z-axis value in the polarization signature. It indicates polarization purity and is related to depolarization. Different types of scattering show different values of pedestal height, so it can be used for classifying forested and deforested areas. Forested areas often display a larger pedestal height value than deforested areas [36].

RFDI was developed by Saatchi and partner to assess the strength of the double-bounce term [37]:

$$\text{RFDI} = \frac{\langle |S_{hh}|^2 \rangle - \langle |S_{hv}|^2 \rangle}{\langle |S_{hh}|^2 \rangle + \langle |S_{hv}|^2 \rangle} \qquad (3)$$

CSI measures the relative importance of vertical versus horizontal structure in the vegetation, and is related to the vertical trunks or stems [38]:

$$\text{CSI} = \frac{\langle |S_{vv}|^2 \rangle}{\langle |S_{hh}|^2 \rangle + \langle |S_{vv}|^2 \rangle} \qquad (4)$$

VSI is a measure of the depolarization of the linearly polarized incident radar signal and is an indicator of canopy thickness or density [38]:

$$\text{VSI} = \frac{\langle |S_{hv}|^2 \rangle}{\langle |S_{hv}|^2 \rangle + BMI} \qquad (5)$$

BMI is an indicator of the relative amount of woody and leafy biomass and is related to the radar wavelength and the size of vegetation components [38]:

$$\text{BMI} = \frac{\langle |S_{hh}|^2 \rangle + \langle |S_{vv}|^2 \rangle}{2} \qquad (6)$$

Ground Data Analysis

The forest ground data were provided by the Central Sub Forest Inventory and Planning Institute, Thua Thien Hue Province, Vietnam (Sub-FIPI). The data were collected over two time periods, January to October in 2007 and 2016. In 2007, 10 measured plots were covered by the PALSAR scene. The size of a plot is 1 km² (1000 × 1000 m). In each plot, 40 subplots of 25 × 20 m were set to measure forest parameters. In 2016, 79 plots were covered by PALSAR-2 data. A sample plot size had a rectangular shape of 30 × 33 m with the longer aspect in an east-west direction and the shorter aspect in the north-south direction. Each plot included four sub-plots 5 × 5 m in size. Diameter at breast height (DBH, in centimeters) was measured for all trees with a diameter over 6 cm, while total tree height (H, in meters) 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. Terrain elevation varied from 10 m to 1400 m and the slope angle was up to 35°. The main ecosystem was dense tropical rainforest with three main levels of vertical stratification.

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.

There are numerous studies on natural forest biomass estimates for the pantropical region and studies for the central coastal region in Vietnam. The biomass estimation formula of Huy et al. [19] based on a single diameter factor shows higher accuracy than the common pantropical formula with three factors. In addition, this formula reduces the error propagation from the height and WSG. DBH was measured directly in the field, thereby reducing the deviation of the sample [8]. The one-factor formula of Huy also showed higher confidence than other local formulas observed in the study area, which were based on the combination of DBH, height, and WSG. The one-factor formula to estimate AGB (in tons ha⁻¹) given by Huy et al. [19] is:

$$\text{AGB} = 0.104189 \text{x } DBH^{2.491453} \qquad (7)$$

For estimating the biomass of bamboo forest, we used the formula given by [39]:

$$\text{AGB} = 0.3002 \, (DBH)^2 + 0.115(DBH) + 1.7632 \quad (8)$$

The uncertainties and bias in the actual AGB calculation were derived from three sub-models: the height, WSG and AGB model. However, we only used the DBH variable in AGB calculation, so errors from height and WSG were eliminated.

These ground data were combined with parameters derived from the SAR image, and the estimated AGB

was evaluated using conventional regression and machine learning algorithms.

Above-Ground Biomass Estimation

For examining the correlation of SAR signals and in-situ data, 479 polygons of forest areas were extracted in the study site. The 2007 polygons were drawn in the size of 25 × 20 m to fit the collected samples. Meanwhile, the 2016 polygons with of 30 × 33 m were derived from the GPS points data. Both PALSAR and PALSAR-2 product data were resampled to 8-m resolution. Therefore, the value of a polygon is approximate to the mean value of 9 pixels that were extracted from SAR data.

For AGB estimation, two approaches were observed based on the parametric and non-parametric methods. The former included linear and polynomial (2 degrees) models. These models are the common conventional way to first explore the correlation of PolSAR data to forest biomass. In addition, in order to improve the biomass estimation we used support vector regression (SVR) and random forest. These are well-known algorithms based on machine learning approaches that have been widely used in many fields.

