ESTIMATION OF NATURAL FOREST BIOMASS USING REMOTE SENSING DATA IN THUA THIEN HUE PROVINCE, VIETNAM

リモートセンシングデータを用いたベトナム､トゥアティエンフエ省における天然林バイオマスの推定

Hokkaido University
Graduate School of Agriculture
Division of Bio-Systems Sustainability
Doctor Course

Truong Thi Cat Tuong

Table of Contents

List of Tables

List of Figures

Acronyms and Abbreviations

AGB	Aboveground biomass
ALOS	The Advanced Land Observing Satellite
BAM	Bamboo forest
BIC	Bayesian information criterion
BMA	Bayesian Model Averaging
BMI	Biomass index
CEOS	Committee on Earth Observation Satellites
CO_2	Carbon dioxide
CSI	Canopy structure index
DBH	Diameter at breast height of tree
DEM	Digital elevation model
DNs	Digital numbers
EORC	The Eastern Ottawa Resource Centre
ESA	The European Space Agency
EVI	The enhanced vegetation index
FAO	The Food and Agriculture Organization
FLAASH	The fast line-of-sight atmospheric analysis of hypercubes
GIS	Geographical inforamtion system
GSO	General Statistics Office
H	Height of tree
H/A/α	Entropy/Anisotropy/alpha
HH polarization	Horizontal transmitting, Horizontal receiving
HV polarization	Horizontal transmitting, Vertical receiving
JAXA	The Japan Aerospace Exploration Agency
Landsat 8 OLI	Landsat 8 Operational Land Imager
Landsat ETM+	Landsat Enhanced Thematic Mapper Plus
Landsat TM	Landsat Thematic Mapper
LOOCV	Leave-one-out cross-validation
MF	Medium forest
MSE	Mean squared error
NASA	National Aeronautics and Space Administration
NIR	Near-infrared

OA	Overall accuracy
PALSAR	The Phased Array type L-band Synthetic Aperture Radar
PF	Poor forest
PolInSAR	The polarimetric interferometry synthetic aperture radar
PolSAR	Polarimetric Synthetic Aperture Radar
r	The correlation coefficient
R^2	The coefficient of determination
ReF	Restoration forest
RF	Rich forest
RFDI	Radar forest degradation index
RMSE	Root mean squared error
SAR	Synthesis aperture radar
SL-KLS	Self learning-Kernel Least Squares
SVM	Support vector machine
SVR	Support vector regression
SWIR	Short-wave infrared
TT-BNNPTNT (MARD)	Circular-Ministry of Agriculture & Rural Development
VH polarization	Vertical transmitting, Horizontal receiving
VIs	Vegetation indices
VSI	Volume scattering index
VV polarization	Vertical transmitting, Vertical receiving
WSG	Wood specific gravity
Y_vol	Yamaguchi volume scattering

Acknowledgments

First of all, I would like to express my deep gratitude to Associate Professor TANI Hiroshi, my research supervisors, for his patient guidance, enthusiastic encouragement and useful critiques of this research work.

My sincere thanks goes to Professor SHIBUYA Masato and Professor SAMESHIMA Ryoji for their insightful reviews and comments.

The advice given by Professor INOUE Takashi and Dr. YAMAMOTO Tadao through seminars has been a great help in building the structure of this research. I would also like to thank Associate Professor WANG Xiufeng, for her advice and assistance in keeping my progress on schedule.

My grateful thanks are also extended to Mr. YAMAYA Yuki for his support in laboratory activities and in adjusting to life in Japan, to all members in the Laboratory of Environmental Informatics for their support and friendship.

Another special thanks go to the scholarship program given by the Ministry of Education, Culture, Sports, Science, and Technology-Japan which provided me a chance to do my research at Hokkaido University. I also wish to acknowledge the help provided by staff members in Graduate school of Agriculture at Hokkaido University.

I am particularly grateful for the assistance given by all colleagues at Mientrung Institute for Scientific Research, who supported favorable conditions for me to study abroad safely.

Finally, I wish to thank my parents for their support and encouragement throughout my studies. My special thanks are extended to my husband and my little daughter for sacrificing their personal life to accompany me in a new environment.

CHAPTER 1. GENERAL INTRODUCTION

1.1. Background

Forests play an essential role in human survival and development. Besides providing habitat for animals and human livelihoods, forests also provide goods (timber and non-timber products) for economic development, watershed protection, prevent soil erosion, conservation of biodiversity, supply social services, and minimize climate change. Forests are special environments that can stabilize atmospheric carbon dioxide (CO_2) through photosynthesis, store CO_2 in organic compounds, and then release CO_2 into the atmosphere through respiration. The amount of CO_2 absorbed by a forest can be indirectly measured by calculating the forest biomass. A natural forest (FAO 2012) is a naturally regenerated forest comprising native species, where there are no clear or clearly visible indications of human activities and the ecological processes are not significantly disturbed. Without disturbance, a natural forest remains relatively stable over a long time, so observed changes in this parameter might prove that the forest is affected by outside factors, including land use transformations. Therefore, biomass changes through time can be used to predict alterations in the net uptake of CO_2 caused by land use change. Studies estimated that the flux of CO_2 released to the atmosphere from land use change, mainly caused by tropical forest deforestation, averaged 1.6 (0.5–2.7) GtC/year during the 1990s (Denman 2007; Houghton *et al*. 2012). Tropical forests account for two-thirds of all terrestrial biomass (Pan *et al*. 2013) but are frequently affected by human activities and climate change, making it essential to observe these regions.

The natural forest has a high level of biodiversity but is frequently affected by human activities, thus it is necessary to observe and manage the characteristics and parameters of natural forest. Therefore, large scale forest inventories should be regularly conducted. In Vietnam, forest inventories occur every 5 years by the conventional approach. Conducting inventory in natural forests is difficult because of the complicated forest structure and accessibility. The conventional approach to the direct measurement of structure parameters provides high accuracy, but it is time-consuming, costly and labor-intensive investigation at large scales. Besides, accessing remote mountainous areas is difficult because of the rugged terrain. The integration of ground measurement and remote sensing has been commonly used because of the cost and time savings; it enables the management of natural resources annually, monthly and even daily.

Remote sensing applications for biomass estimation have been widely applied at both a global and regional scale. Brown *et al.* (1993) estimated the potential carbon densities in Asian tropical forests using databases of climatic, edaphic and geomorphologic indices, and vegetation. The result was recorded with a pixel size of 3.75 km × 3.75 km. A map of biomass carbon density in the tropics indicated that the error increased when the biomass densities were high, with error under 5% for densities under 250 Mg/ha and up to 8% for high densities of 250–400 Mg/ha (Brown *et al*. 1993). In recent years, researches have focused on tropical forest biomass to supplement existing knowledge. Numerous studies have been conducted to improve the formula for estimating aboveground biomass (AGB) from the correlation with factors such as tree height, stem diameter and wood specific gravity (WSG) (Brown *et al.* 1984; Molto *et al.* 2013; Chave *et al*. 2014). The development of methods based on remote sensing technology has also enhanced the ability of people to estimate biomass on a global scale. A major benefit of using remote sensing data is to support calculations of emissions, the establishment of national carbon stock baseline datasets and assessment of change in stocks (Rosenqvist *et al*. 2003).

1.2. Review on the application of remote sensing technique on forest biomass estimation

1.2.1. Optical data

Optical images enable the provision of natural resource observation at both global and local scale from coarse (NOAA-AVHRR, MERIS, MODIS) to medium (Sentinel-2, Landsat) and fine (SPOT, AVNIR, WorldView-2) spatial and temporal resolution. Because of their flexible capability in scale and frequency as well as the diversity of sensors, optical images provide benefits for mapping land use/land cover and change detection. In the context of biomass prediction, optical signals were used to derive a relationship with biomass, which was mainly reflected by the vegetation canopy. Therefore, the reflectance derived from optical sensors has better correlation with aboveground than with belowground biomass of vegetation. Efforts on biomass mapping focused on tropical forests have encountered difficulties in using remotely sensed data. The major causes are local topographic effects (Sader *et al.* 1989), the weather condition, high biomass and the complexity of stratified structure.

Many techniques have been introduced and tested to develop more accurate methods for biomass prediction. Vegetation indices and pattern texture, as extra features extracted from spectral bands, were proposed to evaluate their response to forest biomass. Vegetation indices have been recommended for the removal of variability caused by canopy geometry,

soil background, illumination angles, and atmospheric conditions when estimating biophysical properties such as biomass, leaf area index (LAI) and percent green vegetation cover (Sarker and Nichol 2011). However, the sensitivity of vegetation indices to tropical forest is limited by the dense canopy and saturation with high biomass. The application of texture analysis has become popular for high-resolution images data because of the capability of large additional features based on techniques such as the Grey level co-occurrence matrix (Sarker and Nichol 2011; Dube and Mutanga 2015) or Fourier Transform Textural Ordination (Pargal *et al*. 2017).

Since being launched in the 1970s, Landsat satellites have provided multispectral data that have been commonly used because of the advantage of medium spatial and temporal resolution, and an open satellite imagery source. With the wide range of spectral information, Landsat data have been applied in many areas and especially in sustainable natural resource management. With the release of new product generations, it has constantly improved data quality in terms of spatial resolution, spectral width, and short day repeat. Landsat 8 Operational Land Imager (OLI) quality was improved compared with Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) data with 9 shortwave spectral bands over a 190 km swath, with a 30 m spatial resolution for all bands except a 15 m panchromatic band. The widths of several OLI bands have been refined to avoid atmospheric absorption features within ETM+ bands (Department of the Interior U.S. Geological Survey 2016). With these advantages, many techniques based on Landsat data have been proposed both in classification and biophysical parameters extraction. For forest management, it is well-known that Landsat is an effective source for inventory and change detection using various methodologies. For instance, many supervised methods were proposed and developed such as integrating pixel-based and object-based classifiers (Ceccarelli *et al*. 2013), partial least square to classify Landsat TM (Du H *et al*. 2008), or deep learning, a powerful state-of-the-art technique for landcover (Kussul *et al*. 2017), and an AdaBoost technique to improve classification accuracy with the integration of supervised classifiers (Chen *et al.* 2017). The semi-supervised classification has also received attention for classification as a self-learning approach (Kim *et al.* 2017), and an ensemble of semi-supervised classifiers is proposed for change detection using Landsat 7 (Roy *et al.* 2014).

Landsat has been used extensively to estimate forest biomass stock. For instance, the relationship between continuous estimates of forest structure attributes and Landsat ETM+ was evaluated for modeling forest stand structure attributes (Hall *et al*. 2006); the use of The normalized difference vegetation index (NDVI) seasonal time-series were explored to

estimate AGB (Zhu and Liu 2015); vegetation indices and texture measures were derived for forest AGB estimation in primary tropical rainforest (Phua *et al*. 2017; Dube and Mutanga 2015); and a random forest regression algorithm was used to train data of seasonal time-series Landsat 8 OLI (Chrysafis *et al*. 2017).

1.2.2. Synthesis aperture radar

In contrast to optical sensors, synthesis aperture radar (SAR) is an active remote sensing instrument that can transmit signals itself and receive the echo reflected from the object (Van Zyl and Kim 2011). It is available in all conditions of weather; therefore, it is suitable for areas such as tropical regions where the observation of optical imagery is limited because of heavy rainfall and cloud.

Several radar sensors have been evaluated for forest AGB extraction such as ERS-1 and -2, JERS-1, Envisat ASAR, RADARSAT, and ALOS PALSAR-1 and -2 (Koch 2010; Bwangoy *et al*. 2010; Antropov *et al*. 2017; Cartus and Santoro 2019). Most of the successful studies were documented for boreal and temporal forest. For tropical forest, the sensitivity of radar backscattering varies across forest types because of differences in forest structure and spatial variability (Saatchi *et al*. 2011). This leads to the requirement for a diversity of remote sensing algorithms to estimate AGB for different forest types (Zhao *et al*. 2016; Lucas *et al.* 2015).

Using SAR to estimate biomass is an advantage in a situation that provides highly interpretable imagery of the surface of the Earth even when the ground is obscured by a dense cover of clouds. Successful mapping of AGB has been demonstrated using P- and L-band backscatter data. Whereas P-band data are expected to be sensitive to a larger range of AGB values (Toan *et al*. 2011), only spaceborne L-band data from ALOS PALSAR have been available for studies until now (Mermoz 2014). The relationship between HV (for horizontal transmit and vertical receive) polarization of PALSAR data and AGB is used to derive the function of biomass estimation in a tropical forest. The reason for this particular sensitivity is that while parts of the L-band signal interacts with branches in the crown layer, a significant portion of the long-wavelength signals penetrates through the forest canopy down to the ground, to perform a dihedral reflection on the vertical trunks and the horizontal ground surface (The Japan Institute of Energy 2008). The L-band SAR-2 (PALSAR-2) aboard the DAICHI-2 with its high sensitivity and resolution uses a radio wave of long-wavelength (about 24 cm) that is suitable for observing AGB in tropical areas. Hence, the assessment of signal attenuation in dense tropical forests should be tested using PALSAR-2 data to expand our understanding of radar backscattering behavior.

Many techniques have been developed using SAR data for AGB extraction such as radar backscattering intensity or polarimetry (PolSAR), interferometry (InSAR), polarimetric interferometry (PolInSAR) and tomography (TomoSAR). When it comes to radar backscattering, different polarizations have different sensitivity to forest AGB, thereby, they do not contribute in the same way to the AGB calculation (Carreiras *et al.* 2012). Some studies showed that HV in L-band has better sensitivity than HH (for horizontal transmit and horizontal receive), VH (for vertical transmit and horizontal receive) and VV (for vertical transmit and vertical receive) polarizations (Mitchard *et al*. 2009; Mermoz 2014). Furthermore, many techniques have been introduced to improve the potential of SAR in natural resources management. In the PolSAR technique, the backscattering of a target can be completely described by a 2×2 complex scattering Sinclair matrix (Koeniguer 2014), and decomposed into different scattering components to either enhance or suppress the contribution from a specific scattering mechanism (Van Zyl *et al.* 2011). The interferometry technique analyzes differences in phase information by combining two data observed from almost identical positions of a measurement data satellite in orbit (Koeniguer 2014). PolInSAR is the combination of PolSAR and InSAR, therefore, it overcomes the constraints of temporal decorrelation in InSAR and the inherent high entropy problem in PolSAR (Cloude 2005). These techniques have been commonly applied for AGB prediction (Antropov *et al*. 2017; Thiel and Schmullius 2016; Neumann *et al*. 2012; Hensley *et al*. 2015). In recent years, TomoSAR parameters showed high potential to observe forest stratified structure. TomoSAR allows the conversion of the multi-baseline stack of SAR images into a multilayer stack of SAR images, where each image represents the complex reflectivity associated with a layer at a certain height above the ground (Ho Tong Minh *et al*. 2014). The development of these techniques has enhanced the application of SAR images in global biomass mapping. Despite this, the potential of SAR for AGB prediction should be discussed further to meet the demand for improving accuracy in the estimated function.

1.2.3. The combination of SAR and optical data

Most of the satellite sensors have their own limitations because of the effect of weather conditions, spectral resolution and topographic factors. Generally, saturation is considered a major cause of the sensitivity mitigation of satellite imagery signals to high biomass in tropical forests. Several studies indicated that the variation in saturation value is based on remote sensing data, topography and vegetation of the ecosystem. Zhao *et al*. (2016) discussed the saturation values in Landsat 5 TM for different vegetation types in a subtropical region. An optical sensor cannot penetrate a dense canopy, thereby, it is

insensitive to stem biomass when the forest canopy is closed. Because of this limitation, trees may continuously grow and their biomass continuously increase in volume without changing the spectral signature of the canopy (Zhao *et al*. 2016). This problem of optical data should be solved by SAR data with a long wavelength of L-band or P-band. However, Mermoz *et al*. (2015) examined the behavior of L-band backscatter with biomass of dense forests using both theoretical and experimental approaches. SAR also showed a signal attenuation from the forest canopy as the canopy became denser with increasing biomass.

The use of multi-sources is one approach to reduce the effect of saturation and improve accuracy for AGB mapping. There are various ways to integrate data such as using multi-temporal, multi-frequency, the integration of satellite imagery and geographic information system data or by combining multi-sensors data. For instance, the use of multi-temporal L-band backscatter can achieve accuracies close to the accuracy achieved by a few P-band observations (Cartus and Santoro 2019). The multi-frequency SAR backscatter data with an artificial neural network model showed superior performance for modeling AGB up to 650 t/ha without saturation in the lower biomass ranges in Indonesia's peat swamp forests (Englhart *et al.* 2012). The combination of optical and SAR data is a popular way to enable forest classification and AGB mapping because of the capacity to overcome the constraints in both sensors. Different sensor data have their own characteristics in reflecting land surfaces, and thus integration of different sources of remotely sensed data can enhance the information extraction process. The integration of radar and optical sensor data has the potential to improve AGB estimation because it might reduce the mixed pixels and data saturation problems (Ghasemi *et al.* 2011). In general, two techniques can be used to integrate different source data: data fusion using certain techniques such as wavelet merging, principle component analysis, and partial least squares regression; and combination of different source data as extra bands (Lu *et al*. 2016). The development of many corresponding techniques has been applied, such as the combination of SAR image texture and Landsat TM data (Cutler *et al*. 2012), the combined vegetation index using the weighted optical optimized soil-adjusted vegetation index and microwave HV to estimate forest AGB in native coniferous and broad-leaved forest in Mongolia (Shao and Zhang 2016), and the combination of TM and PALSAR data as extra bands, which can greatly improve AGB estimation performance (Zhao *et al*. 2016).

In general, the accuracy in AGB estimation based on remote sensing data has been enhanced because of the development of satellite sensors as well as the selected algorithms and the variable extraction techniques. The remote sensing techniques provides the potential

for improvement in forest AGB mapping and contributes to carbon stock observation, climate change reduction and forestry management.

1.3. Scientific problem

There are some challenges to monitoring forest biomass mapping and forest transition in tropical areas using remote sensing. First, studies have focused more attention on the change of the pair of forest/nonforest or natural/plantation forests than to the change within forest types. The changes of natural forest areas in the tropics have already been evaluated, but the variation in perspective landscape ecology is not well understood. Second, complex structure and species diversity increases the complexity of remotely sensed based algorithms for AGB estimation in tropical forests when compared with boreal and temporal forest. The use of single-sensor and single methods often do not yield accurate results for classifying and estimating forest biomass, even for optical or SAR images. Even when combining SAR images and optical images for natural forest monitoring, the method for each type of forest has not been specifically determined. Another challenge for monitoring tropical forest resources is influenced by the limited amount of field sample collections because of the rugged terrain, and this makes the assessment of the robustness of estimation models difficult.

The results of this thesis are expected to contribute to solving the above-mentioned problems in developing methods for monitoring natural forest types and biomass changes in a case study in a coastal province in the center of Vietnam. The main outcomes of this study are expected to be follows:

1. A semi-supervised method for forest grades classification in the tropics using the combination of multi-temporal SAR and Landsat data,
2. The spatial pattern changes of forest grades using a high representative set of landscape metrics for forest changes analysis,
3. Estimation of AGB in five forest types using SAR polarimetric parameters, and
4. Evaluation of the performance of remotely sensing data on AGB in bamboo and mixed bamboo forest.

1.4. Objectives

The general goal of this thesis is to estimate the biomass in different tropical forest types using remote sensing. Based on the general goal, the specific objectives were also designed to address the scientific problem mentioned below:

1. Clarify the potential of a semi-supervised model to classify natural forest types by using multi-source remote sensing data,

2. Investigate the process of forest transition from the perspective of landscape ecology by using multi-temporal data,
3. Clarify the contribution of polarimetry to biomass estimation following the individual forest types, and
4. Assess the performance of PALSAR-2 and Landsat 8 OLI on forest biomass estimation to develop the best model for bamboo and mixed bamboo forest.

1.5. Thesis outline

This thesis consists five chapters, each of them from chapter 2 addressing one or two research objectives.

In chapter 2, I addressed objectives 1 and 2, in which the challenges in the classification of natural forest grades of stem volume by developing a semi-supervised model were examined. I demonstrated the approach using multi-temporal and multi-sensor SAR and Landsat data of 2007, 2010 and 2016 which covered the area of Thua Thien Hue province, Vietnam to evaluate the forest transition in the perspective of landscape ecology.

In chapter 3, I addressed objective 3 by developing biomass estimation models based on different characteristics of forest grades of stem volume. The behavior of PolSAR data to biomass differed by forest grades.

In chapter 4, the performance in AGB estimation of bamboo and mixed bamboo forest by PALSAR-2 L-band and Landsat data was evaluated. The linear regression model was chosen and validated for forest biomass estimation in A Luoi district, Thua Thien Hue province, Vietnam.

Finally, chapter 5 summarized the main results of this thesis.

References

1. FAO Forest Resources Assessment 2015: Terms and Definitions. **2012**, 36.
2. Antropov, O.; Rauste, Y.; Häme, T.; Praks, J. Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. *Remote Sens.* **2017**, *9*, 1–24, doi:10.3390/rs9100999.
3. Brown, S.; Iverson, L. R.; Prasad, A.; Liu, D. Geographical distributions of carbon in biomass and soils of tropical asian forests. *Geocarto Int.* 1993, *8*, 45–59.
4. Brown, S.; Lugo, A. E. Biomass of tropical forests: a new estimate based on forest volumes. *Science (80-.).* **1984**, *223*, 1290–3, doi:10.1126/science.223.4642.1290.
5. Bwangoy, J. R. B.; Hansen, M. C.; Roy, D. P.; Grandi, G. De; Justice, C. O. Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices. *Remote Sens. Environ.* **2010**, *114*, 73–86, doi:10.1016/j.rse.2009.08.004.
6. Carreiras, J. M. B.; Vasconcelos, M. J.; Lucas, R. M. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sens. Environ.* **2012**, *121*, 426–442, doi:10.1016/j.rse.2012.02.012.
7. Cartus, O.; Santoro, M. Exploring combinations of multi-temporal and multi-frequency radar backscatter observations to estimate above-ground biomass of tropical forest. *Remote Sens. Environ.* **2019**, *232*, 111313, doi:10.1016/j.rse.2019.111313.
8. Ceccarelli, T.; Smiraglia, D.; Bajocco, S.; Rinaldo, S.; Angelis, A. De; Salvati, L.; Perini, L. Land cover data from Landsat single-date imagery: An approach integrating pixel-based and object-based classifiers. *Eur. J. Remote Sens.* **2013**, *46*, 699–717, doi:10.5721/EuJRS20134641.
9. Chave, J.; Réjou-Méchain, M.; Búrquez, A.; Chidumayo, E.; Colgan, M. S.; Delitti, W. B. C.; Duque, A.; Eid, T.; Fearnside, P. M.; Goodman, R. C.; Henry, M.; Martínez-Yrízar, A.; Mugasha, W. A.; Muller-Landau, H. C.; Mencuccini, M.; Nelson, B. W.; Ngomanda, A.; Nogueira, E. M.; Ortiz-Malavassi, E.; Pélissier, R.; Ploton, P.; Ryan, C. M.; Saldarriaga, J. G.; Vieilledent, G. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Chang. Biol.* **2014**, *20*, 3177–3190, doi:10.1111/gcb.12629.
10. Chen, Y.; Dou, P.; Yang, X. Improving land use/cover classification with a multiple classifier system using AdaBoost integration technique. *Remote Sens.* **2017**, *9*, doi:10.3390/rs9101055.

11. Chrysafis, I.; Mallinis, G.; Gitas, I.; Tsakiri-Strati, M. Estimating Mediterranean forest parameters using multi seasonal Landsat 8 OLI imagery and an ensemble learning method. *Remote Sens. Environ.* **2017**, *199*, 154–166, doi:10.1016/j.rse.2017.07.018.
12. Cloude, S. R. POL-InSAR training course. **2005**, 1–44.
13. Cutler, M. E. J.; Boyd, D. S.; Foody, G. M.; Vetrivel, A. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 66–77, doi:10.1016/j.isprsjprs.2012.03.011.
14. Denman, K. L. et al. Couplings Between Changes in the Climate System and Biogeochemistry. *Clim. Chang. 2007 Phys. Sci. Basis* **2007**, *21*, 499–587.
15. Department of the Interior U.S. Geological Survey Department of the Interior U.S. Geological Survey. *Data Users Handb.* **2016**, *8*.
16. Du H, Q.; Ge H, L.; Liu E, B.; Xu W, B.; Jin, W. A new Classifier for Remote Sensing Data Classification: Partial Least Squares. In *International Workshop on Earth Observation and Remote Sensing Applications*; IEEE: Beijing, China, 2008; p. 302.
17. Dube, T.; Mutanga, O. Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 12–32, doi:10.1016/j.isprsjprs.2015.06.002.
18. Englhart, S.; Keuck, V.; Siegert, F. Modeling aboveground biomass in tropical forests using multi-frequency SAR data-A comparison of methods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 298–306, doi:10.1109/JSTARS.2011.2176720.
19. Ghasemi, N.; Sahebi, M. R.; Mohammadzadeh, A. A review on biomass estimation methods using synthetic aperture radar. *Int. J. Ofgeomatics Geosci.* **2011**, *1*, 776–788.
20. Hall, R. J.; Skakun, R. S.; Arsenault, E. J.; Case, B. S. Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. *For. Ecol. Manage.* **2006**, *225*, 378–390, doi:10.1016/j.foreco.2006.01.014.
21. Handbook, T. J. I. of E. *The Asian Biomass Handbook - A Guide for Biomass Production and Utilization Support*; Yokoyama, Shinya (The University of Tokyo, J., Matsumura, Yukihiko (Hiroshima University, J., Eds.; 2008;
22. Hensley, S.; Van Zyl, J.; Lavalle, M.; Neumann, M.; Michel, T.; Muellerschoen, R.; Pinto, N.; Simard, M.; Moghaddam, M. L-band and P-band studies of vegetation at JPL. *2015 IEEE Radar Conf. - Proc.* **2015**, 516–520, doi:10.1109/RadarConf.2015.7411937.

