（千葉大学審査学位論文）

FOREST AND NON-FOREST MAPPING WITH AN OBJECT-BASED CLASSIFICATION APPROACH USING ALOS PALSAR 50 M MOSAIC DATA

January 2015

Chiba University

Graduate School of Science

Division of Geosystem and Biological Sciences

Department of Earth Sciences

MI LAN

（千葉大学審査学位論文）

ALOS PALSAR 50 m モザイクデータを用いたオブジェクト指向型分類による森林/非森林図の作成

2015 年 1 月

千葉大学大学院理学研究科

地球生命圏科学専攻地球科学コース

弭 兰

（ミ　ラン）

ACKNOWLEDGEMENTS

ALOS PALSAR 50 m mosaic data, the main satellite data used in this study, is provided by JAXA free of charge. I would like to give my appreciation to JAXA for producing the free and useful data to support my PhD study, which marks a great achievement of the six years I spent in Japan, which is an important milestone of my life. Next, I would like to express my sincere gratitude to the people who gave me the greatest help and support during this PhD study.

First of all, I would like to express my deepest sincere appreciation and respect to my supervisor Prof. Ryutaro Tateishi for his selfless helps, patient guidance and valuable suggestions from my master course until now. What I learned from him, not only about remote sensing technique, but also how to become an excellent researcher and kind person.

I also would like to extend my gratefulness to dissertation committee: Prof. Akihiko Kondoh, Assoc. Prof. Chiharu Hongo and Prof. Josaphat Tetuko Sri Sumantyo for their carefully review of this thesis. Specially thanks to Prof. Akihiko Kondoh for his kind support of an important software used in this study, to Assoc. Prof. Chiharu Hongo for her useful advices and to Prof. Josaphat Tetuko Sri Sumantyo for his excellent direction of SAR knowledge.

I appreciate the help from all the members of Tateishi laboratory, for their co-operation in study, for their helps in my life. Especially to Dr. Nguyen Thanh Hoan and Dr. Toshiyuki Kobayashi, who gave me so many valuable comments and suggestions to finish this thesis. Thanks to Mr. Desitika Cahyana for providing the ground truth of land cover photos took by his friend Mr. Rian Nurtyawan in Indonesia.

I wish to express my gratitude to OKAMOTO scholarship foundation and IWATANI NAOJI foundation. I would not have enough free time to focus on research without their financial support. Thank them giving me a chance to make many new friends from different countries and universities.

As the final, I warmly give my appreciation to my dearest parents. Thank them for forgiving me to leave them so far so long. Their selfless love and expectation give me much more power to finish my PhD study.

ABSTRACT

Forests play an important role in ecosystem and environmental management. Especially, in the case of the global warming problem, the benefits of forests reflected in many aspects, such as, the reduction of carbon dioxide in the atmosphere as well as the preparation of basic database by using their distribution in order to estimate the carbon consistency that will control the global warming pollution. However, in contrast to earlier reports about world deforestation prepared by the Food and Agriculture Organization (FAO), where world deforestation caused around 13 million hectares of forests disappeared in 2000s, and forest cover is still declining with a very high rate in some tropical regions. Besides, the gains in forests area have taken place in many countries because of the promotion of tree planting. Therefore, the large scaled and accurate forest maps are indispensable due to better understanding of the information about forest distribution and finally to deal with the constant forest changing.

The objective of this study is to develop a prominent classification method to generate global scale forest and non-forest map with a higher accuracy using ALOS PALSAR 50 m mosaic data. In order to achieve this goal, an object-based approach was used to classify forest and non-forest in South Kalimantan. After selection of an optimal feature combination from 100 kinds of object features, three machine learning classifiers, including J48 (C4.5), Random Forest (RF) and Sequential Minimal Optimization (SMO) were compared.

Finally, the forest and non-forest map in South Kalimantan (2010) was produced with the combinations with RF classifier and J48 classifier. For validation, firstly, forest and non-forest map produced by this study was compared with the New Global 50 m PALSAR Forest/Non-forest map, which was published by Japan

Aerospace Exploration Agency (JAXA), with the generation using a threshold algorithm based on HV backscattering coefficient of 25 m ALOS PALSAR global mosaic data. In this study, the sampling polygons around 300 had been drawn randomly, and these were collected from the different areas within these two products. Finally, 87 sampling polygons were comparably checked by the Google Earth images. There were 59 polygons showed the classification result of this study was correct, while the other 28 polygons showed the classification result of JAXA was correct. The overall accuracy of the forest map produced by this study was 85.43% with kappa coefficient of 0.65.

In addition, the feature combination extracted from the training set that collected in South Kalimantan was applied for forest classification on three areas located in Africa, North America and China. The overall accuracy of these three test areas were 76%, 86% and 70%, respectively. This result indicated that there was a possibility to classify forest and non-forest without collecting new training data globally.

Table of Contents

Chapter 3. Image Correction for ALOS PALSAR 50 m Ortho-rectified Mosaic Data

Chapter 4. Forest Classification Using ALOS PASLAR 50 m Global Scale Mosaic Data

List of Figures

List of Tables

CHAPTER 1
INTRODUCTION

1.1. Background

Forests play an important role in balancing the relationship between human and nature, while a high rate of deforestation had caused astonishment forest disappearing. A loss of forests will result in countless enormous harm on many aspects, not only for the current situation, but also relate with global future implementation.

1) Cause the social problem

According to the statistics data published by the Food and Agriculture Organization (FAO), around 54.2 million people over the world are engaging in the forestry related work, and people in some less developed countries trade wood as their main heater source (FAO, 2014). On the other hand, a large amount of wooden goods are produced every year, like paper, architecture materials, has become the indispensable part of human's living. This means, forest decreasing is a big threat on social unemployment, and the supplication may hardly keep up with the increasing wooden demands in someday.

2) Unbalance ecosystem

The forest ecosystem is the largest ecosystem on land. Millions of plants, animals and microorganisms existing in this natural environment together with sunlight, temperature and the other non-living physical factors. It serves many kinds of animals as a unique habitat with food and shelter, even for more than 350 million indigenous people who are living in rainforests (Sophile Chao, 2012). In addition, forests also have an important impact on soil and water conservation, to keep the balance for a healthy ecosystem. However, forest missing destroys

the forestry environment. Tribal people and the wild animals lose their living home. More and more species disappeared from the world with a rate of 1-10 species gone per year, while deforestation had been proved as the main reason of massive species extinction (Whitmore and Sayer, 1992).

3) Climate change

Since last century, glacier melt has been causing serious concern of global warming. How to prevent the temperature rise of earth surface has become a hot topic between countries. The global warming will not be fixed without decreasing carbon dioxide, because more than half of greenhouse gases is it (Bert Metz *et al*., 2005; IPCC, 2014).

Many reports and researches had shown the sharp increase of carbon emission is caused by the extension of deforestation, since forests are seen as the biggest natural sink, storing one third of the carbon dioxide in the atmosphere (Christopher, 2001; Percy *et al*., 2003). If the tree is cut down, the carbon dioxide absorbed through photosynthesis release back to atmosphere again. This is why forests are so important for the climate change.

Since threats come from forest disappearing has been recognized these years, governments start to focus on forest management and planation. Many countries are joining with forest replanting project in order to extend forest area, like China, India and Thailand (Mead, 2001; Pakkad *et al*., 2001). However, the deforestation is still going on. According to the report of FAO, the highest rate of forest missing is found in Africa and South America. Forest cover was burned or changed to agriculture land use (Annunzio *et al*. 2014).

1.2. The existing forest and non-forest maps

The deforestation continues at a high rate as well as replanting on different areas. The timely

information of forest distribution and its change has been required from region, continent to global. Due to the remote sensing technology has been applying to observe the Earth's surface several decades, using satellite data to provide large scale landscape information within a shorter term had become a uniquely versatile tool.

Last year, the first detailed maps of global forest change from 2000 to 2012 had been published by University of Maryland (Figure 1.1). This production generated from 654,178 Landsat 7 ETM+ images in 30m resolution, with the detailed dataset include tree cover for the year 2000, global forest cover loss (2000-2012), global forest cover gain (2000-2012) and year of gross forest cover loss event (Hansen *et al*., 2013). Google group built a new land observation tool with this database called Global Forest Watch, allows anyone to make use of the forest mapping source for forest management and application.

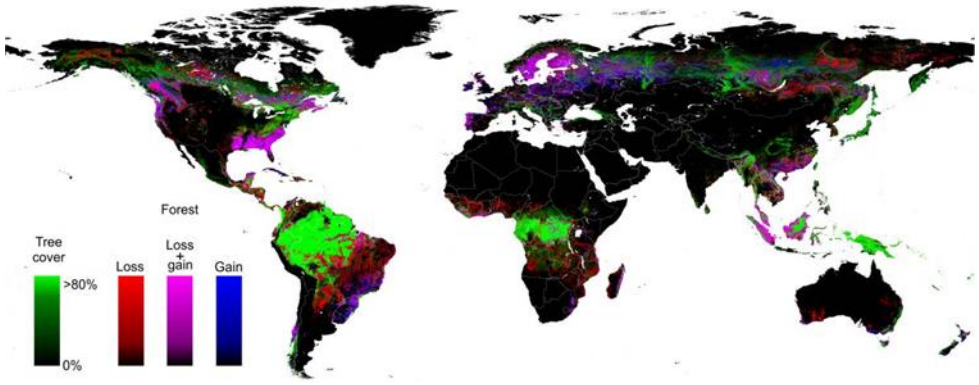


Figure 1.1: First detailed global forest change map made by University of Maryland. (Source: (Hansen *et al*., 2013))

On the other hand, cloud cover results in data missing of optical data, particularly for the rainforest region where is always covered by cloud. In the case of global forest change, the images for the region where there is no data caused by cloud, were replaced by the cloud-free

images took from the nearest year, within the range of 1999-2012 (Hansen *et al*., 2013). In order to avoid the cloud problem of optical data, Synthetic Aperture Radar (SAR) is considered as the best way to observe forests because of its high capability of penetration. In January 2014, Japan Aerospace Exploration Agency (JAXA) took a public of a new 4-year global 50 m Forest/Non-forest map produced by using Phased Array type L-band SAR (PALSAR) data. This map set consists of the forest distribution of 2007, 2008, 2009 and 2010, generated from 25m global mosaic data with the accuracy of about 90% in global scale (Figure 1.2). The methodology of forest extraction begun with segmentation, while only using HV gamma-naught (γ°) to separate forest and non-forest (Shimada *et al*, 2014).

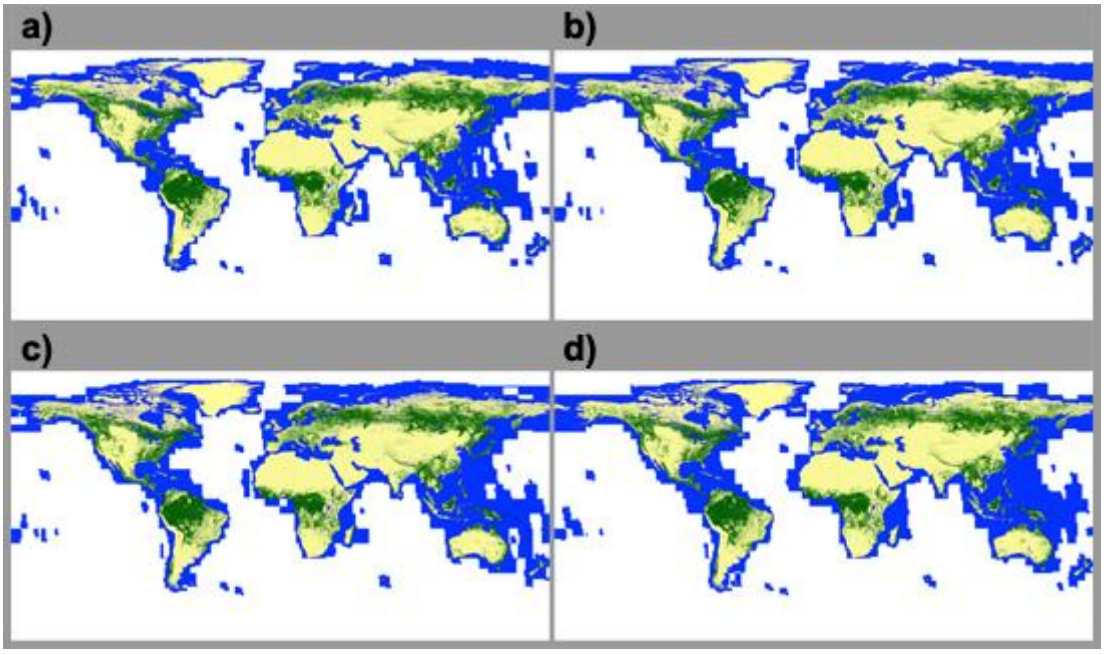


Figure 1.2: New Global 50 m PALSAR Forest/Non-forest map made by JAXA. (a) 2007; (b) 2008; (c) 2009; (d) 2010). Green: forest; Yellow: non-forest; Blue: water body. (Source: JAXA)

1.3. Objectives of this study

As described in the above sections, in order to meet the needs of forest information, many organizations are focusing on mapping forest cover with satellite data. At the same time, a higher requirement of improving the accuracy of classification has become a new challenge. Usually, the accurate land use and land cover information is mainly affected by the factors like data processing and classification methods. In this study, freely-available high resolution 50 m Advanced Land Observing Satellite (ALOS) PALSAR mosaic data over large scale provided by JAXA, was used to extract forest and non-forest area.

The main objective of this study is to develop a classification method for generating global scale forest map with a higher accuracy. In addition, there are two sub-objectives in order to achieve the main purpose. They are:

1) Removing terrain influence from ALOS PALSAR 50 m ortho-rectified mosaic data, which is the unique free large scale SAR data with high resolution from 2008 to 2014.

2) Selecting the best classifier for classifying forest and non-forest from three well-known machine learning classifiers, which are J48 (C4.5), Random Forest (RF) and Sequential Minimal Optimization (SMO).

1.4. Structure of the thesis

Five chapters and conclusion sections are presented in this thesis.

Chapter 1 starts with the description of importance of forests and why forest mapping is so necessary. In this chapter, two existing global forest maps using optical Landsat data and microwave ALOS PALSAR data are introduced. In addition, the objectives and structure of this thesis are explained in this chapter.

Chapter 2 presents data acquisition and the reference global forest map used for comparing with the result of this study.

Chapter 3 explains a new modified model for removing the terrain influence of ALOS PALSAR 50 m ortho-rectified mosaic data. The application of three existing slope correction models also are introduced in this chapter.

Chapter 4 shows the methodology to produce the forest and non-forest map. The main steps of this chapter consist of segmentation, feature selection and comparison of classifiers. A land cover map is generated with eight classes firstly. Then, the forest and non-forest map is produced based on the land cover classification result. In this chapter, combination of different classifiers to extract forests and separating forests with other class one by one are the new points for forest classification.

Chapter 5 compares the generated forest map in this study to the New Global 50 m PALSAR Forest/Non-forest map produced by JAXA. After discussing the result of comparison, the accuracy assessment is carried out by using Google Earth images. The application of the training sites used in this study is also explained.

Conclusion section summarizes all the results of this study.

CHAPTER 2
DATA ACQUISITION

2.1. ALOS PALSAR data

In January 2006, as the succession of Japan Earth Resources Satellite (JERS-1), JAXA successfully launched the Advanced Land Observing Satellite (ALOS) for the purpose of Earth observation, land mapping, and disaster monitoring. Three remote-sensing instruments onboard ALOS are two optical sensors, which are PRISM and AVNIR-2, and Phased Array type L-band Synthetic Aperture Radar (PALSAR).

The PALSAR instrument is used for day-and-night and all-weather land observation with the Fine and ScanSAR modes. Table 2.1 shows the default modes with the polarization over global scale monitoring (Source: JAXA).

Table 2.1: PALSAR default modes over global scale observation.

Sensor mode	**Polarization**	**Off-nadir angle**	**Pass designation**	**Time window**	**Observation frequency**
Fine Beam Single (FBS) polarization	HH	34.3°	Ascending	Dec-Feb	1-2 obs./year
Fine Beam Dual(FBD) polarization	HH+HV	34.3°	Ascending	May-Sept	1-4 obs./year
ScanSAR 5-beam short burst	HH	20.1°-36.5°	Descending	Jan-Dec	1 obs./year

Since 1st July 2008, ALOS Kyoto and Carbon (K&C) Initiative, an international collaborative

project, which led by the Earth Observation Research Center (EORC) of JAXA, opened K&C mosaic homepage. Mosaic products is a special projection aim to develop available basic data for supporting three main themes of K&C Initiative project: observing forests, wetlands and desert & water.

2.1.1. ALOS PALSAR 50 m ortho-rectified mosaic data

ALOS PALSAR 50 m ortho-rectified mosaic data was created with FBD HH and HV polarization from the ascending path globally from 2007 to 2009 by JAXA. JAXA uploaded this data to their homepage after image processing was done. Until 2010, the data was published on this homepage for eleven regions include Japan, Indochina, Sumatra, Borneo/Kalimantan, Philippines, Sulawesi, Jawa, New Guinea, Solomon Islands, Australia and Central Africa (Figure 2.1).

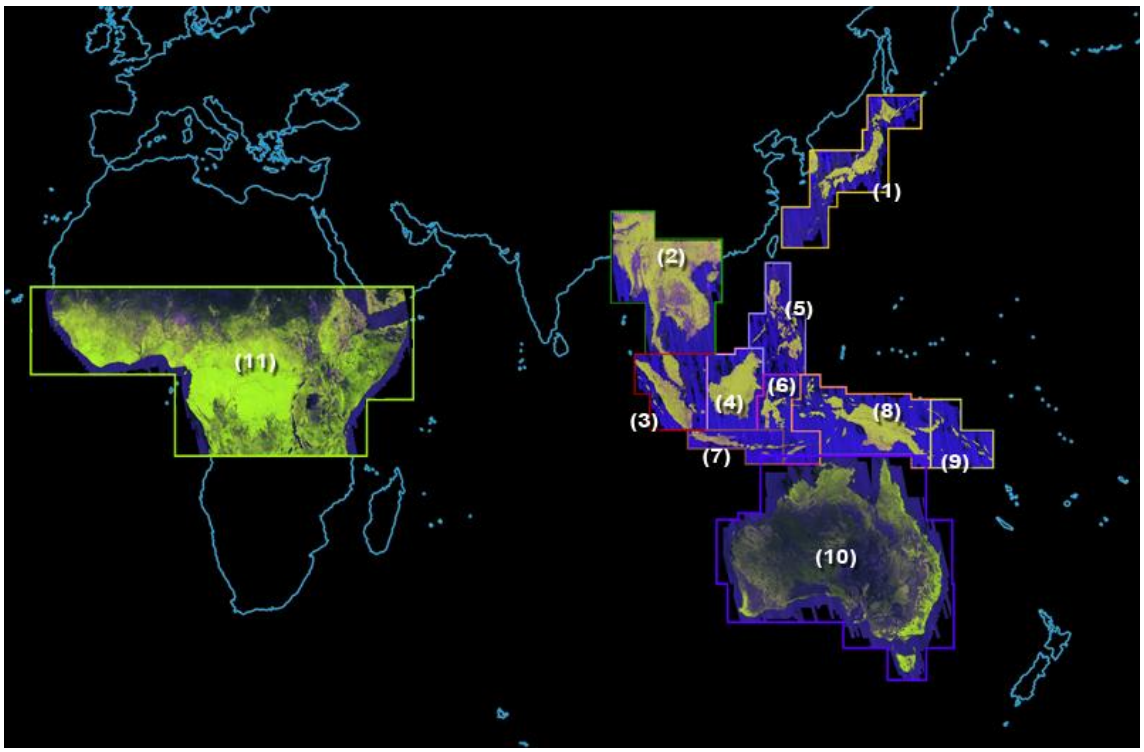


Figure2.1: ALOS PALSAR 50 m ortho-rectified mosaic data. (1) Japan; (2) Indochina; (3) Sumatra; (4) Borne/Kalimantan; (5) Philippines; (6) Sulawesi; (7) Jawa; (8) New Guinea; (9)

Solomon Islands; (10) Australia; (11) Central Africa. (Source: K&C mosaic homepage)

Begun with Sampling Window Start Time (SWST) processing, slant range image is produced after reduce the noise from ground surface. Geometric calibration is applied along with range direction, ortho-rectification is carried out with SRTM 90m and geo-referenced into latitude and longitude coordinate system. In order to keep the characteristic of topographic, slope correction has not applied to this mosaic data (Shimada *et al*., 2008; Longepe *et al*., 2011).

[1]  [2] 

Figure 2.2: An example of 50 m ortho-rectified mosaic data. [1] HH polarization; [2] HV polarization.

Figure 2.2 shows an example of 50 m ortho-rectified mosaic data. Mountaineous area is clearly identified both on HH and HV images because of the lack of terrain correction. The instersting of topograpgic is kept successfully, but the obvious variance of homogeneous backscattering will bring a bad effect on classification result (Bayer *et al*., 1991; Sun *et al*. 2001). Therefore, especially for mapping tropical forests, where most of the area is over mountains, slope correction is nessecessary before classification process.

2.1.2. ALOS PALSAR 50 m global mosaic data

From January 2014 to 31st Oct, JAXA prepared and published a new data source of ALOS PALSAR in 50 m spatial resolution on K&C mosaic homepage. This mosaic data was resampled from 25m mosaic data, which became available from 31st Oct 2014. Four years data including 2007, 2008, 2009 and 2010 were produced on global. Not only ortho-rectification, but also slope correction were applied with SRTM 90m from raw data. Both geometric and radiometric of image correction are described with very high accuracy (Shimada *et al*. 2014; Shimada *et al*. 2009).

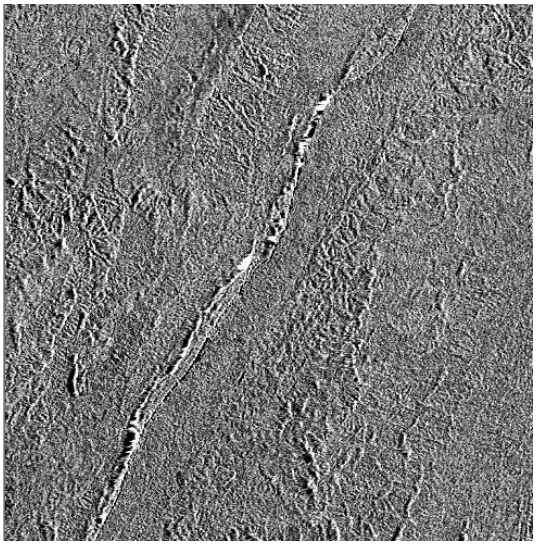
[1]  [2] 

Figure 2.3: An example of the global slope-corrected mosaic data. [1] HH polarization; [2] HV polarization.

Figure 2.3 shows the slope-corrected image of HH and HV polarization for the new ALOS PALSAR 50 m ortho-rectified mosaic data over the same location with Figure 2.2. The terrain change have been removed from obvious mountainous areas.

2.2. Other reference data

The other referenced data used in this study including:

- High-resolution images displayed on Google Earth.

  As it is difficult to collect the ground truth data by field survey, the high-resolution images displayed on Google Earth was used for collecting ground truth data for training, comparison and validation. The use of Google Earth enabled to obtain high-resolution images in inaccessible places.

- SRTM 90m digital elevation database (version 4.1).

  This global Digital Elevation Model (DEM) was generated by NASA originally, and was released by the Consortium for Spatial Information (CGIAR-CSI) of the Consultative Group for International Agriculture Research (CGIAR) after filled void problem as an open data source. In this study, SRTM 90m data was downloaded for slope correction of ALOS PALSAR ortho-rectified mosaic data.