SVR is related to statistical learning theory based on the kernel method to transform a non-linear regression problem into a linear one in a higher dimension's feature space. Given training samples (x_i, y_i), $(i = 1,..., n)$, where x_i is a multivariate input, y_i is a scalar output, and n is the number of training samples, a linear model can then fit this new high-dimensional feature space as follows [40, 41]:

$$y = f(x) = \langle w.\varphi(x) \rangle + b = \sum_{i=1}^{n} w_i \varphi_i(x) + b \tag{9}$$

…where w is the weight vector and b is the bias term. φ denotes a nonlinear mapping function from the input space to the new feature space. Here, instead of determining the explicit form of φ, we used a kernel function as follows:

$$K(x_i, x) = \langle \varphi(x_i).\varphi(x) \rangle \tag{10}$$

Commonly used kernels include linear, polynomial and radial basis function. In this study, a radial basis function kernel was used because of its better performance. This is described as follows using a single parameter γ:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_i\|^2) \tag{11}$$

Next, we need to find w and b by minimizing the regression error. The optimization problem is formulated as follows:

$$\text{minimize} \frac{1}{2} \|W\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$

$$\text{subject to} \begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i^* \\ f(x_i) - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0, i = 1, ..., n \end{cases} \tag{12}$$

…where parameter C determines the tradeoff between the tolerated training error and the model complexity. ξ_i and ξ_i^* are slack variables, which measure the deviation of each training sample point outside the ε-insensitive zone. These sample points are called support vectors, which will be used to develop regression models. Therefore, in this study, SVR is related to find out the parameter C and kernel parameter γ. The SVR model was analyzed using the library of e1071 in the R interface.

The random forest algorithm was provided by the randomForest package in R. In a random forest, n_{tree} bootstrap samples are drawn from the original data. In each bootstrap sample, each node is split using the best among a subset of predictors randomly chosen at that node. New data are predicted by aggregating the predictions of the n_{tree} trees [42]. Random forest for regression is based on two parameters, including the number of trees and the value of the parameter at each node. In this study, 16 variables were used to investigate the behavior of SAR backscattering in different types of natural forest. Therefore, it was necessary to limit the number and complexity of variables that were calculated in a model. A random forest algorithm was performed to select variables that improved the model with a high correlation coefficient and reduced the (RMSE) value. This operation was executed in R with the package randomForestExplainer [43] to help explain which variables were most important in models. The selection depended on various measures of importance for all variables such as the minimal depth of variables, p-value, node purity increase and increasing mean squared error (MSE).

The depth of the decision tree is the length of the longest path from the root (the first variable) to the leaf (a classification result, after several nodes). Tree depth represents the complexity of the model structure. Thus, in a random forest model with 500 trees, the distribution of the min_depth value for each variable can provide information about its presence frequency and the complexity of the rule that it participates in. The higher the min_depth, the more complex the rule. The variables were selected based on the importance of the variable with a p-value <0.05. P-value is the value (1 side) of binomial accreditation based on the binomial distribution. The H_0 hypothesis of this test assumed that a variable joined a node only through randomness. When the significance threshold is set for p = 0.05, if the p-value of the variable is less than 0.05, it plays an important role in the estimation equation. The MSE increase provides the mean increase of the MSE after that variable is permuted. Node purity increase gives

Table 1. Biophysical forest parameters in different forest types.

Type	No of trees/ plot	Height (m)			DBH (cm)			Basal area (m² ha⁻¹)	Stem Volume (m³ ha⁻¹)	AGB (tons ha⁻¹)
		h_{Lorey}	h_{min}	h_{max}	D_g	d_{min}	d_{max}			
RF	75	17.97	15.11	30.92	23.03	6	147.1	32.84	269.85	241.68
MF	53	15.43	13.08	18.86	19.53	6	109.8	20.94	148.88	139.17
PF	44	14.09	11.41	16.66	16.57	6	64.6	11.47	73.99	68.41
ReF	29	11.97	8.71	14.59	12.24	6	80.9	5.42	30.48	26.87
BAM	33	15.36	11.00	17.50	19.85	6	57.3	9.705	67.325	119.15

the mean node purity increase by splits on the variable. Hence, these measures indirectly show the importance of the variable in the model.

To assess the accuracy of regression, an R-squared (R^2) and RMSE were calculated. The ratio of R^2 is called the coefficient of determination, and it 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:

$$R^2 = \frac{\sum(\hat{y}-\bar{y})2}{\sum(y-\bar{y})2} \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{14}$$

A k-fold cross-validation method was used to compare and select a model for the AGB predictive problem. Cross-validation is the most commonly used method for predictive performance evaluation of a model, given beforehand or when it is developed by a modeling procedure [44]. The value of k is the number of groups that will be split from a given sample data. The higher the value of k, the higher the accuracy in cross-validation [44]. In general, the selection of k is in the range of 5-10. In this study, 10-fold cross-validation was used to split data, in which the regression function was estimated using nine groups for training the model and one group for testing.

Results and Discussion

Above-Ground Biomass in the Field Data

Table 1 and Fig. 2 illustrate the forest structure parameters in the five forest types. Lorey's height was used to calculate the average height of plots per forest type. The average DBH per plot is shown as a D_g value. Structural parameters showed a gradual decrease depending on the tree volume levels from the rich to the poor-volume stock forest. The RF had the highest AGB value with the variation in the range of 200-500 tons ha⁻¹. In BAM forest, these plot parameters were performed only for wood trees, but AGB was calculated by combining trees and bamboo. Therefore, although BAM had the lowest number of trees, its AGB was still higher than PF and ReF because a large amount of bamboo contributed to a significant biomass stock.