23. Ho Tong Minh, D.; 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.* **2014**, *52*, 967–979, doi:10.1109/TGRS.2013.2246170.
24. Houghton, R. A.; House, J. I.; Pongratz, J.; Van Der Werf, G. R.; Defries, R. S.; Hansen, M. C.; Le Quéré, C.; Ramankutty, N. Carbon emissions from land use and land-cover change. *Biogeosciences* **2012**, *9*, 5125–5142, doi:10.5194/bg-9-5125-2012.
25. Kim, Y.; Park, N. W.; Lee, K. Do Self-learning based land-cover classification using sequential class patterns from past land-cover maps. *Remote Sens.* **2017**, *9*, doi:10.3390/rs9090921.
26. Koch, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 581–590, doi:10.1016/j.isprsjprs.2010.09.001.
27. Koeniguer, E. C. *Polarimetric radar images*; 2014;
28. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782, doi:10.1109/LGRS.2017.2681128.
29. Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Li, G.; Moran, E. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *Int. J. Digit. Earth* **2016**, *9*, 63–105, doi:10.1080/17538947.2014.990526.
30. Lucas, R. M.; Mitchell, A. L.; Armston, J. Measurement of forest above-ground biomass using active and passive remote sensing at large (Subnational to global) scales. *Curr. For. Reports* **2015**, *1*, 162–177, doi:10.1007/s40725-015-0021-9.
31. Mermoz Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sens. Environ.* **2014**, *155*, 109–119, doi:10.1016/j.rse.2014.01.029.
32. 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.* **2015**, *159*, 307–317, doi:10.1016/j.rse.2014.12.019.
33. Mitchard, E. T. A.; Saatchi, S. S.; Woodhouse, I. H.; Nangendo, G.; Ribeiro, N. S.; Williams, M.; Ryan, C. M.; Lewis, S. L.; Feldpausch, T. R.; Meir, P. Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophys. Res. Lett.* **2009**, *36*, 1–6, doi:10.1029/2009GL040692.
34. Molto, Q.; Rossi, V.; Blanc, L. Error propagation in biomass estimation in tropical forests. *Methods Ecol. Evol.* **2013**, *4*, 175–183, doi:10.1111/j.2041-210x.2012.00266.x.

35. 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.* **2012**, *50*, 714–726, doi:10.1109/TGRS.2011.2176133.
36. 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.* **2013**, *44*, 593–622, doi:10.1146/annurev-ecolsys-110512-135914.
37. Pargal, S.; Fararoda, R.; Rajashekar, G.; Balachandran, N.; Réjou-Méchain, M.; Barbier, N.; Jha, C. S.; Pélissier, R.; Dadhwal, V. K.; Couteron, P. Inverting aboveground biomass-canopy texture relationships in a landscape of forest mosaic in the Western Ghats of India using very high resolution Cartosat imagery. *Remote Sens.* **2017**, *9*, 1–20, doi:10.3390/rs9030228.
38. Phua, M. H.; Johari, S. A.; Wong, O. C.; Ioki, K.; Mahali, M.; Nilus, R.; Coomes, D. A.; Maycock, C. R.; Hashim, M. Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests. *For. Ecol. Manage.* **2017**, *406*, 163–171, doi:10.1016/j.foreco.2017.10.007.
39. Rosenqvist, Å.; Milne, A.; Lucas, R.; Imhoff, M.; Dobson, C. A review of remote sensing technology in support of the Kyoto Protocol. *Environ. Sci. Policy* **2003**, *6*, 441–455, doi:10.1016/S1462-9011(03)00070-4.
40. Roy, M.; Ghosh, S.; Ghosh, A. A novel approach for change detection of remotely sensed images using semi-supervised multiple classifier system. *Inf. Sci. (Ny).* **2014**, *269*, 35–47, doi:10.1016/j.ins.2014.01.037.
41. Saatchi, S.; Marlier, M.; Chazdon, R. L.; Clark, D. B.; Russell, A. E. Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. *Remote Sens. Environ.* **2011**, *115*, 2836–2849, doi:10.1016/j.rse.2010.07.015.
42. Sader, S. A.; Waide, R. B.; Lawrence, W. T. Tropical Forest Biomass and Successional Age Class Relationships to a Vegetation Index Derived from Landsat TM Data. **1989**, *156*, 143–156, doi:https://doi.org/10.1016/0034-4257(89)90112-0.
43. Sarker, L. R.; Nichol, J. E. Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sens. Environ.* **2011**, *115*, 968–977, doi:10.1016/j.rse.2010.11.010.
44. Shao, Z.; Zhang, L. Estimating forest aboveground biomass by combining optical and SAR data: A case study in genhe, inner Mongolia, China. *Sensors (Switzerland)* **2016**, *16*, doi:10.3390/s16060834.

45. Thiel, C.; Schmullius, C. The potential of ALOS PALSAR backscatter and InSAR coherence for forest growing stock volume estimation in Central Siberia. *Remote Sens. Environ.* **2016**, *173*, 258–273, doi:10.1016/j.rse.2015.10.030.
46. Toan, T. Le; 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.* **2011**, *115*, 2850–2860, doi:10.1016/j.rse.2011.03.020.
47. Van Zyl, J. J.; Arii, M.; Kim, Y. Model-based decomposition of polarimetric SAR covariance matrices constrained for nonnegative eigenvalues. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3452–3459, doi:10.1109/TGRS.2011.2128325.
48. Zhao, P.; Lu, D.; Wang, G.; Liu, L.; Li, D.; Zhu, J.; Yu, S. Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *53*, 1–15, doi:10.1016/j.jag.2016.08.007.
49. Zhao, P.; Lu, D.; Wang, G.; Wu, C.; Huang, Y.; Yu, S. Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sens.* **2016**, *8*, doi:10.3390/rs8060469.
50. Zhu, X.; Liu, D. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 222–231, doi:10.1016/j.isprsjprs.2014.08.014.
51. Zyl, J. Van; Kim, Y. Synthetic Aperture Radar Polarimetry (Google eBook). **2011**, 313.

CHAPTER 2.

SEMI-SUPERVISED MODEL AND LANDSCAPE METRICS FOR MAPPING AND SPATIAL PATTERN CHANGES ANALYSIS OF TROPICAL FOREST TYPES

2.1. Introduction

Since the early 1990s, the tropical forests in many countries have been undergoing a transition period from degradation to reforestation (Mather and Needle 1998; Meyfroidt and Lambin 2009; Keenan *et al*. 2015). Forest transition is considered from the perspective of forest area changes and the conversion from other land use/land cover types to forest. With the rapid development of remote sensing technology and the wide application of landscape ecology, they supply effective tools to analyze spatial-temporal changes and related ecological processes. Improved understanding of forest transition provides many benefits, such as global carbon balance or land use and forest policy implementation (Thiha *et al*. 2007; MacDicken 2015). However, there is a need to develop further new methods for forest type classification and forest transition assessment.

Recently, remote sensing combined with the conventional method to supply validation data has been extensively used in forest inventory. The advantages of the remote sensing technique are cost- and labor-saving as well as swift observation of large scale forest changes over the long term. However, the classification accuracy of the remote sensing is affected by many factors, such as the classification techniques, training samples, and the signal reflected from objects.

A natural forest (FAO 2012) is a naturally regenerated forest comprising native species, where there are no clear or clearly visible indications of human activities and the ecological processes are not significantly disturbed. In this study, we classified natural forests based on the timber reserve of standing trees into four main types: rich, medium, poor, and restoration forest. Although these four types differ in species composition and timber reserves, we found that with only a single source of data (optical or radar data) it is often difficult to discriminate between different kinds of natural forest types because of the very similar information on canopy and forest structure captured by remotely sensed data (Li *et al*. 2012). This highlights the need for multi-source remote sensing data to extract more information of interest regarding the objects for classification. By using multi-source data, the classification accuracy is generally improved compared to single data source. This has been shown, for example, with a combination of optical data and synthetic aperture radar (SAR, Congo Basin and Malawi city, Mzimba) (Bwangoy *et al*. 2010; Hirschmugl *et al*. 2018). The fusion of

different frequencies (L– and P–band) of SAR products has also received much attention in recent years (Sahyun Hong *et al*. 2003; Yong-an *et al.* 2007; Shimoni *et al*. 2009).

Another problem is that sampling is limited because of the complexity of ecosystems and inaccessible regions (Chambers *et al*. 2007). In this study, we used semi-supervised classification to overcome the paucity of ground truth samples. Semi-supervised classification focuses on enhancing supervised classification by minimizing errors in the labeled examples, but it must also be compatible with the input distribution of unlabeled instances (Triguero *et al.* 2015). While supervision often provides higher classification accuracy, it requires a good dataset to ensure both the quantity and quality of training samples collected from the field survey. The constraint of field data collection is that it is not always achievable, owing to limitations in finance, terrain, or availability of the data source. To avoid this issue, semi-supervised classification aims at solving the limited number of labeled samples and taking advantage of the abundant unlabeled samples. Many semi-supervised classification algorithms such as expectation-maximization, co-training, and self-training have been developed. The graph-based method has also attracted an increasing amount of interest (Gao *et al*. 2009, 2013; Ma *et al*. 2016; Sawant and Prabukumar 2018). This method works by summarizing base model outputs in a group-object bipartite graph and maximizing the consensus by promoting smoothness of label assignment over the graph and consistency with the initial labeling. Recently, machine learning has received much attention and has been applied to the semi-supervised learning problem. This technology has been successfully developed for binary classification, such as in (Qi *et al.* 2012), where a Laplacian Twin Support Vector Machine was used for semi-supervised classification that can exploit the geometry information of the marginal distribution embedded in unlabeled data to construct a more reasonable classifier-semi-supervised classification with graph convolutional networks (Kipf and Welling 2016) which scales linearly in the number of graph edges and learns hidden layer representations that encode both the local graph structure and the features of nodes.

For land use/land cover, semi-supervised classification has been successfully adopted in the literature. For instance, in (Erkan *et al.* 2010), semi-supervised logistic regression was applied. This is a specific instance of the generalized maximum entropy that finds a probability distribution that minimizes a divergence based on the entropy of the weights of classifiers. In Alok *et al.* (2015), a semi-supervised clustering was presented that is simultaneously optimized using a modern multi-objective optimization technique based on the concepts of simulated annealing. In Silva *et al.* (2017), the weight support vector machine

was used to keep the training effort low with a manually-collected set of pixels of the class of interest and a random sample of pixels. In Cui *et al.* (2018), extended label propagation and rolling guidance filtering that uses superpixel propagation were applied to assign the same labels to all pixels within the superpixels that are generated by the image segmentation method.

In this paper, I present a self-learning approach for forest classification that can propagate labels from labeled samples to unlabeled data to build a large volume of training data. This model does not make any specific assumptions for the input data, but it does accept that its own predictions tend to be correct (Triguero *et al.* 2015). Self-learning, also known as Yarowsky's algorithm, is a rule-based semi-supervised classification. The term "self-learning" is used because the algorithm uses its own prediction to teach itself. Self-learning is very popular, with an initial classifier trained by a small number of training data with given labels, before using this classifier to assign labels to the unlabeled sample. For each unlabeled sample, confidence values are extracted from the probabilistic of learning models (Triguero *et al.* 2015; Yarowsky 1995). The samples that have been labeled with the most confident prediction are then selected to combine with the training data and create a new training set. The classifier is then retrained on that new training set and the procedure repeated. Self-learning has been applied in several text processing tasks in the last few years. Recently, it has been applied with some developed supervisor classifiers to image classification (Silva *et al.* 2017; Dópido *et al*. 2013). This study developed self-learning with a kernel least square classifier for forest types classification. Least squares is a standard approach of statistical analysis and has been well-known for a long time. It was developed by applying kernel functions in high dimensional feature space to solve the problem of a large number of parameters (Saunders *et al.* 1998). Kernel functions are an algorithm with the advantage of being able to flexibly transform an originally non-linear vector into a linear version in feature space. Therefore, they are widely applied in solving classification problems involving multiple features (Du H *et al*. 2008; Laparra *et al*. 2011; Antropov *et al.* 2017).

In this study, I also used time-series remotely sensed data for the evaluation of forest changes by landscape ecology. Landscape ecology can be generally defined as the science and art of studying and improving the relationship between spatial patterns and ecological processes on a multitude of scales and organizational levels (Wu 2008). One fundamental aspect has been its explicit attention to the spatial dimension of ecological processes (Botequilha Leitão and Ahern 2002). Landscape metrics are one of the classical landscape

ecological tools for measurement, analysis, and interpretation of spatial patterns (Antrop *et al*. 2009). The contribution of remote sensing to landscape planning and conservation is mainly in the inventory and determination of objects of interest and in monitoring changes by time-series satellite data (Tlapakova *et al*. 2013). A basic concern in forest management is spatial processes over time, such as deforestation, degradation, or restoration. The analysis of landscape structure is a classic approach for the understanding of spatial processes using various landscape metrics (Botequilha Leitão and Ahern 2002; Lausch and Herzog 2002; Neel *et al.* 2004; Gyenizse *et al.* 2014). Several studies provide evidence of the value of remote sensing and landscape metrics for forest management(Sano *et al*. 2009; Yang *et al*. 2010; Geri *et al.* 2010; Smiraglia *et al*. 2015).

In summary, there are two main objectives in this study. The first objective is to assess the potential of a semi-supervised model to classify natural forest types by using multi-source remote sensing data. The second objective is to assess the process of forest transition from the perspective of landscape ecology by using multi-temporal data.

2.2. Study Area

In Vietnam, the forest plays an important role in the socio-economic system in the mountainous province, where local people have a low income and agroforestry-based livelihoods. Although centralization of forest resource management began in Vietnam very early in the 1950s (Thiha *et al.* 2007), the natural forest experienced a rapid decrease over the long term (FAO, 2019), causing negative impacts to the environment, such as loss of carbon stock, biodiversity degradation, and habitat fragmentation (Meyfroidt and Lambin 2008). Since 2005, however, Vietnam has been experiencing a positive period in the application of forestry policies (Cong Thang 2015), which is contributing to development of the forested area. This dramatic forest transition has resulted in changes in the biophysical, ecological process, as well as in the spatial landscape. However, there is a lack of up-to-date information on forest changes in Vietnam in the period from 2005 to the present, particularly in central Vietnam where the socio-economic dynamics have recently been increasing. To create a reliable forest management strategy, an improved understanding of forest changes is essential. This can be achieved by spatial analysis through multi-temporal remote sensing image processing, combined with landscape metrics assessment.

Thua Thien Hue province, located in central Vietnam (Figure 2.1d), has an area of 5054 km² and the natural forest area accounts for approximately 40% of the total area. According to the General Statistics Office (GSO) in Vietnam, the natural forest area examined in this study slightly decreased from 203,800 ha in 2008 to 202,700 ha in 2010, with the principal

causes of deforestation by the conversion from forest to other land uses (e.g., hydropower, roads, and cultivation) and illegal exploitation of forest products. Conversely, from 2010, there was a significant increase of natural forest area reaching 212,200 ha in 2016. These fluctuations have caused changes not only in the area, but also in the forest landscape structure.

We classified the natural forest into four types based on the specific condition of the study site as well as circular number 34/2009/TT-BNNPTNT of June 10, 2009 (MARD 2009) published by Vietnam Ministry of Agriculture and Rural Development, on the criteria for forest identification and classification in Vietnam:

1. Rich forests are forests with stem volume of between 201 and 300 m^3/hectare;
2. Average forests (or medium forests) are forests with stem volume of between 101 and 200 m^3/hectare;
3. Poor forests are forests with stem volume of between 10 and 100 m^3/hectare;
4. Forests with no reserve ("Restoration forest" in the case of our study site) are forests with a mean diameter of less than 8 cm and stem volume of less than 10 m^3/hectare.

2.3. Data and Methods

2.3.1. Data

We used time-series SAR data and Landsat data acquired in 2007, 2010, and 2016 (Figures 2.1a, 1b, 1c). Two scenes of SAR data were collected per year, which were then used to create a mosaic covering 77% of the study area. The SAR data differed in terms of acquisition mode, which led to a difference in the incidence angle and the size of the range and azimuth. Therefore, preprocessing was necessary to synchronize these data. Two polarization HH (horizontal transmitting, horizontal receiving) and HV (horizontal transmitting, vertical receiving) were used to process the data in this study. Landsat data were also selected to combine with SAR data for forest grade classification. Landsat data were provided by the United States Geological Survey (USGS) with moderate resolution and wide spectral coverage. The swath width of Landsat is 185 km; therefore, it could cover the full study area. The characteristics of these data are described in Table 2.1.

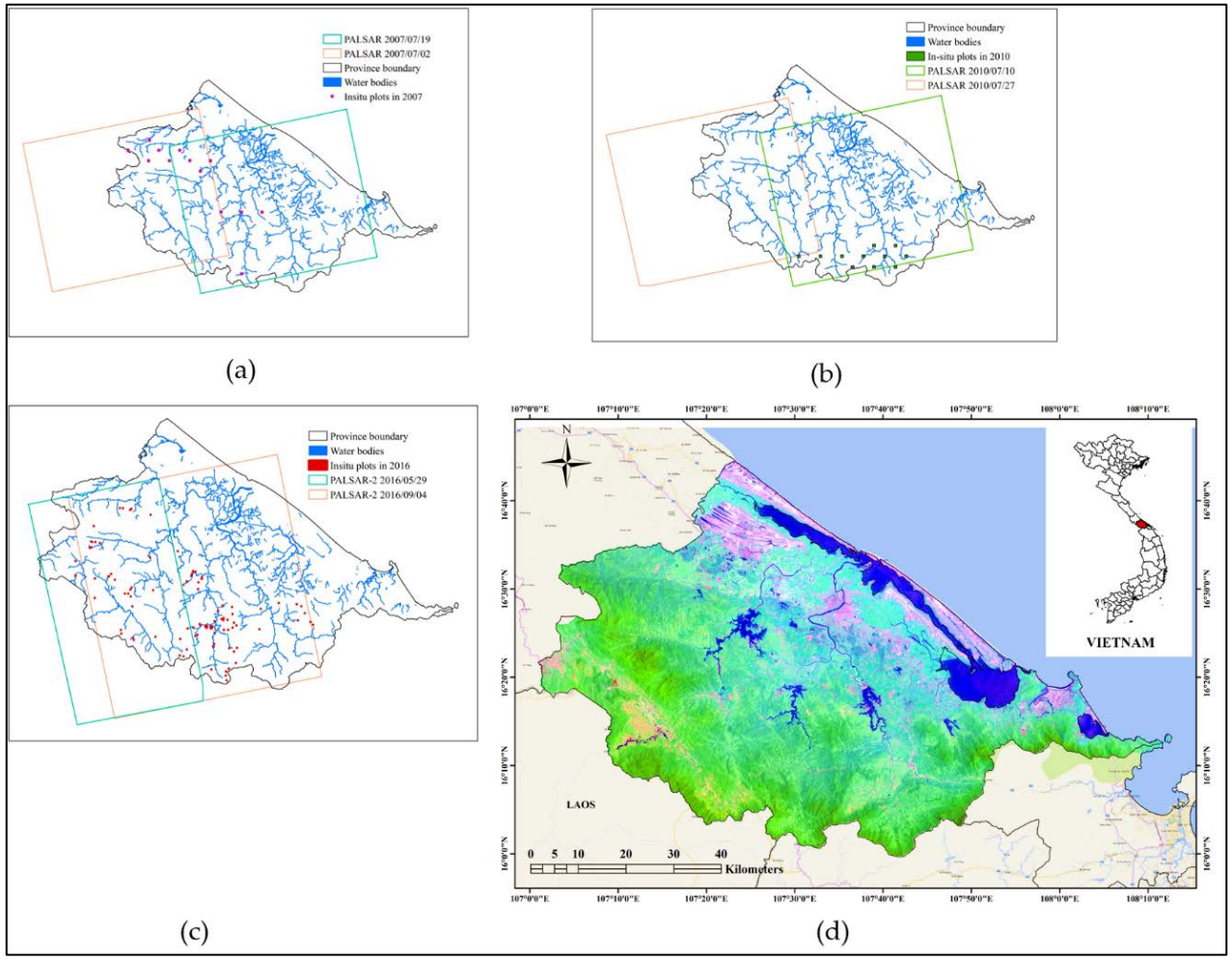


Figure 2.1. Cover of synthetic aperture radar (SAR) images and in-situ data in (**a**) 2007, (**b**) 2010, (**c**) 2016, and (**d**) location map of the study area in Landsat data with pseudo colors (R: SWIR 2, G: near-infrared, B: green)

Table 2.1. Characteristics of satellite image data used in this study

Date	Types	Level	Incidence angle at scene center	Resolution (m)	Polarization/ band
2016/05/29	PALSAR2	1.1	38.99	3.12 × 4.55	HH + HV + VH + VV
2016/09/04	PALSAR2	1.1	40.5	3.4 × 6.6	HH + HV
2010/07/10	PALSAR	1.1	38.7	3.2 × 15	HH + HV
2010/07/27	PALSAR	1.1	38.7	3.2 × 15	HH + HV
2007/07/02	PALSAR	1.5	38.7	12.5	HH + HV
2007/07/19	PALSAR	1.5	38.7	12.5	HH + HV
2007/04/24	Landsat TM	1	-	30	5
2010/02/11	Landsat TM	1	-	30	5
2016/04/16	Landsat 8 OLI	1	-	15, 30	11

A ground sample was collected to support training data and accuracy assessment. These data were provided by the Central Sub Forest Inventory and Planning Institute, Thua Thien Hue province, Vietnam (Sub-FIPI). The data collection was evenly distributed over the entire study area at three time periods—In 2007, 2010, and 2016. The samples were then divided into 80% training data and 20% validation data. In 2007, 13 measured plots were covered by the PALSAR scene, with each plot measuring 1 km^2 (1,000 × 1,000 m), while in 2010, there were 10 such plots. In each plot, 40 subplots of 25 × 20 m were set up to measure forest parameters and describe characteristics. However, not all 40 subplots were measured and selected for classification; only some met the conditions of being natural forests with reserves, not separated by other obstacles such as rivers, streams and roads, and terrain. In 2007, 170 subplots were selected for this study, while in 2010, 115 subplots were selected. In 2016, 106 plots were covered by PALSAR-2 data. Each rectangular plot measured 30 × 33 m with the longer aspect running in an east-west direction and the shorter aspect running north-south. The distribution of samples for the four forest grades is described in Table 2.2.

Table 2.2. Ground data for the four forest types in the study area in 2007, 2010, and 2016

Types	**Number of samples**		
	2007	**2010**	**2016**
Rich forest	17	20	29
Medium forest	68	34	23
Poor forest	48	34	37
Restoration forest	37	27	17
Total	170	115	106

Apart from these samples, a larger amount of unlabeled data was supplied for forest grades classification. A total of 200 unlabeled samples was randomly created over the study area. The proportion of unlabeled samples accounted for approximately 40–60% of the total samples to ensure the accuracy of the classification results. In particular, the number of unlabeled samples was equivalent to 55% for 2007, 64% for 2010, and 65% for 2016.