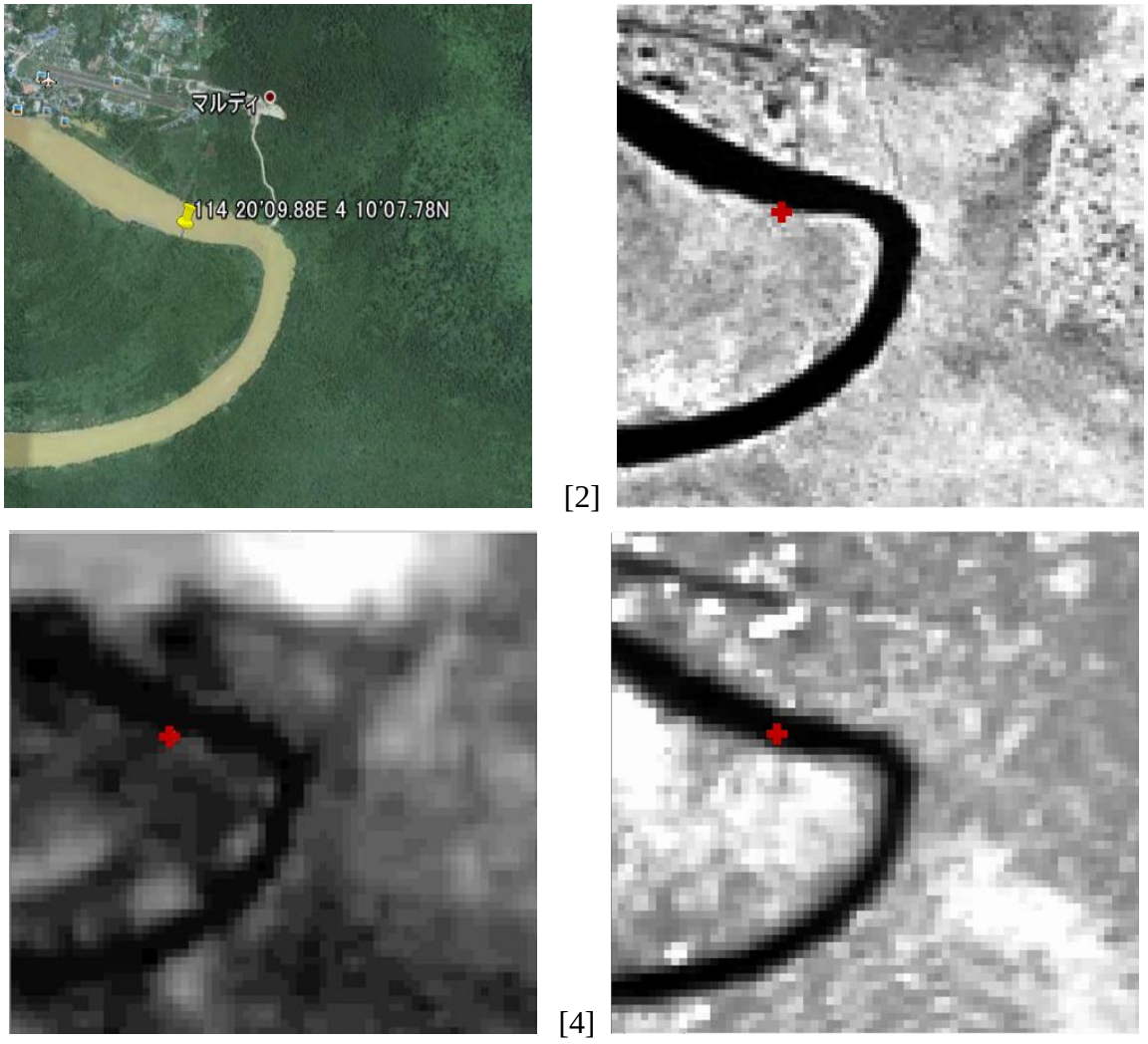
- New Global 50 m PALSAR Forest/Non-forest map in 2010 (version 0).

  The new global 50 m-resolution forest/non-forest map was published on K&C mosaic homepage by JAXA together with the 50 m slope-corrected mosaic data. It was produced by resampling from the original forest map based on PALSAR 25 m mosaic data. The normalized radar cross section with gamma-naught (γ°) of HV polarization was used for deciding the threshold of forest area. In this study, forest distribution extracted using the developed classification method were compared with the new forest/non-forest map produced by JAXA, Google Earth image was used as the reference image.

CHAPTER 3

IMAGE CORRECTION FOR ALOS PALSAR 50 m ORTHO-RECTIFIED MOSAIC DATA

3.1. Geometric correction



[1] [2] [3] [4]

Figure 3.1: Position shift problem. [1] Google Earth image; [2] Landsat image (band 4); [3] SRTM image; [4] ALOS PALSAR HH image.

Due to Shuttle Radar Topography Mission (SRTM) 90 m data was going to be used as digital elevation models (DEM) to reduce the terrain influence of ALOS PALSAR 50 m ortho-

rectified mosaic data, nearest neighbor resampling process was carried out to SRTM data firstly. In addition to this, in order to resolve the position shift problem occurred within these two data, Landsat TM / ETM+ images with a spatial resolution of 30 m were downloaded from Global Land Cover Facility (GLCF) and used as standard images with correct position.

The problem of position shift could be observed in Figure 3.1. The red plus sign represents a position where the coordinate is 114d 20'09.88E, 4d 10'07.78N. Ground truth landscape showed that the red plus sign should be on the edge of river by Google Earth (Figure 3.1 [1]). The detected point of Landsat TM / ETM+ (Figure 3.1 [2]) and resampled SRTM image (Figure 3.1 [3]) corresponded with Google Earth, while that of ALOS PASLAR HH image located in the river (Figure 3.1 [4]).

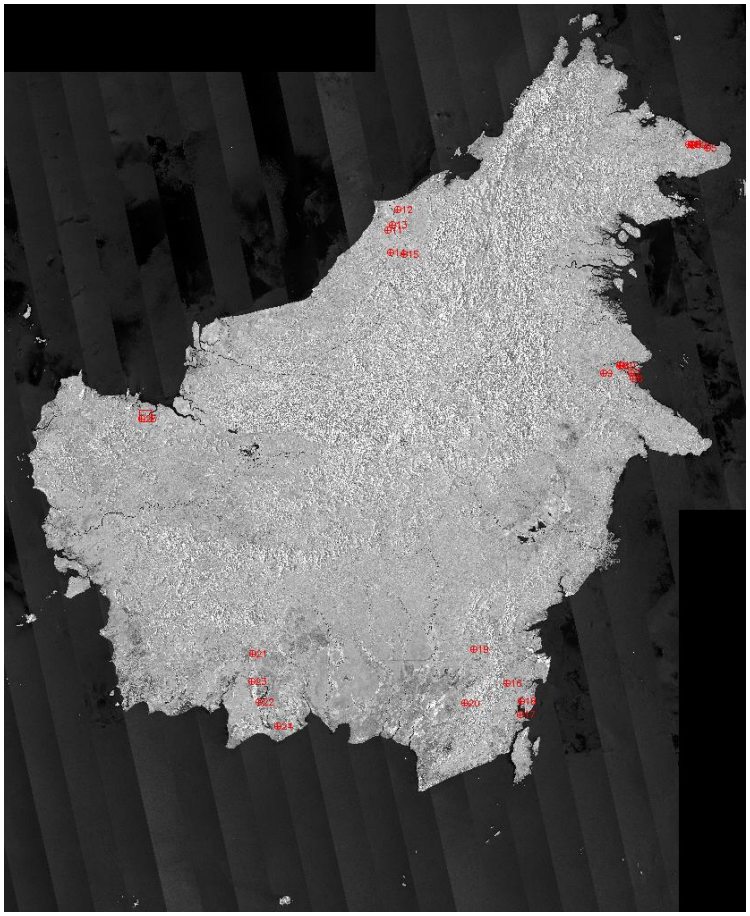


Figure 3.2: Distribution of 25 Grond Control Points.

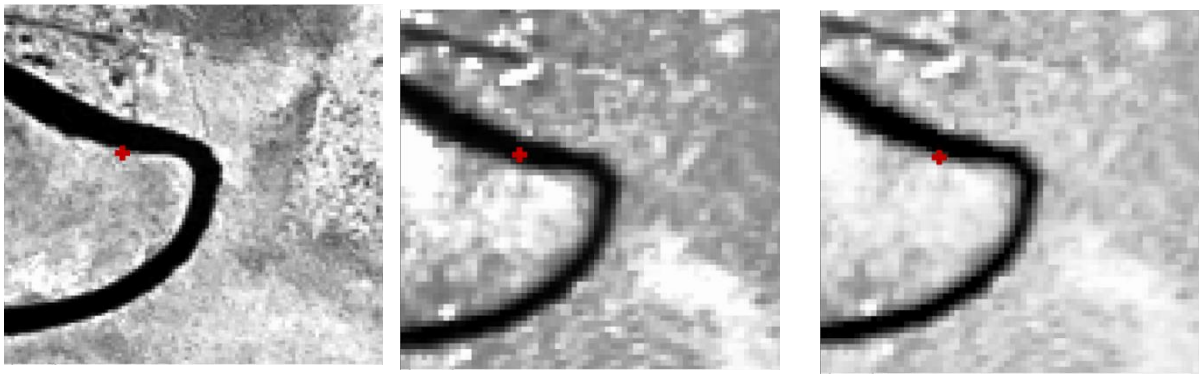
Table 3.1: RMS error of each corrected point.

Point	RMS
1	0.5565
2	0.8537
3	0.4373
4	0.5674
5	0.9368
6	0.797
7	0.6264
8	0.2256
9	0.6061
10	0.5908
11	0.3619
12	0.8411
13	0.9441
14	0.234
15	0.231
16	0.4678
17	0.1987
18	0.7076
19	0.6332
20	0.8091
21	0.5194
22	0.2519
23	0.5773
24	0.6772
25	0.2577

To match the position of ALOS PALSAR HH image with SRTM data, Ground Control Points (GCP) were collected for making geometric correction for ALOS PALSAR 50m ortho-rectified mosaic data. Totally, 25 points were collected around the edge of Kalimantan using six Landsat/ ETM images and shown in Figure 3.2. This processing was conduced by using ENVI 4.3 software tool. The accuracy of translated point is described with Root Mean Square (RMS) Error, which is calculated by:

$$\text{RMSerror} = \sqrt{(x_r - x_i)^2 + (y_r - y_i)^2} \quad (3.1)$$

Where x_i and y_i are the input original coordinates, x_r and y_r are the retransformed coordinates. The RMS error of each point after geometric correction is shown on Table 3.1. A well correction quality could be proved with the maximum RMS error which is less than 1. From visual interpretation of Figure 3.3, PALSAR image is revised about two pixels after geometric correction successfully.



a b c

Figure 3.3: Geometric corrected image. (a) Landsat image; (b) PALSAR image before geometric correction; (c) PALSAR image after geometric correction.

3.2. Slope correction

Some methods of calibration need to be applied to original SAR data before further investigation, because of the amount of distortion that happens on the image (e.g., Speckle filtering, geometric correction and radiometric correction). As an important processing step to reduce the topography influence, different slope correction models had been generated based on the cosine correction method and the scattering area changing method. Many of these models were dealt with the terrain correction with different code-level programming or software tool (Loew and Mauser, 2007; Shimada *et al*., 2014). Therefore, the order of data processing or the processing environment may result in different slope correction effect. In this section, three existing slope correction models were tested to perform their restoration

capability for ALOS PALSAR 50 m ortho-rectified mosaic data.

3.2.1. Test area

The existing formulas were applied with a test area to investigate the slope correction effect for ALOS PALSAR 50 m ortho-rectified mosaic data, where is located within the West Coast Division of Sabah, Malaysia (116d01'55.7685"E, 5d56'25.2183"N and 116d08'43.2737"E, 5d51'19.5894"N). This testing area approximately 12.53km × 9.39km, and DEM (Digital Elevation Model) ranges from 0m to 507m. Figure 3.4 shows the location with the color composite image of PALSAR data (R=HH, G=HV, B=HH-HV). The mountain area is clearly seen, espacially on the bottom-right corner.

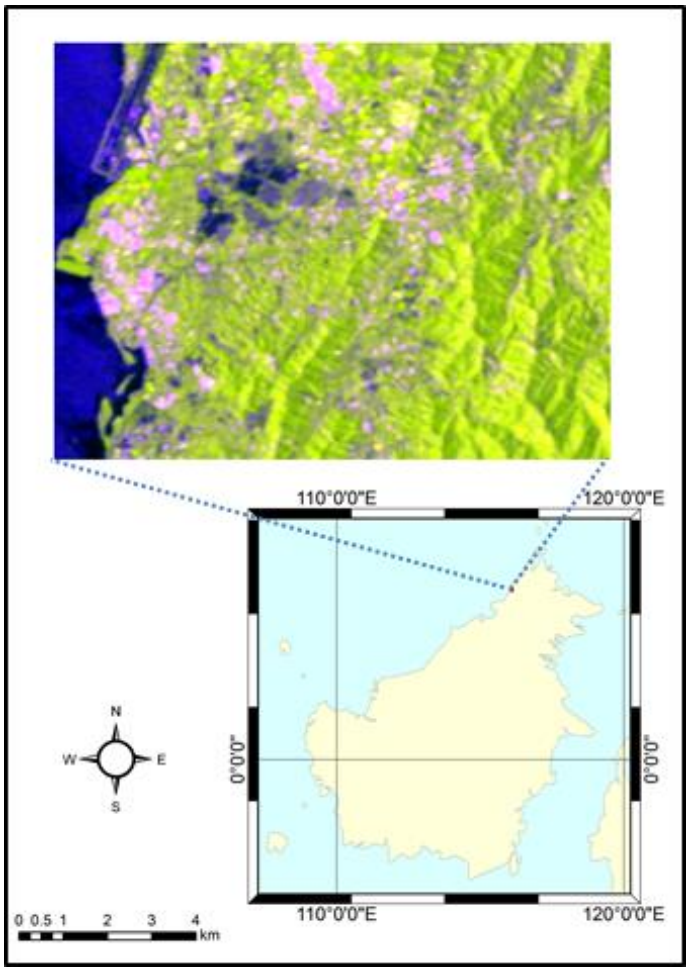


Figure 3.4: Composite RGB image of PALSAR data over study area. (R:HH; G:HV; B:HH-

HV).

3.2.2. Application of the previous slope correction models

Firstly, Digital Number (DN) of HH and HV polarization images were converted to the normalized radar cross section (Sigma-zero) by the following equation :

$$\sigma^\circ = 10*\log_{10} DN^2 + CF \tag{3.2}$$

Where *DN* is the digital number of HH and HV images, Calibration Factor (*CF*) for ALOS PALSAR 50 m ortho-rectified mosaic had been given as (-83) and σ° is the backscattering coefficient (dB).

The brief description of these models are shown in Table 3.2. Model-1 and model-2 were proposed by the cosine correction method (Akatsuka *et al*., 2009; Kellndorfer *et al*., 1998; Rokhmatuloh *et al*., 2012), while model-3 was proposed based on the scattering changing method (Castel *et al*., 2001; Santoro *et al*. 2011). The main calculation steps consist of:

1) Calculation of local incidence angle (θ_{loc})

In this study, θ_{loc} was derived by the following equation which described by Akatsuka et al. (2009):

$$\cos\theta_{loc} = \cos\theta\cos\alpha + \sin\theta\sin\alpha\sin(\phi - \beta) \tag{3.3}$$

Here, the slope α and aspect angle β of SRTM were exported from spatial analyst tools of ArcGIS software. The azimuth angle of PALSAR platform ϕ is 261.84 degree, and θ is equal with the off-nadir angle 34.3 degree.

2) Calculation of local ground scattering area (*A*)

Castel *et al*. (2001) provided a sample equation to describe A over a flat terrain as the following equation:

$$A_{flat} = \frac{r_a r_s}{\sin\theta_{loc}} \tag{3.4}$$

Where r_a and r_s represent the azimuth and slant range pixel spacing respectively. On the other hand, the method for computing A_{slope} was selected from the literature published by Wegmuller (1999) :

$$A_{slope} = \frac{r_a r_s}{\cos\psi} \tag{3.5}$$

Where ψ is the projection angle which defined as the angle between the surface normal and the image plane normal (Ulander, 1996)

$$\cos\psi = \sin\theta\cos\alpha + \cos\theta\sin\alpha\sin\beta \tag{3.6}$$

Here, α, β and θ represent the same meaning within Equation (3.3).

3) Value decision of θ_{ref} and n

θ_{ref} of model-2 and model-3 means a reference incidence angle which was defined as 34.3 degree in this study. Model-3 was applied to correct ALOS PALSAR 50 m ortho-rectified mosaic HH and HV image when n is 0.7.

Table 3.2: Three existing slope correction models.

	Existing slope correction models		
	Model_1	**Model_2**	**Model_3**
Equation expression	$R_c = \frac{R}{\cos^4\theta_{loc} + (1 - \cos^4\theta_{loc})}$	$\sigma^\circ{}_{corr} = \sigma^\circ \frac{\sin\theta_{loc}}{\sin\theta_{ref}}$	$\gamma^\circ = \sigma^\circ \frac{A_{flat}}{A_{slope}} (\frac{\cos\theta_{ref}}{\cos\theta_{loc}})^n$
Authors	Akatsuka et al. (2009); Japan Aerospace Exploration Agency (2009)	Kellndorfer et al. (1998); Rokhmatuloh et al.(2012)	T. Castel et al. (2001); M. Santoro (2011)
Symbol explanation	$\boldsymbol{R_c}$: Corrected digital number of SAR image $\boldsymbol{R}$: Original digital number of SAR image $\boldsymbol{\theta_{loc}}$: Local incidence angle	$\boldsymbol{\sigma^\circ{}_{corr}}$: SAR backscatter coefficient after calibration $\boldsymbol{\sigma^\circ}$: Original SAR backscatter coefficient $\boldsymbol{\theta_{loc}}$: Local incidence angle $\boldsymbol{\theta_{ref}}$: SAR incidence angle at the center of the image	$\boldsymbol{\gamma^\circ}$: SAR backscatter coefficient after calibration $\boldsymbol{\sigma^\circ}$: Original SAR backscatter coefficient ***A :*** Local ground scattering area within a pixel $\boldsymbol{\theta_{loc}}$: Local incidence angle $\boldsymbol{\theta_{ref}}$: Incidence angle at mid-swath ***n:*** 0.7<= n <=1

3.2.3. A modified slope correction model

Based on a sample backscatter terrain correction model, a modified slope correction model for specially calibrating ALOS PALSAR 50 m ortho-rectified mosaic data of this study was generated with the regulation that the homogeneous land cover target should have the similar backscattering property regardless of any topography terrain (Kellndorfer *et al*., 1998). This sample model had been published by Ulaby *et al*. (1996) and Sun *et al*. (2002) as:

$$\sigma^{\circ} = \sigma \cos^{p} \theta_{loc} \tag{3.7}$$

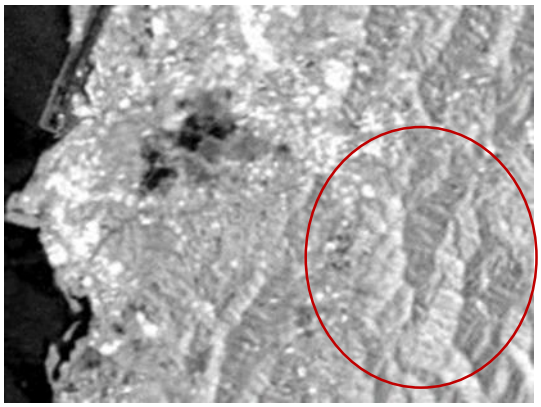
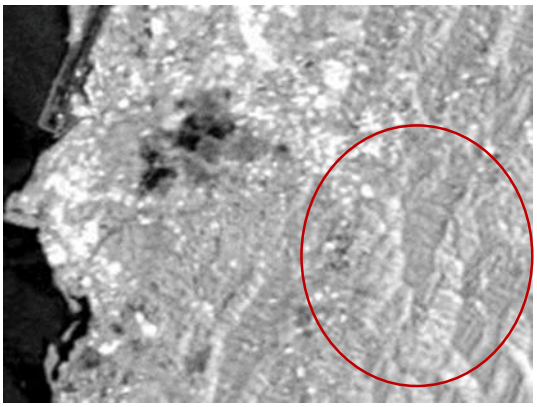
[1]  [2] 

Figure 3.5: Slope corrected result of HH image by using Equation (3.7). [1] Original HH image; [2] Slope corrected HH image.

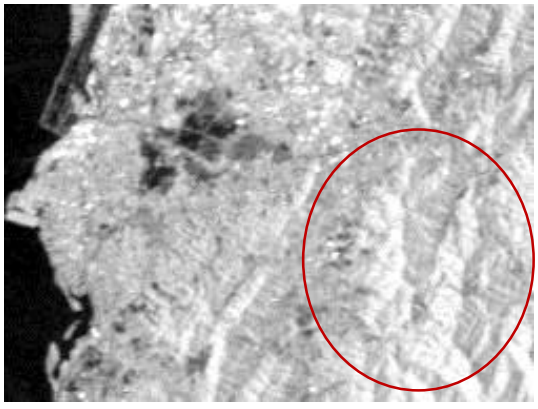
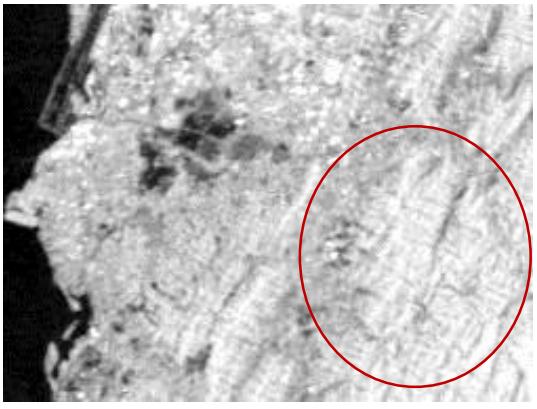
[1]  [2] 

Figure 3.6: Slope corrected result of HV image by using Equation (3.7). [1] Original HV image; [2] Slope corrected HV image.

Where σ and σ° are backscattering coefficient before and after terrain correction, respectively. Sun *et al*. (2002) carried out this model both for HH polarization and HV polarization of L-band wave, and successfully induced the terrain effect with the changing of power *p*, where 1<=*p*<=2.

Figure 3.5 and Figure 3.6 show the corrected images of ALOS PALSAR 50 m mosaic data by using Equation (3.7) when *p* is 1. Slope corrected HV image (Figure 3.6 [2]) shows a more efficient correction on brightness variation than HH image (Figure 3.5 [2]) over the mountain areas. Therefore, the limitation of this model is required to be improved for HH image.

Figure 3.7 shows a sample scattering geometry on the ground surface. Suppose the scattering surface over flat area that has standard backscattering behavior, each target over the tilted area will get an assumptive standard reference. Therefore, in Figure 3.7, A is a real target point (one pixel) of inclined plane face with satellite, the backscattering coefficient of B (σ_B) is considered as A (σ_A)'s standard behavior. Local incidence angle of B is equal to the off-nadir angle (incidence angle at the center of the image,θref). Then, the relationship of A and B is considered with backscattering coefficient and local incidence angle as:

$$\sigma_B \cos\theta_{ref} = \sigma_A \cos\theta_{loc} \tag{3.8}$$

Therefore, the assumptive standard backscattering behavior (σ_B) can be calculated from Equation (3.8):

$$\sigma_B = \sigma_A \frac{\cos\theta_{loc}}{\cos\theta_{ref}} \tag{3.9}$$

Here, we call ($\cos\theta_{loc} / \cos\theta_{ref}$) as the strengthened correction factor for HH polarization.

In addition, according to the geometry theorem, the power of *p* is decided by the relationship of OB and OA:

$$\frac{OB}{OA} = \frac{OC}{OD} = \frac{H}{H-h} \tag{3.10}$$

Where *H* is the satellite's height, and *h* is the DEM. Combing the equations above, leads to the new terrain correction model for ALOS PALSAR 50 m mosaic data following with:

$$\sigma^{\circ}_{corr\_HH} = \sigma^{\circ}_{HH} \cos^{\frac{H}{H-h}} \theta_{loc} \frac{\cos\theta loc}{\cos\theta ref} \tag{3.11}$$

$$\sigma^{\circ}_{corr\_HV} = \sigma^{\circ}_{HV} \cos^{\frac{H}{H-h}} \theta_{loc} \tag{3.12}$$

Where $\sigma^{\circ}_{corr\_HH}$ and $\sigma^{\circ}_{corr\_HV}$ mean the backscatter coefficient after slope correction for HH image and HV image, and σ°_{HH} and σ°_{HV} mean the original backscatter coefficient of HH image and HV image, respectively.