In this study, BAM forest represented a mixed species composition between trees and bamboo, mainly distributed on foothills at an altitude of less than 300 m. This forest type had a two-storey structure with a tree storey and a bamboo storey. In the tree storey, tree density of D1.3≥6 cm was 33 trees/plot (1000 m²). Bamboo density was quite high at 116 individuals/ plot with species such as *Dendrocalamus barbatus*, *Bambusa blumeana*, and *Melocalamus compartiflorus.*

The diameter and stem volume are the important parameters strongly related to the AGB measurement. Therefore, the correlation coefficients among them were examined (Table 2). As expected, AGB had a strong relationship with DBH and stem volume for all forest types. The correlation coefficient was between 0.77 and 0.93 for AGB and DBH, and around 0.99 for AGB and stem volume.

Correlation Between Polarimetric Backscattering and Forest Above-Ground Biomass

The correlation between the actual AGB in the ground data and SAR polarimetric parameters was investigated. In this study, we first measured the natural forest with the combined forests sample data. In the second approach, we divided the forest into five types: RF, MF, PF, ReF, and BAM. The interactions of AGB in each type with the same set of parameters were compared with each other and with the result of the first approach. This correlation was initially surveyed through the simple linear model.

The R^2 showed significant differences between the two approaches: the total samples and separate forest types. In the former, all SAR backscatter coefficients had a very low correlation with AGB. However, in the latter, the correlation noticeably improved in some instances, especially in the coherency matrix elements (for ReF and BAM), co- and cross-pol signal.

In the total samples approach, the correlation between the *in-situ* AGB and PolSAR parameters was
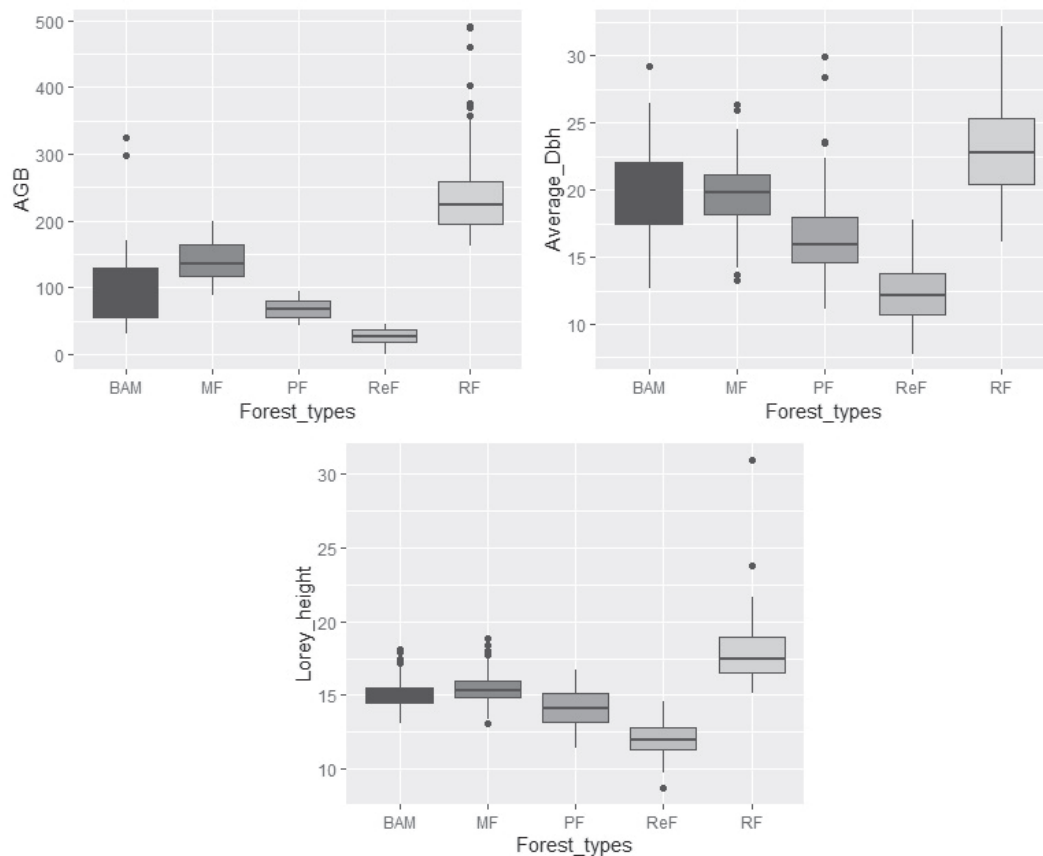
Fig 2. Boxplots of average above-ground biomass (AGB), diameter at breast height (DBH), and Lorey's height (H_{Lorey}) of plots in the five forest types.