2.3.2. Methods

A flowchart of the methodology employed in this study is presented in Figure 2.2

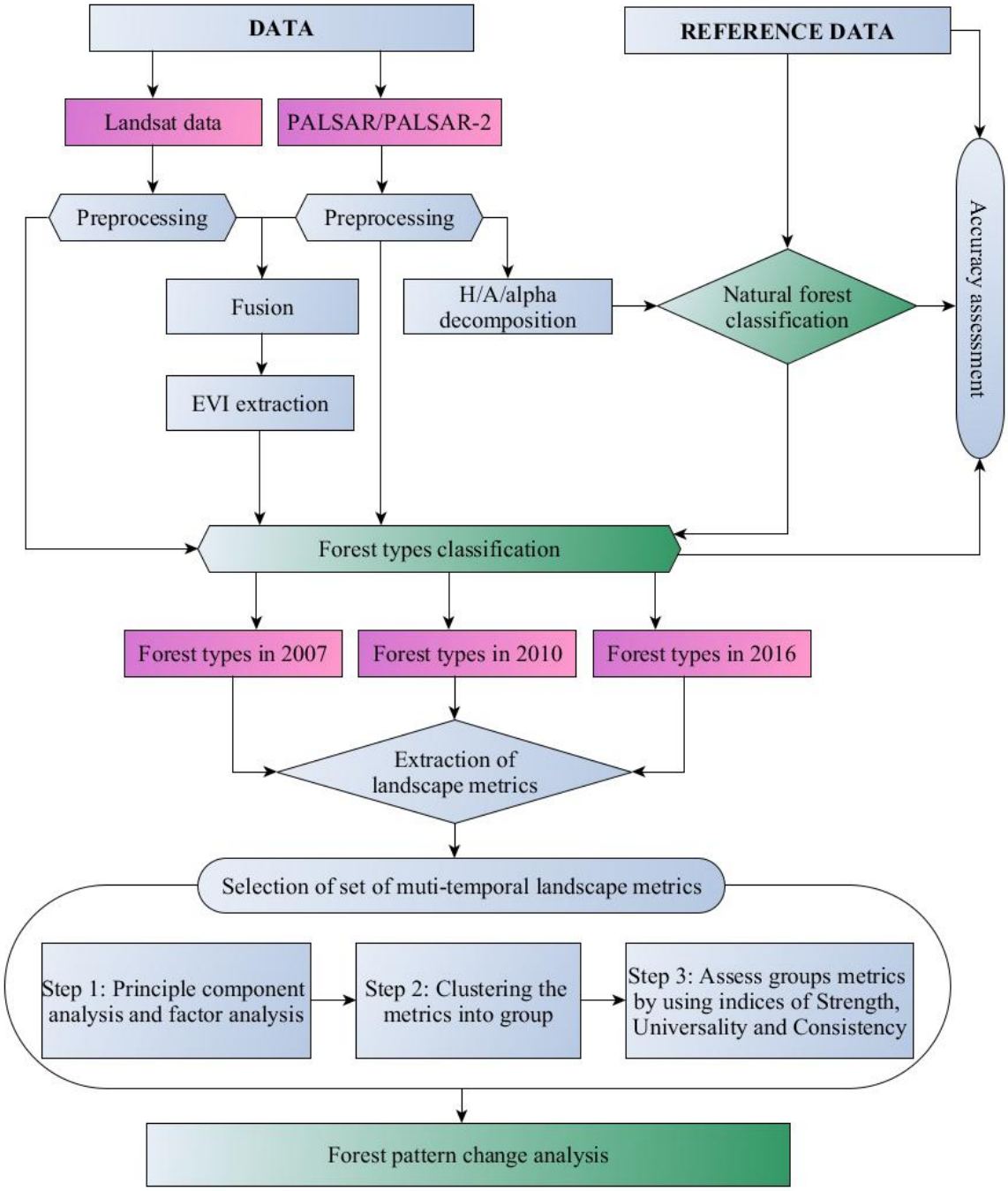


Figure 2.2. Flowchart of the methodology employed in this study

2.3.2.1. Preprocessing

Landsat digital numbers (DNs) were converted to reflectance and atmospheric correction using the fast line-of-sight atmospheric analysis of hypercubes (FLAASH) tool. The enhanced vegetation index (EVI) was then calculated using band near-infrared (0.7–1.1μm), red (0.6–0.7 μm), and blue (0.45–0.52 μm) in accordance with the work of Liu and Huete (Huete *et al*. 1997):

$$EVI = G \times \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + (C_1 \times \rho_{red} - C_2 \times \rho_{blue}) + L} \quad (2.1)$$

where L is a soil adjustment factor, and C_1 and C_2 are coefficients used to correct aerosol scattering in the red band by using the blue band. In general, G = 2.5, C_1 = 6.0, C_2 = 7.5, and L = 1.

In this study, when observing the relationship between reflectance value and the cosine of the solar incidence angle, there was a low correlation coefficient with the value of 0.0075 and 0.0197 for TM and OLI data, respectively. This means that the terrain does not significantly affect this test site. Therefore, topography correction is unnecessary in this case.

For radar data, dual-polarized images (HH, HV polarizations) were created in the single-look complex (SLC) format. The preprocessing data were operated to convert the digital number value into sigma naught (σ^o) values using the following equation:

$$\sigma^o = 10.\log_{10}(I^2 + Q^2) + CF - A \quad (2.2)$$

where I and Q are the real and imaginary parts of the SLC product. A is a conversion factor equal to 32.0. The calibration factor CF is -83.

A refined Lee filter was used with a window size of 7 × 7 to reduce the speckle noise. The topography effect was eliminated using range—Doppler terrain correction with digital elevation model (DEM) from the Shuttle Radar Topography Mission, and all of the product images were resampled to reach 15 meters in pixel spacing.

The preprocessed SAR data were next transformed into covariance matrix elements, and then eigenvalue and eigenvector polarimetric parameters. The cross-pol ratio of HH and HV was also calculated and used as a variable for the classification model. In addition, SAR data and Landsat data were fused and resampled to 15 m. The parameters set for polarimetric SAR (PolSAR) and Landsat data comprise the input features for classification, as detailed in the next section.

To illustrate the polarimetric data, we adopted eigen decomposition of the 2 × 2 covariance matrix for dual polarization data as defined by (Ainsworth *et al*. 2007):

$$\begin{bmatrix} C_{HH,HH} & C_{HH,HV} \\ C_{HV,HH} & C_{HV,HV} \end{bmatrix}$$

H/A/Alpha decomposition was used to decompose the backscatter value into three components: entropy, anisotropy, and alpha (H/A/α). The H/A/α is a polarimetric parameters decomposition based on eigenvalue and eigenvector that was introduced by Cloude and Pottier (Zyl and Kim 2011). In this technique, backscattering is decomposed into entropy (H), anisotropy (A), and alpha angle (α). Entropy is a parameter describing randomness in target scattering, which is defined as:

$$H = -(\bar{\lambda}_1 ln\bar{\lambda}_1 + \bar{\lambda}_2 ln\bar{\lambda}_2)/ln2 \text{ with } \bar{\lambda}_i = \lambda_i/(\lambda_1 + \lambda_2) \quad (2.3)$$

where H_T is target entropy and λ_i (i = 1 to 2) are eigenvalues.

Entropy values vary from 0 for a single scattering mechanism to 1 for pure noise and random targets.

Mean alpha angle is defined as:

$$\alpha = \bar{\lambda}_1 \alpha_1 + \bar{\lambda}_2 \alpha_2 \quad (2.4)$$

The alpha angle varies between 0° for trihedral scattering from a planar surface to 90° for dihedral scattering from a metallic surface. Another element is anisotropy (A), which is a parameter complementary to entropy, which can be employed as a source of discrimination only when H >0.7 owing to the high effect of noise (ESA 2017).

2.3.2.2. Masking Undesirable Areas

In this study, we created a mask to remove undesirable areas before classifying natural forest grades. The classification method of the random forest algorithm was applied based on entropy, alpha, and anisotropy parameters extracted from dual polarization data for images in 2010 and 2016. For the image in 2007, the polarization data of HH, HV, and EVI from Landsat data were used for classification. For other land use/land cover types, samples such as rivers, urban areas, and agricultural land were collected through visual interpretation based on discrimination in color, geometric shapes, and brightness. For the natural forest, 170 samples were collected for 2007 with a plot area of 25 × 40 m, 115 samples for 2010, and 106 samples for 2016 with the area of 30 × 33 m. Polarimetric data were derived from the image for each sample with a window size of 2 × 2 pixels, with a pixel size of 15 m. The classification results create natural forest maps for the study area.

Furthermore, in this study area, because the natural forest is mainly distributed on topography at an elevation above 200 m, a digital elevation map (DEM) was applied to mask out low-altitude forest areas while retaining forests with elevations above 200 m. This DEM

map was downloaded from NASA Shuttle Radar Topography Mission data. The masked forest images were then used for the forest types classification.

2.3.2.3. Self-Learning with the Kernel Least Squares (SL-KLS) Classifier for Forest Types Classification

Kernel Least Squares (KLS)

In this study, the presence of a large number of parameters in the classification problem created computational difficulties due to a high number of dimensions. To solve this problem, we used the KLS technique in the R environment with RSSL package version 0.7. Here, KLS is described as a method using least squares regression as a classification technique with numeric encoding of classes as targets. A detailed description of KLS can be found in various studies (Saunders *et al.* 1998; Ping Sun 2005), with the optimal parameter vector identified by $\Theta = [b\ \alpha_1\ \alpha_2\ ...\ \alpha_n]^{\mathrm{T}}$. The minimized vector has the form $L(\Theta) = \|Y - P\Theta\|^2$,

with $Y = [y_1\ y_2\ ...\ y_n]^{\mathrm{T}}$ and

$$P = \begin{bmatrix} 1 & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ \vdots & & \vdots & \\ 1 & k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix}$$

A radial basis function was used with the form below:

$$k(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{\sigma^2}) \quad (2.5)$$

where x_i: are training data, x_j is a feature vector, and σ is a free parameter. Kernel k has a value in the range of 0 to 1. With α_i as real numbers, the prediction function *f*(x) can be written as follows:

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) + b \quad (2.6)$$

Self-learning with the Kernel Least Squares (SL-KLS) classifier

In this study, a self-learning algorithm was used to turn the KLS classifier into a semi-supervised model to solve the problem of the small amount of labeled data. Based on the training data, KLS was applied to assign labels to unlabeled objects, which were then added to the set of labeled objects for classification. There is a given set of labeled data (L) and a set of unlabeled data (U) (Figure 2.3). By applying a KLS classifier, k number of labels are assigned to unlabeled data. The result of predicted data U then joins with L to create a new training set for classifying the entire segmented images. In this study, we classified the forest into four grades: rich forest, medium forest, poor forest, and restoration forest. The features

of the four grades were extracted from Landsat bands reflectance, EVI, HH, HV signals, covariance elements, and H/A/Alpha decomposition.

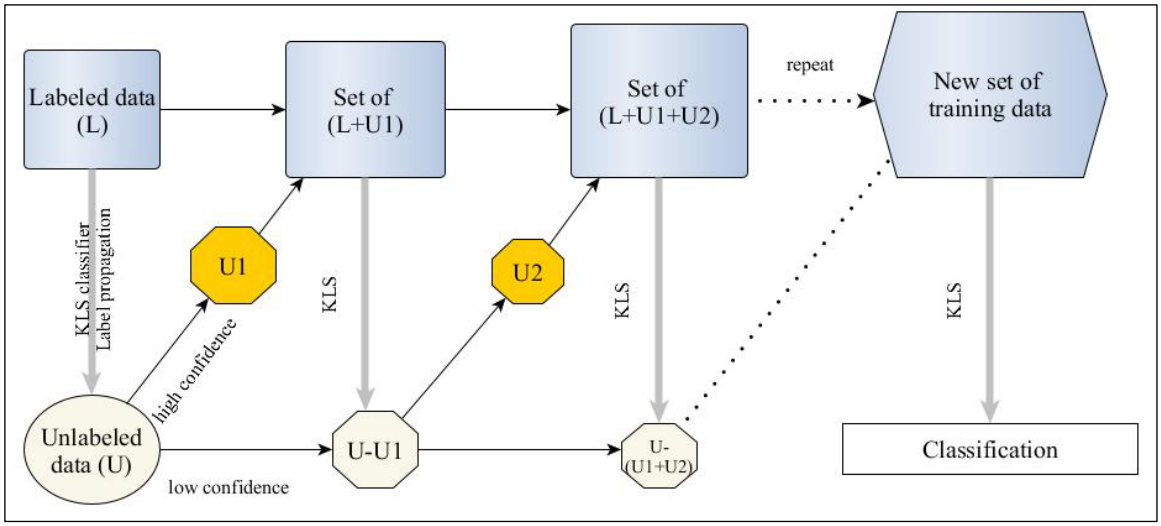


Figure 2.3. Flowchart of classification using the combination of self-learning with kernel least squares classifier in this study

The indicator of overall accuracy (OA), kappa, user's accuracy, and producer's accuracy were used to evaluate classification accuracy by comparing the classified image with the results of ground truth interpretation. The overall accuracy comprises the ratio of the sum of accuracy in an individual class and the number of observed samples, with 100% as the perfect classification. Kappa, user's, and producer's accuracy were proposed by Congalton and have been used widely to date. The function of these indicators is clearly described in Congalton (1991).

2.3.2.4. Forest Pattern Analysis Using Landscape Metrics

Extraction of Landscape Metrics

After the classification step, the forest was divided into four forest grades: rich, medium, poor, and restoration forest in the years 2007, 2010, and 2016. For each year, the classified images were then clipped into 14 non-overlapping sub-landscapes of 2,000 × 2,000 meters. This size was selected to ensure the representativeness of the sample and to reduce computation time. To conduct spatial analysis of the forest pattern, landscape metrics were computed at two levels, class and landscape, for all samples in each year. We calculated 56 metrics at the class level and 63 metrics (Appendix A) at the landscape level for each sub-landscape image using Fragstat version 4.2.1. With a large number of landscape metrics, we then selected the appropriate metrics for analysis of the natural forest process for the study area longitudinally.

Selection of a Set of Landscape Metrics

Principal components analysis was used to identify components and cluster them into various groups. In these groups, the three indices of universality, consistency, and strength were then calculated to select the group of metrics. This operation was conducted using PROC FACTOR in SAS.

Based on the assessment of metrics through the three indices of universality, consistency, and strength, we created a list of selected metrics at the class and landscape levels. At the class level, 11 clusters were created from 56 metrics. Through cluster analysis, two clusters (approximately 16 metrics) were selected with a high level of these three indices at a total percentage >90%, variation explained >7%, and the average in-group correlation >0.8 (Appendix B). Similarly, two clusters (approximately 20 metrics) were selected through analysis of the 10 clusters created from 63 metrics for the landscape level (Table 2.3). The other two metrics—total area (CA in hectares_ha) and percentage of landscape (PLAND_%)—were also added for change analysis of the area in general.

Analysis of Forest Pattern Change

From the set of representative metrics in the study area, we selected various metrics that support the analysis of spatial processes over time, containing aggregation, compactness, and fragmentation. To evaluate the spatial structure change of forest grades in the period 2007–2016, we selected 11 metrics for class level and five metrics for landscape level.

The aggregation is expressed by increasing the size of patches from the combination of small fragments. Therefore, this indicator relates to the recovery of forests from the previously deforested area. The metrics are related to aggregation including the aggregation index (AI), proportion of like adjacencies (PLADJ), and clumpiness index (CLUMPY) for the class level, and Interspersion/juxtaposition index (IJI) for the landscape level. Another term that is strongly involved in the aggregation is forest connectivity, which evidently increases the patch cohesion index (COHESION) that is related to the physical connectedness of the corresponding patch type.

Forest fragmentation is an opposite process to aggregation and occurs when a large contiguous forest is broken down into many small fragments, leading to loss of biodiversity and animal habitat and degradation of forest health and its economic and environmental functions. This process is closely related to the shrinkage ratio of area-weighted mean patch size (AREA_AM) and effective mesh size (MESH) of the landscape over time.

Table 2.3. Set of high representative metrics for analyzing multi-temporal forest types structure at class and landscape level in the study area

No		Metric name	Level	Description
	Aggregation/ Fragmentation	AI	C	Aggregation index
		CLUMPY	C	Clumpiness index
		COHESION	C	Patch cohesion
		NLSI	C	Normalized landscape shape index
		PLADJ	C	Proportion of like adjacencies
		IJI	L	Interspersion/ juxtaposition index
		MESH	L	Effective mesh size
	Area and edge metrics	AREA_AM	L	Area-weighted mean patch size
		AREA_CV	L	Patch size coefficient of variation
		GYRATE_AM	L	Area-weighted radius of gyration
	Core area metrics	CAI_CV	C	Core area coefficient of variation
		CORE_AM	L	Area-weighted mean core area
		CORE_CV	L	Core area coefficient of variation
		DCORE_AM	L	Area-weighted mean disjunct core area
		DCORE_CV	L	Disjunct core area coefficient of variation
	Shape metrics	CIRCLE_AM	C	Area-weighted related circumscribing circle
		CIRCLE_CV	C, L	Circumscribing circle coefficient of variation
		CIRCLE_MN	C, L	Mean related circumscribing circle
		CONTIG_AM	C	Area-weighted contiguity index
		CONTIG_MN	C, L	Mean contiguity index
		CONTIG_CV	L	Contiguity index coefficient of variation
		SHAPE_MN	C, L	Mean shape index
		SHAPE_AM	L	Area-weighted mean shape index
		SHAPE_CV	L	Shape index coefficient of variation
		FRAC_AM	L	Area-weighted mean fractal dimension
		FRAC_MN	C, L	Mean fractal dimension
		FRAC_CV	C, L	Fractal dimension coefficient of variation
		PARA_MN	C, L	Mean perimeter–area ratio
		PARA_AM	C	Area-weighted mean perimeter–area index

Compaction involves the formation of rounded patches in a circular shape that makes them more compact (Aguilera *et al.* 2011). The more closely a patch shape is to a circle, the more it exhibits compaction. While a natural forest has a complex and irregular shape, basic geometry patch shapes show unnatural objects. Therefore, analysis of forest compaction enables us to assess disturbance in the forest using various shape metrics such as the shape index (SHAPE_MN, _AM) and circumscribing circle (CIRCLE_MN, _AM) at the class level. At the landscape level, area-weighted radius of gyration (GYRATE, _AM) is used to analyze compaction. Furthermore, GYRATE_AM also provides the overall characterization of the level of connectivity or subdivision of the landscape (Aguilera *et al.* 2011).

2.4. Results

2.4.1. Forest Grade Classification

For the result of masking undesirable areas, we compared the predicted products with the reference data and evaluated them based on the index of overall accuracy (OA) for each year. The results obtained high accuracies for the images in 2007, 2020 and 2016 with an OA of over 0.87. The 2010 predicted image was the best with an OA of 0.99, followed by 2016 with 0.92 and 2007 with 0.87.

The behavior survey of only Landsat or only PolSAR data on forest objects does not show observable discrimination (Appendix C). For radar images, polarimetric parameters are not used to classify forest objects due to the saturation of entropy throughout the study area. Alpha and anisotropy display slight fluctuations on different forest objects, but they do not create good results in the discrimination. Nor does relying on the polarization of HH and HV signals provide better results. Therefore, with efforts to improve accuracy in forest classification, we have used combined data from optical and SAR data to extract information for forest grades classification.

Another difficulty encountered in the classification process was the limited number of samples collected from the field, particularly in 2016 when only 106 samples were collected for the four forest grades over the entire study area. The small number of samples was inadequate to develop a reliable classification algorithm based on the supervised method. To solve this problem, we used the semi-supervised classification with the addition of information from an unlabeled data source. However, to ensure the accuracy of the classification results, it was necessary to select an appropriate ratio between the number of labelled and unlabeled samples. The higher the percentage of unlabeled samples, the lower the accuracy (Gao *et al*. 2013). To balance the number of unlabeled samples required and

the classification accuracy, 200 random samples were created in the study area to ensure the ratio was approximately 60% for each year.

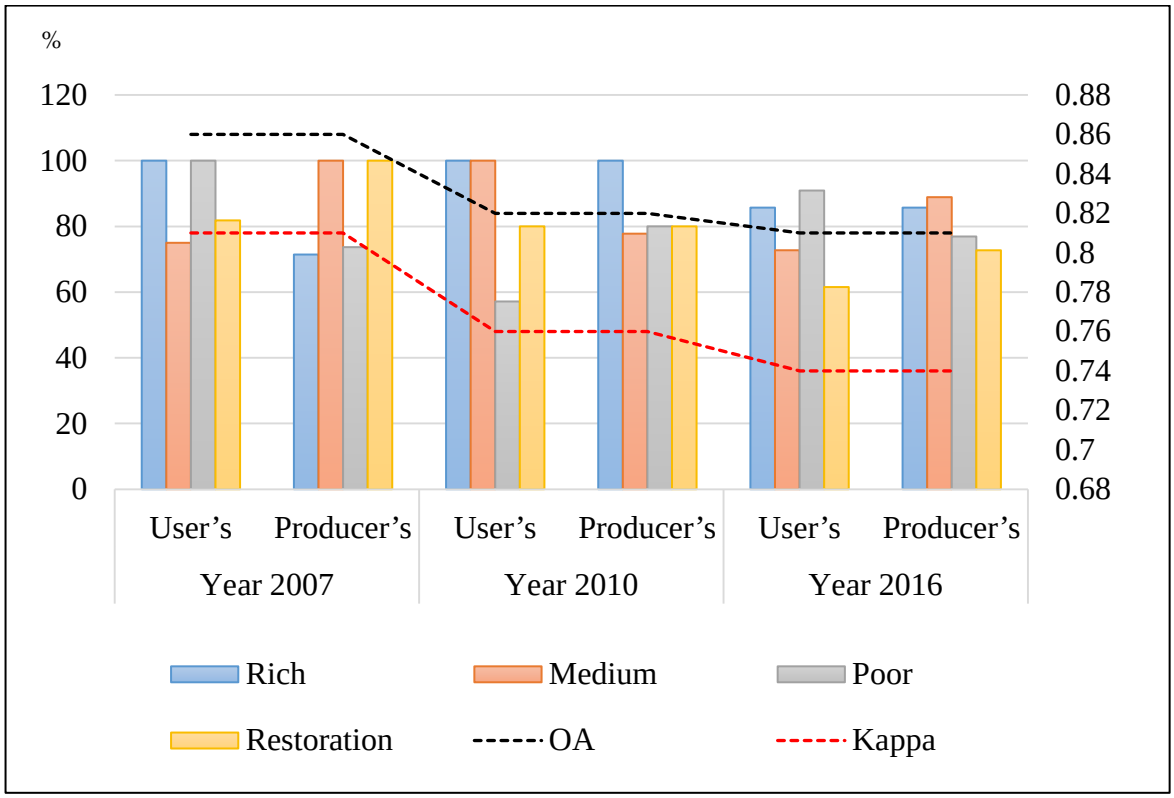


Figure 2.4. Forest types classification accuracies in user, producer (%), overall accuracy (OA), and kappa in the years 2007, 2010, and 2016

Overall, the classification accuracies were relatively high for 2007 with a kappa of 0.81 and OA of 0.86, respectively (Figure 2.4), while they were adequate for 2010 and 2016 with kappas of 0.76 and 0.74, respectively. The accuracies are generally the best for the rich forest over the entire time, with a user's accuracy of 100% in the years 2007 and 2010, and of 86% in 2016. This is followed by medium forest with over 75% in both user's and producer's accuracies, although sometimes it was misclassified as rich or poor forest. On the other hand, the classification accuracies were the lowest in 2010 for poor forest and in 2016 for restoration forest. The confusion matrix in 2010 and 2016 reveals a significant confusion between the poor and restoration forests, and they therefore cause the values of OA and kappa to be reduced at these times (Appendix D).

2.4.2. Forest Pattern Analysis at the Class Level

Based on the metrics of class area (CA) and percentage of landscape (PLAND_%), the natural forest of the study displayed a significant fluctuation within the nine years from 2007 to 2016 (Figure 2.5).

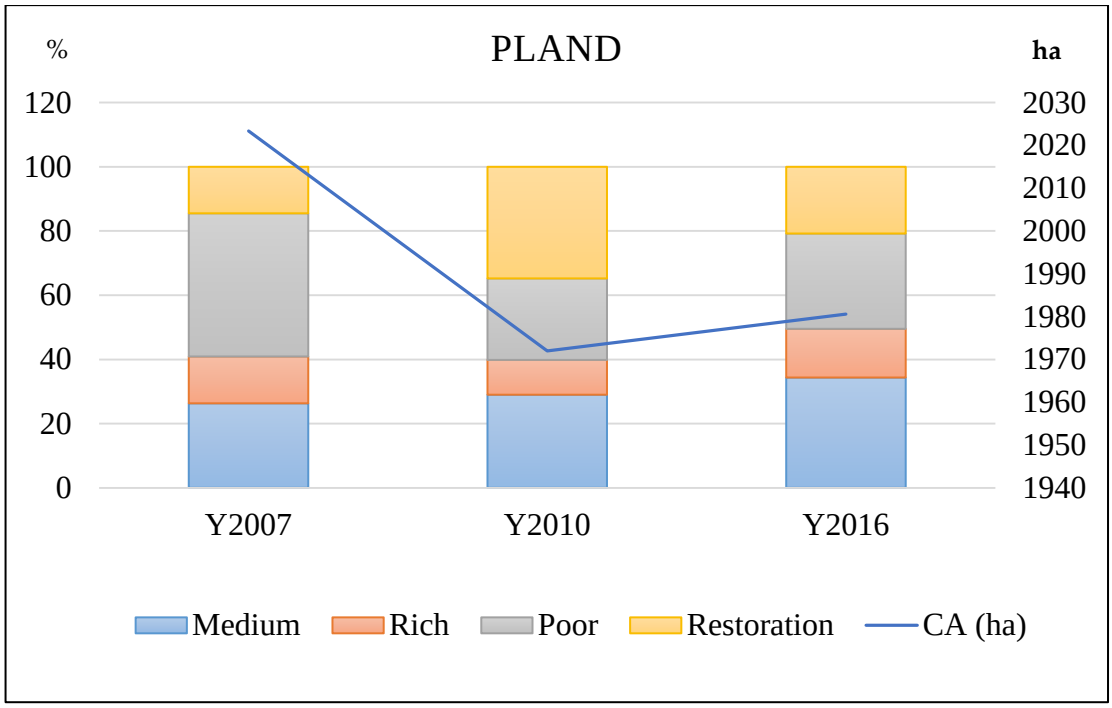


Figure 2.5. Variation of four forest grades in the total class area of natural forest (CA_ha) and percentage of landscape (PLAND_%) for each forest type

In the period 2007–2010, CA decreased quickly with an average loss of 1,713 ha per year. However, in the period 2010–2016, signs of recovery in CA appeared with an average gain of 144 ha each year. Generally, from 2007 to 2016, rich, medium, and restoration forests mainly demonstrated an increase, as shown by the gain of PLAND 1–8%, while PLAND showed reductions for poor forest of up to –15%. Furthermore, to assess the spatial variation of forest patterns, a set of parameters, comprising 11 metrics at the class level and 15 metrics at the landscape level, was selected based on evaluation of the indicators for universality, strength, and consistency. The selection method was based on factor analysis, clustering, and evaluation for the four different forest grades at three different time points. Therefore, this set of metrics ensures the appropriateness and representativeness of forest structure analysis over time for this test site. The changes in each forest grade, based on analyzing landscape metrics from 2007 to 2016, are shown in Table 2.4.