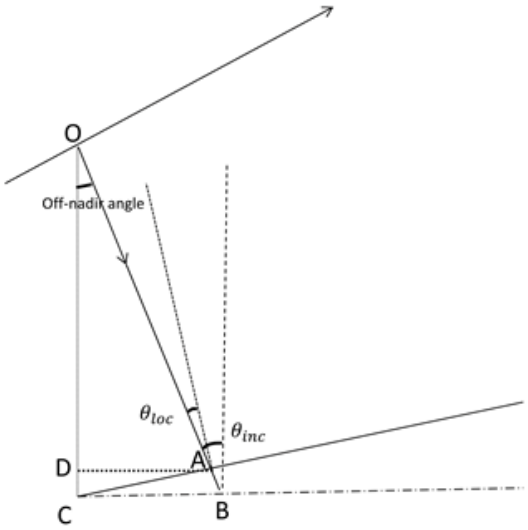
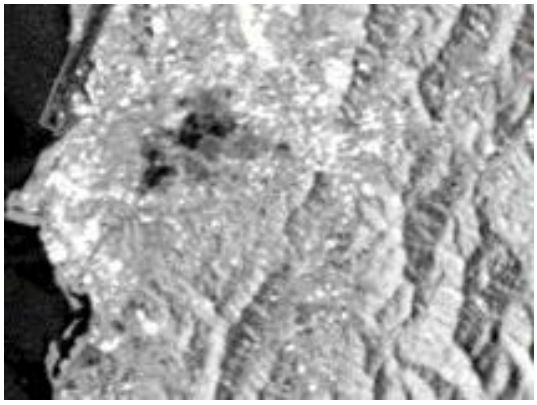
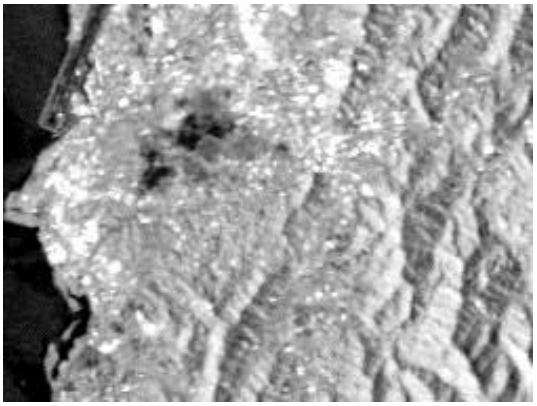


Figure 3.7: Ground scattering geometry. θ_{loc} is the local incidence angle, θ_{inc} is 34.3 degree.

3.2.4. Comparison of each correction result

The visual verification and logical consistency were carried to compare the quality of each slope corrected result. As can be seen, the images from model-1, model-2 and model-3 in Figure 3.8 and Figure 3.9 have less impact on changing the backscattering brightness variation over mountain area, while the corrected image generated using Equation (3.11) and Equation

(3.12) show the mountain area had been changed to flat.

[1]  [2] 

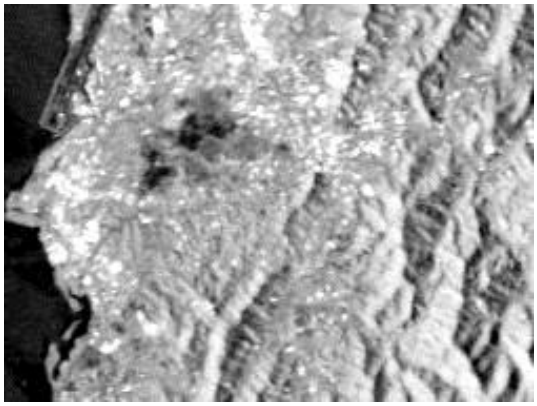
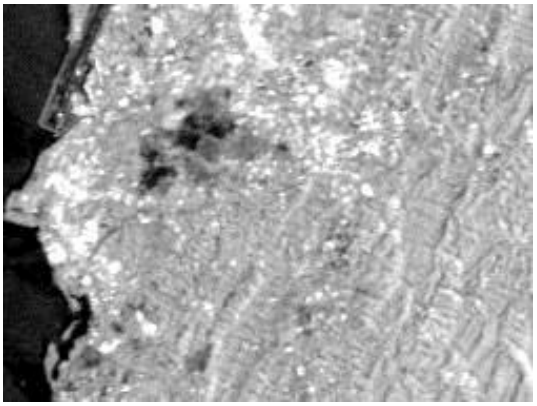
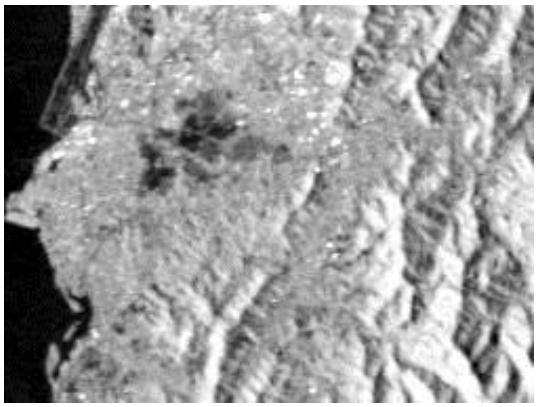
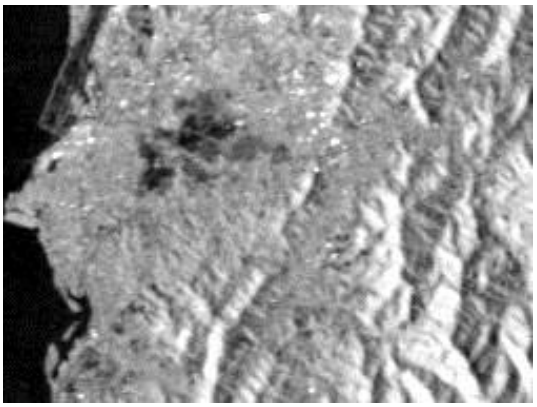
[3]  [4] 

Figure 3.8: Slope correction results of HH image extracted from each slope correction model. [1] Corrected by model-1; [2] Corrected by model-2; [3] Corrected by model-3; [4] Corrected by Equation (3.11).

[1]  [2] 

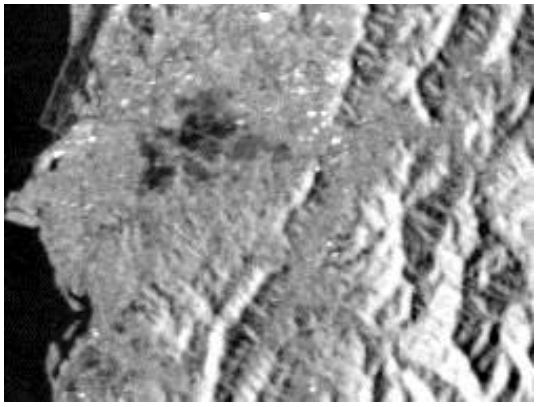
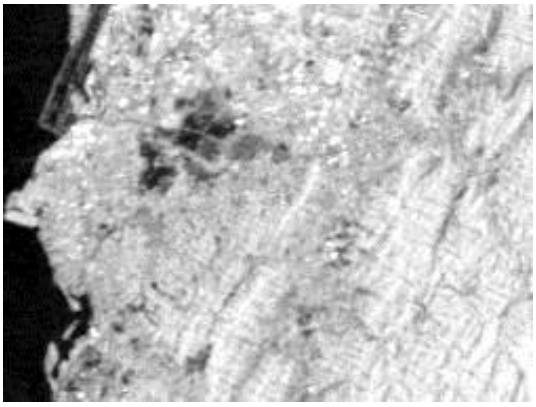
[3]  [4] 

Figure 3.9: Slope correction results of HV image extracted from each slope correction model. [1] Corrected by model-1; [2] Corrected by model-2; [3] Corrected by model-3; [4] Corrected by Equation (3.12).

Taking into account a key factor that smaller brightness variance represents the better terrain correction quality, the backscattering variance over homogeneous land cover of HH polarization and HV polarization were analyzed and represented in Figure 3.10 and Figure 3.11. X-axis means the terrain slope angle, y-axis means the average backscattering coefficient. The difference between the maximum backscattering coefficient and minimum backscattering coefficient of original HH image is about 6.8dB, while the value of 9.6dB, 9.5dB, 11.1dB and 3.1dB were calculated from the result of model-1, model-2, model-3 and the modified equation of this study. In the case of HV polarization, the difference between the maximum and minimum backscattering coefficient of the original image, each existing model and the modified equation of this study are 5.4dB, 12.2dB, 14.2dB, 17.7dB and 2.8dB, respectively. The slope corrected image of Kalimantan are shown in Figure 3.12 and Figure 3.13.

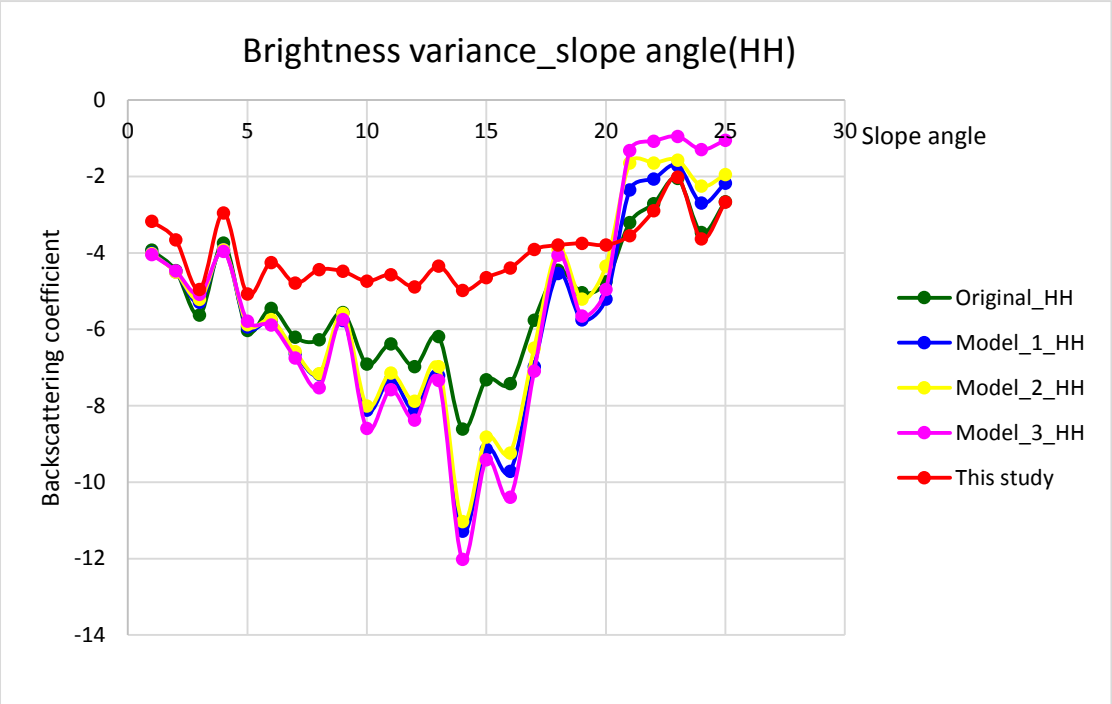


Figure 3.10: Brightness variance of HH image over the homogeneous mountain area.

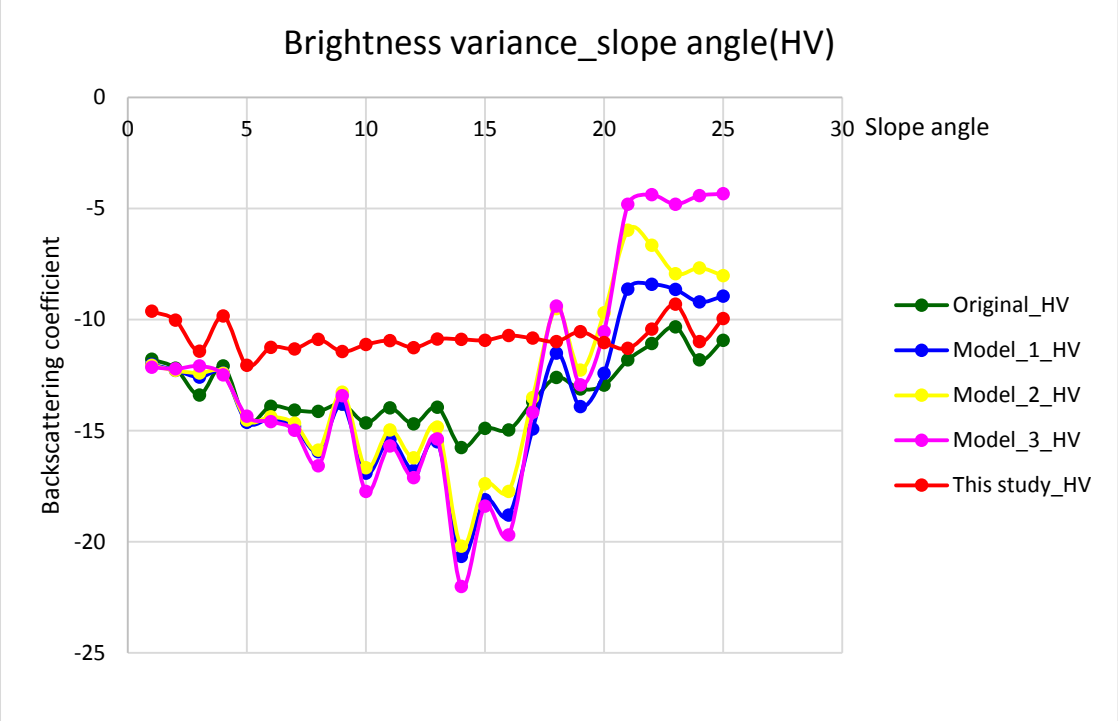
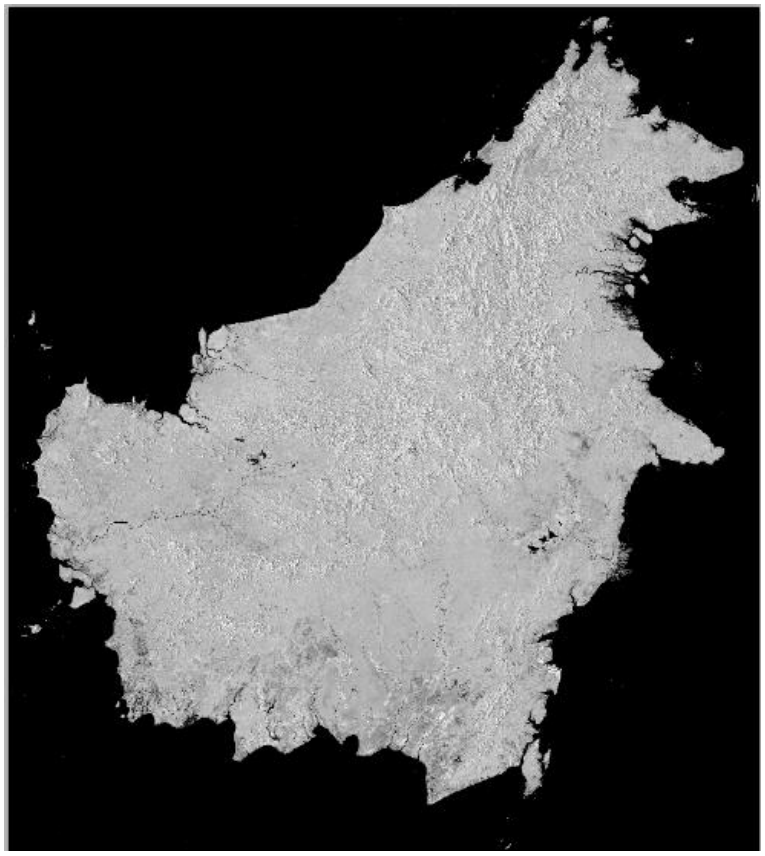
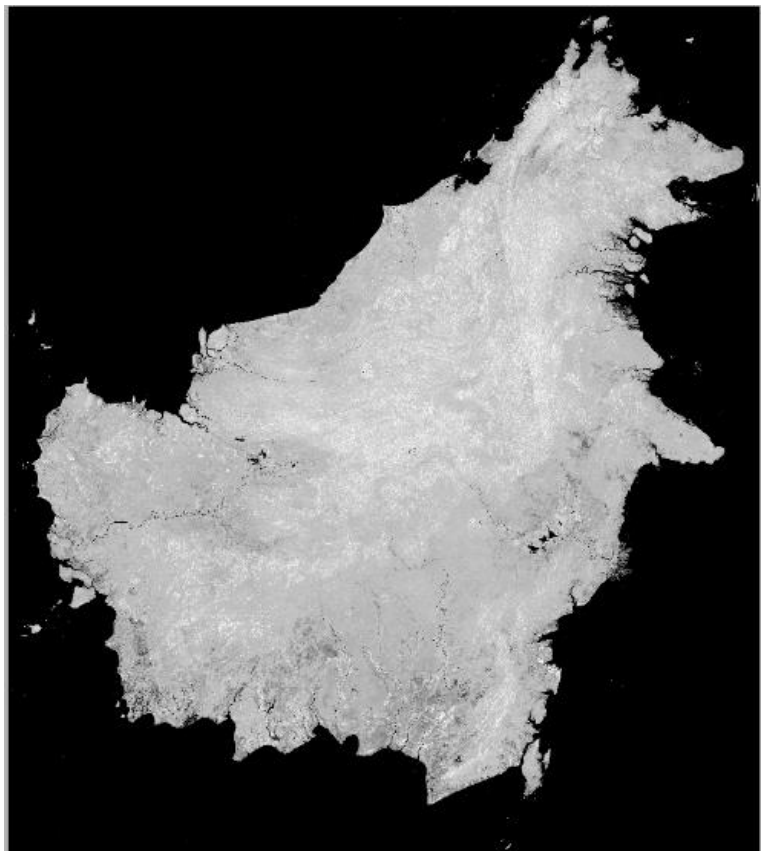


Figure 3.11: Brightness variance of HV image over the homogeneous mountain area.

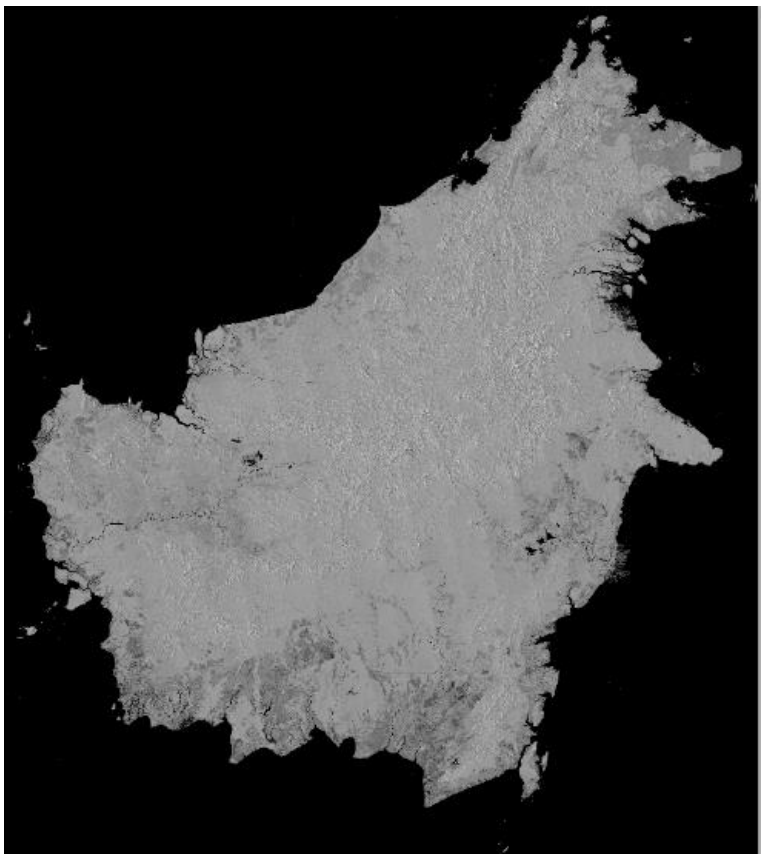


[1] Original HH image

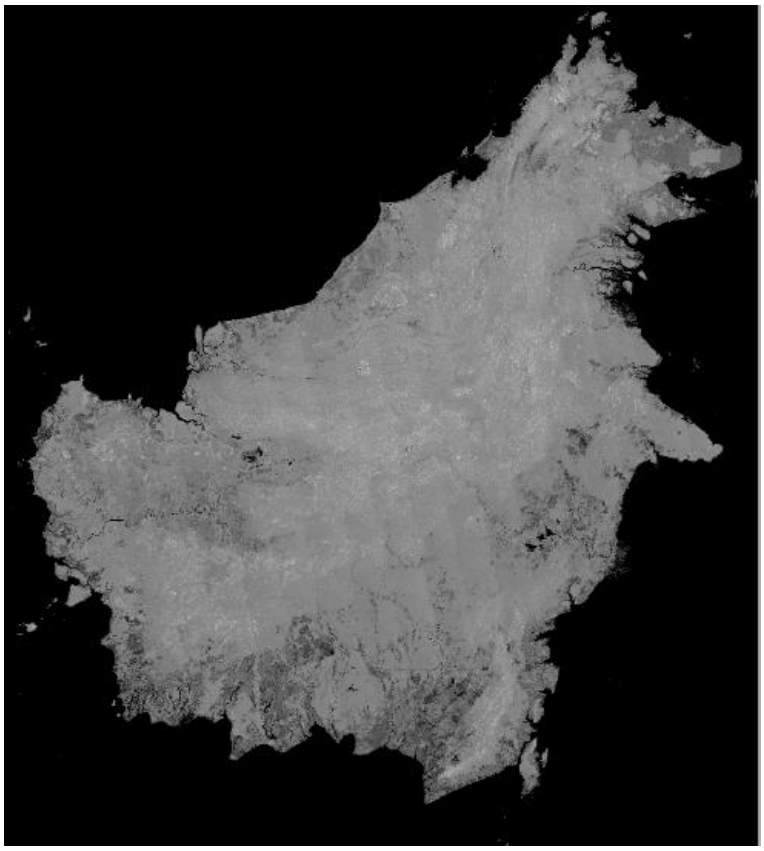


[2] Slope corrected HH image

Figure 3.12: Slope corrected HH image in Kalimantan.



[1] Original HV image



[2] Slope corrected HV image

Figure 3.13: Slope corrected HV image in Kalimantan.

3.2.5. Comparing with ALOS PALSAR 50 m global mosaic data

Slope corrected images generated based on the developed model of this study showed a stronger reduction than the other three existing models both in visual interpretation and backscattering variance analysis. Meanwhile, in January 2014, JAXA published a new global scale forest and non-forest map product within four years period (2007, 2008, 2009, 2010), together with a set of feasible ALOS PALSAR 50 m resolution HH and HV mosaic imagery which have been processed with well-done geometric and slope correction based on the raw data (Shimada *et al*, 2014; JAXA, 2014).

With the purpose of comparing the quality of slope corrected imagery in this study with the new public mosaic data, backscattering behavior analysis was carried out over homogeneous area where is covered by tropical broadleaf forests. In Figure 3.14, the brightness performance of Area 1 and Area 2 should have similar reflection feature of forests, but they represent an obvious different variation of brightness.