very low, with R^2 under 0.0187. The co- and cross-pol ratios did not show a high correlation, as expected. This result indicated that the PolSAR parameters had a low sensitivity to in-situ AGB for the dense natural forest in general. Data points showed a dispersion and did not have a specific relationship. The main reason was caused by the heterogeneous forest, which has a variety of forest types with complex forest structures and mixed tree species. These led to the strongly scattered correlation diagrams, which could not express the correlation between PolSAR data and AGB. Based on the result observed in, it can be seen that AGB in each forest type has a different correlation to PolSAR data. For example, the coherency matrixes T22 and T33, volume scattering or BMI were found to be sensitive to AGB for BAM, while double-bounce scattering and CSI were more sensitive for the ReF. Therefore,

in this study, forests were divided into different forest types with high uniformity in the data structure. We recommend a variety of forest classifications according to species composition and stem volume because of the high correlation with forest biomass, so that forest types have similar characteristics in terms of structure and species composition.

In general, the correlation between AGB and PolSAR signals was improved for the forest types with low tree volume. Among the coherency matrix elements, T22 and T33 improved more than T11 for each forest type because of the relationship with the illuminated medium type such as forest trees. For decomposition compositions, volume scattering showed the best correlation as expected in forest areas. The ratio of co- and cross-polarization did not show a clear improvement, while CSI, BMI, and span indicators were more sensitive than other indicators to the biomass.

In summary, based on the analysis of a mono-linear regression on each forest type, there was an improvement in the correlation between PolSAR and the biomass. Each forest type appeared to be sensitive to some specific indicators. In the hope of improving accuracy for forest AGB estimation, in the next step we explored the association with a multivariate combination based on both linear and non-linear models. Comparison between models was used to find the most suitable

Table 2. Correlation coefficient (R) between in-situ above-ground biomass (AGB), diameter at breast height (DBH) and stem volume in the five forest types.

Type	AGB_{RF}	AGB_{MF}	AGB_{PF}	AGB_{ReF}	AGB_{BAM}
DBH	0.77	0.84	0.90	0.83	0.93
Stem volume	0.99	0.99	0.99	0.99	0.99

Table 3. Set of selected parameters in different forest types.

Forest types	Set of selected variables
Total	Y_volume, CSI, HHVV
Rich	Y_surface, HHVV, VVVH
Medium	VSI, Y_volume, T11
Poor	Y_volume, T22, Span
Restoration	CSI, HHVV, BMI
BAM	Y_volume, T11, T33

model for each forest type. Furthermore, the use of 16 parameters is time-consuming and has a computational cost. This problem will be addressed in the next section using the random forest model for appropriate variable selection.

Selection of Important Parameters

A random forest algorithm was used to assess the importance and correlation of parameters in the regression models. The parameter selection depended on various measures of importance for all variables such as the minimal depth of variables, p-value, node purity increase and increasing mean squared error (MSE).

For analyzing the total data sample, Yamaguchi volume scattering, CSI and HHVV ratio proved to be important. However, when separating data into five types of forest, each different type had a different set of parameters that correlated best with AGB. Polarimetric

indices such as Y_volume, HHVV ratio, CSI and T11 had a high frequency of presence in biomass estimation through their involvement in models for forest types.

Multivariate Regression Analysis

Based on the set of selected predictor variables, we built a multivariate regression of polarimetric parameters and AGB. The regression was based on the parametric models, including linear and polynomial models, and the non-parametric models including SVR and random forest. A 10-fold cross-validation was used for testing the robustness of the model.

Fig. 3 presents the radar graph of (R^2) using training and 10-fold cross-validation in different models. Models using the total sample data had a low value of R^2 in both training and validation results of between 0.067 and 0.29 (excluding the random forest models). The results improved when observing the five forest types separately. BAM forest showed good regression between its AGB and dependent variables with $R^2 > 0.47$ in all models. The R^2 values were between 0.18 and 0.90 for the MF, PF, and ReF. For the RF, unsurprisingly, the regression did not show any significant enhancement compared to the results of total samples data. This matched the result of previous studies, which explained that it is difficult to observe the relationship between radar backscattering and AGB after 100-150 tons ha^{-1} in tropical forest area because it leads to over- or under-estimation [10, 29].