In general, from 2007 to 2016, the forest grades exhibited a relatively stable pattern with no significant changes in the group metrics of aggregation (AI, CLUMPY, PLADJ) but their pattern did show significant changes in patch shape structure (SHAPE, CIRCLE, CONTIG). In particular, rich, medium, and restoration forests had a low level of aggregation with the change percentage of AI and PLADJ ranging from just +1 to +4%. Conversely, the poor forests demonstrated an increased dispersion (AI −7%). However, this period was

expressed by the moderate changes in shape with more compactness (SHAPE −4% to −12%) and contiguity (CONTIG_AM up to 4% excluding the poor forest). The poor forests demonstrated the largest variation and had a trend of disaggregation (nLSI 60%) due to decreasing total area and percentage of the landscape. In summary, when evaluating the subperiods between 2007–2010 and 2010–2016, the forest grades reflect an extreme fluctuation and totally different behavior.

2.4.2.1. Period 2007–2010

This period expressed a growth in the percentage of landscape occupied by medium and restoration forests, as well as a decline in rich and poor forests. Therefore, they exhibit completely different processes in spatial fluctuations (Table 2.4).

Rich forests displayed a moderate decrease (PLAND −4%) and strong disaggregation in this period. The disaggregation is reflected in a decrease in AI −15%, CLUMPY −12%, and PLADJ −15%, and an increase in nLSI (+67%). The patterns show more compactness and less physical connectivity, as a result of reducing complexity in geometric shape (SHAPE_MN −20% and CIRCLE_AM −14%), decreasing contiguity, and continuity (CONTIG_MN –51% and COHESION −12%). The related circumscribing circle coefficient of variation (CIRCLE_CV) with a high value indicates the various changes in patch shapes of rich forests.

Similar to rich forests, poor forests exhibited slightly increased dispersion corresponding to a decrease in clumpiness and aggregation (–7% for both the change of CLUMPY and AI). This is due to shrinkage in the percentage of landscape (PLAND −19%) and the disappearance of like adjacencies in the same patch type (PLADJ −6%). It also coincides with the tendency to increase compactness (SHAPE_MN −9%).

The medium and restoration forests had growth in terms of area (PLAND 3% and 20%, respectively) and demonstrated a different process than rich and poor forests. The patterns display a moderate aggregation, higher connectivity, and compactness. In addition, the growth in area, together with the drop in contiguity index (CONTIG_MN –3% and −24% for the medium and restoration, respectively), reflect the process of creating larger patches from the clumpiness of small adjacencies and the distribution scattered in the landscape.

Table 2.4. Pattern metrics changes in the four forest grades in class level metrics

Grades	Metrics	Change %		
		2007–2010	2010–2016	2007–2016
Rich forest	SHAPE_MN	−20	18	−5
	CIRCLE_MN	−28	27	−9
	CIRCLE_AM	−14	19	2
	CIRCLE_CV	43	−17	18
	CONTIG_MN	−51	58	−22
	CONTIG_AM	−21	29	2
	CLUMPY	−12	11	−3
	PLADJ	−15	16	1
	COHESION	−12	12	1
	AI	−15	16	1
	nLSI	67	−38	4
Medium forest	SHAPE_MN	−16	5	−12
	CIRCLE_MN	−3	−3	−6
	CIRCLE_AM	6	−9	−4
	CIRCLE_CV	−6	6	−1
	CONTIG_MN	−11	−1	−11
	CONTIG_AM	9	−5	4
	CLUMPY	7	1	9
	PLADJ	7	−4	3
	COHESION	5	−8	−3
	AI	6	−2	4
	nLSI	−29	16	−17

	SHAPE_MN	−9	6	−4
	CIRCLE_MN	−19	18	−4
	CIRCLE_AM	−6	−6	−11
	CIRCLE_CV	29	−24	−1
	CONTIG_MN	−32	29	−12
Poor forest	CONTIG_AM	−8	−2	−10
	CLUMPY	−7	6	−1
	PLADJ	−6	−2	−8
	COHESION	−2	−6	−8
	AI	−7	−1	−7
	nLSI	53	4	60
	SHAPE_MN	−8	3	−6
	CIRCLE_MN	−14	9	−6
	CIRCLE_AM	9	−7	2
	CIRCLE_CV	25	−15	6
	CONTIG_MN	−24	24	−6
Restoration forest	CONTIG_AM	8	−7	0
	CLUMPY	−3	−2	−4
	PLADJ	6	−5	1
	COHESION	5	−6	−1
	AI	5	−4	1
	nLSI	−25	33	0

2.4.2.2. Period 2010–2016

In this period, rich forests performed more aggregation than other forest grades. The appearance of new patches (PLADJ +16%) resulted in an increase in spatial connectedness (CONTIG_MN +58%) and improved the continuity of this class in the landscape (COHESION +12%). This also meant a gain in the aggregation process (AI +16%, CLUMPY +11%, and nLSI –38%). The growth in PLAND coincided with a higher area-weighted mean contiguity of each patch (CONTIG_AM +29%), indicating the appearance of larger patch sizes.

Medium and poor forests demonstrated less area variation than in the previous period with a slight increase. However, there was a negligible decrease in tendency of aggregation (AI –1 to –2%), continuity (COHESION –6 to –8%), and connectedness (CONTIG_AM –2% to –5%) for both grades. In poor forests, there was a different tendency of the mean index and area-weighted mean index in CIRCLE and CONTIG due to measuring the patch-centric and landscape-centric perspectives. The increase in the related circumscribing circle shows a trend of elongation based on evaluating entire patches (CIRCLE _MN +18%), but displays the opposite trend based on evaluating an arbitrary patch selected randomly from the landscape (CIRCLE_AM –6%).

Similarly, CONTIG_MN demonstrated a significant increase (+24%) and performed a higher level of spatial connectedness in poor forests. However, the drop in CONTIG_AM (–2%) together with the expansion of area partly revealed the presence of new small patches.

An extreme decline in the restoration area was recorded during this period, resulting in increasing dispersion (nLSI +33%), a higher level of complexity in shape structure (SHAPE_MN +3% and CIRCLE_MN +9%), and less contiguity (CONTIG_AM –7%).

2.4.3. Forest Pattern Analysis in Landscape Level

This period was marked by a rapid decrease in the total landscape area of natural forests, from 202,300 ha in 2007 to 197,200 ha in 2010, and to 198,100 ha in 2016. This caused a reduction in the percentage of the landscape and a sharp decline in patch size distribution (AREA_AM –60%) (Table 2.5). In addition, there was a decrease in symmetry in the patch distribution in the landscape (IJI –8%). The moderate decline in SHAPE (–20%) and GYRATE (–26%) demonstrated more compactness and less complexity in spatial patterns. The continuity and connectedness of the forest pattern also tended to decrease (CONTIG_MN –21%)). In general, the natural forest experiences increased fragmentation

over the entire landscape, which involved an increase in landscape area with shrinkage of patch size and disproportionate distribution of patches.

Table 2.5. Pattern metrics changes in landscape level metrics

Metrics	Change %		
	2007–2010	2010–2016	2007–2016
AREA_AM	−49	−22	−60
GYRATE_AM	−7	−20	−26
SHAPE_AM	7	−25	−20
CONTIG_MN	−30	14	−21
IJI	−5	−2	−8

2.5. Discussion

To assess the changes in natural forest area in the study area, we compared the results with those in global and tropical regions, as well as in Vietnam overall, in the same period. Keenan *et al*. (Keenan *et al*. 2015) reviewed the dynamics of global forest area between 1990 and 2015 based on statistics from the FAO global forest resources assessment 2015. Worldwide, the natural forest area declined by 2% between 2005 and 2015, with the majority of the losses occurring in the tropics where the rate of loss fell by 7.2 million $ha.y^{-1}$. Compared to the trend of forest transition worldwide and in Vietnam overall, the status of forest loss in the study area is similar in the period from 2007–2010. This status is confirmed by the findings of Quy Van Khuc *et al*. (Van Khuc *et al*. 2018) that degradation mainly occurred in natural forest at the rate of 3–4%. Cochard's (Roland Cochard *et al*. 2017) review also demonstrated a slow increase in natural forests in the period 2000–2013 in Thua Thien Hue province. From 2010–2016, the natural forest in the study area demonstrated the opposite trend, while there was a significant decrease in the natural forest worldwide and in the tropics generally. It is difficult to compare fluctuations in forest grades of rich, medium, poor, and restoration grades because there are few documented records for the study area, and Vietnam in general, particularly in the period 2010–2016. Therefore, the findings of this study contribute to the understanding of the transition of natural forest grades in recent years, particularly in the ecological processes in terms of spatial patterns that have still not received adequate attention.

Analysis of the reflectance behavior on some bands on Landsat and backscatter on SAR images (Appendix C) demonstrates on histograms the overlap of all four forest grades. In

image data from 2007 and 2010, rich forest exhibited better distinctions than the other forest grades. Histogram analysis of forest grades in 2016 shows little separation, so its accuracy was lower than that of 2007 and 2010. This low separation is due to the characteristic of natural forest, with its combination of various canopy stories and species diversity. Sparser wood trees have more vines, which cover the whole canopy. Therefore, it is difficult to detect the difference among forest grades based on optical images. Despite having a long wavelength that can supposedly penetrate the canopy and reach the trunk, L-band signals still demonstrate a low difference between polarization signals. In this study, to enhance the differences between classes, a multivariate model was essential to observe objects under multidimensional space and provide more information and attributes for objects.

The change of certain forest grades between any two periods comprises the net effect (Keenan *et al*. 2015) of conversion from any one forest grade to another or non-forested area and natural regeneration or restoration. A conversion matrix was used to clearly illustrate transition areas between forest grades in this study area in the period 2007–2016 (Table 2.6). In this table, the cross cells demonstrated no change values in terms of percentage of forest grades' area. The rows demonstrate an increase in the proportion converted from other grades. The columns demonstrate a decrease in the proportion converted into other grades. The net area change is the total effect of increase and decrease in the area of specific forest grades.

Table 2.6. Conversion matrix of forest grades between 2007 and 2016 in percentage (%) and area (ha)

2007 \ **2016**	**Rich**	**Medium**	**Poor**	**Restoration**	**Others**	**Area increase (ha)**
Rich	19	16	13	1	1	24,569
Medium	36	29	30	13	4	52,749
Poor	21	23	26	29	3	34,970
Restoration	14	18	15	26	2	33,585
Others	11	14	15	31	90	
Area decrease (ha)	−24,046	−37,955	−66,409	−21,732		
Net area change (ha)	**523**	**14,794**	**−31,439**	**11,853**		

From the conversion matrix of forest grades between 2007 and 2016 in this study area, we considered three main findings. First, the net change of natural forest comprised a small loss of area, but this was due to two opposite trends of an area increase in one place and a decrease elsewhere. Second, the levels of forest restoration and deforestation were nearly equal (total of 145,873 ha and 150,143 ha, respectively) and occurred simultaneously in the study area during period. Third, there was a strong internal transition between forest grades and an external transition between them and other land use/land cover grades. Medium forest had the highest gain of its area, followed by restoration forest, at 14,794 ha and 11,852 ha, respectively. Poor forest showed a sharp loss, while rich forest had an adequate increase. When considering the percentage of conversion area, the most dramatic transformation was in rich forest, which changed to medium forest at a rate of 36% but was compensated by medium (16%) and poor forests (13%). However, when considering the area change, medium and poor forest had areas of both high increase and decrease. Changes from natural forest to other grades were the strongest in restoration forest at 31% of its area. Restoration forest is the most vulnerable forest grade because it is often distributed in areas that are easily accessible and affected by human and agricultural activities.

The spatial changes in natural forests presented in Figure 2.6 showed two directions: increase and decrease. The area loss of natural forest occurred throughout the study area, but it was mainly distributed near water bodies such as rivers and streams. The local population distribution is often concentrated in the downstream of rivers where conditions for agriculture are developing. Therefore, natural forests near rivers are easily deforested and degraded due to human activity. In the other direction, the expansion of rich forest created larger fragments and scattered distribution in the study area, resulting in increasing compactness, less connectivity, and higher isolation. The expansion in other grades occurred more evenly and therefore with greater connectivity.

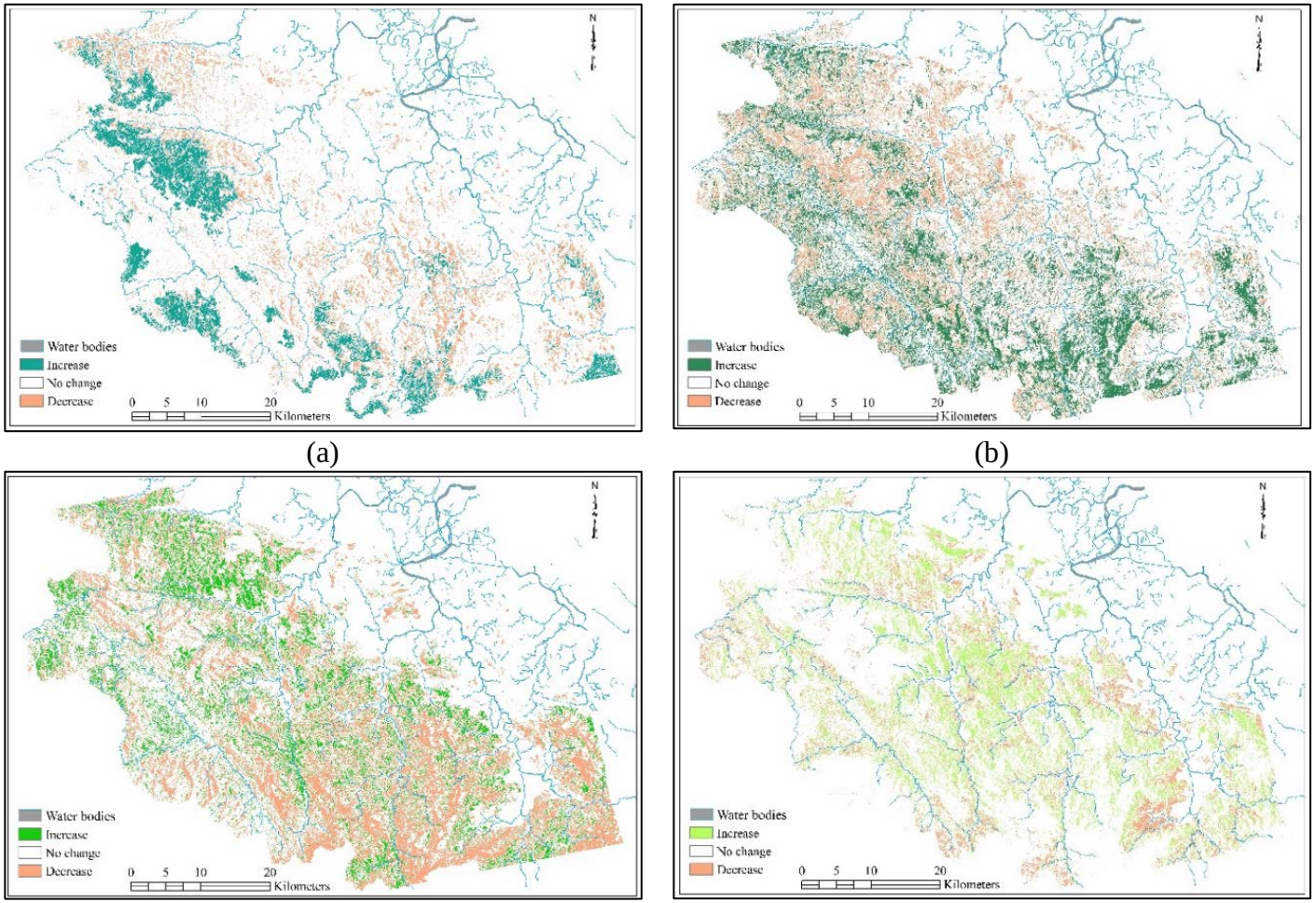


Figure 2.6. Changes in (a) rich forest, (b) medium forest, (**c**) poor forest, and (**d**) restoration forest between 2007 and 2016

During this period, there were many factors affecting forest dynamics. The policies of prohibiting logging in natural forests and enhancing forest protection and restoration are considered to be the correct policies in terms of reducing natural forest degradation, which was implemented by the Vietnamese government since the early 1990s (Vu and Pham 2001). However, illegal logging still occurred (Phuc and Junior 2006) due to the increasing demand for woods from population pressure, which is the main reason for the continued decline of natural forests in the period 2005–2010. There were also many other causes, such as poverty, forest resources, population density, agricultural production, and province-level governance (Van Khuc *et al*. 2018). In parallel with the logging ban policy in natural forest, Vietnam has successfully socialized forestry organization, calling for public participation in afforestation and forest protection, and resulting in reduced deforestation and degradation and improved long-term income for people in rural mountainous areas. The speed of loss of natural forests has also decreased slightly and there have been signs of increase from 2010 to the present day. In 2016, Vietnam began to introduce bans on wood exploitation from natural forests into the law on forest protection and development, which is the most powerful law in forestry. Simultaneously, it maximized the closure of natural forests, did not convert natural forests to other purposes, and especially did not convert poor natural forests to industrial crops. This is the driving force behind reductions in degradation and prevention of illegal logging, and allows us to predict recovery and increase in the quality of natural forests in the future.

Generally, this study provides information on the dynamics and spatial processes of natural forest in the study site between 2007 and 2016. The result obtained demonstrates the general trend of forest grades conversion and provides useful information for sustainable forest planning.

2.6. Conclusion

There is an essential requirement for forest management and protection to classify natural forests into the stem volume grades and assess their fluctuations over time. However, classifying the natural forest grades in tropical areas using remote sensing images is challenging because of the very similar information captured by remotely sensed data on many forests as well as the constraint of samples data. Furthermore, there is a lack of research assessing forest transition in the natural forest from the perspective of landscape ecology, which can be used for forest structure management, and to quantitatively characterize the

spatial patterns of forest landscapes. In this study, we addressed these issues by applying semi-supervised classification for data integration of optical and SAR data.

The combination of Landsat and PolSAR data resulted in improved discrimination of forest grades. The using of multi-source remotely sensed data can provide more information about the object, as well mitigate the disadvantages of Landsat images (cloud, lower spatial resolution), and limited information regarding objects in PALSAR/PALSAR-2 image (only two polarization HH and HV).

In this study, we assessed the potential of a proposed semi-supervised model developed and validated for mapping forest grades and assessed the process of forest transition in a tropical natural forest in Vietnam. The model produced high accuracies in the classified images in 2007, 2010, and 2016 with over 0.74 for kappa, and over 0.8 for OA. Additionally, landscape metrics were used to evaluate the forest changes based on the spatial processes, such as aggregation, fragmentation, and compaction. At the class level, the poor forest demonstrated the largest variation with more dispersed patterns, while other grades had a low level of aggregation. At the landscape level, the natural forest experiences increased fragmentation, which involved an increase in landscape area with shrinkage of patch size and disproportionate distribution of patches.

We recommend that future research includes comparison of different models to estimate the improvement resulting from the proposed model. Another important study that should be conducted is testing of the proposed methods on larger areas.

References

1. Aguilera, F.; Valenzuela, L.M.; Botequilha-Leitão, A. Landscape metrics in the analysis of urban land use patterns: A case study in a Spanish metropolitan area. *Landsc. Urban Plan.* **2011**, *99*, 226–238, doi:10.1016/j.landurbplan.2010.10.004.
2. Ainsworth, T.L.; Preiss, M.; Stacy, N.; Nord, M.; Lee, J. Analysis of Compact Polarimetric SAR Imaging Modes Compact Polarimetry-Enhancing Dual-Pol Imagery. In Proceedings of the POLinSAR Workshop 2007, Frascati, Italy, 22–26 January 2007.
3. Alok, A.K.; Saha, S.; Ekbal, A. Pixel Classification of Remote Sensing Satellite Image using Semi-supervised Clustering. In Proceedings of the 2014 9th International Conference on Industrial and Information Systems *(ICIIS)*, Gwalior, India, 15–17 December 2015.
4. Antrop, M.; Mander, Ü.; Marja, R.; Roosaare, J.; Uuemaa, E. Landscape Metrics and Indices: An Overview of Their Use in Landscape Research. *Living Rev. Landsc. Res.* **2009**, *3*, 1–28.
5. Antropov, O.; Rauste, Y.; Häme, T.; Praks, J. Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. *Remote Sens.* **2017**, *9*, 1–24, doi:10.3390/rs9100999.
6. Botequilha Leitão, A.; Ahern, J. Applying landscape ecological concepts and metrics in sustainable landscape planning. *Landsc. Urban Plan.* **2002**, *59*, 65–93, doi:10.1016/S0169-2046(02)00005-1.
7. Bwangoy, J.R.B.; Hansen, M.C.; Roy, D.P.; Grandi, G. De; Justice, C.O. Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices. *Remote Sens. Environ.* **2010**, *114*, 73–86, doi:10.1016/j.rse.2009.08.004.
8. Chambers, J.Q.; Asner, G.P.; Morton, D.C.; Anderson, L.O.; Saatchi, S.S.; Espírito-Santo, F.D.B.; Palace, M.; Souza, C. Regional ecosystem structure and function: Ecological insights from remote sensing of tropical forests. *Trends Ecol. Evol.* **2007**, *22*, 414–423, doi:10.1016/j.tree.2007.05.001.
9. Cochard, R.; Tri Ngo, D.; Waeber ETH Zurich, P.; Waeber, P.O.; Kull, C.A.; Cochard, R.; Ngo, D.; Waeber, P.; Kull, C. Extent and causes of forest cover changes in Vietnam's provinces 1993–2013: A review and analysis of official data Alaotra Resilience Landscape (AlaReLa) Madagascar View project Global Forest Watch in Madagascar View project REVIEW Extent and causes of forest cover changes in Vietnam's

provinces 1993–2013: A review and analysis of official data. *Environ. Rev.* **2017**, *25*, 199–217, doi:10.1139/er-2016-0050.

10. Cong Thang, T. *Review of Vietnam Forestry Policies*; 2015. Food and Fertilizer Technology Center for the Asian and Pacific Region, Vietnam.
11. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* **1991**, *37*, 35–46, doi:10.1016/0034-4257(91)90048-B.
12. Cui, B.; Xie, X.; Hao, S.; Cui, J.; Lu, Y. Semi-Supervised Classification of Hyperspectral Images Based on Extended Label Propagation and Rolling Guidance Filtering. *Remote Sens.* **2018**, *10*, 515, doi:10.3390/rs10040515.
13. Dópido, I.; Member, S.; Li, J.; Marpu, P.R.; Plaza, A.; Member, S.; Dias, J.M.B.; Benediktsson, J.A. Semisupervised Self-Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4032–4044, doi:10.1109/TGRS.2012.2228275.
14. Du, H.Q.; Ge, H.L.; Liu, E.B.; Xu, W.B.; Jin, W. A new Classifier for Remote Sensing Data Classification: Partial Least Squares. In *International Workshop on Earth Observation and Remote Sensing Applications*; IEEE: Beijing, China, 2008; p. 302.
15. Erkan, A.N.; Camps-Valls, G.; Altun, Y. Semi-supervised remote sensing image classification via maximum entropy. In Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing, Kittilä, Finland, 29 August–1 September 2010; pp. 313–318, doi:10.1109/MLSP.2010.5589199.
16. ESA Tutorial on SAR polarimetry. Available online: https://earth.esa.int/documents/653194/656796/ Polarimetric_Decompositions.pdf (6th October 2017).
17. FAO FAOSTAT. Available online: http://www.fao.org/faostat/en/#compare (accessed on 25 April 2019).
18. FAO Forest Resources Assessment 2015: Forest Resources Assessment Working Paper: Terms and Definitions.
19. Gao, J.; Liang, F.; Fan, W.; Sun, Y.; Han, J. A graph-based consensus maximization approach for combining multiple supervised and unsupervised models. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 15–28, doi:10.1109/TKDE.2011.206.
20. Gao, J.; Liang, F.; Fan, W.; Sun, Y.; Han, J. Graph-based Consensus Maximization among Multiple Supervised and Unsupervised Models. *Adv. Neural Inf. Process. Syst.*

2009, *22*, 585–593.