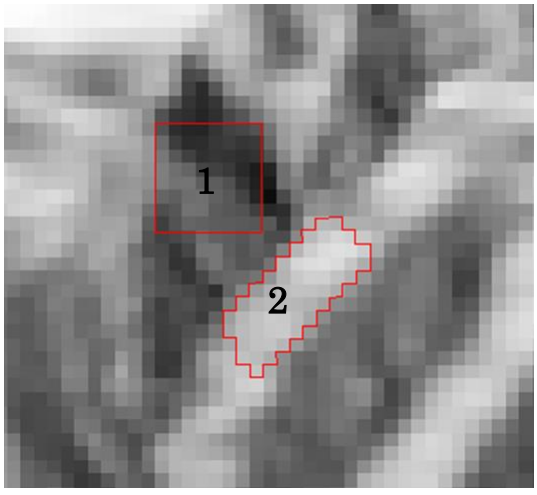


Figure 3.14: Brightness variance on the homogeneous land cover.

The extracted backscattering coefficient of Area 1 and Area 2 were shown in Figure 3.15 ~ Figure 3.17. These three figures express the distribution of backscattering coefficient of the

original HH image, slope corrected HH image by using the proposed modified model (Equation 3.11), and the new public HH image produced by JAXA, respectively. *X*-axis means the number of extracted pixels, *y*-axis means the backscattering coefficient of each pixel.

In Figure 3.15, the distribution of Area 1 and Area 2 were separated with each other like two different clustering. This is out of the ground truth that they belong to the same class. After applied with Equation 3.11, the values of Area 1 were changed closer to Area 2 in Figure 3.16, while mixed together well in Figure 3.17.

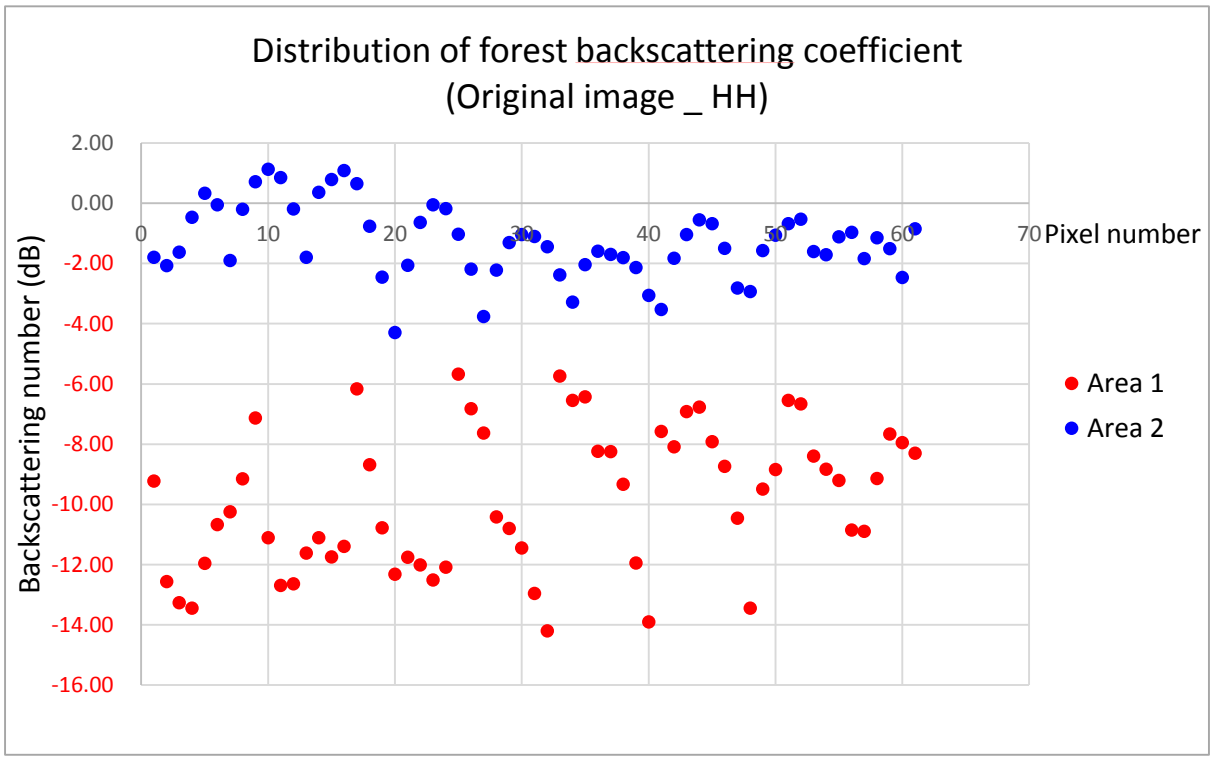


Figure 3.15: Distribution of backscattering coefficient of the original HH image.

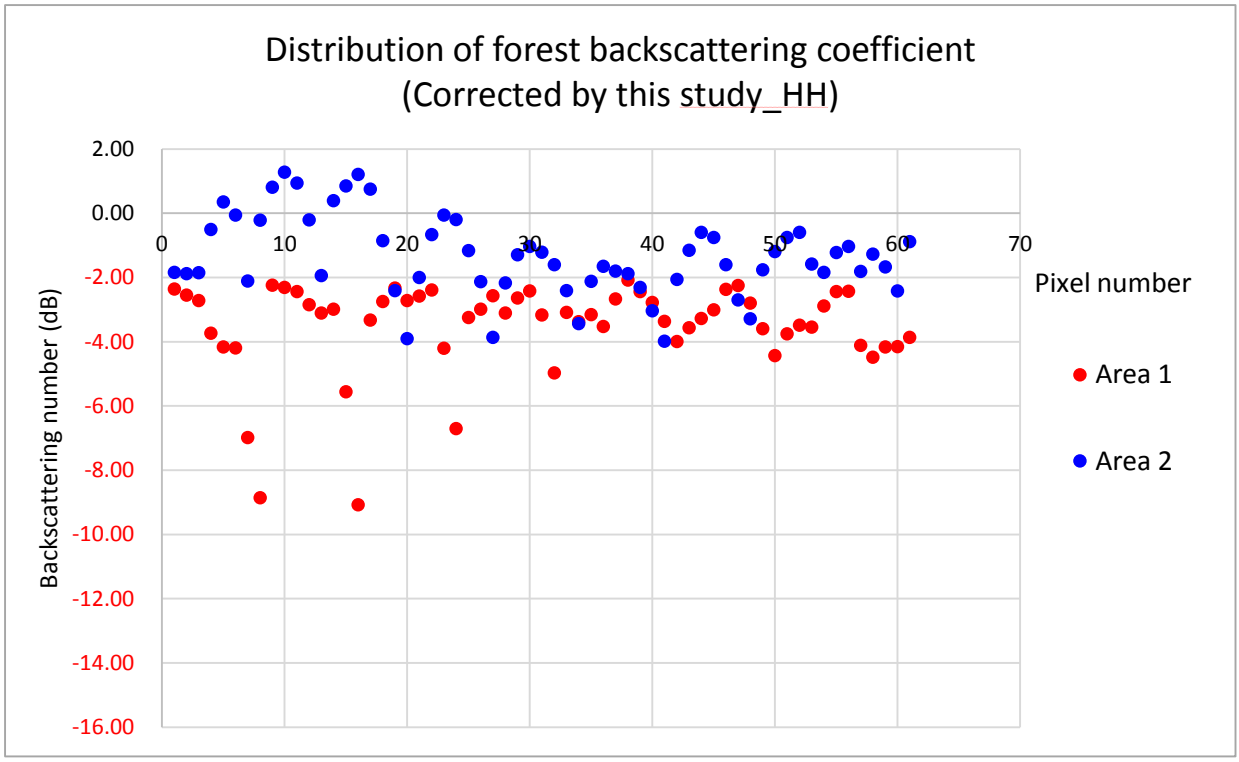


Figure 3.16: Distribution of backscattering coefficient of the slope corrected HH image.

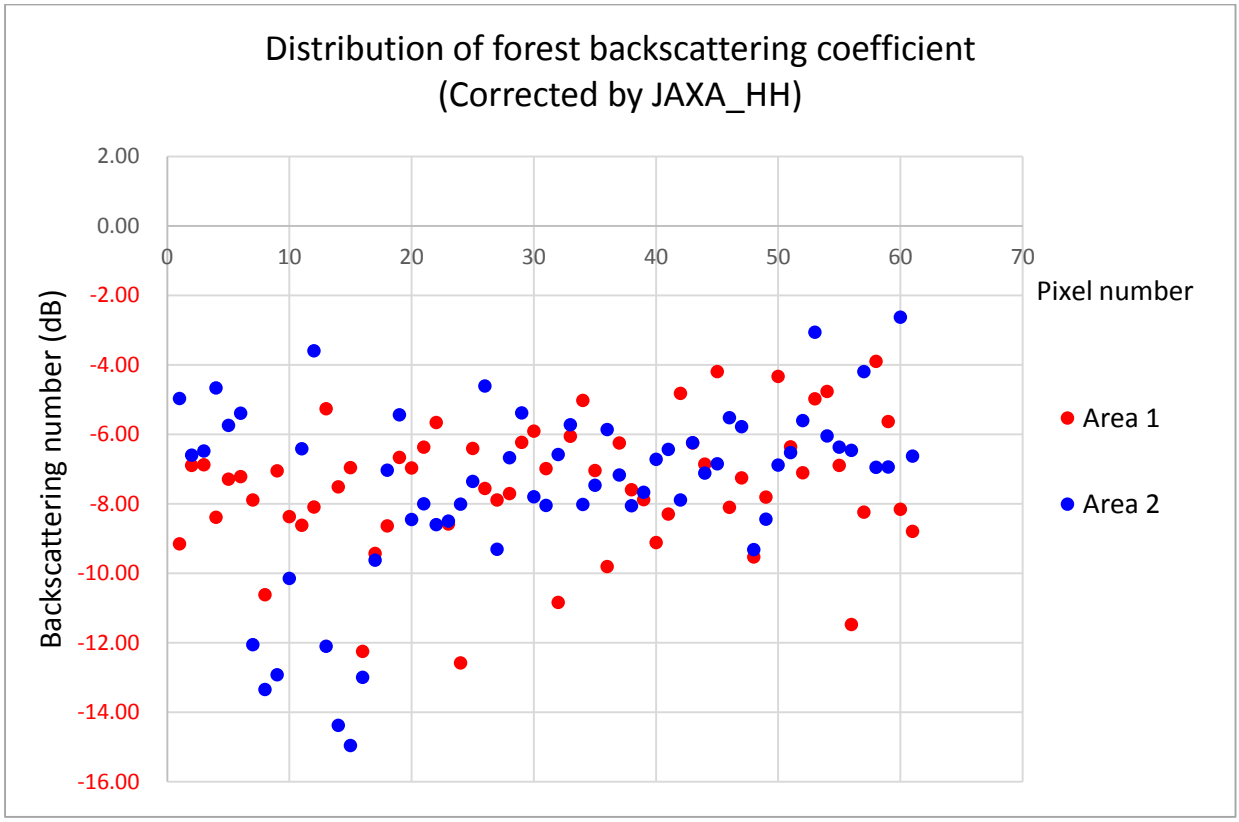


Figure 3.17: Distribution of backscattering coefficient of the new public HH image.

Table 3.3: Backscatter variance over the homogenous land cover class of Figure 3.11.

	Original	**This study**	**New public data**
Variance	3.21	2.51	2.6

The distribution of backscattering coefficient in Figure 3.16 and Figure 3.17 indicate the improved slope corrected imagery over homogeneous area both from the proposed model in this study and the new public data. Terrain influence of backscattering variance had been successfully reduced about 0.7 dB and 0.61 dB from the images of this study and JAXA, respectively (Table 3.3). However, over the mountain ridge and the valley between mountains, the public data (Figure 3.18 [1]) shows smoother impression than the correction result of 50 m ortho-rectified mosaic data (Figure 3.18 [2]). Additionally, the new mosaic data extend to the global scale, forest classification in next section was chosen to use the new PALSAR global mosaic data.

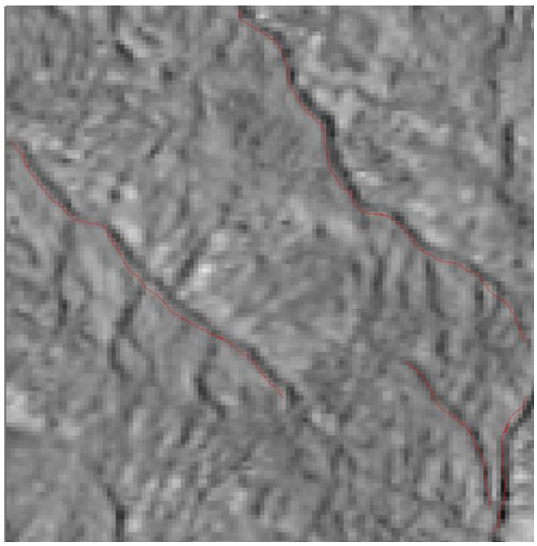
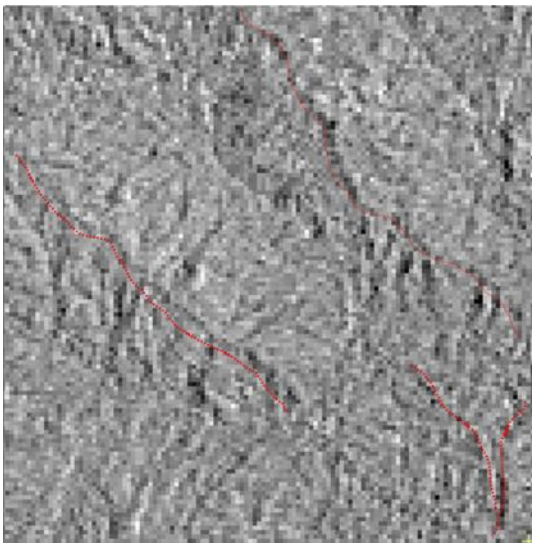
[1]  [2] 

Figure 3.18: Slope corrected image. [1] Image corrected by the proposed model in this study; [2] New public global mosaic data.

3.3. Conclusions

In this chapter, image calibration of ALOS PALSAR 50 m ortho-rectified mosaic data had been done with a well quality. Firstly, geometric correction of HH and HV images were applied by using GCPs method with the average RMS error less than 1. In terms of slope correction, the application of three previous slope correction models could not reduce the terrain influence of ALOS PALSAR 50 m ortho-rectified mosaic data. This might be caused by the using of different software tool (model-1), different code-level programming (model-2) and different determination of a number of parameter factors (model-3). On the other hand, a modified model which developed based on a sample assumption of ground scattering geometry showed the best performance in the case of comparing with the previous formulas. The corrected image was carried out to compare with the public of global ALOS PALSAR 50 m mosaic data, which processing image correction from raw data. Except for the visual interpretation over the areas of mountain ridge and valley, both of these two products shown very well terrain quality on variance analysis of homogenous area. However, due to this study is aiming to classify forests globally in the future, the global 50 m mosaic data was chosen for forest classification in next chapter.

CHAPTER 4

FOREST CLASSIFICATION USING ALOS PALSAR 50 m GLOBAL SCALE MOSAIC DATA

4.1. Study area

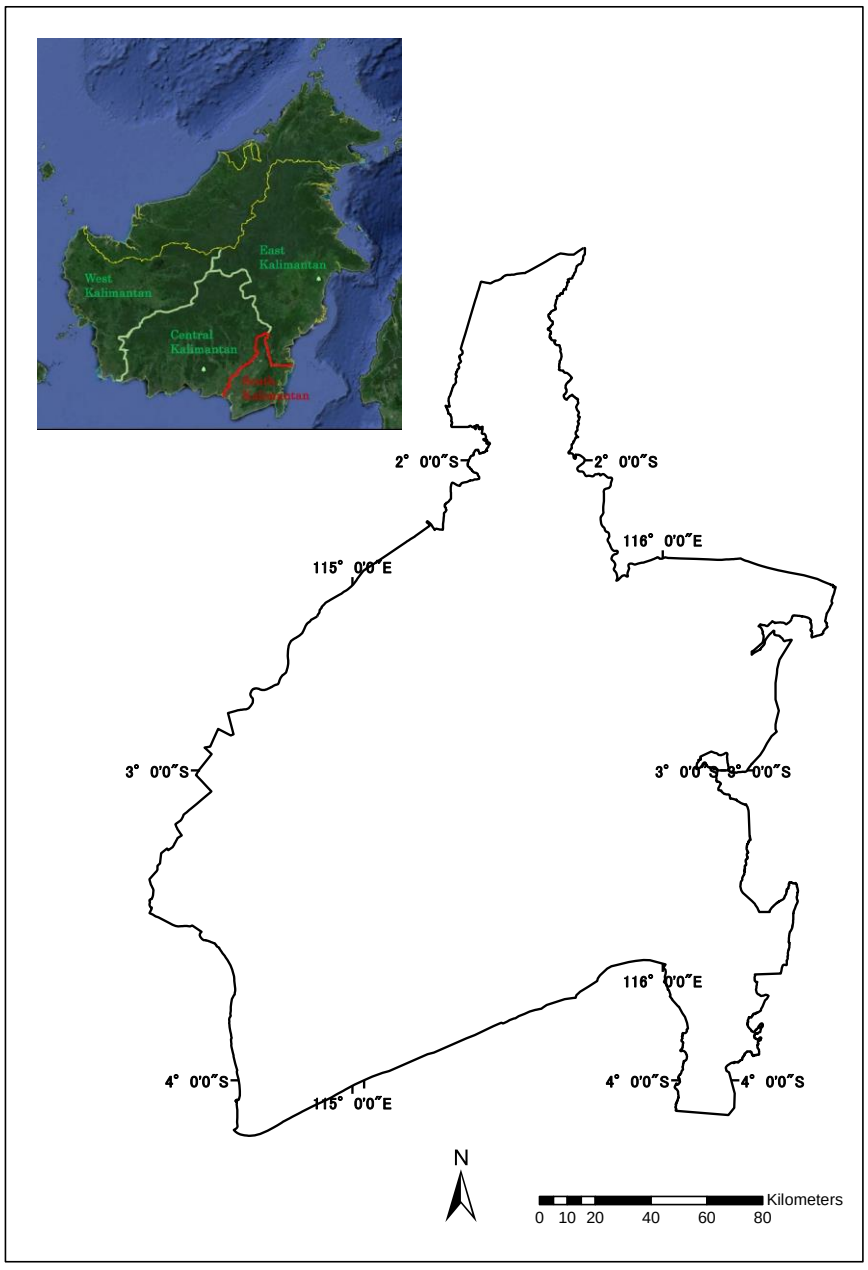


Figure 4.1: Location of study area.

The study area for developing the methodology of forest classification using ALOS PALSAR 50 m global scale mosaic data (2010) is located in South Kalimantan, Indonesia, between the range (114d00'00.0000"E, 1d15'39.2000"S) and (117d00'01.5892"E, 5d00'01.5865"S). It is situated adjacent to Central Kalimantan and East Kalimantan, and meets the Jawa Sea to the south (Figure 4.1). The total area in South Kalimantan is approximately 38,744.23 km^2, the Meratus Mountains, that the highest peak is 1,892 meters, across this area from the south-western part to the north-eastern part. About 120 days every year is raining, the amount of annual rainfall is ranging from 2,000 to 3,700 mm (Indrabudi Hermawan, 2002).

4.1.1. Land cover types

The knowledge of land cover classes of South Kalimantan was collected from the clear satellite images and uploaded photos of Google Earth. Woody vegetation including natural forest, rubber plantation, oil palm, coconut and mangrove are shown in Figure 4.2. According to the forest definition described in Shimada 2014, forest should include natural forest and rubber plantation, while the other woody plantation is trained as non-forest.

[1] Natural forest

[2] Rubber plantation

[3] Oil palm

[4] Coconut plantation

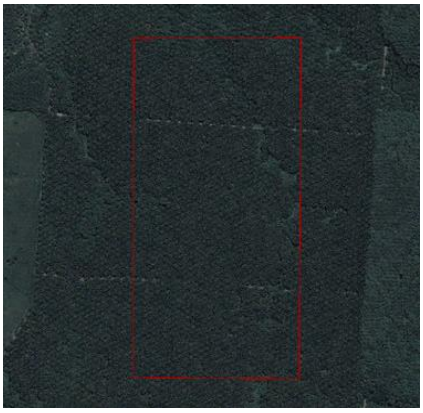
[5] Mangrove

Figure 4.2: Woody plantation (Source: Google Earth).

Training data of non-forest also selected for bare area (where includes coal mine and the open space), herbaceous, urban (where the built-up area is more than 50%), natural wetland (e.g. mangrove), artificial wetland (where includes all plantation over water area, e.g. paddy), and water body (Figure 4.3).

Forest

Forest (rubber)

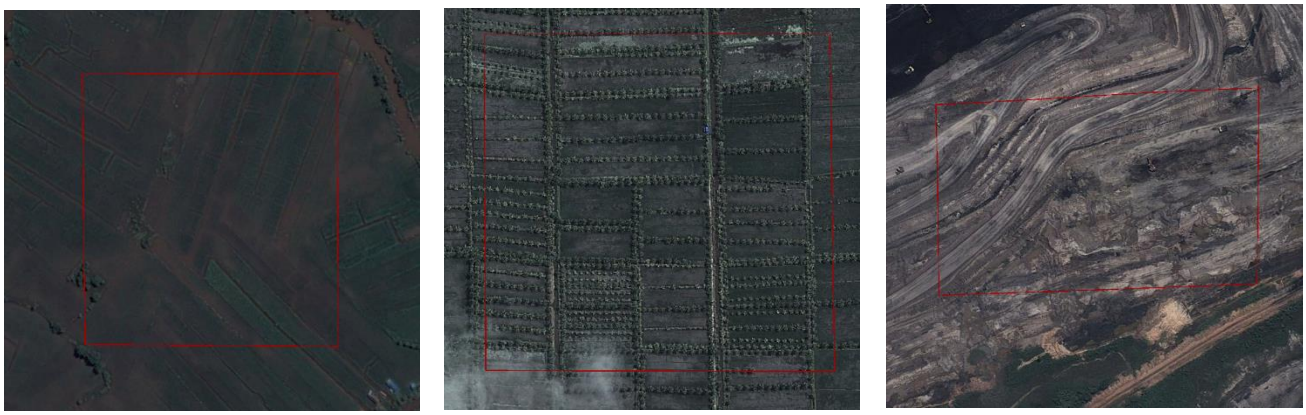


Non-forest (oil palm)

Forest

Non-forest (urban)

Non-forest (natural wetland)



Non-forest (artificial wetland) Non-forest (artificial wetland) Non-forest (bare land)

Non-forest (bare land) Non-forest (herbaceous) Non-forest (water body)

Figure 4.3. Training data collected by Google Earth image.

4.1.2. ALOS PALSAR Digital Number analysis

The relationship between forest and other land cover classes on ALOS PALSAR 50 m global mosaic data was described by Digital Number (DN) of HH and HV image. Figure 4.4 shows the DN distribution extracted from different land cover classes using the training data. Except for bare area and water body, forest distribution is covered by the other classes. This means, in the case of only using HH and HV images, it is difficult to separate forest with other land covers based on the DN value.