In addition, based on the low R^2 value of the linear model compared to others, we found that AGB has a
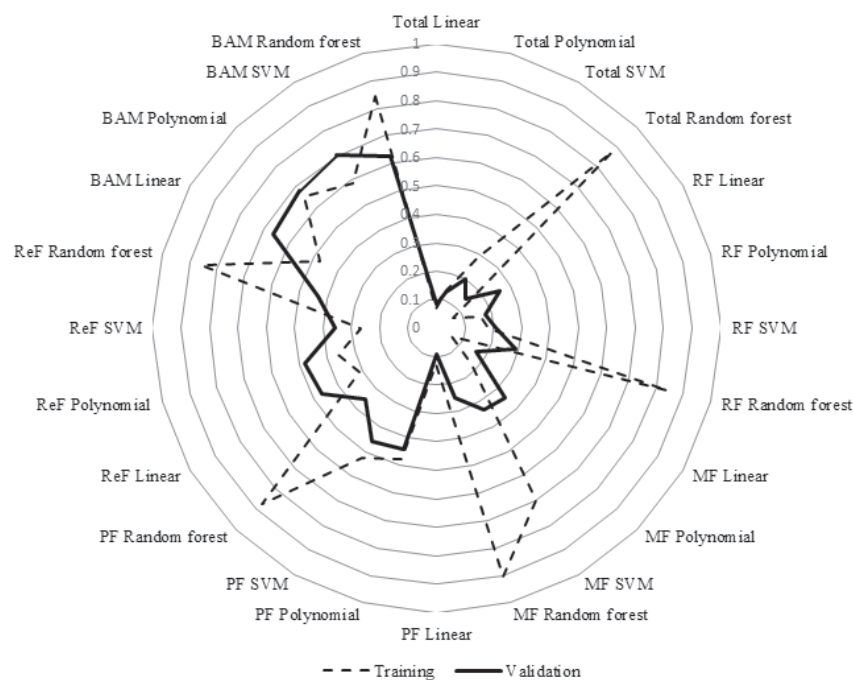

Fig. 3. Result of regression models in the coefficient of determination (R^2) using training data and 10-fold cross-validation following forest types.

Table 4. Root mean squared error (RMSE) of regression models using training and 10-fold cross-validation for different forest types.

Types	Linear	Polynomial	SVM	Random forest
Total	78.18/ 77.29	75.61/ 75.33	69.03/ 74.26	34.36/ 75.78
RF	63.79/ 64.18	62.53/ 68.24	64.39/ 61.28	31.62/ 62.40
MF	13.97/ 14.28	13.44/ 14.81	8.27/ 13.73	6.17/ 12.98
PF	17.13/ 17.50	13.57/ 14.06	12.38/ 14.58	7.04/ 15.24
ReF	9.911/ 10.38	10.11/ 15.76	10.43/ 10.40	5.03/ 10.14
BAM	44.04/ 84.25	41.09/2729.98	35.49/ 47.02	23.18/ 44.08

positive non-linear correlation with PolSAR signals. Random forest showed good performance in most of the observation. However, its robustness was lower than others when observing the difference between training and validation results. The significant difference in random forest models, especially for high biomass data, suggests it is not reliable for use for AGB measurement (except for BAM). Generally, parametric models showed stronger robustness than non-parametric models.

The RMSE of regression models (Table 4) showed that multivariate regression considerably increased the accuracy in all forest types. The RMSE was around 70 tons ha^{-1} for the total sample approach. For the RF, the RMSE improved up to 62.40 tons ha^{-1} and up to 10-13 tons ha^{-1} in random forest models for the MF, PF, and ReF. Generally, with observations in particular forest types, the enhancement of RMSE in AGB measurement rose by 9-18% for RF, and 80-85% for the remaining forest types.

The boxplot graph in Fig. 4 illustrates the variation of RMSE in 10-fold cross-validation using different models with the total sample and five forest types. The RMSE showed a considerable variation in all models for the RF with the range of 33.52-115.68 tons ha^{-1}.

A similar pattern was also noticed in the linear model for BAM. The highest variance was in the polynomial model for BAM of between 43.34-4016.66 tons ha^{-1}, but we could not illustrate this in this figure. Conversely, a low variance of RMSE was shown for the remaining forest types.

In conclusion, following the analysis of R^2 and RMSE, different models were suitable for different forest types to measure AGB. The random forest models could be chosen to estimate AGB for BAM forest because they obtained the best R^2 and RMSE with the value of 0.85 and 23.18 tons ha^{-1}, respectively. For the remaining forest types, random forest was not reliable for use based on the comparison between training and validation results. For the ReF, the polynomial model displayed good correlation between AGB and with PolSAR variables with R^2 of 0.47 and RMSE improved up to 10.11 tons ha^{-1}. The SVR should be selected for MF and PF with RMSE between 8.27 and 12.38 tons ha^{-1}. For the RF, the random forest exhibited a high correlation in training data with RMSE decrease to 31.62 tons ha^{-1}. However, the overfitting, which occurred between training and validation, needs to be solved to ensure the reliability of the model. To solve
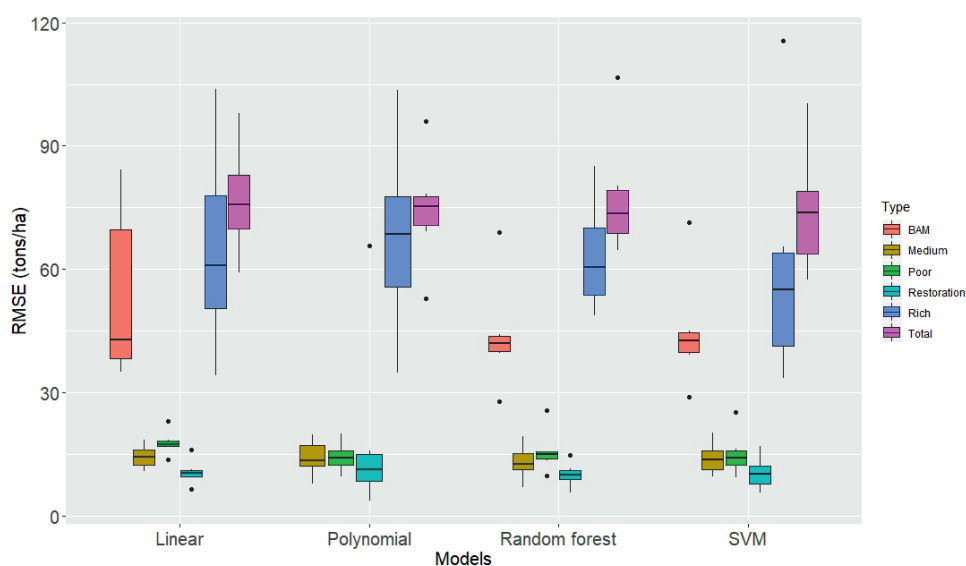