21. Geri, F.; Rocchini, D.; Chiarucci, A. Landscape metrics and topographical determinants of large-scale forest dynamics in a Mediterranean landscape. *Landsc. Urban Plan.* **2010**, *95*, 46–53, doi:10.1016/j.landurbplan.2009.12.001.
22. Gyenizse, P.; Bognár, Z.; Czigány, S. Landscape shape index , as a potential indicator of urban development in Hungary. *Landsc. Environ.* **2014**, *8*, 78–88.
23. Hirschmugl, M.; Sobe, C.; Deutscher, J.; Schardt, M. Combined Use of Optical and Synthetic Aperture Radar Data for REDD + Applications in Malawi. *Land* **2018**, *7*, 116, doi:10.3390/land7040116.
24. Hong, S.; Moon, W.M.; Paik, H.Y. Gi-Hyuk Choi Data fusion of multiple polarimetric SAR images using discrete wavelet transform (DWT). *IEEE Int. Geosci. Remote Sens. Symp.* **2003**, *6*, 3323–3325, doi:10.1109/igarss.2002.1027170.
25. Huete, A.R.; Liu, H.Q.; Batchily, K.V.; van Leeuwen, W. A comparsion of vegetation indices over a Global set of TM images for EO-MODIS. *Remote Sens. Environ.* **1997**, *59*, 440–451, doi:10.1016/S0034-4257(96)00112-5.
26. Keenan, R.J.; Reams, G.A.; Achard, F.; de Freitas, J.V.; Grainger, A.; Lindquist, E. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *For. Ecol. Manag.* **2015**, *352*, 9–20, doi:10.1016/j.foreco.2015.06.014.
27. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2016**, arXiv:1609.02907.
28. Laparra, V.; Gómez-Chova, L.; Malo-López, J.; Camps-Valls, G.; Muñoz-Marí, J. A Review of Kernel Methods in Remote Sensing Data Analysis. In *Optical Remote Sensing*; Springer: Berlin, Germany, 2011; pp. 171–206 ISBN 3496354402.
29. Lausch, A.; Herzog, F. Applicability of landscape metrics for the monitoring of landscape change: Issues of scale, resolution and interpretability. *Ecol. Indic.* **2002**, *2*, 3–15, doi:10.1016/S1470-160X(02)00053-5.
30. Li, G.; Lu, D.; Moran, E.; Dutra, L.; Batistella, M. A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 26–38, doi:10.1016/j.isprsjprs.2012.03.010.
31. Ma, L.; Ma, A.; Ju, C.; Li, X. Graph-based semi-supervised learning for spectral-spatial hyperspectral image classification. *Pattern Recognit. Lett.* **2016**, *83*, 133–142,

doi:10.1016/j.patrec.2016.01.022.

32. MacDicken, K.G. Global Forest Resources Assessment 2015: What, why and how? *For. Ecol. Manag.* **2015**, *352*, 3–8, doi:10.1016/j.foreco.2015.02.006.
33. Martinez del Castillo, E.; García-Martin, A.; Longares Aladrén, L.A.; de Luis, M. Evaluation of forest cover change using remote sensing techniques and landscape metrics in Moncayo Natural Park (Spain). *Appl. Geogr.* **2015**, *62*, 247–255, doi:10.1016/j.apgeog.2015.05.002.
34. Mather, A.S.; Needle, C.L. The forest transition: A theoretical basis. *Area* **1998**, *30*, 117–124, doi:10.1111/j.1475-4762.1998.tb00055.x.
35. Meyfroidt, P.; Lambin, E.F. Forest transition in Vietnam and displacement of deforestation abroad. *PNAS* **2009**, *106*, 16139–16144.
36. Meyfroidt, P.; Lambin, E.F. Forest transition in Vietnam and its environmental impacts. *Glob. Chang. Biol.* **2008**, *14*, 1319–1336, doi:10.1111/j.1365-2486.2008.01575.x.
37. Ministry of Agriculture and Rural development (MARD) in Vietnam. Circular No. 34/2009/TT-BNNPTNT of June 10, 2009, on criteria for forest identification and classification.
38. Neel, M.C.; McGarigal, K.; Cushman, S.A. Behavior of class-level landscape metrics across gradients of class aggregation and area. *Landsc. Ecol.* **2004**, *19*, 435–455, doi:10.1023/B:LAND.0000030521.19856.cb.
39. Phuc, X.; Junior, T.S. Illegal timber logging in Vietnam: Who profits from forest privatization connected with a logging ban? In *Survival of the Commons: Mounting Challenges and New Realities.* In Proceedings of the Eleventh Conference of the International Association for the Study of Common Proverty, Bali, Indonesia, 19–23 June 2006.
40. Qi, Z.; Tian, Y.; Shi, Y. Laplacian twin support vector machine for semi-supervised classification. *Neural Netw.* **2012**, *35*, 46–53, doi:10.1016/j.neunet.2012.07.011.
41. Sano, M.; Miyamoto, A.; Furuya, N.; Kogi, K. Using landscape metrics and topographic analysis to examine forest management in a mixed forest, Hokkaido, Japan: Guidelines for management interventions and evaluation of cover changes. *For. Ecol. Manag.* **2009**, *257*, 1208–1218, doi:10.1016/j.foreco.2008.10.005.
42. Saunders, C.; Gammerman, A.; Vovk, V. Ridge Regression Learning Algorithm in Dual Variables. In Proceedings of the Fifteenth International Conference on Machine

Learning, Madison, WI, USA, 24–27 July 1998; Volume 37, pp. 515–521, doi:10.1111/j.0033-0124.1985.00197.x.

43. Sawant, S.S.; Prabukumar, M. A review on graph-based semi-supervised learning methods for hyperspectral image classification. *Egypt. J. Remote Sens. Space Sci.* **2018**, 1–6, doi:10.1016/j.ejrs.2018.11.001.
44. Shimoni, M.; Borghys, D.; Heremans, R.; Perneel, C.; Acheroy, M. Fusion of PolSAR and PolInSAR data for land cover classification. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 169–180, doi:10.1016/j.jag.2009.01.004.
45. Silva, J.; Bacao, F.; Caetano, M. Specific land cover class mapping by semi-supervised weighted support vector machines. *Remote Sens.* **2017**, *9*, 1–16, doi:10.3390/rs9020181.
46. Smiraglia, D.; Ceccarelli, T.; Bajocco, S.; Perini, L.; Salvati, L. Unraveling Landscape Complexity: Land Use/Land Cover Changes and Landscape Pattern Dynamics (1954–2008) in Contrasting Peri-Urban and Agro-Forest Regions of Northern Italy. *Environ. Manag.* **2015**, *56*, 916–932, doi:10.1007/s00267-015-0533-x.
47. Sun, P. Sparse Kernel Least Squares Classifier. In Proceedings of the Fourth IEEE International Conference on Data Mining, Houston, TX, USA, 27–30 November 2005; pp. 539–542, doi:10.1109/icdm.2004.10054.
48. Tlapakova, L.; StejskalovA, D.; Karasek, P.; Podhrazka, J. Landscape metrics as a tool for evaluation landscape structure-case study hustopece. *Eur. Ctry.* **2013**, *5*, 52–70, doi:10.2478/euco-2013-0004.
49. Triguero, I.; García, S.; Herrera, F. Self-labeled techniques for semi-supervised learning: Taxonomy, software and empirical study. *Knowl. Inf. Syst.* **2015**, *42*, 245–284, doi:10.1007/s10115-013-0706-y.
50. Van Khuc, Q.; Tran, B.Q.; Meyfroidt, P.; Paschke, M.W. Drivers of deforestation and forest degradation in Vietnam: An exploratory analysis at the national level. *For. Policy Econ.* **2018**, *90*, 128–141, doi:10.1016/j.forpol.2018.02.004.
51. Van Zyl, J.; Kim, Y. *Synthetic Aperture Radar Polarimetry*; California Institute of Technology, the United States. Google eBook; 2011; Volume 2, p. 313.
52. Vu, H.T.; Pham, X.P. Impacts and effectiveness of logging bans in natural forests in Vietnam. In *Asia-Pacific Forestry Commission Forests out of Bounds: Impacts and Effectiveness of Logging Bans in Natural Forests in Asia-Pacific*; Durst, P.B., Waggener, T.R., Enters, T., Cheng, T.L., Eds.; RAP Publication, Bangkok, Thailand: 2001; pp. 185–207.

53. Webb, E.L.; Honda, K. Biophysical and policy drivers of landscape change in a central Vietnamese district. *Environ. Conserv.* **2007**, *34*, 164–172, doi:10.1017/S037689290700389X.
54. Wu J. Landscape ecology. In: Jorgensen SE, editor. *Encyclopedia of ecology*. Oxford: Elsevier; **2008**. 2103–2108, doi: http://dx.doi.org/10.1016/B978-008045405-4.00864-8.
55. Yang, Y.; Zhou, Q.; Gong, J.; Wang, Y. Gradient analysis of landscape spatial and temporal pattern changes in Beijing metropolitan area. *Sci. China Technol. Sci.* **2010**, *53*, 91–98, doi:10.1007/s11431-010-3206-2.
56. Yarowsky, D. Unsupervised word sense disambiguation rivaling supervised methods. In Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics, Cambridge, MA, USA, 26–30 June 1995; pp. 189–196.
57. Yong-an, Z.; Wen-ming, Z.; Rui-hua, W. False Color Fusion for Multi-band SAR Images Based on Contourlet Transform. *Acta Autom.* **2007**, 33, 337–341, doi:10.1360/aas-007-0337.

CHAPTER 3.

COMBINATION OF SAR POLARIMETRIC PARAMETERS FOR ESTIMATING TROPICAL FOREST ABOVE-GROUND BIOMASS

3.1. Introduction

Forest planning plays an important role in sustainable management of forest and land resources, and provide protection of areas identified as significant for conservation (FAO 1999). In forest planning, it is important to observe and predict the change of forest land, as well as forest biophysical parameters. The changes of forests can be effectively observed by remote sensing techniques. For instances, Landsat multitemporal data was used to assess the land cover changes and CO_2 emissions in the tropical forest (Askar *et al*. 2019), or to obtain the construction information (Liu *et al.* 2019). Besides, the integration of remote sensing and geographical information system (GIS) provides a useful tool to develop a forest planning. A recent study was 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 planning (Cetin *et al*. 2018).

Biomass is one of important forest biophysical parameters which 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 (Toan *et al*. 2011). Tropical forest accounts for two-thirds of all terrestrial biomass (Pan *et al*. 2013) but is frequently affected by human activities and climate change. In recent years, research for tropical forest areas has focused on biomass 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) (Brown and Lugo 1984; Molto *et al*. 2013; Chave *et al*. 2014). 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 demonstrates high accuracy of the estimation 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 initial European Space Agency's 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 (Mermoz *et al*. 2015; Ho Tong Minh *et al*. 2016). In some instance (eg. 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 (Neumann *et al*. 2012) 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 (Brown *et al*. 1993). 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 (FAO 2015). 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 (Cochard *et al*. 2017). Such variations resulted in a reduction in the carbon existing amount 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 (Gibbs *et al*. 2007; Phuong *et al*. 2012; Huy *et al*. 2016; Huy *et al*. 2016; Kralicek *et al*. 2017). 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 to map tropical forest biomass (Viet Nguyen *et al*. 2016). In addition, forest disturbances and

regrowth can be found through HH and HV backscattering from the L-band ALOS PALSAR mosaic (Mermoz and Le Toan 2016).

SAR is one of the most promising remote sensors for mapping forest carbon (Mermoz *et al*. 2015). In the past decade, the increase in SAR techniques has resulted in an improvement in the 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 a superior performance because it combines the advantages of PolSAR and InSAR and is an important technique for determining the height of reflectors (Sauer *et al*. 2009; Neumann *et al*. 2010). In recent years, a new technique—SAR tomography—has been proposed for forest height measurement which has enhanced the accuracy of biomass prediction, especially in tropical forest areas (Reigber and Moreira 2000; Ho Tong Minh *et al*. 2016).

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 the estimations. The scattering components obtained from polarimetric decomposition are used to estimate AGB (Chowdhury *et al*. 2013; Sai Bharadwaj *et al*. 2015). Some indices such as the biomass index, canopy structure, and volume scattering index, that can be retrieved from SAR data, are also expected to show a linear regression with AGB (dos Santos *et al*. 2014). The PolSAR approach only performs well for AGB which is less than 200–300 tons/ha because of saturation in signal intensity (Sai Bharadwaj *et al*. 2015; Ho Tong Minh *et al*. 2014, 2016). However, the PolSAR 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 coherency in a pair of images. Therefore, the PolSAR approach is still being used in parallel with PolInSAR for monitoring and measuring vegetation.

Many studies have indicated the high power of PolSAR for predicting biomass. In this paper, we focused on investigating the behavior of SAR backscattering over multiple forest grades 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 types based on the species composition: tropical evergreen broad-leaved forests and bamboo forests. Tropical evergreen broad-leaved forests can be divided into four levels of stem volume including rich, medium, poor and restoration forest. The aim of this study was to show the contribution of polarimetry to biomass estimation following to the individual forest grades in which previous assessments were not recorded or were incomplete. The study also selected a suitable function for AGB estimation in different forest grades 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.

3.2. Materials and Methods

3.2.1. Study area

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

In an effort to reduce climate change and improve livelihoods for people living near forests, the Vietnamese government has implemented programs on green house gas 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 storage and carbon emission reductions is a challenge and is essential to be additional.

The total area of the province is 5,033 km^2, including 2,026 km^2 of natural forest. The natural vegetation mainly consists of lowland tropical forest with characteristic species including *Rhodomyrtus tomentisa, Melastoma candidum,* and *Imperata cylindrica* as dominant species. The natural forest links to the Annamite range providing high forest continuity and integrity. The protected forest in Thua Thien Hue Province includes Bach Ma National Park and Phong Dien Natural Reserve.

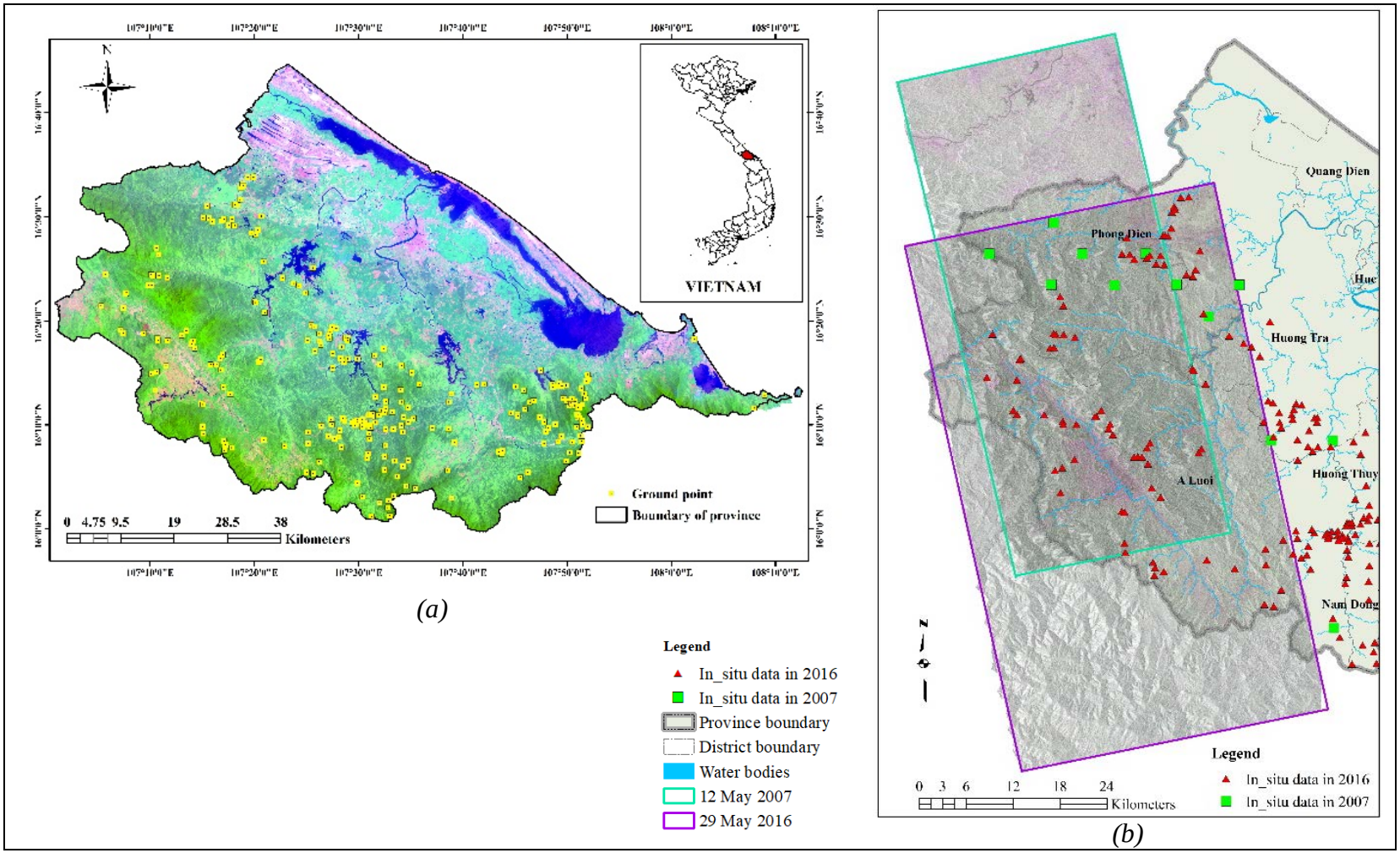


Figure 3.1. Location map of the study area in Landsat TM in pseudo-colors band 7,5,2 (a), and the cover of SAR images and in-situ data in 2007 and 2016 (b)

According to Circular Number 34/2009/TT-BNNPTNT of June 10, 2009 (MARD 2009) 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 grades, including:

- Rich forests (RF), which are forests with stem volume of between 201 and 300 m^3/ha;
- Medium forests (MF) (or average forest), which are forests with stem volume of between 101 and 200 m^3/ha;
- Poor forests (PF), which are forests with stem volume 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 trees with a mean diameter of less than 8 cm and stem volume of less than 10 m^3/ha.
- Bamboo forests (BAM): in our study area, BAM is 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. Those data showed that poor forest often had stem volume between 50 and 100 m^3/ha. All forest with stem volume lower than 50 m^3/ha was then assigned as the restoration forest. We called this forest grade “restoration” instead of “no reserve forest” because it still contained some wood stocks.

3.2.2. 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 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 continued to launch 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^o with a size of 3.55

× 9.3 meters in azimuth and range, respectively. The 2016 scene has the incidence angle at the image center was 38.99^{o} 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 meters 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 transpose of the complex conjugate as $\mathbf{T} = K_p \times K_p^{*T}$ where * represents the conjugate and denotes the Hermitian transpose (Bagan *et al*. 2012). The simple form of the coherency matrix is:

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \quad (3.1)$$

T11, T22, T33 are $S_{HH}+S_{VV}$, $S_{HH}-S_{HV}$, $S_{HV,}$ 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 (Neumann *et al*. 2012). $S_{HH}+S_{VV}$ is related to single bounce scattering on a rough surface, $S_{HH}-S_{VV}$ is an indicator of double bounce while S_{HV} is an indicator of volume scattering.

In this study, the general Yamaguchi four-component decomposition with unitary transformation model was used. 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 (Yamaguchi *et al*. 2005; 2006). In 2013, Singh and colleagues developed the generalized four-component

decomposition by a unitary transformation of the coherency matrix. 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 (Singh *et al*. 2013).

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**]) (Antropov *et al*. 2017).

The pedestal height of a polarization signature plot is the lowest Z-axis value in the polarization signature. It indicates the polarization purity and is related to depolarization. Different grades 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 (Durden *et al.* 1989).

The RFDI was developed by Saatchi and partner to assess the strength of the double bounce term (Mitchard *et al*. 2012):

$$\text{RFDI} = \frac{\langle|S_{hh}|^2\rangle - \langle|S_{hv}|^2\rangle}{\langle|S_{hh}|^2\rangle + \langle|S_{hv}|^2\rangle} \tag{3.2}$$

The CSI is a measure of the relative importance of vertical versus horizontal structure in the vegetation, and is related to the vertical trunks or stems (Pope *et al*. 1994):

$$\text{CSI} = \frac{\langle|S_{vv}|^2\rangle}{\langle|S_{hh}|^2\rangle + \langle|S_{vv}|^2\rangle} \tag{3.3}$$

The VSI is a measure of the depolarization of the linearly polarized incident radar signal and is an indicator of canopy thickness or density (Pope *et al*. 1994):

$$\text{VSI} = \frac{\langle|S_{hv}|^2\rangle}{\langle|S_{hv}|^2\rangle + BMI} \tag{3.4}$$

The 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 (Pope *et al*. 1994):

$$\text{BMI} = \frac{\langle |S_{hh}|^2 \rangle + \langle |S_{vv}|^2 \rangle}{2} \quad (3.5)$$

3.2.3. 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, there were ten measured plots covered by the PALSAR scene. The size of a plot is of 1 km^2 (1,000 × 1,000 m). In each plot, 40 subplots with a size of 25 × 20 m were set to measure forest parameters. In 2016, there were 79 plots 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. In each plot, there were four sub-plots 5 m × 5 m in size. Diameter at breast height (DBH, in centimeters) was measured for all trees with a diameter of 6 cm and over, while tree height (H, in meters) was measured for five normal-shaped 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 1,400 m, and the slope angle was up to 35^o. The main ecosystem was dense tropical rainforest with three layers 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 for the central coastal region in Vietnam. The biomass estimation formula by Huy *et al.* (2016) based on a single diameter shows higher accuracy than the common pantropical formula with three dimensiosns. In addition, this formula reduces the error propagation from the height and Wood specific gravity (WSG). DBH was measured directly in the field, thereby reducing the deviation of the sample (Molto *et al.* 2013). The allometry of Bao Huy et al. 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 allometry to estimate AGB (in tons/ha) given by Huy *et al.* (2016) is:

$$\text{AGB} = 0.104189 \text{x } DBH^{2.491453} \quad (3.6)$$

For estimating the biomass of bamboo forest, we used the formula given by Ly *et al*. (2012):

$$AGB = 0.3002\ (DBH)^2 + 0.115(DBH) + 1.7632 \quad (3.7)$$

With the allometry with three factors of DBH, height, and WSG, the uncertainties and bias in the estimated AGB 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.

3.2.4. 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 a size 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 which were extracted from SAR data.

For AGB estimation, two approaches were employed 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, to improve the biomass estimation, support vector regression (SVR) and random forest were used. These are well-known algorithms based on machine learning approaches and have been widely used in many fields.

Support vector regression (SVR) is related to statistical learning theory based on Kernel method to transform a non-linear regression problem into a linear one in a higher dimensions 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; then a linear model can fit this new high-dimensional feature space as follows (Smola and Schölkopf 2004; G. Chen and Hay 2011):

$$y = f(x) = \langle w.\varphi(x)\rangle + b = \sum_{i=1}^{n} w_i\varphi_i(x) + b \quad (3.8)$$

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 follow:

$$K(x_i, x) = \langle \varphi(x_i). \varphi(x) \rangle \quad (3.9)$$

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

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_i\|^2) \quad (3.10)$$

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

$$\text{minimize } \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (3.11)$$

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

where the 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. 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 (Liaw and Wiener 2002). 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 grades 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 (Paluszynska and

Biecek 2017) 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 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:

$$R^2 = \frac{\sum(\hat{y}-\bar{y})2}{\sum(y-\bar{y})2} \tag{3.12}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.13}$$

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 (Yadav and Shukla 2016). The value of k is the number of groups will be split from a given samples data. The higher the value of k, the higher the accuracy in cross-validation (Yadav and Shukla 2016). 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.

3.3. Results and Discussion

3.3.1. Above-ground biomass in the field data

Table 3.1 and Figure 3.2 illustrate the forest parameters in the five forest grades. The Lorey's height was used to calculate the average height of plots per forest grade. The average DBH per plot is shown as a D_g value. The parameters showed a gradual decrease depending on the stem 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. In BAM forest, parameters other than AGB 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 AGB of bamboos contributed significantly to AGB.

In this study, BAM forest represented a mixed species composition between trees and bamboo, mainly distributed on foothills with an altitude less than 300 m. This forest grade had a two-storey structure with a tree storey and a bamboo storey. In the tree storey, tree density of DBH $\geq$ 6 cm was 33 trees/plot (1,000 m^2). Bamboo density was quite high at 116 clumps/ plot with species such as *Dendrocalamus barbatus*, *Bambusa blumeana*, and *Melocalamus compartiflorus*.

The diameter and stem volume are the important parameters closely relating with the AGB, therefore, the correlation coefficients among them were examined (Table 3.2). As expected, AGB had a close relationship with DBH and stem volume for all forest grades. The correlation coefficient was between 0.77 and 0.93 for AGB and DBH, and around 0.99 for AGB and stem volume.