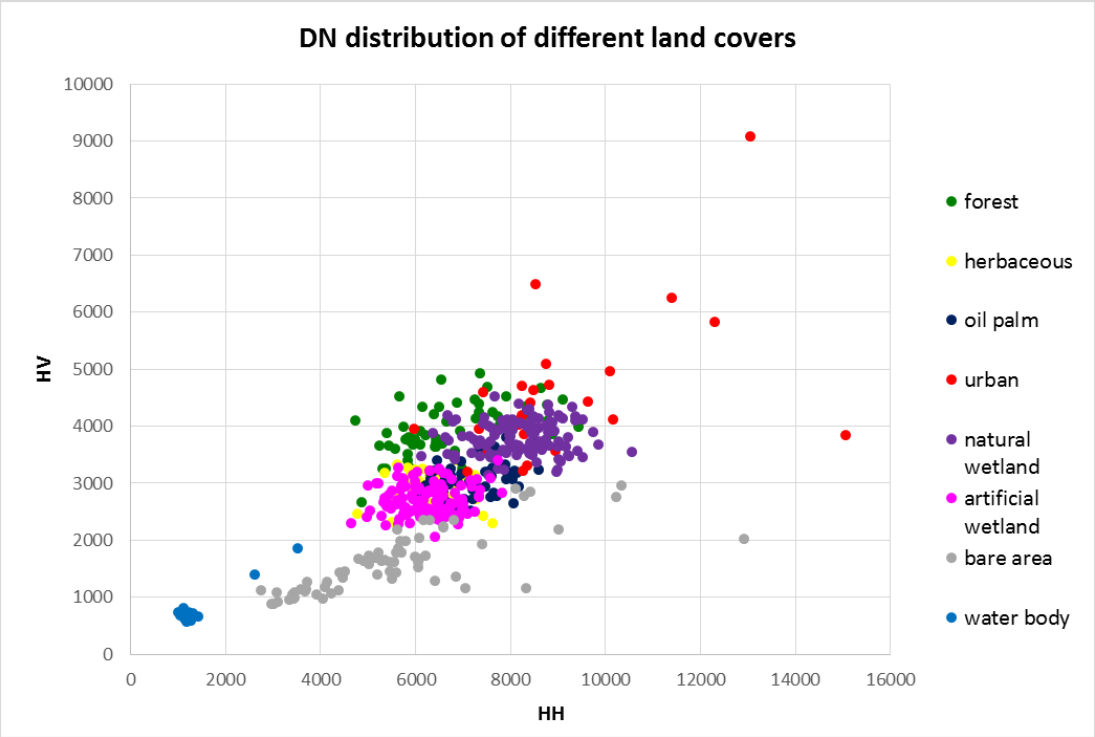


Figure 4.4: DN distribution of different land cover classes extracted from HH and HV images.

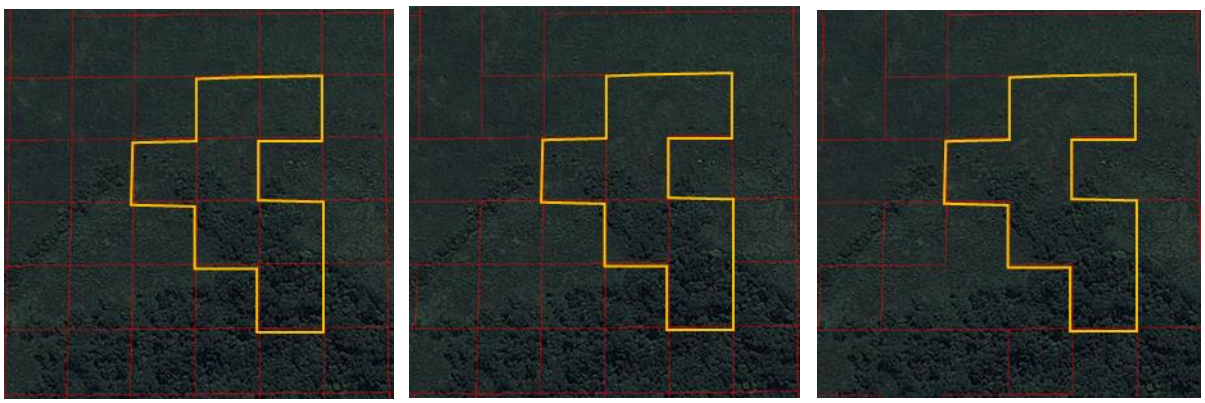
4.2. Object-based land cover classification

Recently, many scientific literatures have pointed out that the object-based approaches would improve classification accuracy when compared to traditional pixel-based approaches (Hussain *et al*., 2013; Blaschke *et al*., 2010). Especially for the single band SAR data, image segmentation may provide lots of object information, not only about spectral, but also the spatial or shape features, which are seen as the limitations of pixel-based technique. Hence, in this study, object-based approach was going to extract forest and non-forest class using ALOS PALSAR HH and HV images.

4.2.1. Image segmentation

As the first step of an object-based approach, image segmentation plays an important role on the accuracy of classification. If the scale parameter, which decide the size of object is not

considering carefully, the heterogeneous pixels will be merged within one segmented object. In this study, multi-resolution algorithm in eCognition tool was used to divide the input image as homogeneous regions. Three different scale values were tested to find the suitable segmented object (Figure 4.5). The input layers for segmentation consist of HH, HV, and the additional images (HH-HV, HH/HV, HH+HV). In Figure 4.5 [1], most of the segmented objects are divided by each pixel. This is the safest way to avoid combining the heterogeneous pixels into one object, but at the same time, it also ignored many homogeneous pixels need to be combined with each other. Thus, the scale was increased to 15 in Figure 4.5 [2], which shows some pixels of herbaceous had successfully been merged together. However, when testing with a larger scale is 20, forest is mixed with herbaceous area on the yellow polygon (Figure 4.5 [3]). After the comparison of different scale parameter, scale is 15 was chosen to generate the first segmentation layer. Then, special difference segmentation algorithm with the maximum spectral difference is 200 was applied to produce the final segmentation image. Forest is defined as the areas where the tree crown cover is more than 50%.



[1] Scale is 10 [2] Scale is 15 [3] Scale is 20

Figure 4.5: Segmentation with different scale parameters.

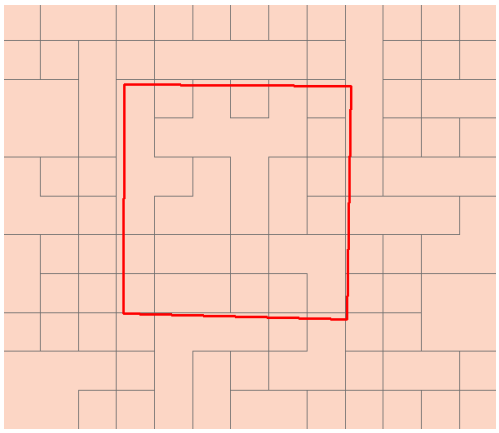
4.2.2. Training data selection

Google Earth can provide the visual global geographic information based on the high resolution satellite images and aerial photos, but only very few clear images uploaded over South Kalimantan from 2002 to 2010 (Table 4.1). The total number of 282 training polygons were collected from Google Earth where is viewable in 2010, including forest, herbaceous, oil palm, urban, bare area, natural wetland, artificial wetland and water body. HH image, HV image and the additional images were used as the input layers to generate segmentation object.

Table 4.1: Clear satellite images of Google Earth in South Kalimantan.

Time	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Number	5	1	4	4	1	2	5	19	11

Figure 4.6 [1] shows a training polygon of oil palm is overlaying with the segmentation image. Many objects were produced in this oil palm area look like this segmentation algorithm is not successfully merge the homogeneous region. However, small object would avoid the misclassification caused by the multi-classes existing in a single object. In addition, the number of training data may make an impact on the classification accuracy (Shiraishi *et al*. 2014), more training data were generated using the method in Figure 4.6 [2]. Table 4.2 shows the number of training data collected from Google Earth and the segmentation image.

[1] 

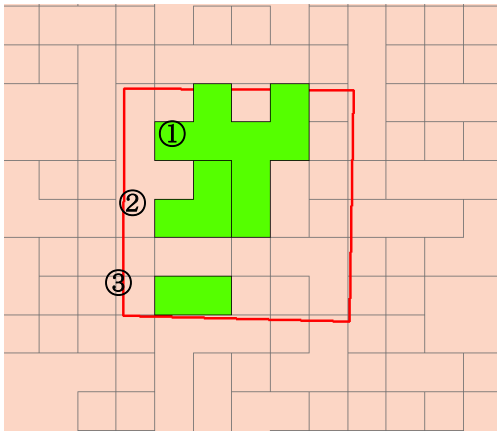
[2] 

Figure 4.6: Training data generation based on segmentation. [1] Original training polygon drew

on Google Earth; [2] Example of generation of new training data.

Table 4.2: Number of training data.

Land cover classes	**Number of training data**	
	From Google Earth	From segmented image
Forest	65	256
Herbaceous	30	93
Oil palm	30	91
Urban	30	63
Bare area	30	99
Natural wetland	25	46
Artificial wetland	32	77
Water body	40	65
Total number	282	790

4.2.3. Classifiers

Classification by using machine learning algorithms have been used well in the field of remote sensing more than several decades. The machine leaning algorithms can be divided into supervised learning and unsupervised learning according to the use of training data. Some automatic classifiers have received increased recognition on their classification abilities in the previous researches, for example, the linear classier Support Vector Machine (SVM) and various Decision Trees (DT) (Mountrakis *et al.*, 2011; Lan *et al*., 2011; Pal and Mather, 2005).

In this study, three well-known automated machine learning classifiers provided by an open software tool called Weka were selected to evaluate and compare their capability in the case of

multiple classes' classification. They are: J48 (C4.5), Random Forest (RF) and Sequential Minimal Optimization (SMO).

Both J48 and RF are the derivation classifiers from DT algorithm. The main difference between J48 and RF are the number of constructed trees. For J48, there is only one decision tree, while multitude trees are built on RF. The advantages of these two classifiers are demonstrated by the published literatures, that J48 is better than all the other DT approaches when considering both accuracy and processing speed (Zhao and Zhang, 2007), while the overfitting problem occurred easily on the other DT algorithms and it would be avoided when using RF classifier (Ali *et al*., 2012).

On the other hand, as an improved modification of SVM, SMO fixed up the quadratic programming (QP) problem when running SVM processing with a set of training data (Ruben, 2007). So far, there are some researches focusing on the comparison of different classifiers (e.g. various DT approaches comparison, SMO and J48, SVM and DT) (Zhao and Zhang, 2007; Cufoglu *et al*., 2009; Huang *et al*. 2002; Shiraishi *et al*. 2014)). However, there is a lack of the assessment analysis among J48, RF and SMO classifiers. Hence, the objective of this section is, to compare these three well-known classifiers to find the best classification algorithm, when ALOS PALSAR 50 m global scale mosaic data was used for classification.

4.2.4. Feature selection

Table 4.3: List of generated features exported from eCognition Developer.

Layer value feature	**Texture feature**
(1) Mean	(9) GLCM Homogeneity
(2) Mode	(10) GLCM Contrast
(3) Quantile	(11) GLCM Dissimilarity
(4) Standard Deviation	(12) GLCM Entropy
(5) Skewness	(13) GLCM Ang.2nd moment
(6) Circular Mean	(14) GLCM Mean
(7) Circular StdDev	(15) GLCM StdDev
(8)CircularStdDev/Mean	(16) GLCM Correlation
	(17) GLDV Ang.2nd moment
	(18) GLDV Entropy
	(19) GLDV Mean
	(20) GLDN Contrast

20 kinds of features based on layer value and texture extracted from eCognition 9.0 are shown on Table 4.3, while they would apply with five input layers consist of HH, HV, (HH-HV), (HH/HV) and (HH+HV). Thus, the total number of attributes is 100($20\ \text{features} \times 5\ \text{layes}$).

Figure 4.7 shows the result of accuracy assessment with Kappa statistic of the classification result, which based on the cross-validation (10 folds) by using different feature combinations selected by three attribute evaluations. The attribute evaluation are: ① Correlation-based Feature Selection, which evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them

(Hall, 1999); ② Chi-Square Evaluator, which evaluates the worth of an attribute by computing the value of the Chi-Square statistic with respect to the class (Jin *et al*., 2006); ③ Wrapper Subset Evaluation, which evaluates attribute sets by using a learning scheme (Kohavi *et al*., 1997).

In Figure 4.7, the best performance of classification is shown by RF classifier with the feature subset selected by Correlation-based Feature Selection, with the Kappa statistic is 0.7152 while the others is less than 0.7. Fifteen features were selected as the optimal feature subset for RF classifier (Table 4.4).

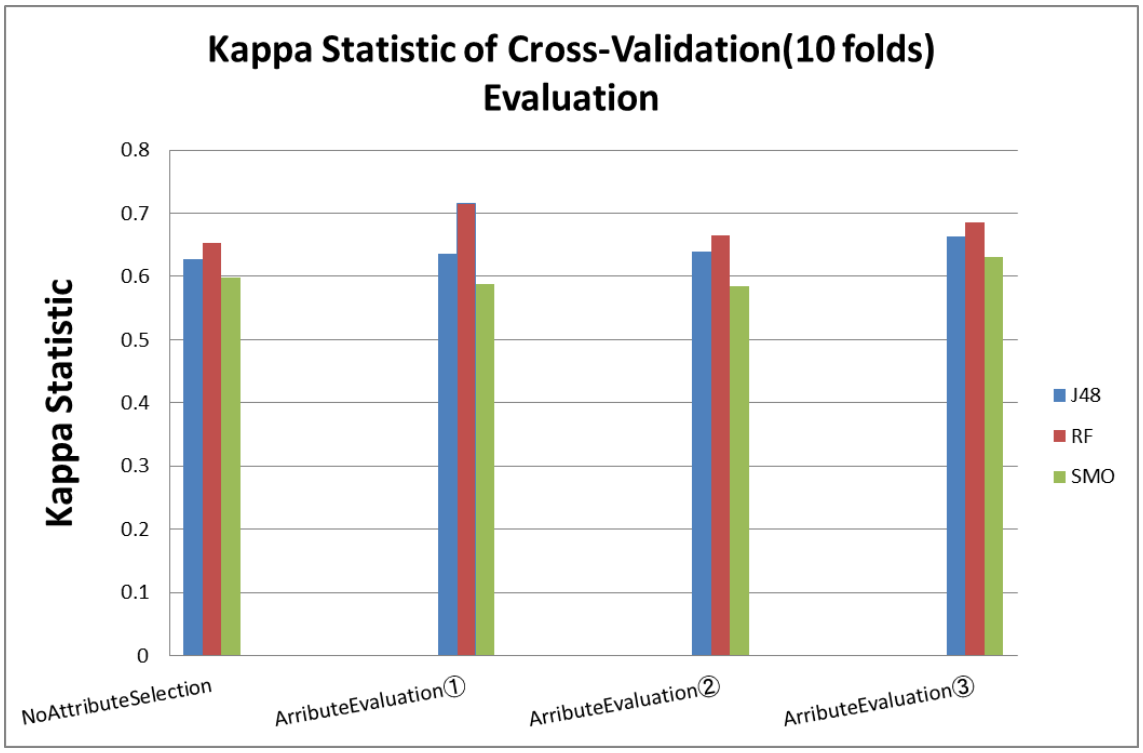


Figure 4.7: The accuracy of classifying training data by using J48, RF and SMO classifier. The first set of column is without attribute selection, the others used different attribute evaluation algorithms: ① Correlation-based Feature Selection; ② Chi-Square Evaluator; ③ Wrapper Subset Evaluation.

Table 4.4: Optimal feature subset for RF classifier.

Classifier	Optimal feature subset
RF	Mode_ HV Mode_ (HH/HV) Quantile_ HV Quantile_ (HH+HV) Circular Mean_ HH Circular Mean_ HV Circular Mean_ (HH-HV) Circular Mean_ (HH/HV) Circular Mean_ (HH+HV) Circular StdDev_ HV Circular StdDev_ (HH/HV) CircularStdDev/Mean_ HV CircularStdDev/Mean_ (HH/HV) CircularStdDev/Mean_ (HH+HV) GLCM Contrast_ (HH-HV)

4.2.5. Land cover classification result

Random Forest (RF) classifier was chosen to produce the land cover map over South Kalimantan with the optimal feature combination. 900 validation points were collected based on the classification result by random sampling approach. Here, 200 points were used to verify the result of forest, while for the other classes, each result had 100 validation points.

Table 4.5 shows the accuracy assessment of land cover classification. The producer’s accuracy

and user's accuracy of forest class were 88% and 58.90%, respectively. Urban represented the lowest accuracy with the producer' accuracy of 14% and user's accuracy of 25.50%. Half of the validation points took from the classification result of urban were forest as a matter of fact. In the case of the classes which the user's accuracy is lower than 60%, there are 12 validation points of forest were misclassified to herbaceous, 8 points of forest were misclassified to bare land, 13 points of forest were misclassified to artificial wetland, and 15 points of forest were misclassified to natural wetland.

The land cover confusion matrix indicates that only using Random Forest (RF) classifier based on object analysis may not solve the problem discussed in Chapter 4.1.2, which is difficult to separate forest with other classes based on the pixel DN value. Hence, in order to achieve the purpose of a higher accuracy forest map, the forest that misclassified to the other land cover classes need to be extracted.

Table 4.5: Confusion matrix of land cover classification.

Google Earth / Result	**Forest**	**Oil palm**	**Herbaceous**	**Urban**	**Bare**	**Artificial wet area**	**Natural wet area**	**Water**	**Total**	**User's accuracy (%)**
Forest	**176**	1	12	3	1	4	1	2	200	88.00%
Oil palm	14	**66**	10	1	1	8	0	0	100	66.00%
Herbaceous	12	0	**56**	0	3	23	1	5	100	56.00%
Urban	50	5	18	**14**	0	4	6	3	100	14.00%
Bare	8	0	48	4	**19**	12	6	3	100	19.00%
Artificial wet area	13	16	32	3	5	**27**	3	1	100	27.00%
Natural wet area	15	1	39	0	0	19	**26**	0	100	26.00%
Water	11	0	5	0	1	4	3	**76**	100	76.00%
Total	299	89	220	25	30	101	46	90	**900**	
Producer's accuracy (%)	58.90%	74.20%	25.50%	56%	63.30%	26.73%	56.50%	84.40%		
Overall accuracy	51.10%									

4.3. Forest reclassification

The purpose of this section is to extract forest land cover. First, forest and non-forest map was generated using the land cover classification result by merging all the land cover classes, except forest into the non-forest class. Second, all the classes in the land cover map which accuracy was less than 60%, were reclassified into forest and non-forest. For example, the areas using the training data of forest and urban. The same procedure was conducted for herbaceous, bare area, artificial wetland and natural wetland result. Water body was replaced by the water mask made by JAXA.

From Table 4.5, the most negative class was urban, in which 50 forest points were misclassified. Therefore, three classifiers, J48, RF and SMO were compared again for separating only forest and urban.

Figure 4.8 shows the Kappa statistic of J48, RF, SMO classifier in the case of classifying the training data of forest and urban. The kappa coefficient of these three classifiers were calculated as 0.7764, 0.7758 and 0.7239 for J48, RF and SMO, respectively.

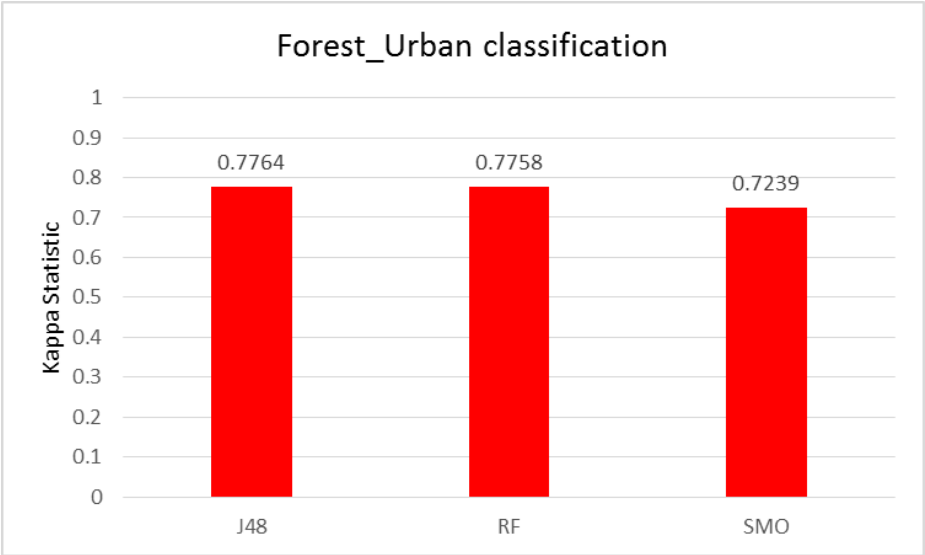


Figure 4.8: Comparison of three classifiers in the case of classifying forest and urban.

J48 was chosen to extract forest from urban class, while it was also tested on the other classes. In Figure 4.9, accuracy assessment were carried out with J48 classifier for classifying forest and herbaceous (F_H), forest and bare area (F_B), forest and artificial wetland (F_A), forest

and natural wetland (F_N). All the kappa coefficient represented higher performance more than 0.8 on different classes by J48 classifier.

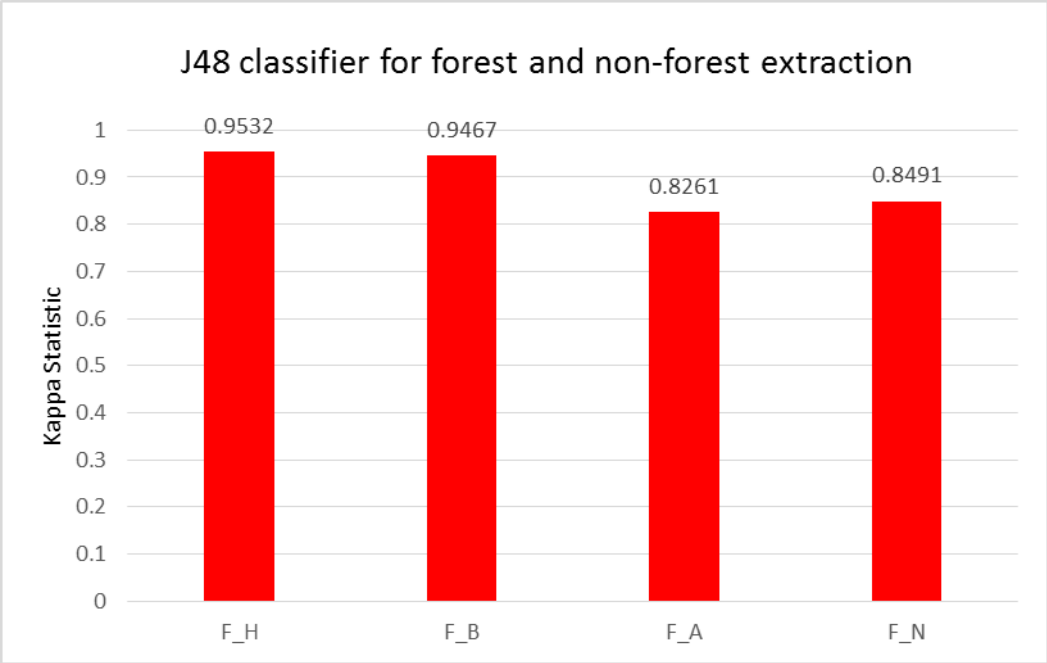


Figure 4.9: Accuracy assessment of forest and non-forest extraction by J48 classifier.

Pruned tree generated by J48 classifier for separating forest and other class are shown in Figure 4.10 ~ Figure 4.14. The feature combination selected by using Correlation-based Feature Selection for forest and non-forest classification include: (1) F_U: Circular Mean _ (HH-HV), Circular StdDev _HH, Circular StdDev _ (HH-HV), Circular StdDev _ (HH+HV), CircularStdDev/Mean _ HV; (2) F_H: Circular Mean _HV, Circular Mean _ (HH/HV), CircularStdDev/Mean _ HV; (3) F_B: Mean _ HV, Circular Mean _ HV, Circular StdDev _(HH/HV); (4) F_A: Quantile _ HV, Circular Mean _ HV, Circular Mean _ (HH/HV), Circular StdDev _ HV; (5) F_N: Mean _ (HH/HV), Circular Mean _ HH, Circular Mean _ HV.