Fig. 4. Root mean squared error (RMSE) of regression models in tons/ha using 10-fold cross-validation for different forest types.

this problem, further study should be carried out only focusing on RF. RF has a highly complex structure caused by the combination of the great diversity of species composition and many forest storeys. Because of a dense canopy, the L-band backscattering only reaches to the canopy and the higher storey, leading to poor correlation between backscattering signals and lower canopy structure. Although the P-band radar or Lidar signals data are supposed to solve this limitation of L-band, they currently are not available for application in large areas because of the cost and accessibility of the data source. It is recommended that we divide the RF following vertical stratification and observe the sensitivity of backscattering to the various storeys. This requires an increase in data samples for the RF, which was beyond the scope of this study.

Conclusions

In this study, we determined the models of AGB measurement in different forest types based on correlation with PolSAR parameters extracted by L-band in the PALSAR-2 image. Significant variables were selected for each forest type to improve biomass estimation accuracy. Regression models included linear, polynomial, SVR and random forest. The data were analyzed in two ways, including the total sample data and data for different forest types. We also analyzed the correlation of biomass to individual PolSAR signals in specific forest types with *in situ* biomass to find the best model for each forest type.

Forests were categorized into five different types. Based on natural forest species, it was divided into broad-leaved forest and bamboo forest (mix of broad-leaved species and bamboo species). Based on volume levels, the broad-leaved forest was divided into rich, medium, poor, and restoration. *In-situ* AGB values were up to 500 tons ha^{-1}. To improve the accuracy of the biomass estimation, we divided the data and assessed the correlation using the multivariate equations in different forest types.

The correlation analysis of individual PolSAR parameters to forest biomass revealed different sensitivities in different forest types. Therefore, the different sets of parameters were proposed for use for particular forest types. The result also indicated that the volume scattering composition (Y_vol) and the coherency elements mainly contributed to the AGB estimation function.

There was a significant improvement in the correlation coefficients of the model. The RMSE values were enhanced by 9-18% for the rich forest, and 80-85% for the remaining forest types when compared to the total sample. Among all models, SVR and random forest regression exhibited better performance in R^2 and RMSE in training data. SVR provided the best result for MF and PF with RMSE 8.27 tons ha^{-1} and 12.38 tons ha^{-1}, respectively. Although random

forest displayed a good R^2, it was only chosen to calculate AGB for BAM because it had low reliability with the problem of underfitting with high values in both training and validation data for the remaining forest types. The polynomial function proved to be a suitable model for estimating biomass for the ReF with R^2 of 0.47 and RMSE of 10.11 tons ha^{-1}. In this study, the low correlation of AGB with PolSAR parameters in the RF type was because of the reduction of radar signals in high biomass areas. The noise in the ground data caused by heterogeneous samples also affected the correlation results. The method has been mentioned by observing the sensitivity of backscattering to various storeys following vertical stratification.

In general, the results showed that multivariate analysis combined with the selection of significant variables provided a satisfactory result in biomass estimation in different forest types (except RF). This study also confirmed the effectiveness and role of PolSAR data in calculating above-ground biomass in tropical forests. Furthermore, the strong correlation of PolSAR signals to the bamboo forest was exhibited in this study. This verified its potential for estimating AGB in bamboo biomass using SAR data, which was not documented in previous studies, as well as creating a foundation for further study in various bamboo species.

Conflict of Interest

The authors declare no conflict of interest.