Table 3.1**.** Biophysical forest parameters in different forest grades

Grade	No of trees/ plot	**Height** (m)			**DBH** (cm)			**Basal area** (m^2 ha^{-1})	**Stem Volume** (m^3 ha^{-1})	**AGB** (tons/ha)
		h_{Lorey}	h_{min}	h_{max}	D_g	d_{min}	d_{max}			
Rich forest (RF)	75	17.97	15.11	30.92	23.03	6	147.1	32.84	269.85	241.68
Medium forest (MF)	53	15.43	13.08	18.86	19.53	6	109.8	20.94	148.88	139.17
Poor forest (PF)	44	14.09	11.41	16.66	16.57	6	64.6	11.47	73.99	68.41
Restoration (ReF)	29	11.97	8.71	14.59	12.24	6	80.9	5.42	30.48	26.87
Bamboo (BAM)	33	15.36	11.00	17.50	19.85	6	57.3	9.705	67.325	119.15

Table 3.2. Correlation coefficient (R) between in-situ above-ground biomass (AGB), diameter at breast height (DBH) and stem volume in the five forest grades

Grade	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

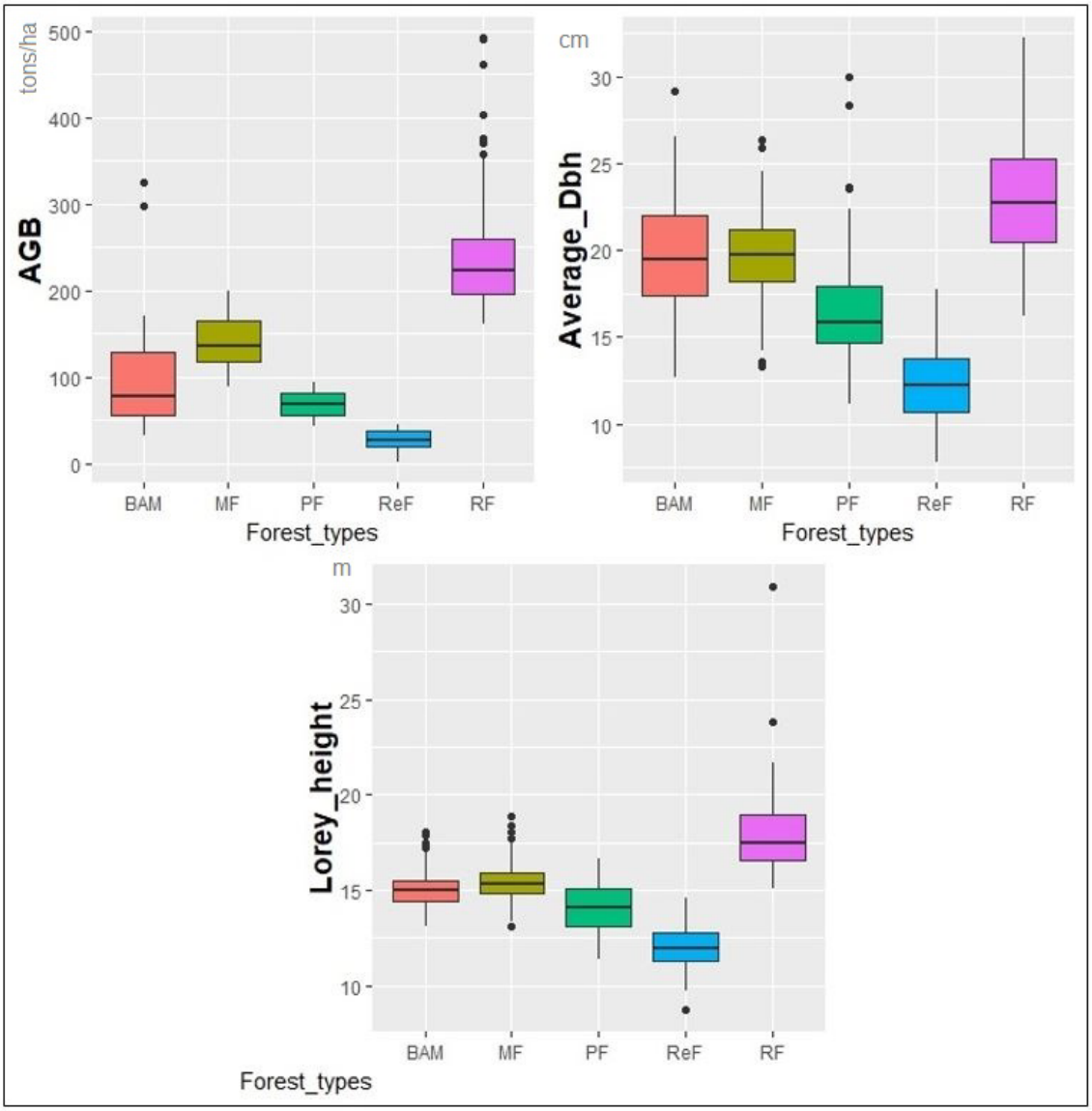


Figure 3.2. Boxplots of average above-ground biomass (AGB), Diameter at breast height (DBH), and Lorey's height ($H_{Lorey)}$ of plots in the five forest grades

3.3.2. Correlation between polarimetric backscattering and forest above-ground biomass

The correlation between the AGB estimated from the ground data (DBH) and SAR polarimetric parameters was investigated. In this study, there were two approaches to estimate AGB, including the total samples and the separate samples of five forest grades. In the second approach, we divided the forest into five grades: RF, MF, PF, ReF, and BAM. The interactions of AGB in each grade 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 grades. 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 AGB and PolSAR parameters was 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 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 grades 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 this study, it can be seen that AGB in each forest grade 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 of BAM, while double-bounce scattering and CSI were more sensitive to AGB of the ReF. Therefore, in this study, forests were divided into different forest grades for the uniformity of the forest structure. We recommend a variety of forest classifications according to species composition and stem volume because of their high correlation with forest biomass.

In general, the correlation between AGB and PolSAR signals was improved for the forest grades with small stem volume. Among the coherency matrix elements, T22 and T33 improved more than T11 for each forest grade because of the relationship with the illuminated medium grade 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 sensitive more than other indicators to the biomass.

In summary, based on the analysis of a mono-linear regression on each forest grade, there was an improvement in the correlation between PolSAR and the AGB. Each forest grade appeared 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 model for each forest grade. However, 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.

3.3.3. The 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).

Table 3.3. Set of selected parameters in different forest grades

Forest grades	**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

For analyzing the total data sample, the Yamaguchi volume scattering, CSI and HHVV ratio proved to be important. However, when separating data into five grades of forest, each different grade had a different set of parameters which 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 grades.

3.3.4. 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.

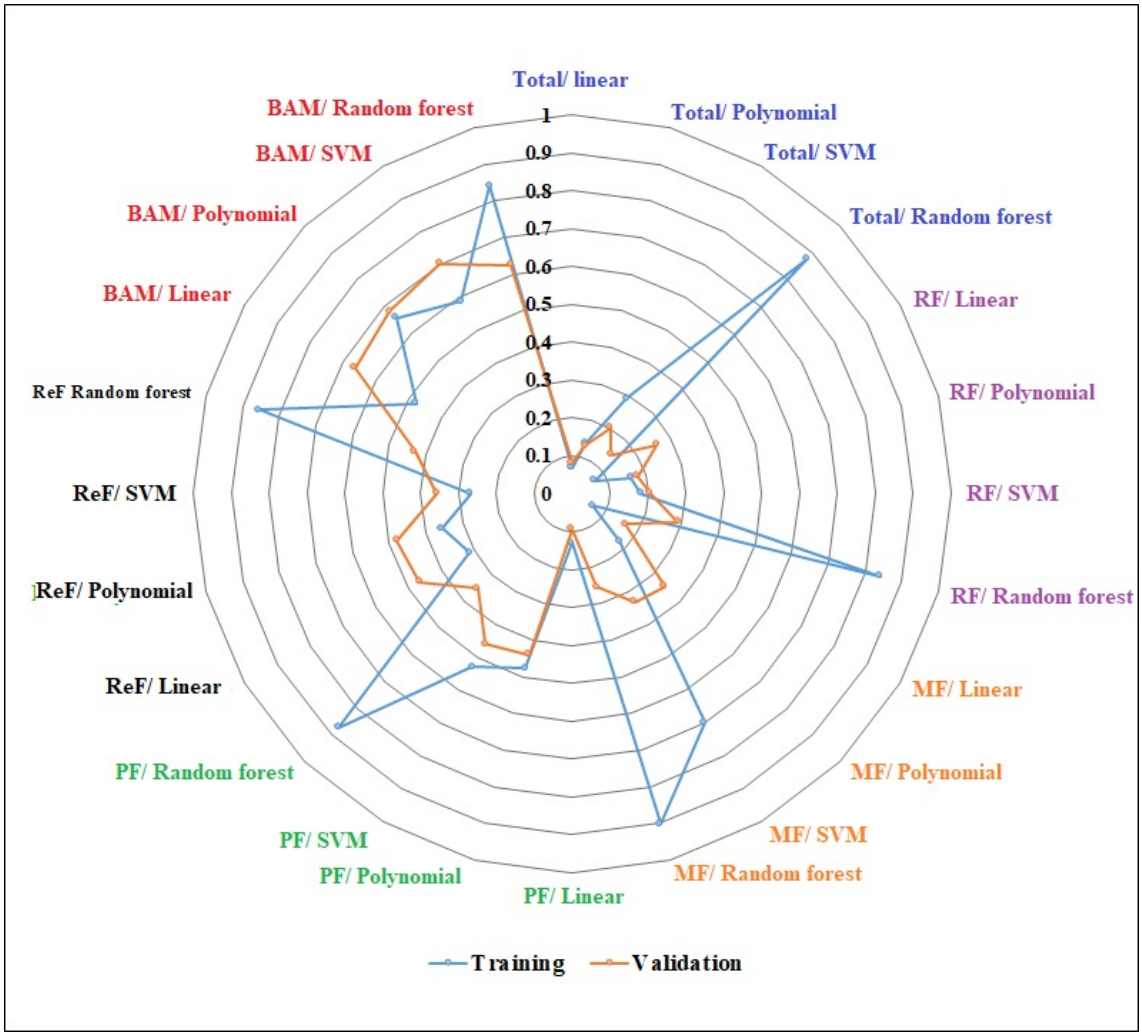


Figure 3.3. Result of regression models in the coefficient of determination (R^2) using training data and 10-fold cross-validation following forest grades

Figure 3.3 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 grades separately. BAM forest showed a 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 of 100–150 tons/ha in tropical forest area because it leads to over or under-estimation (Mermoz *et al*. 2015; Ho Tong Minh *et al*. 2014).

In addition, based on the low R^2 value of the linear model compared to the others, we found that AGB has a positive non-linear correlation with PolSAR signals. Random forest showed a good performance in most of the observation. However, its robustness was lower than the others when observing the difference between training and validation result. 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 3.4) showed that multivariate regression considerably increased the estimation accuracy in all forest grades. The RMSE was around 70 tons/ha for the total sample approach. For the RF, the RMSE improved up to 62.40 tons/ha and up to 10–13 tons/ha in random forest models for the MF, PF, and ReF. Generally, with observations in particular forest grades, the enhancement of RMSE in AGB measurement rose by 9%–18% for RF, and 80%–85% for the remaining forest grades.

The boxplot graph in Figure 3.4 illustrates the variation of RMSE in 10-fold cross-validation using different models with the total sample and five forest grades. The RMSE showed a considerable variation in all models for the RF with the range of 33.52–115.68 tons/ha. A similar pattern was also noticed in the linear model for BAM. The highest variance was in the polynomial model for BAM between 43.34–4016.66 tons/ha, but we could not illustrate this in this figure. Conversely, a low variance of RMSE was shown for the remaining forest grades.

Table 3.4. Root mean squared error (RMSE) of regression models using training and 10-fold cross-validation for different forest grades

Grades	Linear	Polynomial	SVM	Random forest
Total	78.18/ 77.29	75.61/ 75.33	69.03/ 74.26	34.36/ 75.78
Rich	63.79/ 64.18	62.53/ 68.24	64.39/ 61.28	31.62/ 62.40
Medium	13.97/ 14.28	13.44/ 14.81	8.27/ 13.73	6.17/ 12.98
Poor	17.13/ 17.50	13.57/ 14.06	12.38/ 14.58	7.04/ 15.24
Restoration	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

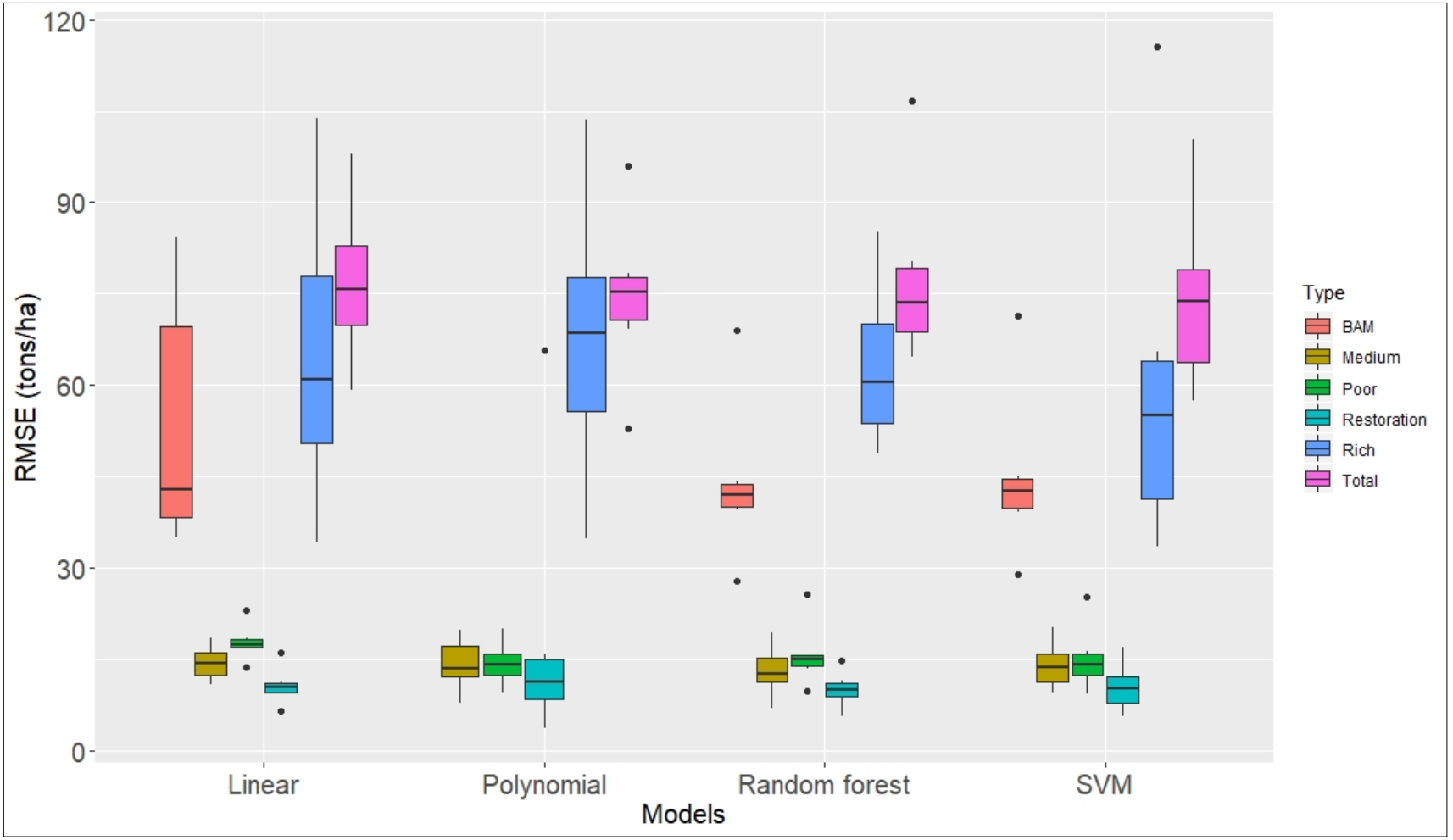


Figure 3.4. Root mean squared error (RMSE) of regression models in tons/ha using 10-fold cross-validation for different forest grades

In conclusion, following the analysis of R^2 and RMSE, different models were suitable for different forest grades to estimate 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, respectively. For the remaining forest grades, random forest was not reliable for use based on the comparison between training and validation results. For the ReF, the polynomial model displayed the good correlation between AGB and PolSAR variables with R^2 of 0.47 and RMSE improved up to 10.11 tons/ha. The SVR should be selected for MF and PF with RMSE between 8.27 and 12.38 tons/ha. For the RF, the random forest exhibited a high correlation in training data with RMSE decrease to 31.62 tons/ha. However, the overfitting, which occurred between training and validation, needs to be solved to ensure the reliability of the model. To solve this problem, further study should be carried out only focusing on RF. RF has a highly complex structure caused by the combination by the high diversity of species composition and many vertical storeys. Because of a dense canopy, the L-band backscattering only reaches to the canopy and the higher layers, leading to the poor correlation between backscattering signals and lower layers. The P-band radar or Lidar signals data are supposed to solve this limitation of L-band, however, 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.

3.4. Conclusions

In this study, we determined the models of AGB estimation in different forest grades based on correlation with PolSAR parameters extracted by L-band in the PALSAR-2 image. Significant variables were selected for each forest grade 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 each forest grades. We also analyzed the correlation of biomass to individual PolSAR signals in specific forest grades with AGB estimated from allometries to find the best model for each forest grade.

Forest was categorized into five forest grades. Based on species composition, forest was divided into broad-leaved forest and bamboo forest (mix of broad-leaved species and bamboo species). Based on stem volume levels, the broad-leaved forest was divided into rich, medium, poor, and restoration. The AGB was up to 500 tons/ha. To improve the

accuracy of the biomass estimation, we divided the data and assessed the correlation using the multivariate equations in different forest grades.

The correlation analysis of individual PolSAR parameters to forest biomass revealed different sensitivity in different forest grades. Therefore, the different sets of parameters were proposed for use for particular forest grades. 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 grades 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 and 12.38 tons/ha, 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 AGB in both training and validation data for the remaining forest grades. The polynomial function proved to be a suitable model for estimating AGB for the ReF with R^2 of 0.47 and RMSE of 10.11 tons/ha. In this study, the low correlation of AGB with PolSAR parameters in the RF grade 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 grades (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 AGB was exhibited in this study. This verified its potential for estimating AGB of bamboos using SAR data, which was not documented in previous studies, as well as creating a foundation for further study in various bamboo species.

References

1. FAO *Code of Practice for Forest Harvesting in Asia-Pacific*; Food and Agriculture Organization of the United Nations: Thailand, 1999; ISBN 9748666948.
2. Antropov, O.; Rauste, Y.; Häme, T.; Praks, J. Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. *Remote Sens.* **2017**, *9*, 1–24, doi:10.3390/rs9100999.
3. Askar, A.; Nuthammachot, N.; Sayektiningsih, T.; Hermudananto, H. Assessing Land Cover Changes and CO2 Emissions in Tropical Forests, 1998-2016: A Case Study of the Sungai Wain Protection Forest. *Polish J. Environ. Stud.* **2019**, *28*, 3597–3604, doi:10.15244/pjoes/94845.
4. Bagan, H.; Kinoshita, T.; Yamagata, Y. Combination of AVNIR-2, PALSAR, and Polarimetric Parameters for Land Cover Classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1318–1328, doi:10.1109/TGRS.2011.2164806.
5. Brown, S.; Iverson, L. R.; Prasad, A.; Liu, D. Geographical distributions of carbon in biomass and soils of tropical asian forests. *Geocarto Int.* 1993, *8*, 45–59.
6. Brown, S.; Lugo, A. E. Biomass of tropical forests: a new estimate based on forest volumes. *Science (80-.).* **1984**, *223*, 1290–3, doi:10.1126/science.223.4642.1290.
7. 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.* **2018**, *190*, doi:10.1007/s10661-018-6534-5.
8. Chave, J.; Réjou-Méchain, M.; Búrquez, A.; Chidumayo, E.; Colgan, M. S.; Delitti, W. B. C.; Duque, A.; Eid, T.; Fearnside, P. M.; Goodman, R. C.; Henry, M.; Martínez-Yrízar, A.; Mugasha, W. A.; Muller-Landau, H. C.; Mencuccini, M.; Nelson, B. W.; Ngomanda, A.; Nogueira, E. M.; Ortiz-Malavassi, E.; Pélissier, R.; Ploton, P.; Ryan, C. M.; Saldarriaga, J. G.; Vieilledent, G. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Chang. Biol.* **2014**, *20*, 3177–3190, doi:10.1111/gcb.12629.
9. 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.* **2011**, *77*, 733–741, doi:10.14358/PERS.77.7.733.
10. 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.* **2013**, *5*, 5725–5756, doi:10.3390/rs5115725.

11. 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.* **2017**, *25*, 199–217, doi:10.1139/er-2016-0050.
12. dos Santos, J. R.; Gama, F. F.; da ConceiçãO Bispo, P. Estimating forest biomass by remote sensing radar data in brazil. *Drewno* **2014**, *57*, 120–132, doi:10.12841/wood.1644-3985.S01.08.
13. 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.* **1989**, *27*, 290–301, doi:10.1109/36.17670.
14. FAO *Global Forest Resources Assessment 2015: Desk Reference*; 2015; ISBN 9789251088265.
15. 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.* **2007**, *2*, 045023, doi:10.1088/1748-9326/2/4/045023.
16. Ho Tong Minh, D.; 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.* **2014**, *52*, 967–979, doi:10.1109/TGRS.2013.2246170.
17. Ho Tong Minh, D.; Le Toan, T.; 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.* **2016**, *175*, 138–147, doi:10.1016/j.rse.2015.12.037.
18. Huy, B.; Kralicek, K.; Poudel, K. P.; Phuong, V. T.; Khoa, P. Van; Hung, N. D.; Temesgen, H. Allometric equations for estimating tree aboveground biomass in evergreen broadleaf forests of Viet Nam. *For. Ecol. Manage.* **2016**, *382*, 193–205, doi:10.1016/j.foreco.2016.10.021.
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.* **2016**, *376*, 276–283, doi:10.1016/j.foreco.2016.06.031.
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.*

2017, *390*, 147–156, doi:10.1016/j.foreco.2017.01.030.

21. Liaw, a; Wiener, M. Classification and Regression by randomForest. *R news* **2002**, *2*, 18–22, doi:10.1177/154405910408300516.
22. 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.* **2019**, *28*, 1277–1286, doi:10.15244/pjoes/86221.
23. 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.* **2012**, *149*, 80–90, doi:10.1016/j.agee.2011.12.013.
24. MARD *Circular No. 34/2009/TT-BNNPTNT of June 10, 2009, on criteria for forest identification and classification*; 2009; Vol. 2009;.
25. 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.* **2016**, *8*, 1–22, doi:10.3390/rs8030217.
26. 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.* **2015**, *159*, 307–317, doi:10.1016/j.rse.2014.12.019.
27. Mitchard, E. T. A.; Saatchi, S. S.; White, L. J. T.; Abernethy, K. A.; 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* **2012**, *9*, 179–191, doi:10.5194/bg-9-179-2012.
28. Molto, Q.; Rossi, V.; Blanc, L. Error propagation in biomass estimation in tropical forests. *Methods Ecol. Evol.* **2013**, *4*, 175–183, doi:10.1111/j.2041-210x.2012.00266.x.
29. 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.* **2010**, *48*, 1086–1104.
30. 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.* **2012**, *50*, 714–726, doi:10.1109/TGRS.2011.2176133.
31. Paluszynska, A.; Biecek, P. Explaining and Visualizing Random Forests in Terms of

Variable Importance (V 0.9). **2017**, doi:10.1198/jasa.2009.tm08622>.

32. 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.* **2013**, *44*, 593–622, doi:10.1146/annurev-ecolsys-110512-135914.
33. 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 - Introduction and Background of the Study*; 2012;
34. 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.* **1994**, *48*, 205–219, doi:10.1016/0034-4257(94)90142-2.
35. Reigber, A.; Moreira, A. First Demonstration of Airborne SAR Tomography Using Multibaseline L-Band Data. *Trans. Geosci. Remote Sens.* **2000**, *38*, 2142–2152.
36. 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* **2015**, *83–84*, 187–195, doi:10.1016/j.pce.2015.09.003.
37. 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.* **2009**, *6*, 408–412.
38. Singh, G.; Yamaguchi, Y.; Park, S.-E. General Four-Component Scattering Power Decomposition With Unitary Transformation of Coherency Matrix. *IEEE Geosci. Remote Sens. Lett.* **2013**, *51*, 3014–3022, doi:10.1109/LGRS.2015.2451369.
39. Smola, A. J.; Schölkopf, B. A Tutorial on Support Vector Regression. *Stat. Comput.* **2004**, *14*, 199–222, doi:Doi 10.1023/B:Stco.0000035301.49549.88.
40. Toan, T. Le; 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.* **2011**, *115*, 2850–2860, doi:10.1016/j.rse.2011.03.020.
41. Viet Nguyen, L.; Tateishi, R.; Kondoh, A.; Sharma, R.; Thanh Nguyen, H.; Trong To, T.; Ho Tong Minh, D. Mapping Tropical Forest Biomass by Combining ALOS-2, Landsat 8, and Field Plots Data. *Land* **2016**, *5*, 31, doi:10.3390/land5040031.
42. 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 **2016**, 78–83, doi:10.1109/IACC.2016.25.

43. Yamaguchi, Y.; Moriyama, T.; Ishido, M.; Yamada, H. Four-component scattering model for polarimetric SAR image decomposition. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1699–1706, doi:10.1109/TGRS.2005.852084.
44. Yamaguchi, Y.; Yajima, Y.; Yamada, H. A four-component decomposition of POLSAR images based on the coherency matrix. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 292–296, doi:10.1109/LGRS.2006.869986.

CHAPTER 4.