Forest area extracted from the result of urban, herbaceous, bare land, artificial wetland and natural wetland is shown in Figure 4.15. Figure 4.16 shows the final forest map produced by the combination of RF classifier and J48 classifier.

```
Circular StdDev _ (HH+HV) <= 2254.181218
|   Circular Mean _ (HH-HV) <= 4488.1875
|   |   CircularStdDev/Mean _HV <= 0.188182: forest
|   |   CircularStdDev/Mean _HV > 0.188182
|   |   |   Circular StdDev _HH <= 1253.701643: urban
|   |   |   Circular StdDev _HH > 1253.701643
|   |   |   |   Circular StdDev _(HH-HV) <= 1123.38011: forest
|   |   |   |   Circular StdDev _(HH-HV) > 1123.38011: urban
|   Circular Mean _(HH-HV) > 4488.1875: urban
Circular StdDev _(HH+HV) > 2254.181218: urban
```

Figure 4.10: J48 pruned tree for separating forest and urban (F_U).

```
Circular Mean _HV <= 3176.9375
|   Circular Mean _(HH/HV) <= 1.4375
|   |   CircularStdDev/Mean _HV <= 0.254439: forest
|   |   CircularStdDev/Mean _HV > 0.254439: herbaceous
|   Circular Mean _(HH/HV) > 1.4375: herbaceous
Circular Mean _HV > 3176.9375: forest
```

Figure 4.11: J48 pruned tree for separating forest and herbaceous (F_H).

```
Circular Mean _HV <= 3086.75: bare
Circular Mean _HV > 3086.75
|   Circular StdDev _(HH/HV) <= 0.496078: forest
|   Circular StdDev _(HH/HV) > 0.496078
|   |   Mean _HV <= 2427: bare
|   |   Mean _HV > 2427: forest
```

Figure 4.12: J48 pruned tree for separating forest and bare area (F_B).

Circular Mean _HV <= 3194.3125

| Quantile _HV <= 3292: **artificial wetland**

| Quantile _HV > 3292

| | Quantile _HV <= 4323: **artificial wetland**

| | Quantile _HV > 4323: **forest**

Circular Mean _HV > 3194.3125

| Circular Mean _HV <= 3489.8125

| | Circular StdDev _HV <= 451.906085: **artificial wetland**

| | Circular StdDev _HV > 451.906085: **forest**

| Circular Mean _HV > 3489.8125

| | Circular Mean _(HH/HV) <= 1.5625: **forest**

| | Circular Mean _(HH/HV) > 1.5625

| | | Quantile _HV <= 4191: **artificial wetland**

| | | Quantile _HV > 4191: **forest**

Figure 4.13: J48 pruned tree for separating forest and artificial wetland (F_A).

Circular Mean _HV <= 3143.3125: **natural wetland**

Circular Mean _HV > 3143.3125

| Mean _(HH/HV) <= 1.5: **forest**

| Mean _(HH/HV) > 1.5

| | Circular Mean _HH <= 8100.25

| | | Circular Mean _HH <= 6919.6875: **natural wetland**

| | | Circular Mean _HH > 6919.6875: **forest**

| | Circular Mean _HH > 8100.25: **natural wetland**

Figure 4.14: J48 pruned tree for separating forest and natural wetland (F_N).

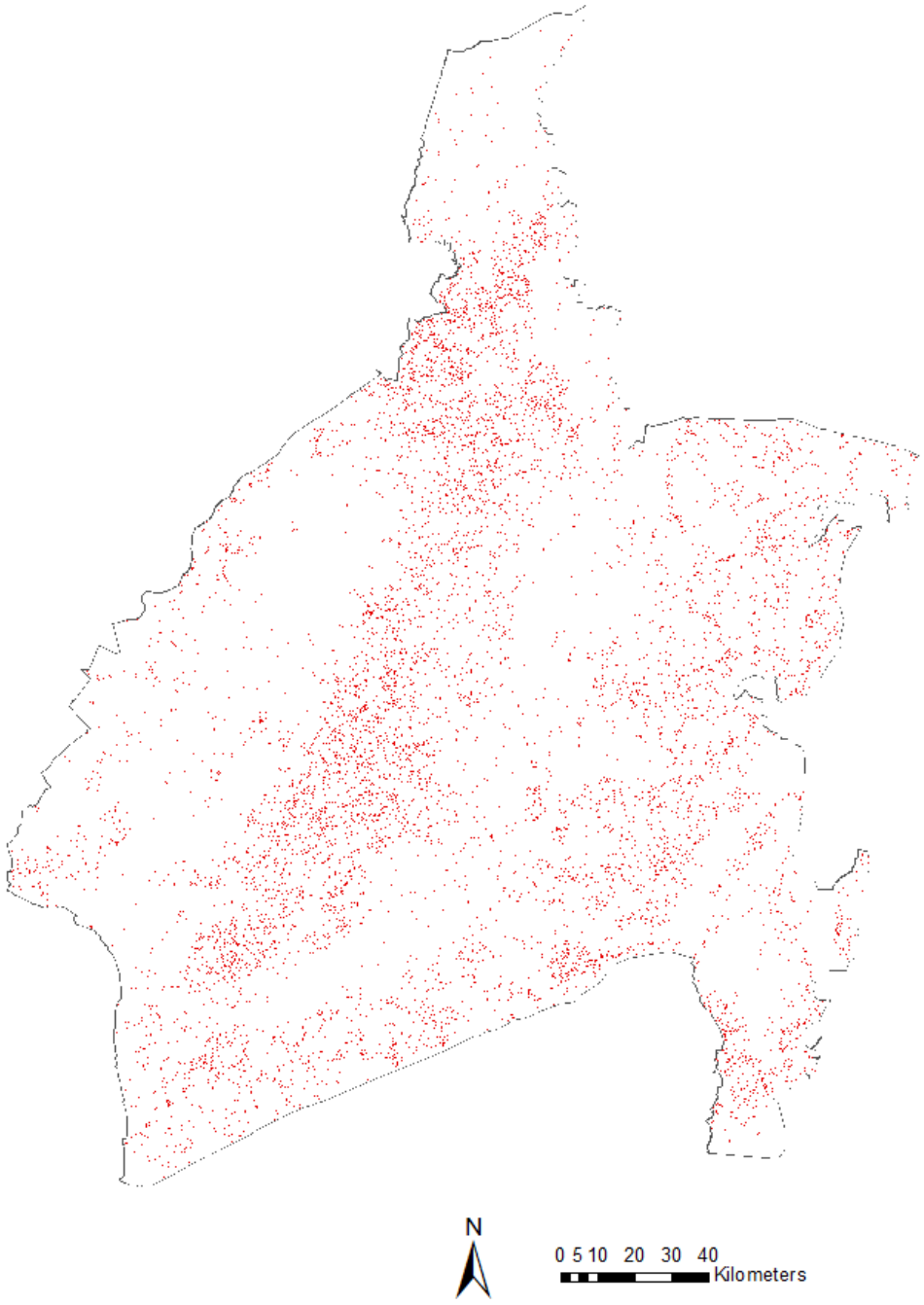


Figure 4.15: Reclassification result. Red color: reclassified forest.

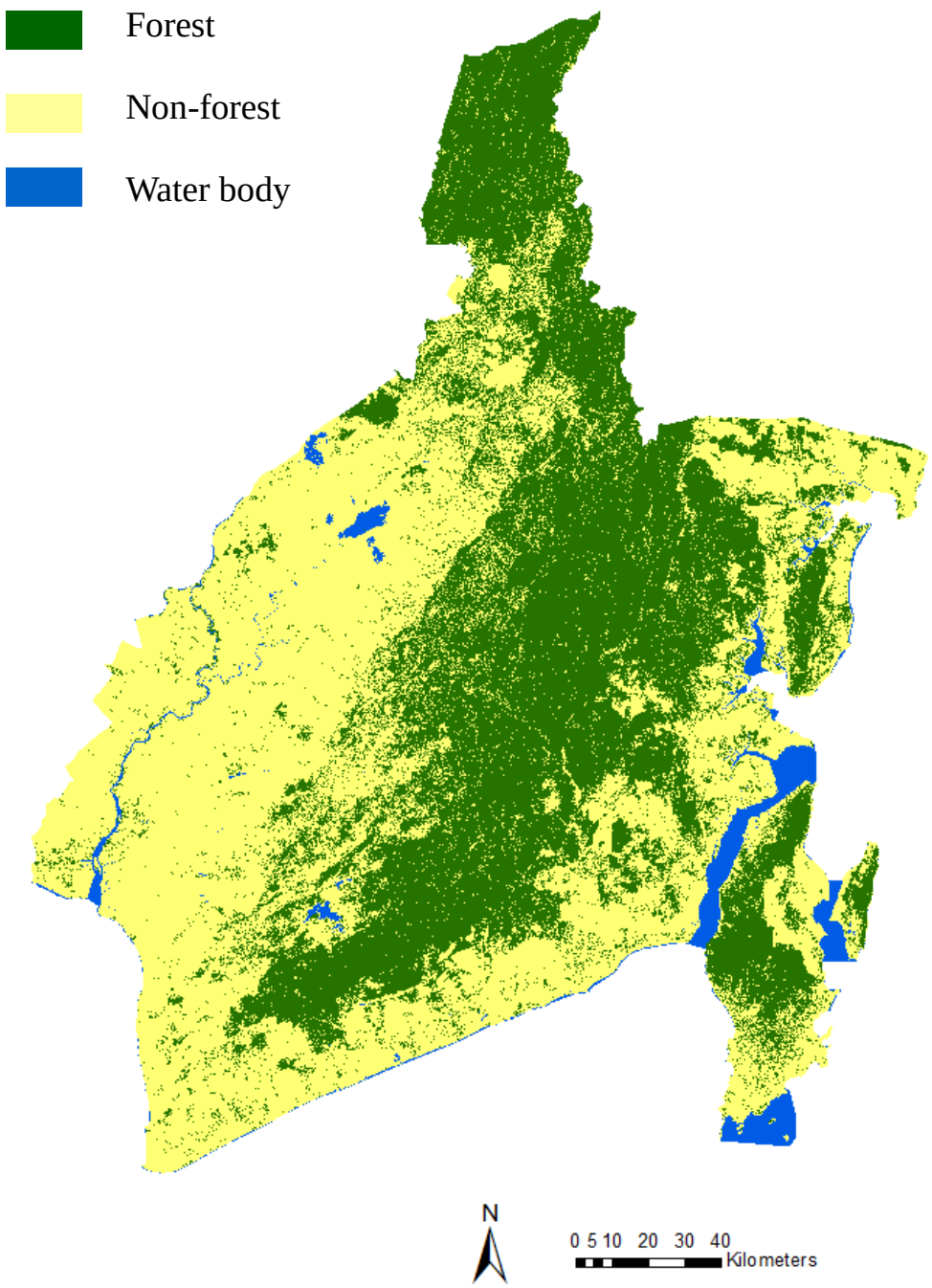


Figure 4.16: Final forest and non-forest map combined with RF and J48 classifiers.

CHAPTER 5
VALIDATION AND DISCUSSIONS

5.1. Compare with the New Global 50 m PALSAR Forest/Non-forest map

The New Global 50 m Forest/Non-forest map (2007, 2008, 2009, 2010), which generated with ALOS PALSAR 25 m special resolution mosaic data, was published by JAXA in January 2014. The average accuracy of these four years' maps had been demonstrated as 91.25% by identifying 4114 random points using Google Earth images (Shimada *et al*, 2014).

The different classification result of this study and the public of JAXA is shown in Figure 5.1. Black color represents the common area, red color is the forest class in JAXA but was classified as non-forest in this study, and green color is non-forest class in JAXA but was classified as forest in this study.

The method of comparing these two products is to evaluate the area where has the different classification result by using Google Earth image. For instance, in Figure 5.2, the green color area inside the polygon, that represents forest class in this study while non-forest class in JAXA was proved as forest when contrast with Google Earth image. While in Figure 5.3, the forest area classified in JAXA (red color) was identified as wetland. The size of the polygons are approximately (660×1070) m (Figure 5.2) and (5.64×7.65) km (Figure 5.3), respectively.

For the validation of the whole different result area, 300 points which were collected randomly is shown in Figure 5.4. In addition, in consideration of avoiding the position shift on Google Earth, the polygons that including 4 pixels at least but above same classification result were drawn around the random point and were used for validation (Figure 5.5). Except that images are not clear be confirmed, 106 polygons which observed from 2009 to 2012 were identified.

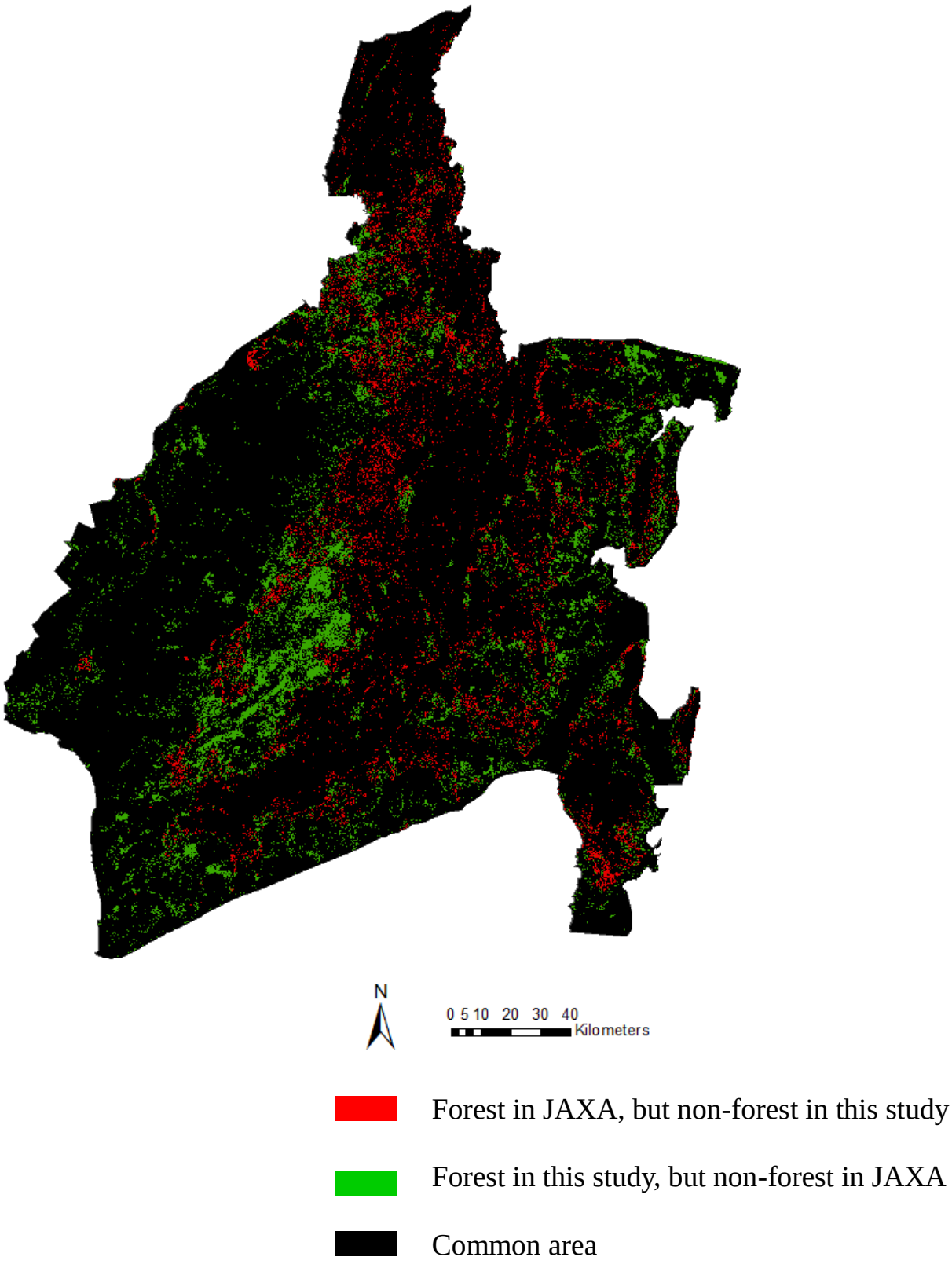
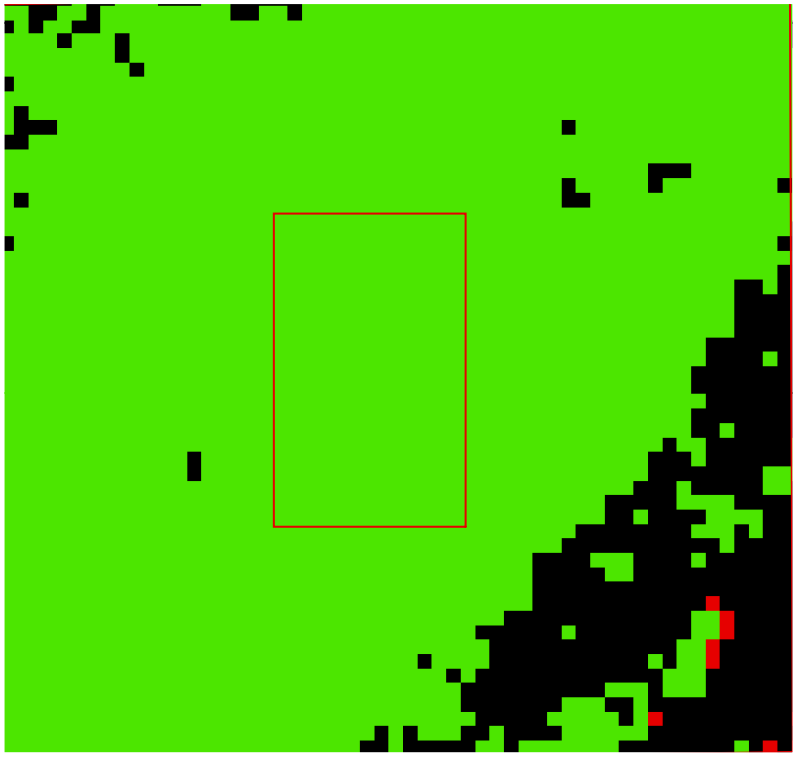


Figure 5.1: Difference of forest/non-forest map between JAXA and this study.

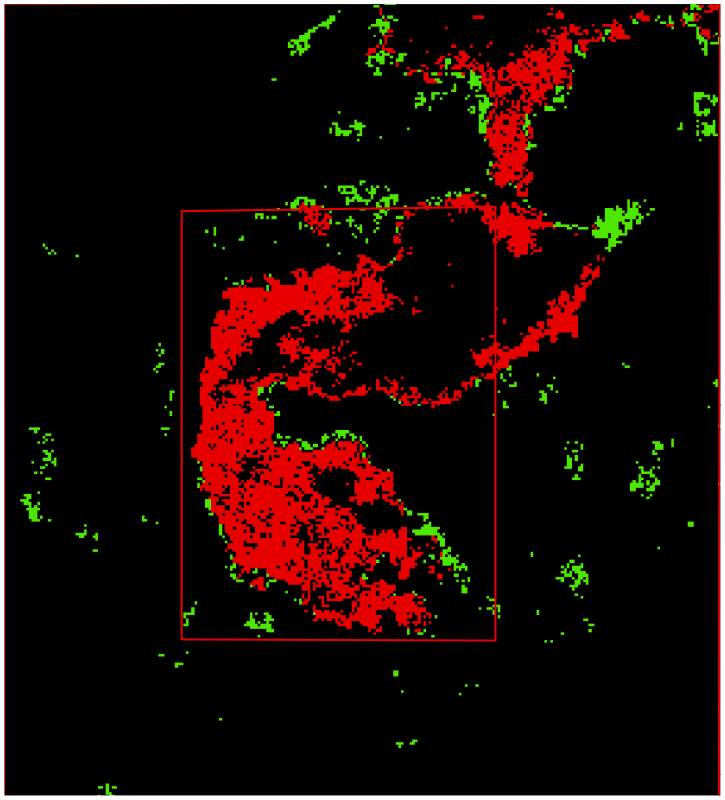
[1]



[2]

Figure 5.2: Forest result in this study while non-forest in JAXA. [1]: Different area; [2]: Google Earth image.

[1]



[2]

Figure 5.3: Forest result in JAXA while non-forest in this study. [1]: Different area; [2]: Google Earth image.

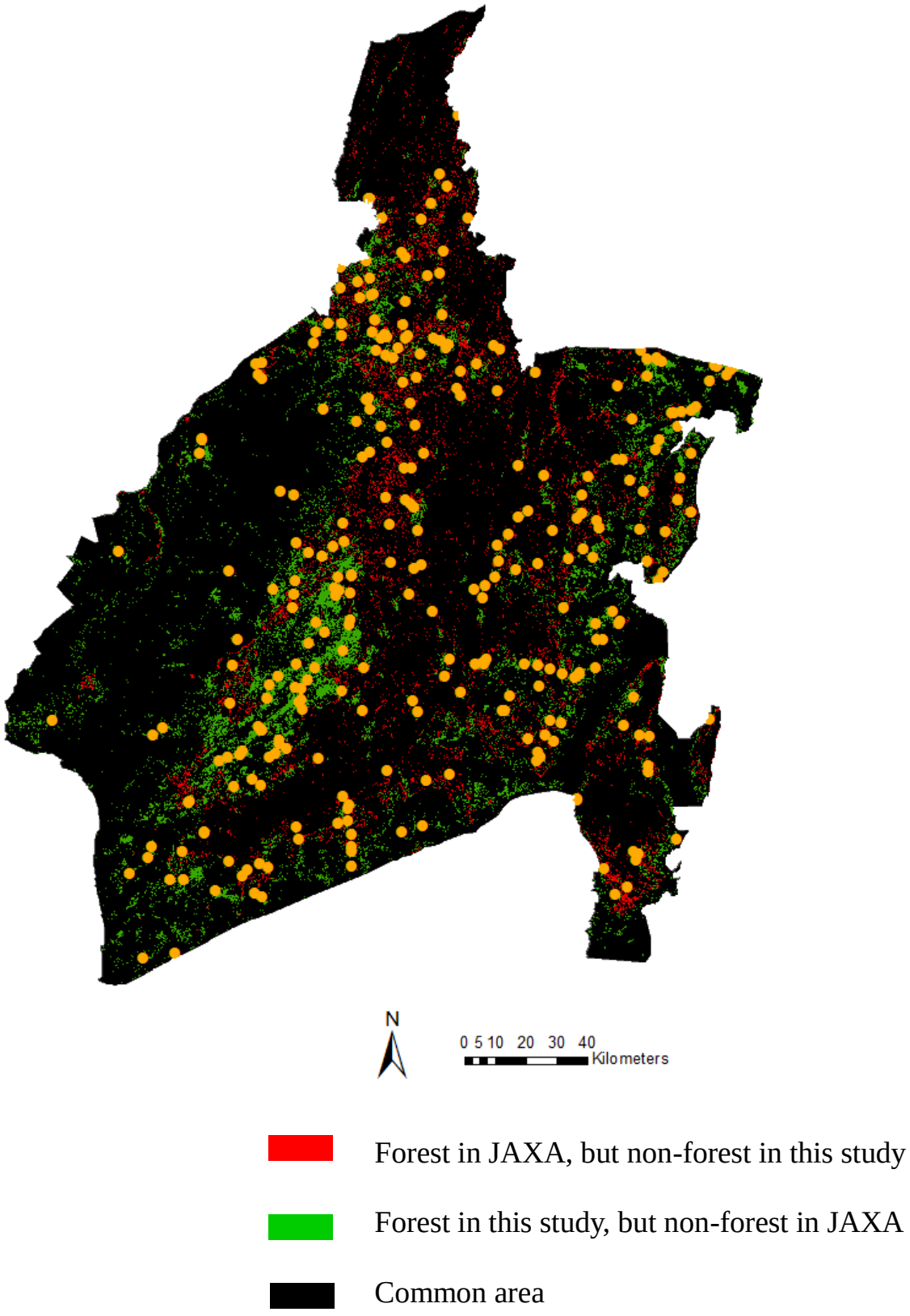


Figure 5.4: Distribution of random points over different area.