References

1. CETIN, M. (2019). Using GIS analysis to assess urban green space in terms of accessibility: case study in Kutahya. Int. J. Sust. Dev. World **22** (5), 420, **2015**.
2. ASKAR A., NUTHAMMACHOT N., SAYEKTININGSIH T., HERMUDANANTO H. Assessing Land Cover Changes and CO_2 Emissions in Tropical Forests, 1998-2016: A Case Study of the Sungai Wain Protection Forest. Polish J. Environ. Stud. **28** (5), 3597, **2019**.
3. LIU H., ZHANG Y., ZHANG X. Using the RS method to analyse construction land changes in tongren during 2002 and 2016. Polish J. Environ. Stud. **28** (3), 1277, **2019**.
4. CETIN M., ZEREN I., SEVIK H., CAKIR C., AKPINAR H. A study on the determination of the natural park's sustainable tourism potential. Environ. Monit. Assess. **190** (3), **2018**.
5. TOAN T.L., QUEGAN S., DAVIDSON M.W.J., BALZTER H., PAILLOU P., PAPATHANASSIOU K., PLUMMER, S., ROCCA F., SAATCHI S., SHUGART H., ULANDER, L. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. Remote Sens. Environ. **115** (11), 2850, **2011**.
6. PAN Y., BIRDSEY R.A., PHILLIPS O.L., JACKSON R. B. The Structure, Distribution, and Biomass of the World's Forests. Annu. Rev. Ecol. Evol. Syst. **44** (1), 593, **2013**.
7. BROWN S., LUGO A. E. Biomass of tropical forests: a new estimate based on forest volumes. Science **223** (4642), 1290, **1984**.

8. MOLTO Q., ROSSI V., BLANC L. Error propagation in biomass estimation in tropical forests. Methods Ecol. Evol. **4** (2), 175, **2013**.

9. CHAV, J., RÉJOU-MÉCHAIN M., BÚRQUEZ A., CHIDUMAYO E., COLGAN M.S., DELITTI W.B., DUQUE A., EID T., FEARNSIDE P.M., GOODMAN R.C. AND HENRY M., Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Chang. Biol. **20** (10), 3177, **2014**.

10. MERMOZ S., RÉJOU-MÉCHAIN,M., VILLARD L., LE TOAN T., ROSSI V., GOURLET-FLEURY S. Decrease of L-band SAR backscatter with biomass of dense forests. Remote Sens. Environ. **159** (January), 307, **2015**.

11. MINH D.H.T., TOAN T.L., ROCCA F., TEBALDINI S., VILLARD L., RÉJOU-MÉCHAIN M., PHILLIPS O.L., FELDPAUSCH T.R., DUBOIS-FERNANDEZ P., SCIPAL K., CHAVE J. SAR tomography for the retrieval of forest biomass and height: Cross-validation at two tropical forest sites in French Guiana. Remote Sens. Environ. **175**, 138, **2016**.

12. NEUMANN M., SAATCHI S.S., ULANDER L.M.H., FRANSSON J.E.S. Assessing Performance of L- and P-Band Polarimetric Interferometric SAR Data in Estimating Boreal Forest Above-Ground Biomass. IEEE Trans. Geosci. Remote Sens. **50** (3), 714, **2012**.

13. BROWN S., IVERSON L.R., PRASAD A., LIU D. Geographical distributions of carbon in biomass and soils of tropical asian forests. Geocarto International **8** (4), 45, **1993**.

14. Food and Agriculture Organization of the United Nations. Global Forest Resources Assessment 2015: Desk Reference, **2015**. doi:10.1002/2014GB005021.

15. COCHARD R., NGO D. T., WAEBER P.O., KULL C.A. Extent and causes of forest cover changes in Vietnam's provinces 1993-2013: A review and analysis of official data. Environ. Rev. **25** (2), 199, **2017**.

16. GIBBS H.K., BROWN S., NILES J.O., FOLEY J.A. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environ. Res. Lett. **2** (2007), 045023, **2007**.

17. PHUONG V.T., INOGUCHI A., BIRIGAZZI L., HENRY M., SOLA G. Tree allometric equation development for estimation of forest above-ground biomass in Viet Nam Part A. UN-REDD Programme, Hanoi, Viet Nam., **2012**.

18. HUY B., KRALICEK K., POUDEL K.P., PHUONG V.T., VAN KHOA P., HUNG N.D., TEMESGEN H. Allometric equations for estimating tree aboveground biomass in evergreen broadleaf forests of Viet Nam. For. Ecol. Manage. **382** (December 2016), 193, **2016**.

19. HUY B., POUDEL K.P., TEMESGEN H. Aboveground biomass equations for evergreen broadleaf forests in South Central Coastal ecoregion of Viet Nam: Selection of eco-regional or pantropical models. For. Ecol. Manage. **376**, 276, **2016**.

20. KRALICEK K., HUY B., POUDEL K.P., TEMESGEN H., SALAS C. Simultaneous estimation of above- and below-ground biomass in tropical forests of Viet Nam. For. Ecol. Manage. **390** 147, **2017**.

21. VIET L.N., TATEISHI R., KONDOH A., SHARMA R., THANH H.N., TRONG T.T., MINH D.H.T. Mapping Tropical Forest Biomass by Combining ALOS-2, Landsat 8, and Field Plots Data. Land **5** (4), 31, **2016**.

22. MERMOZ S., LE TOAN T. Forest disturbances and regrowth assessment using ALOS PALSAR data from 2007 to 2010 in Vietnam, Cambodia and Lao PDR. Remote Sens. **8** (3), 1, **2016**.