ESTIMATING ABOVEGROUND BIOMASS OF BAMBOO AND MIXED BAMBOO FOREST IN THUA THIEN-HUE PROVINCE, VIET NAM USING PALSAR-2 AND LANDSAT 8 OLI DATA

4.1. Introduction

Recently, using multi-source data has become increased to evaluate the forest biomass because of the potential improvement of the estimation accuracy. Various methods have been developed for combining different data, for example, the remotely sensed and ground data (Badreldin and Sanchez-Azofeifa 2015; Zhang *et al*. 2019), or fusion technique between satellite images (Cutler *et al*. 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 *et al*. 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 forest biomass estimation through the relationship between Landsat signals and forest structure parameters such as tree height or stand volume (Hall *et al*. 2006; Chrysafis et al. 2017). The sensitivity of spectral bands to the biomass estimation was mentioned in Powell *et al*. (2010) and 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 by Sarker and 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 *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 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 with high-resolution, wide swath width and image quality until now (EORC). At L-band, the penetration depth exceeds the crown layer and backscatter mechanism involving the lower part of the canopy presented a strong relationship with biomass from the major contribution of branches and trunk (Toan *et al*. 1992). Much successful application of PALSAR-1/-2 data for forest biomass mapping was recognized based on backscatter intensities (Suzuki *et al*. 2013; Mermoz 2014) or different techniques such as polarimetry and interferometry (Neumann *et al*. 2012; Chowdhury *et al*. 2013; Thiel and Schmullius 2016). A recent approach of tomography showed a high potential for biomass estimation through its correlation to forest height (Tebaldini and Rocca 2012; Ho Tong Minh *et al*. 2016).

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, 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 for tropical forests (Häme *et al*. 2013). For Landsat data, a wide range of saturation levels was examined for different vegetation grades in a subtropical region (Zhao *et al*. 2016). Elimination of the effects of saturation have been discussed in a limited number of studies for Landsat images (Avitabile *et al*. 2012; Phua *et al*. 2017) and for SAR images (Carreiras *et al*. 2012; Mermoz *et al*. 2015). The combination of Landsat and SAR products is also possible to reduce the effect in some instances (Cutler *et al*. 2012; Basuki *et al*. 2013; Zhao *et al*. 2016).

Another issue in forest biomass calculation is the divergence in species composition and stand structure in different vegetation grades. Therefore, different vegetations have different data saturation values in Landsat or radar data (Zhao *et al*. 2016), and require the diversity of remote sensing algorithms and datasets that have been used to estimate 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.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.ha^{-1}, helping to mitigate the climate change (Yuen *et al*.

2017). Besides, because of its economic and environmental benefits, the bamboo forest is being considered as an alternative sustainable land-use strategy particular in the upland region in Vietnam (Ly *et al*. 2012). Therefore, there is a demand to evaluate the performance of satellite source data for estimation bamboo forest AGB, which provides useful information for our understanding to plan a sustainable development strategy in Vietnam.

In this study, we focused on evaluating the performance of ALOS/PALSAR-2 and Landsat 8 OLI on forest biomass estimation 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 4.2 describes the Study area and field data; Section 4.3 explains the methods to process data and to estimate forest biomass; Section 4.4 showed the performance of Landsat 8 OLI and PALSAR-2 for aboveground biomass estimation and the best-fit model selection; Our findings are discussed in Section 4.5; Finally, a conclusion is showed in Section 4.6.

4.2. Study area and field data

4.2.1. Description of the study area

Bamboo forest has a unique morphological structure that is easily identifiable from a far 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 bamboos. 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 in AGB. Therefore, the bamboo forest is an important resource to minimize the green house effect of climate change.

A Luoi is a mountainous district of Thua Thien Hue province in Central Vietnam (Figure 4.1). This region covers about 1,224.6 km² (accounting for 24.17% of the natural land area of Thua Thien Hue province) with the altitude of 600-800 m above sea level and the slope of 2^o-25^o. 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, and the annual precipitation is an average of about 3,500 mm. In general, A Luoi has a cool climate throughout the year, and the local weather in A Luoi has favorable for developing bamboo forest.

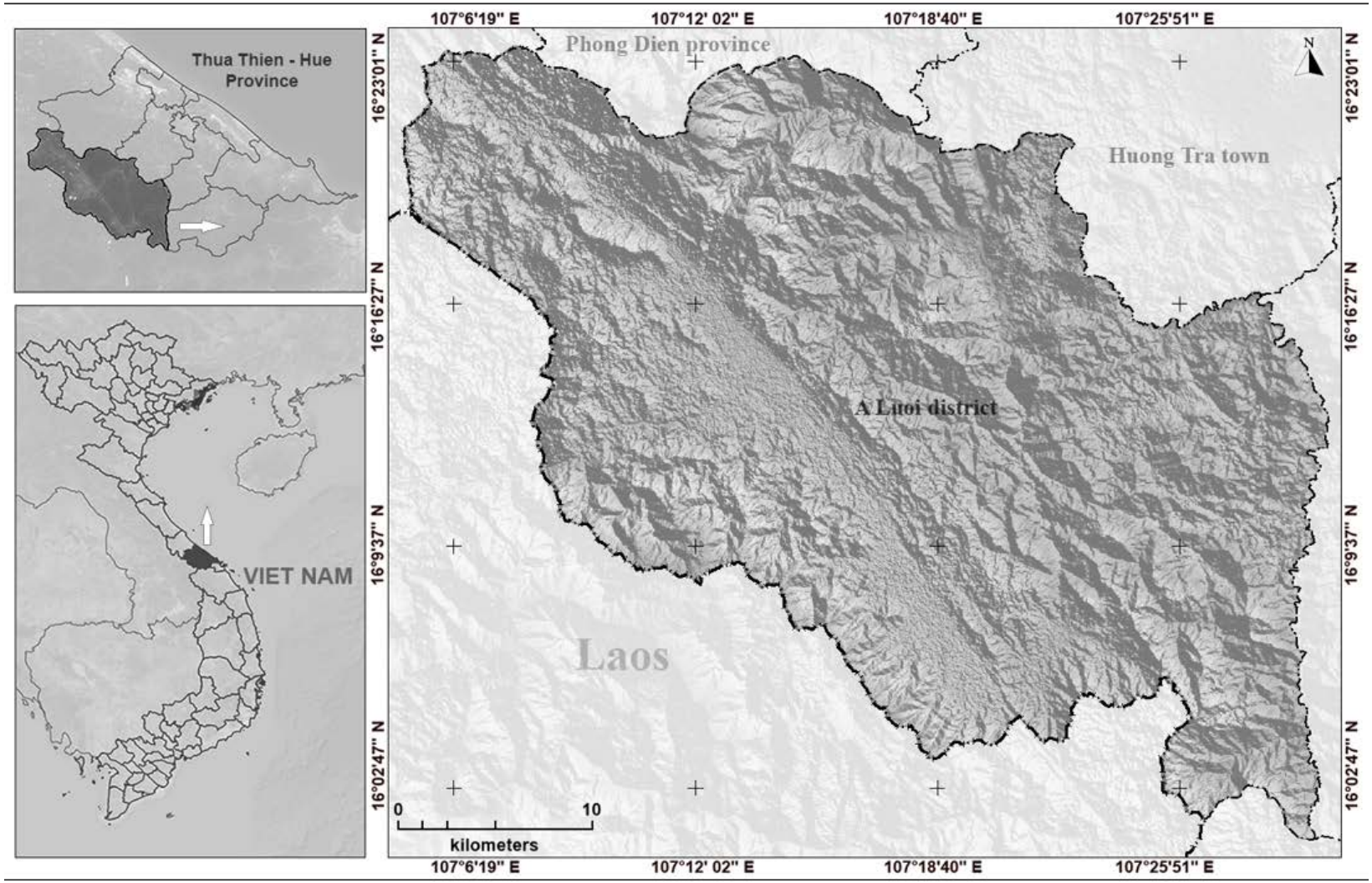


Figure 4.1. Location of the study areas in Thua Thien Hue province, Vietnam

4.2.2. 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 ground 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 forests. 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 of 6 cm and over, while tree height (H) was measured for five normal-shaped 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 mean 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 AGB of bamboo forest, we used the formula given by Ly *et al*. (2012):

$$AGB = 0.3002\ DBH^2 + 0.115\ DBH + 1.7632 \quad (4.1)$$

For estimating the biomass of broadleaved trees, the one-factor formula to estimate AGB (in tons/ha) given by Huy *et al.* (2016) is:

$$AGB = 0.104189x\ DBH^{2.491453} \quad (4.2)$$

4.3. Methodology

4.3.1. 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^o) 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 8 OLI was acquired in April 2016 and covered 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 4.1).

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 and Huete 1991) and successfully used for estimating biophysical properties (Anderson and Hanson 1993; Sarker and Nichol 2011).

In this study, we used vegetation indices calculated using band Near-Infrared (0.7-1.1 μm), red (0.6-0.7 μm), blue (0.45-0.52 μm) and green band (0.52-0.60 μm) in Landsat 8 OLI data. For the Enhanced vegetation index (EVI) formula (Table 4.1), L is a soil adjustment factor, and C_1 and C_2 are coefficients used to correct aerosol scattering in the red band by the use of the blue band. In general, G=2.5, C_1=6.0, C_2=7.5, and L=1 (Huete *et al*. 1997).

For TSAVI, the equation was developed by (Baret and 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.

4.3.2. 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 8 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 AGB estimated from allometries 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 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 models 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 4.1. Predictor variables from PALSAR-2 and LANDSAT 8 OLI used in this study

Indepent variables	Explanations/Name	Describe/Wavelength	Formula/Resulution
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.

4.3.3. Accuracy Assessment

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 lear+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 as 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:

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

4.4. Result

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

Figure 4.2 shows the correlation between forest AGB and various parameters derived from satellite images. For Landsat 8 OLI data, the weak correlation with AGB 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 whose linear combination can fulfill their contribution to models. Therefore, nine out of eleven VIs were retained to evaluate the 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 generally have a 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 obsolute 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. The sigma value of HV (|r| of 0.7) polarization has a better correlation than HH (|r| of 0.63).

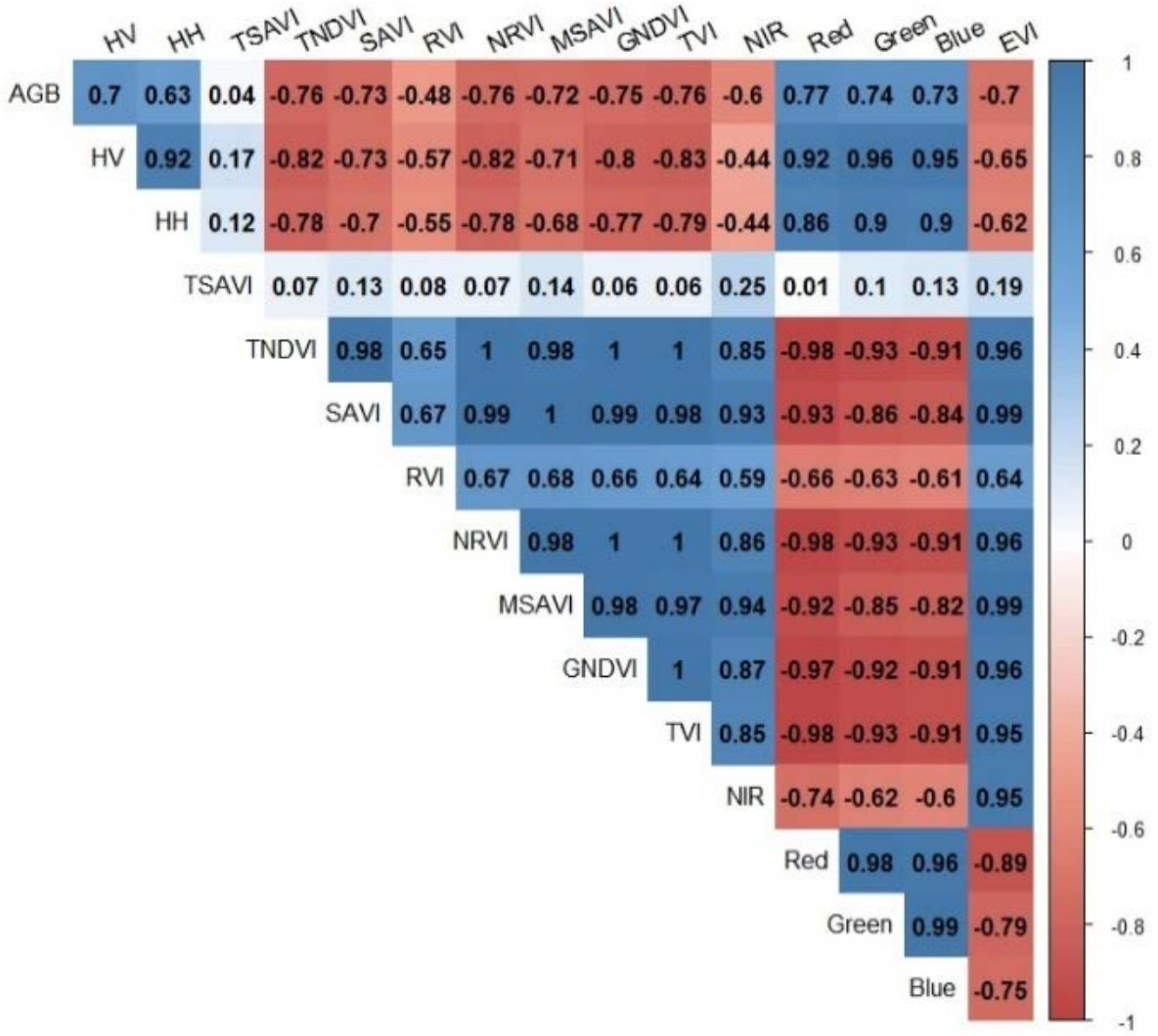


Figure 4.2. Pearson's correlation coefficient (R) between aboveground biomass and different remotely sensed parameters

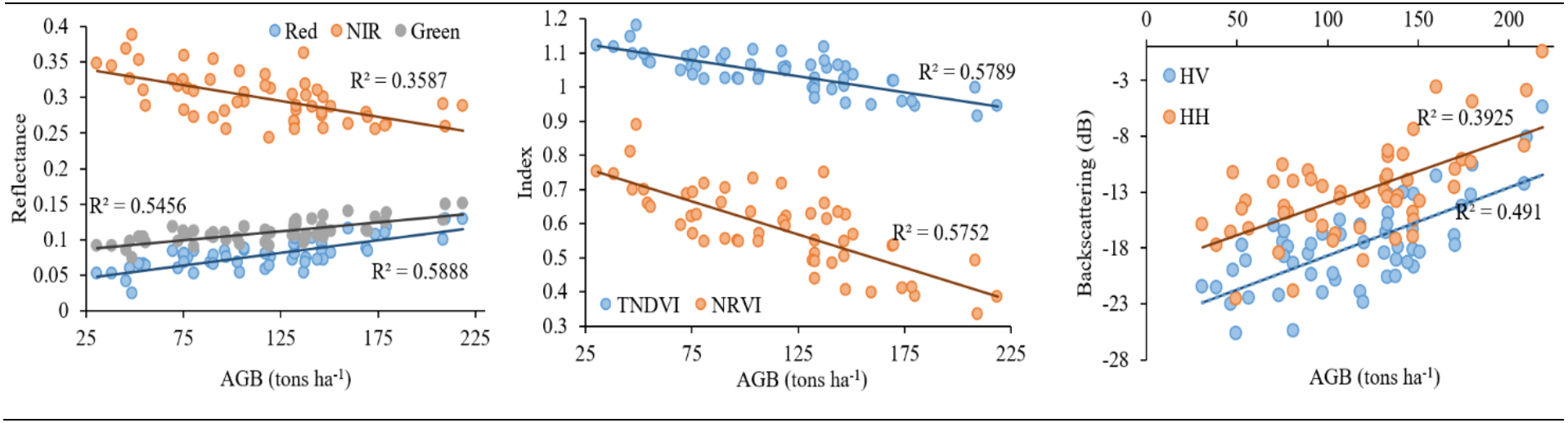


Figure 4.3. Linear regression between aboveground biomass and Landsat reflectance in Green, Red, and NIR (left panel); Vegetation indices in TNDVI and NRVI (middle panel); and PALSAR-2 backscattering in HH and HV (right panel)

Figure 4.3 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.

4.4.2. 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 4.2). Comparatively, multiple linear models are slightly better than univariate models with improving R^2 values from 0.58 to 0.60. This table shows that the increasing number of variables in models performed higher R^2, however it led to increase value of BIC and therefore, increase the risk of overfitting.

Table 4.3 provides a comparative analysis of models using only SAR data, only Landsat data and the combination of SAR and Landsat based on R^2 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 R^2 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 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 R^2 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 for the entire data and 30.06 tons/ha for LOOCV.

Table 4.2. The result of selecting the best fit model using Bayesian model averaging

	p!=0	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Intercept	100	-11.73	680.23	757.24	289.46	292.50	372.61	288.09
HV	16.5	.	.	.	.	4.88	6.27	4.10
TNDVI	19.7	.	.	-616.53	.	.	.	.
SAVI	9.7	.	.	.	.	.	.	-263.55
NRVI	16.7	.	.	.	-296.00	.	.	.
TVI	21.1	.	-447.71	.	.	.	.	.
NIR	9.9	.	.	.	.	.	-486.61	.
Red	28.2	1635.97	.	.	.	.	.	.
EVI	8.2	.	.	.	.	-123.15	.	.
nVar		1.00	1.00	1.00	1.00	2.00	2.00	2.00
R^2		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 4.3. 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

4.4.3. Aboveground biomass prediction and mapping

Figure 4.4 illustrates the relationship between AGB estimated from allometries and AGB predicted using LOOCV in some linear regression models. The distribution of AGB values by LOOCV in different models are quite similar. Analyzing the density distribution of AGB by LOOCV showed that overestimation occurs in the AGB range from 75 to 130 tons/ha, however, the remaining showed an underestimation. For the AGB by LOOCV, a high density is considered in values from 85-150 tons/ha, and the highest density concentrates in the range of 100-125 tons/ha. For the AGB by allometries, the density distributes more equally and mainly concentrates between 50-170 tons/ha, with the highest density in 125-150 tons/ha.

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 4.5 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 for the entire data and 30 tons/ha 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. 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. 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

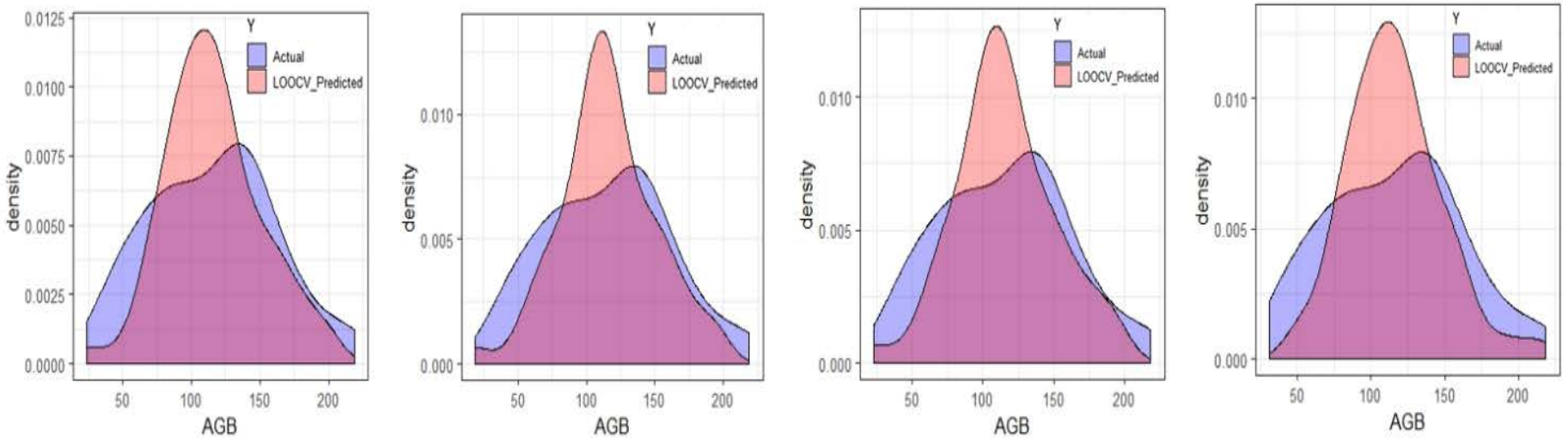
combination of HV and EVI or NIR enables to decrease the portion of negative values, however, as analyzed above, the increasing number of variables can increase the risk of overfitting. The model using the Red band has the best performance for AGB estimation in the bamboo and mixed bamboo forest with the good regression evaluation in R^2 and RMSE, but it still has the presence of negative AGB with the ratio 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.

4.5. Discussion

This study focused on evaluating the performance of Landsat 8 OLI and PALSAR-2 on forest AGB estimation. The linear regression was selected to train the data and was developed in the bamboo and mixed bamboo forest.

The results indicated that Landsat 8 OLI has a slightly better relationship to AGB of the bamboo and mixed bamboo forest than PALSAR-2 did. 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 (Suzuki *et al*. 2013; Mermoz *et al*. 2015), while the reflectance of Red band did not show a clear saturation to biomass and was sensitive to greater AGB. Visually, the saturation threshold of imagery signals for bamboo was found to be greater than other ecosystems with 100 tons/ha for the subtropical area (Häme *et al*. 2013), or 150 tons/ha for dense rainforest and savanna (Mermoz *et al*. 2015). This discrepancy confirms that saturation threshold widely fluctuates in different types of forest, and therefore, biomass estimation accuracy depends on forest types. 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 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 replaced by EVI in this case study). Although the application of VIs for biomass estimations is considered a solution in reducing saturation in some studies (Gasparri *et al*. 2010; Zhu and Liu 2015) but needs to be careful, because Sarker and Nichol (2011) have demonstrated the inefficiency in applying VIs for biomass estimations. An example indicated that the sensitivity of NDVI to variation in land surface properties varies in space and time (Foody *et al*. 2003).



(a) Red **(b) HV+NIR** **(c) HV+ EVI** **(d) HV+HH**

Figure 4.4. 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

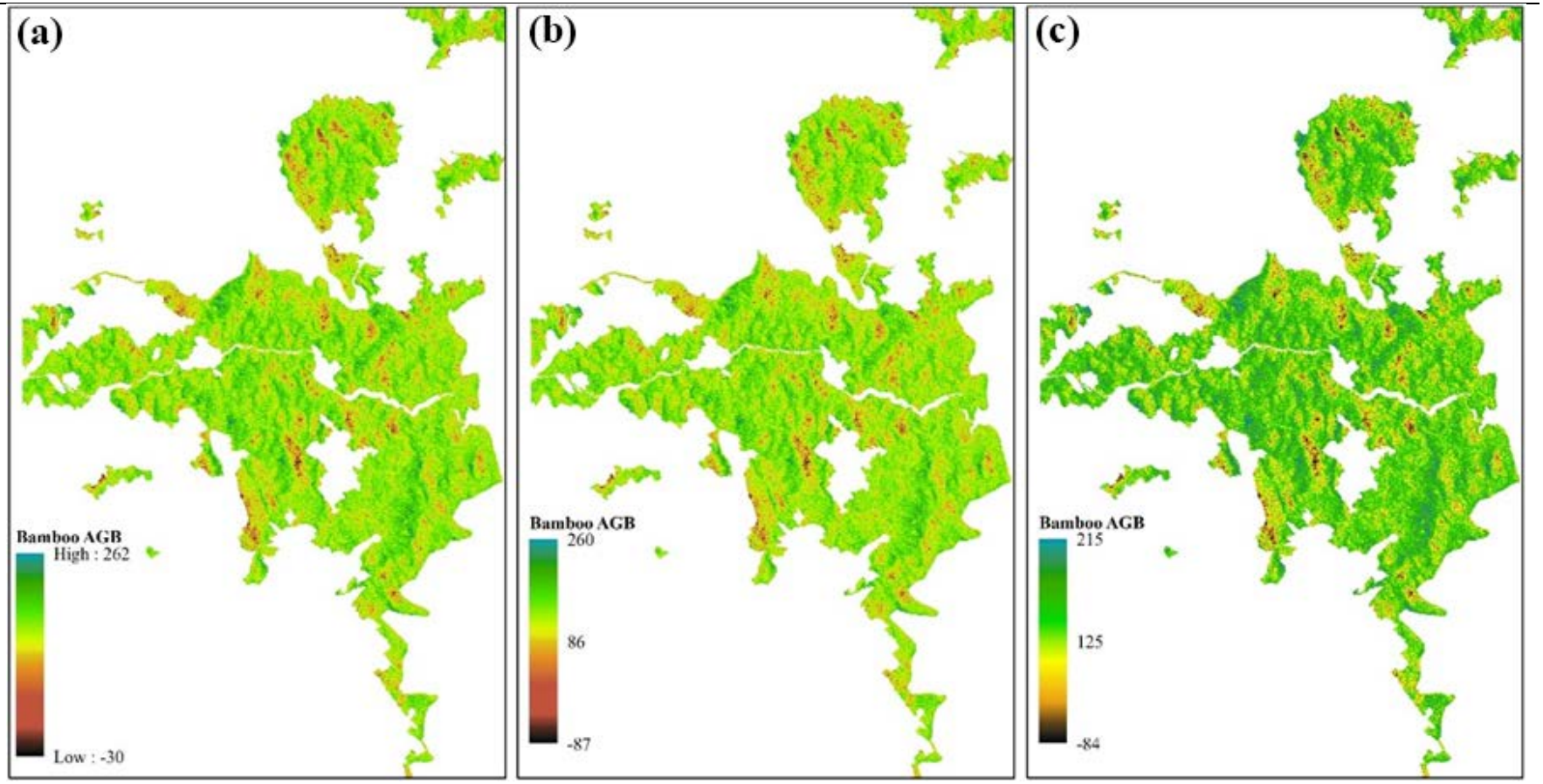


Figure 4.5. The predicted biomass using different models with sets of variables: (a) Red, (b) HV+EVI, and (c) HV+NIR

Previous studies pointed out the improvement of the combination of optical and SAR data for biomass estimation compared to single data sources (Hoan *et al*. 2011; Häme *et al*. 2013; Shao and Zhang 2016; Zhao *et al*. 2016).