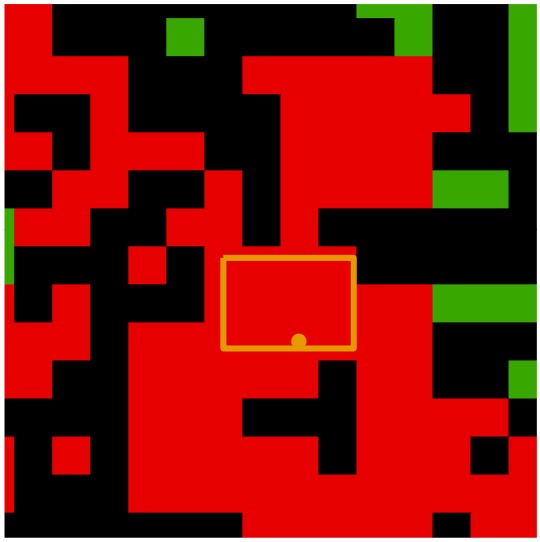
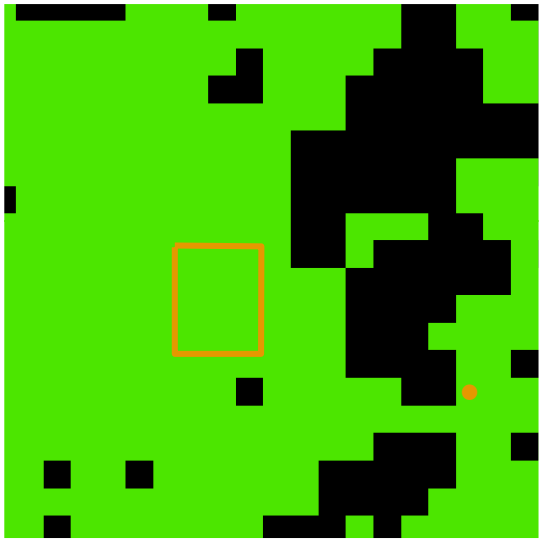
[1]  [2] 

Figure 5.5: Drawn validation polygons. [1] Around the random point; [2] Near to the random point.

Due to forest of this study is defined as the areas where the tree crown cover is more than 50%, while where the tree crown cover is more than 10% is defined as forest in the case of the public of JAXA, the percent of tree crown cover inside the identified polygons was taken into consideration to determine the ground truth is forest or non-forest. Therefore, for Google Earth image where the tree crown cover is more than 50% is forest, while less than 10% is non-forest. In such a case if the tree crown cover is more than 10% but less than 50% will be evaluated individually, but not be taken as the use of comparison.

Table 5.1: Validation result of different area for this study.

Result \ Reference	**Forest**	**Non-forest**	**Total**	**User's accuracy**
Forest	**54**	26	80	67.50%
Non-forest	2	**5**	7	71.43%
Total	56	31	**87**	
Producer's accuracy	96.43%	16.13%		
Overall accuracy	67.82%			

Table 5.2: Validation result of different area for the public of JAXA.

Result \ Reference	**Forest**	**Non-forest**	**Total**	**User's accuracy**
Forest	**26**	54	80	32.50%
Non-forest	5	**2**	7	28.57%
Total	31	56	**87**	
Producer's accuracy	83.87%	3.57%		
Overall accuracy	32.18%			

The number of the polygons that include tree crown cover is more than 50% and less than 10% were 87. The validation results were shown on Table 5.1 and Table 5.2 for this study and the map generated by JAXA, respectively. There are 59 polygons are correctly classified in this study, while 28 polygons are correct for JAXA's map. The validation result indicated the forest and non-forest map produced by this study has a higher accuracy with an overall accuracy of 67.82%.

In addition, there are 19 polygons were included in the case of the ground truth where tree crown cover is more than 10% but less than 50%. If refer to the definition of JAXA, the correct number of polygon was 14. If refer to this study, there are 12 polygons were correctly classified.

The detaild information of each validation polygon and their classification results are described on Table 5.3 and Table 5.4.

Table 5.3: The imformation of each validation polygon in the case of the ground truth tree crown cover is more than 50% and less than 10%.

ID number	Coordinate		Pixels	Google Earth	Classification result		Y/M/D
	Upper left	Lower right			This study	JAXA	
1	115d30'55.9482"E, 2d03'01.6040"S	115d30'59.4482"E, 2d03'04.7560"S	7	F	NF	F	2010/9/20
2	115d34'40.5363"E, 2d06'19.6480"S	115d34'46.0505"E, 2d06'22.6651"S	10	F	NF	F	2010/9/20
3	115d17'32.1788"E, 2d14'49.9430"S	115d17'37.7326"E, 2d14'55.0851"S	13	F	F	NF	2011/11/27
4	115d15'22.0068"E, 2d16'09.8706"S	115d15'26.2475"E, 2d16'12.3789"S	7	F	NF	F	2010/9/25
5	115d25'06.7408"E, 2d16'04.2344"S	115d25'11.1662"E, 2d16'08.7464"S	9	F	NF	F	2010/9/15
6	115d38'04.3385"E, 2d18'19.4841"S	115d38'09.3885"E, 2d18'25.7658"S	13	F	F	NF	2010/9/15
7	115d33'30.6807"E, 2d20'11.0528"S	115d33'37.4614"E, 2d20'14.9425"S	13	F	F	NF	2010/9/15
8	115d27'28.8785"E, 2d20'34.3294"S	115d27'34.2762"E, 2d20'38.9978"S	11	NF	NF	F	2010/9/15
9	115d28'40.9488"E, 2d20'42.1257"S	115d28'45.1634"E, 2d20'45.1065"S	7	F	NF	F	2010/9/15
10	116d15'37.6980"E, 2d21'21.3513"S	116d15'45.1939"E, 2d21'30.3360"S	19	F	F	NF	2009/5/15
11	115d34'14.7015"E, 2d37'25.4284"S	115d34'17.2736"E, 2d37'29.1307"S	7	F	NF	F	2012/9/23
12	114d55'05.4077"E, 2d37'36.6527"S	114d55'14.5631"E, 2d37'46.1130"S	21	F	F	NF	2012/5/11
13	115d32'05.5374"E, 2d40'13.1730"S	115d32'09.3069"E, 2d40'15.5487"S	7	F	NF	F	2012/9/23
14	115d29'18.2389"E, 2d50'05.3054"S	115d29'26.4608"E, 2d50'10.1622"S	16	F	NF	F	2012/9/23
15	115d21'27.4230"E, 2d58'50.1595"S	115d21'32.9065"E, 2d58'55.3632"S	13	F	F	NF	2012/9/23
16	115d11'31.8595"E, 2d59'38.3883"S	115d11'35.0809"E, 2d59'44.7250"S	11	F	F	NF	2012/7/23
17	116d15'05.9863"E, 2d59'47.3433"S	116d15'10.1422"E, 2d59'50.0294"S	7	F	F	NF	2011/1/8
18	115d18'32.4695"E, 3d02'14.9824"S	115d18'45.6613"E, 3d02'29.3691"S	34	F	F	NF	2012/9/23
19	115d20'53.5352"E, 3d07'31.6435"S	115d21'01.8921"E, 3d07'36.3026"S	15	F	F	NF	2010/10/19
20	115d12'02.4607"E, 3d14'09.5610"S	115d12'10.4834"E, 3d14'15.5913"S	17	F	F	NF	2012/5/5
21	115d00'38.3740"E, 3d14'46.3761"S	115d00'41.5610"E, 3d14'49.3823"S	6	F	NF	F	2012/5/5
22	115d14'58.3193"E, 3d14'58.8811"S	115d15'03.5926"E, 3d15'04.4068"S	13	F	F	NF	2012/5/5
23	115d08'38.1834"E, 3d16'24.6374"S	115d08'41.0651"E, 3d16'29.3323"S	7	F	F	NF	2012/5/5
24	115d11'40.7311"E, 3d18'31.4393"S	115d11'46.4215"E, 3d18'37.1229"S	13	F	F	NF	2012/5/5
25	115d06'44.3231"E, 3d20'22.0514"S	115d06'50.9364"E, 3d20'28.7655"S	17	F	F	NF	2012/5/5
26	115d05'52.9424"E, 3d25'54.2410"S	115d05'57.8160"E, 3d26'01.8461"S	14	F	F	NF	2004-2013
27	114d59'44.8108"E, 3d30'42.4451"S	114d59'49.5875"E, 3d30'47.1188"S	9	F	F	NF	2009-2013
28	115d24'37.1796"E, 2d06'54.4089"S	115d24'42.6663"E, 2d07'01.0554"S	15	F	F	NF	2010/9/20
29	115d25'26.9619"E, 2d14'06.7342"S	115d25'30.8428"E, 2d14'09.1228"S	7	F	NF	F	2009/3/18

30	115d29'29.5755"E, 2d14'40.0379"S	115d29'34.8231"E, 2d14'48.2570"S	17	F	NF	F	2009-2012
31	115d27'14.6725"E, 2d16'43.9639"S	115d27'21.3222"E, 2d16'53.5657"S	20	F	F	NF	2010/9/15
32	115d30'53.2071"E, 2d17'18.6862"S	115d30'55.8148"E, 2d17'22.9676"S	7	F	NF	F	2010/9/15
33	115d37'20.8344"E, 2d17'55.6233"S	115d37'25.7739"E, 2d18'00.7585"S	13	F	NF	F	2010/9/15
34	115d15'09.8357"E, 2d17'50.4653"S	115d15'13.3986"E, 2d17'55.9457"S	11	F	NF	F	2010/9/15
35	115d04'46.3029"E, 2d21'56.6322"S	115d04'54.6308"E, 2d22'02.8431"S	17	NF	NF	F	2010/9/15
36	115d05'13.2836"E, 2d23'57.4832"S	115d05'20.4834"E, 2d24'06.1305"S	19	NF	NF	F	2010/9/15
37	115d05'54.6628"E, 2d24'19.7039"S	115d05'58.5694"E, 2d24'24.4619"S	9	NF	NF	F	2010/9/15
38	115d30'08.4962"E, 2d25'06.1256"S	115d30'13.2980"E, 2d25'11.1254"S	12	F	F	NF	2012/9/23
39	115d31'21.1763"E, 2d28'28.1111"S	115d31'25.5500"E, 2d28'34.3184"S	11	F	NF	F	2012/9/23
40	115d30'15.7758"E, 2d39'43.1079"S	115d30'20.2536"E, 2d39'47.1151"S	9	F	NF	F	2012/9/23
41	115d27'51.0125"E, 2d56'21.0825"S	115d27'54.4447"E, 2d56'24.8161"S	9	F	NF	F	2012/9/23
42	114d59'32.2994"E, 2d57'29.9238"S	114d59'37.0019"E, 2d57'33.7954"S	9	NF	F	NF	2012/5/11
43	115d18'22.3012"E, 3d00'41.4265"S	115d18'29.4805"E, 3d00'49.0222"S	17	F	F	NF	2012/9/23
44	115d15'08.4362"E, 3d07'06.3640"S	115d15'15.4623"E, 3d07'15.6995"S	19	F	F	NF	2009/4/10
45	116d04'04.6660"E, 3d07'14.1764"S	116d04'10.1425"E, 3d07'18.6435"S	11	F	F	NF	2010/12/17
46	116d08'21.3204"E, 3d07'15.2196"S	116d08'23.6338"E, 3d07'26.0337"S	15	F	NF	F	2010/12/17
47	115d19'55.7729"E, 3d11'46.3854"S	115d20'02.7323"E, 3d11'55.3898"S	19	F	F	NF	2012/5/5
48	115d06'54.0677"E, 3d18'15.5000"S	115d07'03.0607"E, 3d18'20.7657"S	17	F	F	NF	2012/5/5
49	115d00'06.4120"E, 3d21'40.1238"S	115d00'12.1762"E, 3d21'45.7765"S	13	F	F	NF	2009-2013
50	114d28'48.6048"E, 3d23'53.8393"S	114d28'52.0811"E, 3d23'59.1880"S	11	NF	F	NF	2005-2012
51	114d48'13.9389"E, 3d24'56.5501"S	114d48'17.8496"E, 3d25'02.3802"S	11	F	F	NF	2009-2011
52	115d05'32.8755"E, 3d25'57.2838"S	115d05'36.6289"E, 3d26'03.0362"S	11	F	F	NF	2004-2013
53	115d02'02.3907"E, 3d30'06.7919"S	115d02'09.0693"E, 3d30'12.1764"S	15	F	F	NF	2004-2013
54	115d01'59.3736"E, 3d30'23.4887"S	115d02'08.4001"E, 3d30'33.6760"S	23	F	F	NF	2004-2013
55	114d58'10.3241"E, 3d31'18.9007"S	114d58'17.2158"E, 3d31'26.7089"S	17	F	F	NF	2009-2014
56	115d00'48.9852"E, 3d35'30.5015"S	115d00'55.6614"E, 3d35'37.4409"S	17	F	F	NF	2009-2013
57	114d42'32.5246"E, 3d50'51.1933"S	114d42'40.2346"E, 3d50'59.1480"S	17	F	F	NF	2012/7/28
58	114d50'32.5170"E, 4d04'53.0874"S	114d50'35.0028"E, 4d04'58.1373"S	9	F	F	NF	2010/2/10
59	114d45'04.8306"E, 4d05'48.3646"S	114d45'06.6846"E, 4d05'55.8320"S	11	F	F	NF	2010/2/10
60	115d24'04.1456"E, 2d03'42.6109"S	115d24'09.0815"E, 2d03'47.7471"S	13	F	F	NF	2010/9/20
61	115d24'58.1710"E, 2d09'51.5322"S	115d24'59.9652"E, 2d09'57.0785"S	9	F	F	NF	2010/9/20

62	115d20'11.1288"E, 2d14'44.4159"S	115d20'19.0999"E, 2d14'49.8732"S	15	F	F	NF	2010/9/15
63	115d31'21.6513"E, 2d16'47.8027"S	115d31'25.4232"E, 2d16'53.7638"S	11	F	NF	F	2010/9/15
64	115d27'43.9400"E, 2d17'18.1839"S	115d27'48.4579"E, 2d17'23.2874"S	11	F	F	NF	2010/9/15
65	115d35'43.6451"E, 2d16'52.5081"S	115d35'47.1918"E, 2d16'56.8732"S	9	F	NF	F	2010/9/15
66	115d26'43.7391"E. 2d17'23.2921"S	115d26'48.3660"E, 2d17'27.5578"S	9	F	F	NF	2010/9/15
67	116d12'38.0969"E, 2d20'12.3302"S	116d12'42.7499"E, 2d20'17.9650"S	11	F	F	NF	2009/5/17
68	116d14'42.8124"E, 2d20'43.2395"S	116d14'48.6425"E, 2d20'49.6591"S	15	F	F	NF	2009/5/17
69	116d13'07.7158"E, 2d20'54.6917"S	116d13'11.5765"E, 2d21'01.4583"S	12	F	F	NF	2009/5/17
70	116d14'51.9600"E, 2d21'00.4049"S	116d14'55.7978"E, 2d21'04.9906"S	9	F	F	NF	2009/5/17
71	116d14'59.8558"E, 2d21'13.6224"S	116d15'03.4237"E, 2d21'17.5954"S	9	F	F	NF	2009/5/17
72	115d05'52.7595"E, 2d21'36.7482"S	115d05'57.3333"E, 2d21'43.8769"S	13	NF	NF	F	2010/9/15
73	115d38'28.9471"E, 2d26'41.7774"S	115d38'37.0192"E, 2d26'46.7090"S	17	F	NF	F	2010/10/19
74	114d55'17.5724"E, 2d34'56.7526"S	114d55'23.6628"E, 2d35'01.9908"S	13	F	F	NF	2012/5/11
75	114d55'15.9667"E, 2d35'09.5432"S	114d55'24.1779"E, 2d35'17.6091"S	21	F	F	NF	2012/5/11
76	115d27'33.2245"E, 2d45'07.1063"S	115d27'36.2312"E, 2d45'12.3980"S	9	F	NF	F	2012/9/23
77	115d16'17.8553"E, 2d55'31.6020"S	115d16'20.1930"E, 2d55'35.5278"S	7	F	NF	F	2012/9/23
78	115d21'31.5122"E, 2d58'33.2381"S	115d21'35.1504"E, 2d58'42.3295"S	15	F	F	NF	2012/9/23
79	115d11'03.1842"E, 3d02'23.0002"S	115d11'07.5587"E, 3d02'29.6764"S	13	F	NF	F	2012/9/23
80	116d06'11.3105"E, 3d04'36.7092"S	116d06'19.5403"E, 3d04'43.7718"S	18	F	F	NF	2010/12/17
81	115d21'08.0879"E, 3d06'38.0610"S	115d21'11.9449"E, 3d06'44.8452"S	13	F	F	NF	2010/10/19
82	115d16'56.0730"E, 3d08'53.5547"S	115d17'04.4702"E, 3d09'01.2632"S	19	F	F	NF	2009/4/10
83	115d05'26.7450"E, 3d35'45.1826"S	115d05'34.3501"E, 3d35'50.1662"S	15	F	F	NF	2004-2013
84	115d25'16.1388"E, 2d09'33.4359"S	115d25'21.4941"E, 2d09'37.8061"S	11	F	F	NF	2010/9/20
85	115d36'53.2215"E, 2d17'37.2061"S	115d36'59.0058"E, 2d17'43.3690"S	13	F	NF	F	2010/9/15
86	115d25'43.0295"E, 2d19'35.4574"S	115d25'48.3457"E, 2d19'39.7033"S	11	F	F	NF	2010/9/15
87	115d32'23.4057"E, 2d24'15.3772"S	115d32'28.5481"E, 2d24'22.6561"S	15	F	NF	F	2010/9/15

Table 5.4: The imformation of each validation polygon in the case of the ground truth tree crown cover is more than 10% but less than 50%.

ID number	Coordinate		Pixels	Classification result		Y/M/D
	Upper left	Lower right		This study	JAXA	
1	116d08'29.9047"E, 3d06'40.9883"S	116d08'34.5859"E, 3d06'46.1135"S	11	F	NF	2010/12/17
2	115d30'53.3740"E, 2d10'53.5240"S	115d30'57.1415"E, 2d11'00.1170"S	13	NF	F	2010/9/20
3	115d38'02.2457"E, 2d19'02.7250"S	115d38'06.0350"E, 2d19'06.5242"S	9	NF	F	2010/9/15
4	115d31'30.9851"E, 2d32'02.4731"S	115d31'36.8909"E, 2d32'08.8925"S	15	F	NF	2012/9/23
5	115d14'08.8131"E, 2d54'11.7001"S	115d14'16.1444"E, 2d54'16.2174"S	13	NF	F	2012/9/23
6	115d18'52.9345"E, 2d59'31.8020"S	115d18'57.0513"E, 2d59'36.4539"S	9	NF	F	2012/9/23
7	115d19'11.3452"E, 3d01'22.9598"S	115d19'18.0273"E, 3d01'30.6580"S	17	NF	F	2012/9/23
8	115d36'53.2889"E, 2d05'53.1809"S	115d36'57.5927"E, 2d05'57.8837"S	9	NF	F	2010/9/20
9	116d08'45.2085"E, 2d25'39.5063"S	116d08'47.4988"E, 2d25'43.6496"S	7	NF	F	2009/3/27
10	115d27'38.1417"E, 2d35'24.0542"S	115d27'41.9172"E, 2d35'30.0658"S	11	F	NF	2012/9/23
11	115d20'03.9327"E, 2d52'57.1904"S	115d20'07.8372"E, 2d53'03.1718"S	11	F	NF	2012/9/23
12	115d18'10.4639"E, 2d53'45.3229"S	115d18'14.5038"E, 2d53'53.8050"S	15	NF	F	2012/9/23
13	115d21'20.7244"E, 3d02'00.8983"S	115d21'25.0045"E, 3d02'06.8972"S	11	F	NF	2012/9/23
14	115d58'59.3768"E, 3d03'47.6909"S	115d59'05.2115"E, 3d03'56.2419"S	17	NF	F	2011/1/8
15	115d13'46.1271"E, 3d10'36.8745"S	115d13'51.2548"E, 3d10'43.2733"S	14	NF	F	2009/4/10
16	116d24'58.1412"E, 3d23'55.7908"S	116d25'03.2875"E, 3d24'03.1370"S	15	NF	F	2009/6/11
17	114d45'04.7279"E, 3d46'48.5154"S	114d45'08.0241"E, 3d46'53.5141"S	11	NF	F	2012/7/28
18	115d19'21.5392"E, 2d16'58.1892"S	115d19'27.1944"E, 2d17'04.2403"S	13	NF	F	2010/9/15
19	115d38'22.0816"E, 3d34'05.4128"S	115d38'28.0394"E, 3d34'16.8998"S	21	NF	F	2010/11/12

5.2. Validation

199 random samplings were successfully identified by using Google Earth images. The distribution of random samplings, only were collected from areas where there are clear satellite images or photos in 2010, are shown in Figure 5.6.

Finally, the overall accuracy of this forest map is 85.43% with kappa coefficient of 0.65. The producer's accuracy of forest and non-forest are 84.91% and 85.62%, respectively. The user's accuracy of forest and non-forest are 68.18% and 93.98%, respectively (Table 5.5).

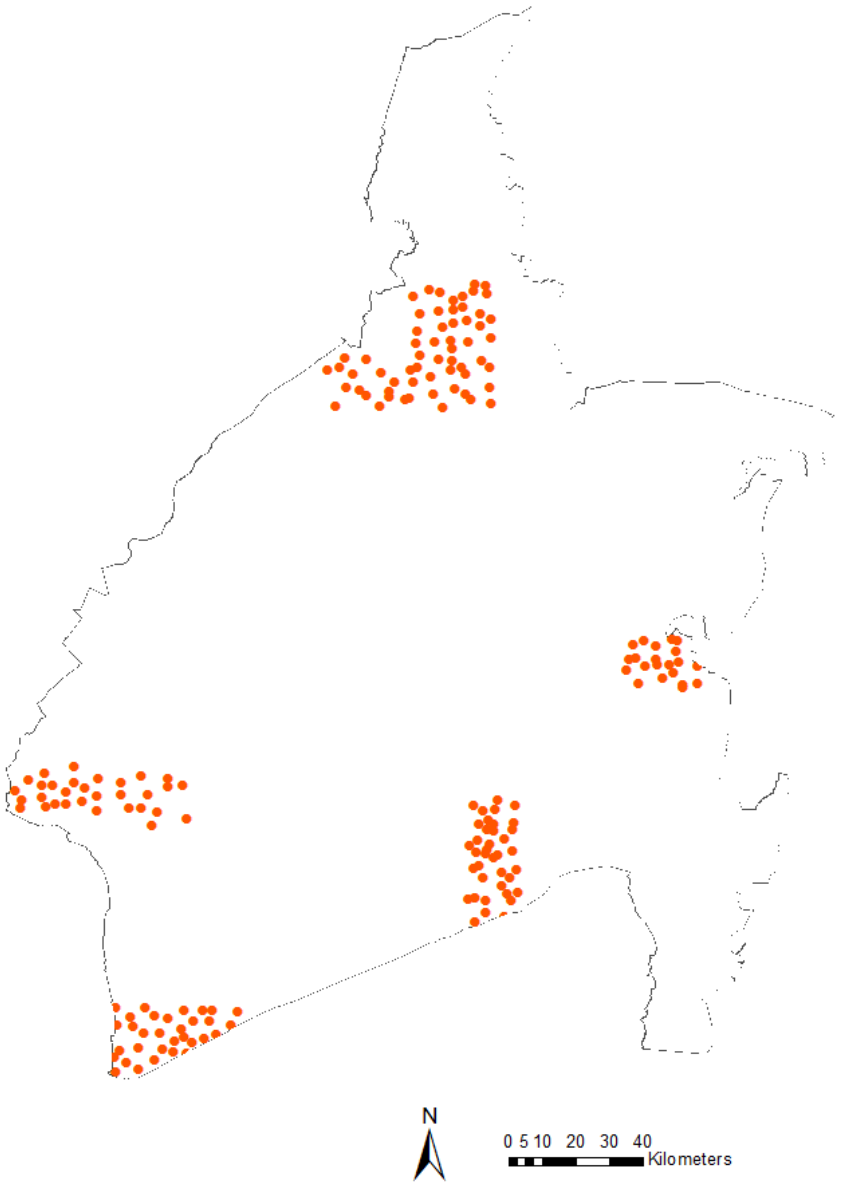


Figure 5.6: Distribution of validation points.

Table 5.5: Accuracy of the final forest and non-forest map.

Reference / Result	**Forest**	**Non-forest**	**Total**	**User's accuracy (%)**
Forest	**45**	8	53	84.91%
Non-forest	21	**125**	146	85.62%
Total	66	133	**199**	
Producer's accuracy (%)	68.18%	93.98%		
Overall accuracy	85.43%			
Kappa coefficient	0.65			

5.3. Discussions

5.3.1. Limitation of forest classification result

According to the accuracy assessment result, in the total number of forest ground truth, there are still 8 forest validation polygons are misclassified as non-forest, while 21 non-forest are misclassified as forest. The erroneous validation polygons occurred on:

- Where the trees are growing on the water land (mangrove, peat forest) (Figure 5.7).