23. SAUER S., FERRO-FAMIL L., REIGBER A., POTTIER E., MEMBER S. Polarimetric Dual-Baseline InSAR Building Height Estimation at L-Band. Geosci. Remote Sens. Lett. **6** (3), 408, **2009**.

24. NEUMANN M., NEUMANN M., MEMBER S., FERRO-FAMIL L. Estimation of Forest Structure , Ground , and Canopy Layer Characteristics From Multibaseline Polarimetric Interferometric SAR Data. Trans. Geosci. Remote Sens. **48** (3), 1086, **2010**.

25. REIGBER A., MOREIRA A. First Demonstration of Airborne SAR Tomography Using Multibaseline L-Band Data. Trans. Geosci. Remote Sens. **38** (5), 2142, **2000**.

26. CHOWDHURY T.A., THIEL C., SCHMULLIUS C., STELMASZCZUK-GÓRSKA M. Polarimetric parameters for growing stock volume estimation using ALOS PALSAR L-Band data over siberian forests. Remote Sens. **5** (11), 5725, **2013**.

27. SAI BHARADWAJ P., KUMAR S., KUSHWAHA S.P.S., BIJKER W. Polarimetric scattering model for estimation of above ground biomass of multilayer vegetation using ALOS-PALSAR quad-pol data. Phys. Chem. Earth **83-84** 187, **2015**.

28. SANTOS J.R.D., GAMA F.F., BISPO P.C. Estimating forest biomass by remote sensing radar data in brazil. Drewno **57** (192), 120, **2014**.

29. MINH D.H.T., LE TOAN T., ROCCA F., TEBALDINI S., D'ALESSANDRO M.M., VILLARD L. Relating P-band synthetic aperture radar tomography to tropical forest biomass. IEEE Trans. Geosci. Remote Sens. **52** (2), 967, **2014**.

30. MARD. Circular No. 34/2009/TT-BNNPTNT of June 10, 2009, on criteria for forest identification and classification. **2009** (341).

31. BAGAN H., KINOSHITA T., YAMAGATA Y. Combination of AVNIR-2, PALSAR, and Polarimetric Parameters for Land Cover Classification. IEEE Trans. Geosci. Remote Sens. **50** (4), 1318, **2012**.

32. YAMAGUCHI Y., MORIYAMA T., ISHIDO M., YAMADA H. Four-component scattering model for polarimetric SAR image decomposition. IEEE Trans. Geosci. Remote Sens. **43** (8), 1699, **2005**.

33. YAMAGUCHI Y., YAJIMA Y., YAMADA H. A four-component decomposition of POLSAR images based on the coherency matrix. IEEE Geosci. Remote Sens. Lett. **3** (3), 292, **2006**.

34. SINGH G., YAMAGUCHI Y., PARK S.E. General Four-Component Scattering Power Decomposition With Unitary Transformation of Coherency Matrix. IEEE Geosci. Remote Sens. Lett. **51** (5), 3014, **2013**.

35. ANTROPOV O., RAUSTE Y., HÄME T., PRAKS J. Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. Remote Sens. **9** (10), 1, **2017**.

36. DURDEN S.L., VAN ZYL J.J., ZEBKER H.A. Modeling and Observation of the Radar Polarization Signature of Forested Areas. IEEE Trans. Geosci. Remote Sens. **27** (3), 290, **1989**.

37. MITCHARD E.T., SAATCHI S.S., WHITE L., ABERNETHY K., JEFFERY K.J., LEWIS S.L., COLLINS M., LEFSKY M.A., LEAL M.E., WOODHOUSE I.H., MEIR P. Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: Overcoming problems of high biomass and persistent cloud. Biogeosciences **9** (1), 179, **2012**.

38. POPE K.O., REY-BENAYAS J.M., PARIS J.F. Radar remote sensing of forest and wetland ecosystems in the

Central American tropics. Remote Sens. Environ. **48** (2), 205, **1994**.

39. LY P., PILLOT D., LAMBALLE P., DE NEERGAARD A. 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. Agric. Ecosyst. Environ. **149**, 80, **2012**.

40. SMOLA A.J., SCHÖLKOPF B. A Tutorial on Support Vector Regression. Stat. Comput. **14** (3), 199, **2004**.

41. CHEN G., HAY G.J. A Support Vector Regression Approach to Estimate Forest Biophysical Parameters at the Object Level Using Airborne Lidar Transects and QuickBird Data. Photogramm. Eng. Remote Sens. **77** (7), 733, **2011**.

42. LIAW A., WIENER M. Classification and Regression by randomForest. R news **2**, 18, **2002**.

43. PALUSZYNSKA A., BIECEK P. Explaining and Visualizing Random Forests in Terms of Variable Importance (V 0.9). **2017**.

44. YADAV S., SHUKLA S. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. Proc. - 6th Int. Adv. Comput. Conf. IACC 2016 (Cv), 78, **2016**.