On the contrary, although this study indicated the combination of multisource remotely sensed data has a better performance at forest biomass estimation compared to SAR data, the combination makes no significant improvement compared to using the single optical band. The reason for this result is maybe the fusion technique of Landsat and SAR data cannot effectively incorporate radar information into the newly fused data (Zhao *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. However, the problem of linear regression is that the extrapolation of very low values of an explanatory variable (Neumann *et al*. 2012) occurs negative biomasses. 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 to −87 tons/ha. Furthermore, the overestimation and underestimation in the predicted AGB indicated that the linear regression may be an insufficient approach to the AGB estimation. However, the area proportion of negative biomass values is small and the Red band has an important potential for AGB estimation in the bamboo and mixed bamboo forest.

4.6. 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. The biomass sensitivity for optical bands is better than L-band SAR data. Furthermore, the combination of optical band and SAR data did not indicate better performance than a single

spectral of Red band in the biomass estimation. Besides, the attempt using multivariate regression was not successful because the increasing number of explanatory 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 8 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 in Vietnam. More studies should be further analyzed by adding more samples with other approaches, such as nonlinear and nonparametric models.

References

1. Badreldin, N.; Sanchez-Azofeifa, A. 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 Sens.* **2015**, *7*, 2832–2849, doi:10.3390/rs70302832.
2. Anderson, G. L.; Hanson, J. D. 1-S2.0-0034425793900405-Main.Pdf. **1993**, *175*, 165–175.
3. Antropov, O.; Rauste, Y.; Häme, T.; Praks, J. Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. *Remote Sens.* **2017**, *9*, 1–24, doi:10.3390/rs9100999.
4. Avitabile, V.; Baccini, A.; Friedl, M. A.; Schmullius, C. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sens. Environ.* **2012**, *117*, 366–380, doi:10.1016/j.rse.2011.10.012.
5. Baret, F.; Guyot, G. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sensing of. *Environment* **1991**, *35*, 161–173.
6. Basuki, T. M.; Skidmore, A. K.; Hussin, Y. A.; van Duren, I. Estimating tropical forest biomass more accurately by integrating ALOS PALSAR and Landsat-7 ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 4871–4888, doi:10.1080/01431161.2013.777486.
7. Carreiras, J. M. B.; Vasconcelos, M. J.; Lucas, R. M. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sens. Environ.* **2012**, *121*, 426–442, doi:10.1016/j.rse.2012.02.012.
8. CEOS *최종보고서 Table of Contents*; 2018; Vol. 28;.
9. 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.* **2013**, *5*, 5725–5756, doi:10.3390/rs5115725.
10. Chrysafis, I.; Mallinis, G.; Gitas, I.; Tsakiri-Strati, M. Estimating Mediterranean forest parameters using multi seasonal Landsat 8 OLI imagery and an ensemble learning method. *Remote Sens. Environ.* **2017**, *199*, 154–166, doi:10.1016/j.rse.2017.07.018.
11. Cutler, M. E. J.; Boyd, D. S.; Foody, G. M.; Vetrivel, A. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 66–77, doi:10.1016/j.isprsjprs.2012.03.011.

12. EORC, J. ALOS-2 Project- PALSAR-2 Available online: https://www.eorc.jaxa.jp/ALOS-2/en/about/palsar2.htm (accessed on Oct 1, 2019).
13. Fayad, I.; Baghdadi, N.; Guitet, S.; Bailly, J. S.; Hérault, B.; Gond, V.; El Hajj, M.; Tong Minh, D. H. Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 502–514, doi:10.1016/j.jag.2016.07.015.
14. Foody, G. M.; Boyd, D. S.; Cutler, M. E. J. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sens. Environ.* **2003**, *85*, 463–474, doi:10.1016/S0034-4257(03)00039-7.
15. Foody, G. M.; Cutler, M. E.; Mcmorrow, J.; Pelz, D.; Tangki, H.; Boyd, D. S.; Douglas, I. Mapping the Biomass of Bornean Tropical Rain Forest from Remotely Sensed Data Published by : Blackwell Publishing Stable URL : http://www.jstor.org/stable/2665383. *Glob. Ecol. Biogeogr.* **2001**, *10*, 379–387.
16. Gasparri, N. I.; Parmuchi, M. G.; Bono, J.; Karszenbaum, H.; Montenegro, C. L. Assessing multi-temporal Landsat 7 ETM+ images for estimating above-ground biomass in subtropical dry forests of Argentina. *J. Arid Environ.* **2010**, *74*, 1262–1270, doi:10.1016/j.jaridenv.2010.04.007.
17. Hall, R. J.; Skakun, R. S.; Arsenault, E. J.; Case, B. S. Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. *For. Ecol. Manage.* **2006**, *225*, 378–390, doi:10.1016/j.foreco.2006.01.014.
18. Häme, T.; Rauste, Y.; Antropov, O.; Ahola, H. A.; Kilpi, J. Improved mapping of tropical forests with optical and sar imagery, part ii: Above ground biomass estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 92–101, doi:10.1109/JSTARS.2013.2241020.
19. Ho Tong Minh, D.; Le Toan, T.; 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.* **2016**, *175*, 138–147, doi:10.1016/j.rse.2015.12.037.
20. Hoan, N. T.; Tateishi, R.; Bayan, A.; Ngigi, T. G.; Lan, M. Improving tropical forest mapping using combination of optical and microwave data of ALOS. *Int. Geosci. Remote Sens. Symp.* **2011**, *34*, 736–739, doi:10.1109/IGARSS.2011.6049235.
21. Huete, A. R.; Liu, H. Q.; Batchily, K.; J., L. van W. A comparsion of vegetation indices

over a Global set of TM images for EO -MODIS. *Remote Sens. Environ.* **1997**, *59*, 440–451, doi:10.1016/S0034-4257(96)00112-5.

22. Jackson, R. D.; Huete, A. R. Interpreting vegetation indices. *Prev. Vet. Med.* **1991**, *11*, 185–200, doi:10.1016/S0167-5877(05)80004-2.
23. Lu, D.; Chen, Q.; Wang, G.; Moran, E.; Batistella, M.; Zhang, M.; Vaglio Laurin, G.; Saah, D. Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *Int. J. For. Res.* **2012**, *2012*, 1–16, doi:10.1155/2012/436537.
24. Lucas, R. M.; Mitchell, A. L.; Armston, J. Measurement of forest above-ground biomass using active and passive remote sensing at large (Subnational to global) scales. *Curr. For. Reports* **2015**, *1*, 162–177, doi:10.1007/s40725-015-0021-9.
25. 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.* **2012**, *149*, 80–90, doi:10.1016/j.agee.2011.12.013.
26. Mahyoub, S.; Fadil, A.; Mansour, E. M.; Rhinane, H.; Al-Nahmi, F. Fusing of optical and synthetic aperture radar (SAR) remote sensing data: A systematic literature review (SLR). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* **2019**, *42*, 127–138, doi:10.5194/isprs-archives-XLII-4-W12-127-2019.
27. Mermoz Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sens. Environ.* **2014**, *155*, 109–119, doi:10.1016/j.rse.2014.01.029.
28. 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.* **2015**, *159*, 307–317, doi:10.1016/j.rse.2014.12.019.
29. 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.* **2012**, *50*, 714–726, doi:10.1109/TGRS.2011.2176133.
30. Phua, M. H.; Johari, S. A.; Wong, O. C.; Ioki, K.; Mahali, M.; Nilus, R.; Coomes, D. A.; Maycock, C. R.; Hashim, M. Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests. *For. Ecol. Manage.* **2017**, *406*, 163–171, doi:10.1016/j.foreco.2017.10.007.
31. Powell, S. L.; Cohen, W. B.; Healey, S. P.; Kennedy, R. E.; Moisen, G. G.; Pierce, K. B.; Ohmann, J. L. Quantification of live aboveground forest biomass dynamics with

Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sens. Environ.* **2010**, *114*, 1053–1068, doi:10.1016/j.rse.2009.12.018.

32. Propastin, P. Modifying geographically weighted regression for estimating aboveground biomass in tropical rainforests by multispectral remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 82–90, doi:10.1016/j.jag.2011.12.013.
33. Raftery, A. A.; Hoeting, J.; Volinsky, C.; Painter, I.; Yeung, K. Y.; Sevcikova, M. H. Package ' BMA ' 2018.
34. Sarker, L. R.; Nichol, J. E. Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sens. Environ.* **2011**, *115*, 968–977, doi:10.1016/j.rse.2010.11.010.
35. Shao, Z.; Zhang, L. Estimating forest aboveground biomass by combining optical and SAR data: A case study in genhe, inner Mongolia, China. *Sensors (Switzerland)* **2016**, *16*, doi:10.3390/s16060834.
36. Suzuki, R.; Kim, Y.; Ishii, R. Sensitivity of the backscatter intensity of ALOS/PALSAR to the above-ground biomass and other biophysical parameters of boreal forest in Alaska. *Polar Sci.* **2013**, *7*, 100–112, doi:10.1016/j.polar.2013.03.001.
37. Tebaldini, S.; Rocca, F. Multibaseline polarimetric SAR tomography of a boreal forest at P- and L-Bands. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 232–246, doi:10.1109/TGRS.2011.2159614.
38. Thiel, C.; Schmullius, C. The potential of ALOS PALSAR backscatter and InSAR coherence for forest growing stock volume estimation in Central Siberia. *Remote Sens. Environ.* **2016**, *173*, 258–273, doi:10.1016/j.rse.2015.10.030.
39. Tian, X.; Li, Z.; Chen, E.; Yan, M.; Han, Z.; Liu, Q. Modeling of forest above-ground biomass dynamics using multi-source data and incorporated models: A case study over the Qilian mountains. *Int. Geosci. Remote Sens. Symp.* **2017**, *2017-July*, 5770–5773, doi:10.1109/IGARSS.2017.8128319.
40. Toan, T. Le; Beaudoin, A.; Riom, J.; Guyon, D. Relating Forest Biomass to SAR Data.pdf. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 403–411, doi:10.1109/36.134089.
41. Torgo, A. L. DMwR package in R 2015.
42. Wong, T. T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48*, 2839–2846, doi:10.1016/j.patcog.2015.03.009.

43. Yuen, J. Q.; Fung, T.; Ziegler, A. D. Carbon stocks in bamboo ecosystems worldwide: Estimates and uncertainties. *For. Ecol. Manage.* **2017**, *393*, 113–138, doi:10.1016/j.foreco.2017.01.017.
44. Zhang, R.; Zhou, X.; Ouyang, Z.; Avitabile, V.; Qi, J.; Chen, J.; Giannico, V. Estimating aboveground biomass in subtropical forests of China by integrating multisource remote sensing and ground data. *Remote Sens. Environ.* **2019**, *232*, 111341, doi:10.1016/j.rse.2019.111341.
45. Zhao, P.; Lu, D.; Wang, G.; Liu, L.; Li, D.; Zhu, J.; Yu, S. Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *53*, 1–15, doi:10.1016/j.jag.2016.08.007.
46. Zhao, P.; Lu, D.; Wang, G.; Wu, C.; Huang, Y.; Yu, S. Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sens.* **2016**, *8*, doi:10.3390/rs8060469.
47. Zhu, X.; Liu, D. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 222–231, doi:10.1016/j.isprsjprs.2014.08.014.

CHAPTER 5. SUMMARY

1. For forest classification, we addressed this issue by applying semi-supervised classification for data integration of optical and SAR data. The combination of Landsat and PolSAR data resulted in improved discrimination of forest grades. The use of multisource remotely sensed data can provide more image information of the forest grades, as well mitigate the disadvantages of Landsat images (cloud, lower spatial resolution), and limited information regarding objects in PALSAR/PALSAR-2 image (only two polarization HH and HV). In this study, we assessed the potential of the proposed semi-supervised model developed and validated for mapping forest grades and assessed the process of forest transition in a tropical natural forest in Vietnam. The model produced high accuracies in the classified images in 2007, 2010, and 2016 with over 0.74 for kappa, and over 0.8 for overall accuracy.

2. Additionally, we created a set of high representative metrics for analyzing multi-temporal forest grades structure at class and landscape level in the study area. Landscape metrics were used to evaluate the forest changes based on the spatial processes, such as aggregation, fragmentation, and compaction. At the class level, the poor forest demonstrated the largest variation with more dispersed patterns, while other grades had a low level of aggregation. At the landscape level, the natural forest experiences increased fragmentation, which involved an increase in landscape area with shrinkage of patch size and disproportionate distribution of patches.

3. For forest aboveground biomass (AGB) estimation, the contribution of PolSAR data to AGB estimation was evaluated using two approaches: the total samples data, and the five forest grades. The correlation analysis of individual PolSAR parameters to the forest biomass revealed different sensitivity of the parameters to different forest grades. Therefore, the different sets of parameters were proposed for use for particular forest grades. The result also indicated that the volume scattering composition (Y_vol) and the coherency elements mainly contributed to the AGB estimation function.

The RMSE values were enhanced by 9%–18% for the rich forest and 80%–85% for the remaining forest grades when compared to the total sample. Among all models, Support vector regression (SVR) and random forest regression exhibited better performance in R^2 and RMSE in training data. SVR provided the best result for Medium forest and Poor forest with RMSE 8.27 tons/ha and 12.38 tons/ha, respectively. Although random forest displayed a good R^2, it was only chosen to calculate AGB for Bamboo forest because it had low

reliability with the problem of underfitting with high values in both training and validation data for the other forest grades. The polynomial function proved to be a suitable model for estimating biomass for the Restoration forest with R^2 of 0.47 and RMSE of 10.11 tons/ha. In this study, the low correlation of AGB with PolSAR parameters in the Rich forest grade was due to 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 grades (except Rich forest). This study also confirmed the effectiveness and role of PolSAR data in calculating aboveground biomass in tropical forests. Furthermore, the strong correlation of PolSAR signals to the Bamboo forest was exhibited in this study. This verified the potential for estimating AGB in bamboo biomass using SAR data, that was not documented in previous studies, as well as creating a foundation for further study in various bamboo species.

4. For the bamboo and mixed bamboo forest, 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. The biomass sensitivity for optical bands is better than L-band SAR data. Furthermore, the combination of optical band and SAR data did not indicate better performance than a single spectral of the Red band. Besides, the attempt using multivariate regression was not successful because the increasing number of variables leads 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 8 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 in Vietnam.

In general, the highlights of the thesis are below:

1. Developing a proposed semi-supervised model for mapping forest grades and assessing forest transition.
2. Assessing forest transition from the perspective of landscape ecology by using multi-temporal data.

3. Improving the accuracy of biomass estimation by selecting the best regression models based on observations of the contribution of radar signals to aboveground biomass in five forest grades.
4. Evaluating the performance of by PALSAR-2 L-band and Landsat data for aboveground biomass estimation of bamboo and mixed bamboo forest in Vietnam.

Appendix A.

Table A1. List of 56 metrics in Class level (C) and 63 metrics in Landscape level (L)

Number	Variable		Description
1	CA	C	Total class area
2	CLUMPY	C	Clumpiness index
3	CPLAND	C	Core area percentage of *landscape*
4	NLSI	C	Normalized Landscape shape index
5	AI	C,L	Aggregation index
6	AREA_AM	C,L	Area-weighted mean patch size
7	AREA_CV	C,L	Patch size coefficient of variation
8	AREA_MN	C,L	Mean patch size
9	CAI_AM	C, L	Area-weighted mean core area index
10	CAI_CV	C,L	Core area coefficient of variation
11	CAI_MN	C, L	Mean core area index
12	CIRCLE_AM	C, L	Area-weighted mean circumscribing circle
13	CIRCLE_CV	C,L	circumscribing circle coefficient of variation
14	CIRCLE_MN	C,L	Mean coefficient of variation
15	COHESION	C, L	Patch cohesion
16	CONNECT	C,L	Connectance index
17	CONTIG_AM	C,L	Area-weighted contiguity index
18	CONTIG_CV	C,L	Contiguity index coefficient of variation
19	CONTIG_MN	C,L	Mean coefficient of variation
20	CORE_AM	C,L	Area-weighted mean core area
21	CORE_CV	C,L	Core area coefficient of variation
22	CORE_MN	C, L	Mean core area
23	DCAD	C,L	Disjunct core area density
24	DCORE_AM	C, L	Area-weighted mean disjunct core area
25	DCORE_CV	C,L	Disjunct core area coefficient of variation
26	DCORE_MN	C, L	Mean disjunct core area
27	DIVISION	C,L	Division index
28	ED	C,L	Edge density
29	ENN_AM	C,L	Area-weighted mean nearest neighbor distance
30	ENN_CV	C, L	Nearest neighbor distance coefficient of variation
31	ENN_MN	C, L	Mean nearest neighbor distance
32	FRAC_AM	C,L	Area-weighted mean fractal dimension
33	FRAC_CV	C,L	Fractal dimension coefficient of variation
34	FRAC_MN	C,L	Mean fractal dimension

35	GYRATE_AM	C, L	Mean radius of gyration
36	GYRATE_CV	C, L	Radius of gyration coefficient of variation
37	GYRATE_MN	C, L	Mean radius of gyration
38	IJI	C,L	Interspersion/ juxtaposition index
39	LPI	C, L	Largest patch index
40	LSI	C, L	Landscape shape index
41	MESH	C, L	Mesh index
42	NDCA	C, L	Number of Disjunct Core Areas
43	NP	C,L	Number of patch
44	PAFRAC	C,L	Perimeter-Area Fractal Dimension
45	PARA_AM	C,L	Area-weighted mean perimeter-area ratio
46	PARA_CV	C,L	Perimeter-area ratio coefficient of variation
47	PARA_MN	C,L	Mean perimeter-area ratio
48	PD	C,L	Patch density
49	PLADJ	C,L	Proportion of like adjacencies
50	PLAND	C, L	Proportion of landscape
51	SHAPE_AM	C, L	Area-weighted mean shape index
52	SHAPE_CV	C, L	Shape index coefficient of variation
53	SHAPE_MN	C, L	Mean shape index
54	SPLIT	C,L	Splitting index
55	TCA	C, L	Total core area
56	TE	C,L	Total edge
57	CONTAG	L	Contagion
58	MSIDI	L	Modified Simpson's Diversity Index
59	MSIEI	L	Modified Simpson's evenness index
60	PR	L	Patch Richness
61	PRD	L	Patch richness density
62	RPR	L	Relative patch richness
63	SHDI	L	Shannon's diversity index
64	SHEI	L	Shannon's Evenness Index
65	SIDI	L	Simpson’s patch density
66	SIEI	L	Simpson’s patch evenness
67	TA	L	Total landscape area

Appendix B.

Table B1. The result of universality, strength, and consistency at class level

Cluster	Members	%Total	Eigenvalue	% Variation Explained	Average in group correlation
1	9	96	4.01	7.29	0.81
2	4	83	2.13	3.87	0.97
3	6	93	2.35	4.27	0.71
4	8	100	4.05	7.36	0.92
5	4	75	1.88	3.43	0.85
6	3	92	1.00	1.83	0.61
7	2	83	0.71	1.29	0.65
8	7	89	2.62	4.77	0.68
9	3	100	1.22	2.22	0.74
10	4	83	1.69	3.07	0.76
11	5	85	2.15	3.91	0.78

Table B2. The result of universality, strength, and consistency at the landscape level

Cluster	Members	%Total	Eigenvalue	Variation Explained	Average in group correlation
1	10	100	5.27	8.50	0.85
2	7	100	4.17	6.72	0.96
3	10	100	4.85	7.83	0.78
4	8	100	3.55	5.73	0.72
5	8	100	4.03	6.49	0.81
6	5	100	2.13	3.43	0.69
7	2	100	1.22	1.98	0.99
8	2	100	1.24	2.00	1.00
9	8	100	3.64	5.87	0.73
10	2	100	1.18	1.91	0.95

Appendix C.

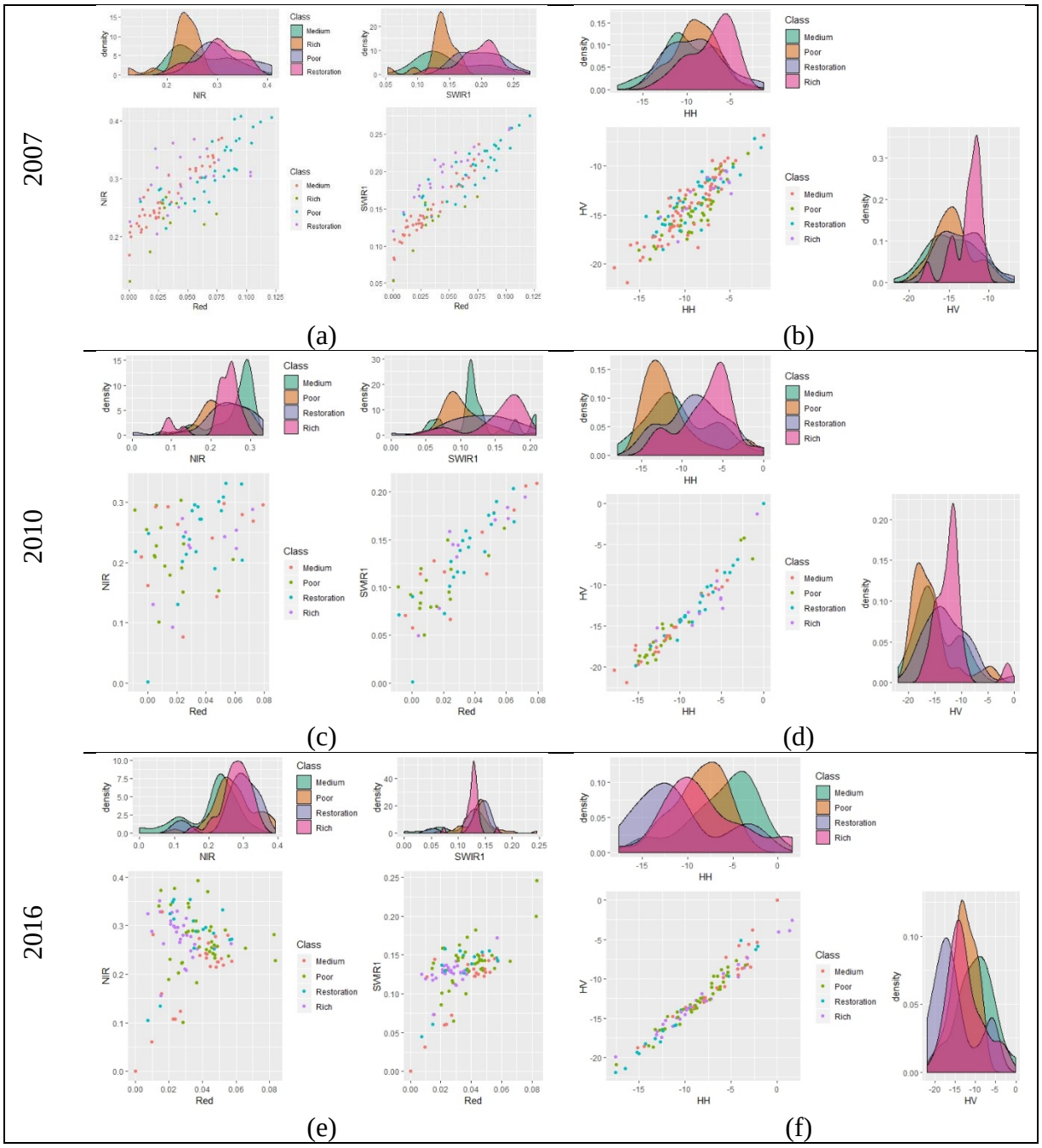


Figure C1. Distribution and density of some parameters (HH and HV signals in decibels for SAR data, and red, near-infrared, and shortwave infrared 1 in reflectance for Landsat data) in four forest grades in three years 2007 (**a,b**), 2010 (**c,d**), and 2016 (**e,f**)

Appendix D.

Table D1. Confusion matrix of classification in 2007

Prediction	Medium	Rich	Poor	Restoration	User's
Medium	15	1	4	0	75.00
Rich	0	5	0	0	100.00
Poor	0	0	14	0	100.00
Restoration	0	1	1	9	81.82
Producer's	100.00	71.43	73.68	100.00	
Overall accuracy			0.86		
Kappa			0.81		

Table D2. Confusion matrix of classification in 2010

Prediction	Medium	Rich	Poor	Restoration	User's
Medium	7	0	0	0	100.00
Rich	0	4	0	0	100.00
Poor	2	0	4	1	57.14
Restoration	0	0	1	4	80.00
Producer's	77.78	100.00	80.00	80.00	
Overall accuracy			0.82		
Kappa			0.76		

Table D3. Confusion matrix of classification in 2016

Prediction	Medium	Rich	Poor	Restoration	User's
Medium	7	0	2	1	70.00
Rich	0	12	2	0	85.71
Poor	0	0	12	1	92.31
Restoration	1	2	0	5	62.50
Producer's	87.50	85.71	75.00	71.43	
Overall accuracy			0.81		
Kappa			0.74		