Figure 5.7: Misclassification occurred on wetland or mangrove.

Figure 5.7 shows non-forest result of classification while peat forest on Google Earth image. The reason of this misclassification is considering from the training data of nature wetland, due to the difficulty of separating mangrove with peat forest by using Google Earth visually alone. Therefore, the low accuracy produced from peat forest can be improved with new training data collected from high resolution aerial photography.

- Where forest is misclassified as urban.

Given the kappa coefficient of separating forest and urban in Figure 4.8, using J48 classifier together with the selected features is better than SMO and RF, but it is not perfect. It has a good performance where urban area is the big city, while a low accuracy happen on the place mixed with trees and buildings. Thus, more and more object features and different classifiers are

considering to be tried in the case of forest and urban classification when using ALOS PALSAR alone, while carrying out the Normalized Difference Vegetation Index (NDVI) with adding optical data in the future.

5.3.2. Application of training site

Collecting ground truth data is the first task of using supervised classification method, its quality will directly affect the accuracy of classification result. The most commonly-used tool to take training data is Google Earth, which provides time-series images over the world. However, there is a big problem that the clear image uploaded on Google Earth is not including every year. For example, in this study, the clear images over South Kalimantan are far from enough with the number is eleven. On the other hand, the high quality aerial photos is good for selecting training data, but with a high cost in global scale.

Table 5.6: Forest types on test areas.

Time/Area	**Forest type**
2007/Africa	broadleaf evergreen forest
2008/North America	broadleaf deciduous forest needleleaf evergreen forest
2009/China	broadleaf deciduous forest needleleaf evergreen forest

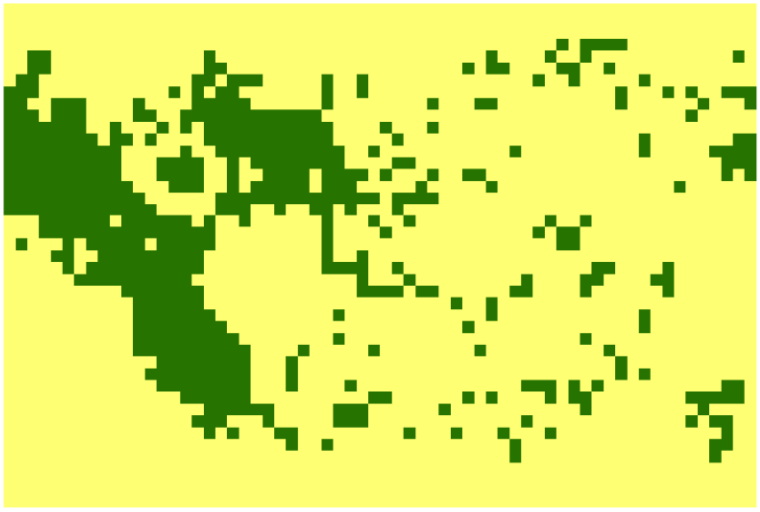
In order to solve the inconvenience of collection training data, the training site used in this study were tested to classify forest and non-forest on different regions and years. All the training data collected in chapter 4 were separated as two classes only including forest and non-forest. RF classifier was carried out to extract forest with the optimal feature combination determined in Table 4.4. The test areas were chosen from Africa (2007). North America (2008), and China (2009), where the forest type including broadleaf evergreen forest, broadleaf deciduous forest and needleleaf evergreen forest (Table 5.6). The classification results are shown in Figure 5.8-Figure 5.10.

1) Africa _2007

(9d27'24.8000"E, 0d26'35.2000"N; 9d29'07.2000"E, 0d25'28.0000"N).

[1]

[2]



[3]

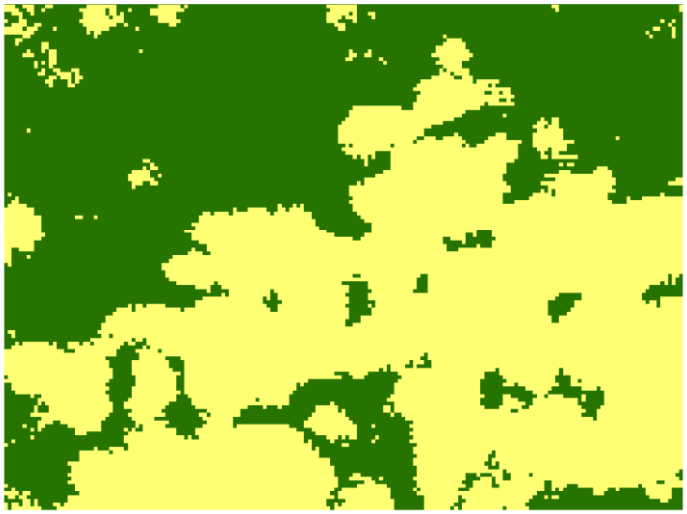
Figure 5.8: Forest/non-forest classification on Africa (2007) region. [1] Google Earth image (2007); [2] Test result; [3] Result of JAXA. Green: forest; Yellow: non-forest.

2) North America _2008

(77d14'24.2500"W, 41d28'12.7000"N; 77d09'33.8500"W, 41d24'34.3000"N)

[1]

[2]



[3]

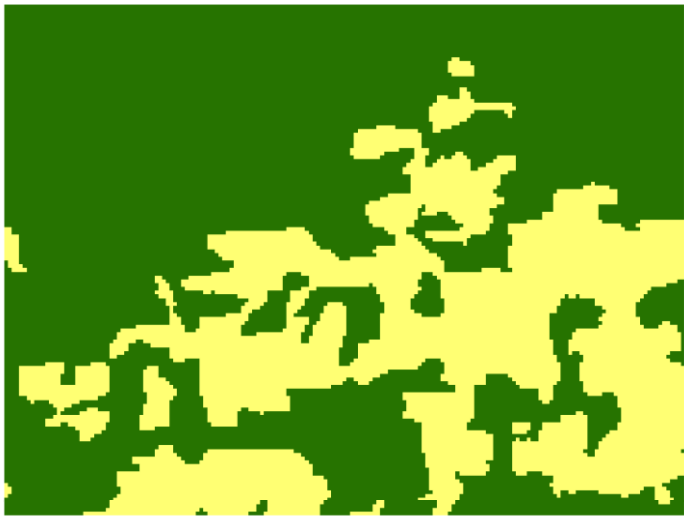


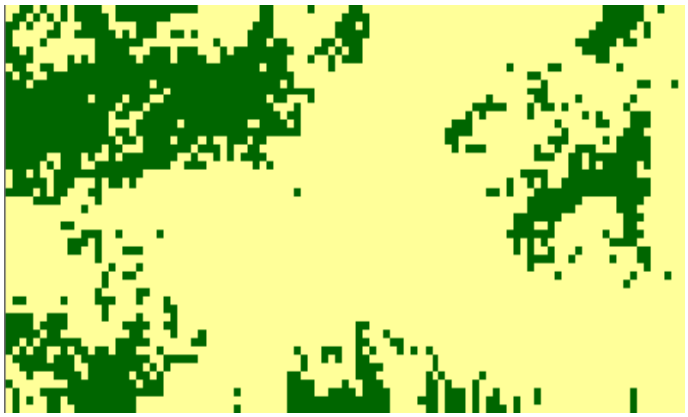
Figure 5.9: Forest/non-forest classification on North America (2008) region. [1] Google Earth image (2008); [2] Test result; [3] Result of JAXA. Green: forest; Yellow: non-forest.

3) China _2009

(125d30'20.8000"E, 43d46'59.2000"N; 125d32'59.2000"E, 43d45'40.8000"N)

[1]

[2]



[3]

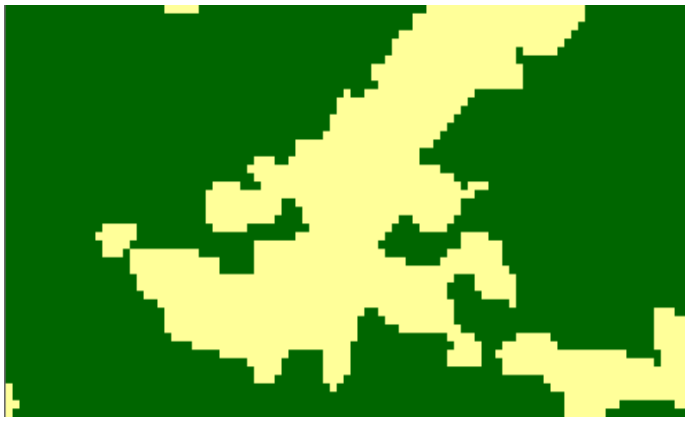


Figure 5.10: Forest/non-forest classification on China (2009) region. [1] Google Earth image (2009); [2] Test result; [3] Result of JAXA. Green: forest; Yellow: non-forest.

According to the visual interpretation, forest and urban on Africa (2007), forest and agriculture both for North America (2008) and China (2009), were successfully classified as forest and non-forest.

Each of the 50 points were collected randomly for validating the results of three different test regions, the user's accuracy of forest class and producer's accuracy of non-forest from North American and China were achieved to 100%. The overall accuracy were 76%, 86% and 70%, respectively (Table 5.7 (a), Table 5.8 (a) and Table 5.9 (a)), while the overall accuracy of global forest/non-forest map generated by JAXA were 38%, 90% and 78%, respectively (Table 5.7 (b), Table 5.8 (b) and Table 5.9 (b)). In the case of classifying forest and urban in Africa (2007), overall accuracy of this study showed the accuracy of 28% is higher than JAXA, while in North America (2008) and China (2009), 4% and 8% is lower than JAXA.

Three different types of forests locating on the different locations were separated from non-forest area based on the same training data features. More and more different type of forest is considering to be tested in the future. However, the acceptable result in this study point out that backscattering characteristic is changing with different forest type and different location, but they may have a common characteristic from the object features. If the object features is enough accurate, it is possible to map forest and non-forest regardless of time, location and types.

Table 5.7 (a): Accuracy assessment of forest/non-forest map tested on Africa 2007.

Reference / This study	**Forest**	**Non-forest**	**Total**	**User's accuracy**
Forest	**7**	9	16	43.75%
Non-forest	3	**31**	34	91.18%
Total	10	40	**50**	
Producer's accuracy	70%	77.50%		
Overall accuracy	76%			

Table 5.7 (b): Accuracy assessment of JAXA's global forest/non-forest map.

Reference / JAXA	**Forest**	**Non-forest**	**Total**	**User's accuracy**
Forest	**16**	0	16	100.00%
Non-forest	31	**3**	34	8.82%
Total	47	3	**50**	
Producer's accuracy	34.04%	100.00%		
Overall accuracy	38%			

Table 5.8 (a): Accuracy assessment of forest/non-forest map tested on North America 2008.

Reference / This study	**Forest**	**Non-forest**	**Total**	**User's accuracy**
Forest	**24**	7	31	77.42%
Non-forest	0	**19**	19	100.00%
Total	24	26	**50**	
Producer's accuracy	100%	73.08%		
Overall accuracy	86%			

Table 5.8 (b): Accuracy assessment of JAXA's global forest/non-forest map.

Reference / JAXA	**Forest**	**Non-forest**	**Total**	**User's accuracy**
Forest	**29**	2	31	93.55%
Non-forest	3	**16**	19	84.21%
Total	32	18	**50**	
Producer's accuracy	90.63%	88.89%		
Overall accuracy	90%			

Table 5.9 (a): Accuracy assessment of forest/non-forest map tested on China 2009.

Reference / This study	**Forest**	**Non-forest**	**Total**	**User's accuracy**
Forest	**14**	15	29	48.28%
Non-forest	0	**21**	21	100.00%
Total	14	36	**50**	
Producer's accuracy	100%	58.33%		
Overall accuracy	70%			

Table 5.9 (b): Accuracy assessment of JAXA's global forest/non-forest map.

Reference / JAXA	**Forest**	**Non-forest**	**Total**	**User's accuracy**
Forest	**26**	3	29	89.66%
Non-forest	8	**13**	21	61.90%
Total	34	16	**50**	
Producer's accuracy	76.47%	81.25%		
Overall accuracy	78%			

CONCLUSIONS

In this thesis, a new modified slope correction model was produced for ALOS PALSAR 50 m ortho-rectified mosaic data, and the main aim of this study that developing a prominent classification method had been achieved with a higher accuracy forest map in South Kalimantan which generated by a classifiers combination.

In terms of data calibration,

1) In order to match with the SRTM image, Ground Control Points (GCPs) geometric correction approach was applied to ALOS PALSAR 50 m ortho-rectified mosaic data by using the Landsat images. The correction quality had shown with the maximum RMS error is 0.9441;
2) The application of three existing slope correction models were not able to meet a requirement of reducing the terrain influence from ALOS PALSAR 50 m ortho-rectified mosaic data, while the modified model in this study, which produced based on the scattering geometry had shown a stronger ability is both on visual verification and logical analysis;
3) The reasons of less available of the existing slope models were considered as, that might be caused of using of different software tool (model-1), different code-level programming (model-2) and different determination of a number of parameter factors (model-3);
4) With the visual interpretation, the new open source global ALOS PALSAR 50 m mosaic data had shown a smoother correction effectiveness over the mountain ridge and the valley between mountains than the use of the modified slope correction model in this study, which was considering to be caused by the different process starting. The new open source data was corrected based on raw data, while ALOS PALSAR 50 m ortho-rectified mosaic data is only leaving with slope correction undone.

In the following cases, in terms of forest classification,

1) Classification performance of the three well-known machine learning classifiers, J48 (C4.5), Sequential Minimal Optimization (SMO) and Random Forest (RF) which applied with object-based approach were compared in the case of classifying forest, oil palm, herbaceous, urban, bare area, artificial wetland, natural wetland and water body within South Kalimantan. In this case, RF classifier had shown the best performance with the most suitable combination consisting of 15 features which was chosen by correlation-based evaluation;

2) Forest reclassification was done based on the classified land cover results consist of herbaceous, urban, bare land, artificial wetland and natural wetland, which the user's accuracy is lower than 60%. The accuracy assessment of land cover classification indicated that the urban class was easier to confuse with forest than the other non-forest classes;
3) The final forest mapping in South Kalimantan was generated with the combination of RF and J48 classifier. The overall accuracy was 85.43% and kappa coefficient was 0.65. The accomplished production had shown a better classification performance in compare with the public forest and non-forest map of JAXA over the different classified area with the overall accuracy of 67.82% and 32.18%, respectively;
4) The accuracy assessment of other test site of forest mapping on Africa, North America and China indicated that the using of optimal feature combination extracted from the training data collected on South Kalimantan might be used for forest/non-forest classification on the different regions and years without collecting new training data.

Finally, in order to improve the accuracy of this classification methodology which developed in this study, the following work is expected to be practiced in the future.

1) Adding some correction on the unsure training data;
2) Using Normalized Difference Vegetation Index (NDVI) to separate urban individually;
3) Percent tree coverage data which was produced before (e.g. Percent Tree Cover produced by Maryland University) is considering to map sparse forests where the tree crown cover is less than 50%;
4) The application of training site will continue to be tested and verified on the extraction of forest class globally, particularly for the new PALSAR data which is producing by ALOS-2 in near future.

REFERENCES

Akatsuka, S., Takeuchi, W., Rakwatin, P., Sawada, H., 2009. Evaluation of Slope Correction Effects on ALOS PALSAR Mosaic Data Set in Forest Mapping in Indonesia and Malaysia. 30th *Asian Conference on Remote Sensing*, Beijing, 19-23 October 2009, TS11-03

Ali, J., Khan, R., Ahmad, N., Maqsood, I., 2012. Random forest and decision trees. *International Journal of Computer Science Issues*. Vol. 9, No. 3, pp. 272-278.

ALOS/PALSAR Mosaic Product Format (2008). Available online at: http://www.eorc.jaxa.jp/ALOS/en/kc_mosaic/kc_mosaic.htm

Annunzio, D.R., Lindquist, E.J., MacDicken, K.G., 2014, Global forest land-use change from 1990 to 2010: an update to a global remote sensing survey of forests. *The Global Remote Sensing Survey*, FAO. Available online at: http://www.lafranceagricole.fr/var/gfa/storage/fichiers-pdf/Docs/2014/forets-FAO.pdf

Bayer, T., Winter, R., Schreier, G., 1991. Terrain Influences in SAR Backscatter and Attempts to their Correction. *IEEE Transactions on Geoscience and Remote Sensing*. Vol. 29, pp. 451-462.

Bert Metz, Ogunlade Davidson, Heleen de Coninck, Manuela Loos, Leo Meyer, 2005. *Carbon dioxide capture and storage* (digital book). IPCC.

Blaschke, T., 2010. Object based images analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing.* Vol. 65, pp. 2-16.

Castel, T., Beaudoin, A., Stach, N., Stussi, N., Le Toan, T., Durand, P., 2001. Sensitivity of Space-Borne SAR Data to Forest Parameters over Sloping Terrain. Theory and Experiment. *International Journal of Remote Sensing*. Vol. 22, pp. 2351-2376.

Christopher, B.F., 2001. Plant physiology of the "Missing" carbon sink. *American Society of Plant Biologists*. Vol. 125, pp. 25-28.

Cufoglu, A., Lohi, M., Madani, K., 2009. A comparative study of selected classifiers with classification accuracy in user profiling. *World Congress on Computer Science and Information Engineering*. pp. 708-712.

FAO, 2014, State of the World's Forests. Available online at:
http://www.fao.org/forestry/sofo/en/.

Japan Aerospace Exploration Agency, 2009. SAR Image Processing Using GRASS in GIS Knoppix Ver. 1.2. Available online at:
http://www.safe.iis.u-tokyo.ac.jp/pdf/SAR_image_processing_ver1.1.pdf

JAXA, 2014. New global 25m-resolution PALSAR mosaic and forest/non-forest map (2007-2010), Version 0.

Jin, X., Xu, A., Bie, R., Guo, P., 2006. Machine Learning Techniques and Chi-Square Feature Selection for Cancer Classification Using SAGE Gene Expression Profiles. *Lecture Notes in Computer Science.* Vol. 3916, pp. 106-115.

Hall, M.A., 1999. Correction-based feature selection for machine learning. PhD Thesis, University of Waikato.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., et al., 2013. High-resolution global maps of 21st-centure forest cover change. *Science*. Vol. 342, pp. 850-853.

Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*. Vol. 23, No. 4, pp.725-749.

Hussain, M., Chen D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: From pixel based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*. Vol. 80, pp. 91-106.

Indrabudi, H., 2002. Forestland: Its dynamics, disorganised uses and planning in South Kalimantan, Indonesia. Thesis Wageningen University. Available online at:
http://library.wur.nl/WebQuery/wurpubs/319836

IPCC, 2014.*Climate change 2014 synthesis report*. Available online at:
http://www.ipcc.ch/report/ar5/syr/

Kellndorfer, J.M., Pierce, L.E., Dobson, M.C., Ulaby, F.T., 1998. Toward Consistent Regional-to-Global-Scale Vegetation Characterization Using Orbital SAR Systems. *IEEE Transactions on Geoscience and Remote Sensing*. Vol. 36, pp. 1396-1411.

K&C mosaic homepage: http://www.eorc.jaxa.jp/ALOS/en/kc_mosaic/kc_mosaic.htm

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. *ScienceDirect ELSEVIER*. Vol. 97, pp. 273-324.

Lan, H.W., Frank, E., Hall, M.A., 2011. Data Mining: Practical machine learning tools and techniques, 3rd edition (digital book). *ELSEVIER*.

Longepe, N., Rakwatin, P., Isoguchi, O., Shimada, M., Uryu, Y., Yulianto, K., 2011. Assessment of ALOS PALSAR 50 m Orthorectified FBD Data for Regional Land Cover Classification by Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing*. Vol. 49, pp. 2135-2150.

Loew, A., Mauser, W., 2007. Generation of Geometrically and Radiometrically Terrain Corrected SAR Image Products. *Remote Sensing of Environment*, Vol. 106, pp. 337-349

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. Vol. 66. pp. 247-259.

Mead, D.J., 2001. *Forest plantations thematic papers: Protecting plantations from pests and diseases.* Rome, Italy: FAO.

Pakkad, G., Elliott, S., Anusarnsunthorn, V., 2001. Forest restoration planting in Northern Thailand. *FORSPA Publication (FAO)*. NO. 31. Available online at: http://www.fao.org/docrep/005/ac648e/ac648e0e.htm#TopOfPage

Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*. Vol. 26, No. 5, pp. 1007-1011.

Percy, K.E., Jandl, R., Hall, J.P. Lavigne, M., 2003. The role of forests in carbon cycles, sequestration, and storage. *International Union of Forest research organizations.* Newsletters, NO.1. Available online at:

http://iufro.boku.ac.at/taskforce/hptfcs.htm “Newsletters”

Ruben Ramirez-Padron, 2007. A roadmap to SVM sequential minimal optimization for classification, a tutorial. *School of Electrical Engineering and Computer Science University of Central Florida.*

Rokhmatuloh, Murfi, H., Tateishi, R., 2012. Support Vector Machine (SVM) for Classification of Forest and Non Forest Derived from ALOS PALSAR Data. *Report and Proceedings of ALOS Application and Verification Project in Indonesia*. pp. 18-25.

Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmuller, U., Wiesmann, A., 2011. Retrieval of Growing Stock Volume in Boreal Forest Using Hyper Temporal Series of Envisat ASAR Scan SAR Backscatter Measurements. *Remote Sensing of Environment*, Vol. 115, pp. 490-507.

Shiraishi, T., Motohka, T., Thapa, R.B., Watanabe, M., Shimada, M., 2014. Comparative assessment of supervised classifiers for land use – land cover classification in a tropical region using time-series PALSAR mosaic data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Vol. 7, No. 4, pp. 1186-1199.

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., Lucas, R., 2014. New global forest/non-forest maps from ALOS PALSAR data (2007-2010). *Remote Sensing of Environment*. Vol. 155, pp. 13-31.

Shimada, M., Isoguchi, O., Rosenqvist, A., 2008. Continent Scale PALSAR Mosaic Products for Kyoto and Carbon Project. *Institute of Electronics. Information and Communication Engineers*. pp. 81-86.

Shimada, M., Isoguchi, O., Tadono, T., Isono, K., 2009. PALSAR radiometric and geometric calibration. IEEE *Transactions on Geoscience and Remote Sensing*. Vol. 47, No. 2, pp. 3915-3932.

Sophie Chao, 2012. *Forest peoples: Numbers across the world*. Forest Peoples Programm (FPP).

Sun, N., Ranson, K.J., 2001. Terrain Effect on Forest Radar Backscatter: Modeling and Correction. *Committee on Earth Observation Satellites (CEOS) SAR Workshop*, CEOS-

SAR01-074.

Sun, G., Ranson, K.J. and Kharuk, V.I., 2002. Radiometric Slope Correction for Forest Biomass Estimation from SAR Data in the Western Sayani Mountains, Siberia. *Remote Sensing of Environment*. Vol. 79, pp. 279-287.

Ulaby, F.T., Moore, R.K. and Fung, A.K., 1982. Microwave Remote Sensing, Active and Passive. Artech House, Norwood.

Ulander, L.M.H., 1996. Radiometric Slope Correction of Synthetic-Aperture Radar Images. *IEEE Transactions on Geoscience and Remote Sensing*. Vol. 34, pp. 1115-1122.

Wegmuller, U., 1999. Automated Terrain Corrected SAR Geocoding. *IEEE Geoscience and Remote Sensing Symposium*. Hamburg, 28 June -2 July 1999, pp. 1712-1714.

Whitmore, T.C., Sayer, J.A., 1992. *Tropical deforestation and species extinction* (digital book). London: Chapman & Hall.

Zhao, Y., Zhang, Y., 2007. Comparison of decision tree methods for finding active objects. *Advances of Space Research*. Available online at:
http://arxiv.org/pdf/0708.4274.pdf