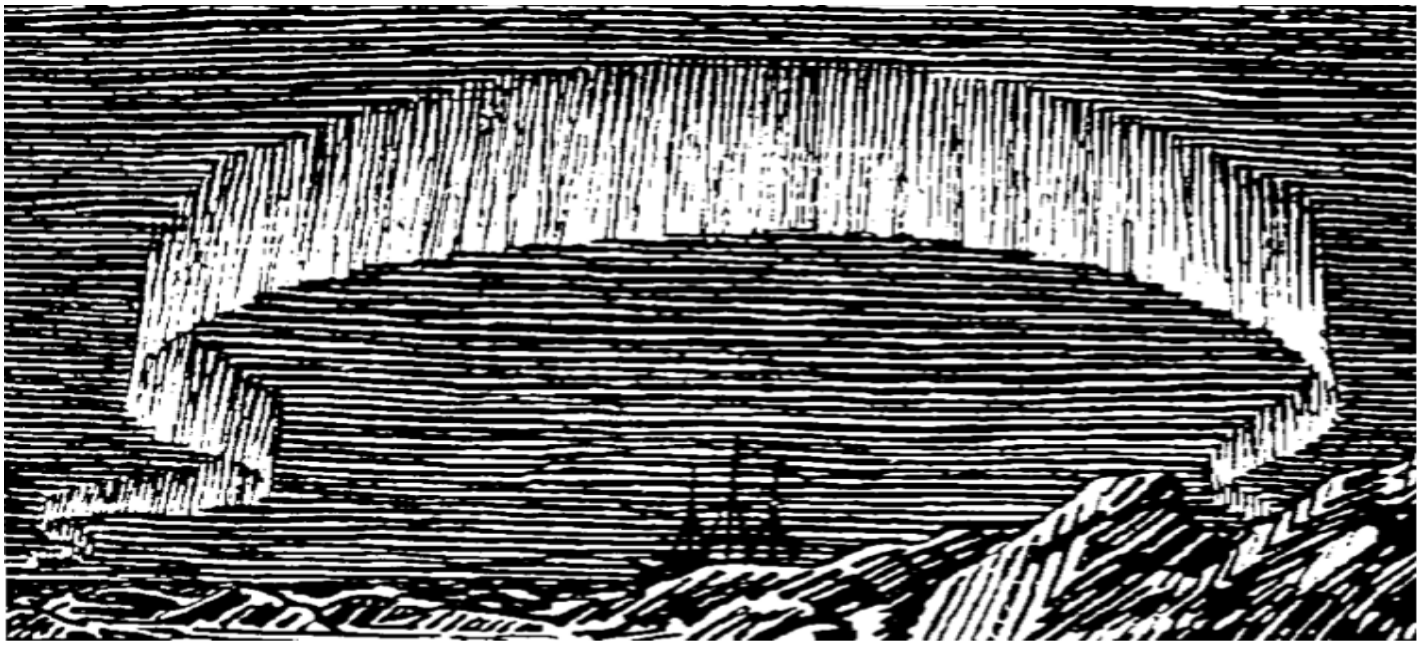
UNIVERSITY OF TROMSØ UIT

FACULTY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF PHYSICS AND TECHNOLOGY

Evaluation of Polarimetric SAR Decomposition methods for tropical forest analysis



Original by F. Nansen

Stein Erik Sakshaug

EOM-3901 - Master Thesis in Energy, Climate and Environment

June 2013

Abstract

Information about tropical forest can be obtained by remote sensing, using either optical instruments or an active radar like synthetic aperture radar(SAR). Polarimetric decomposition theorems break polarimetric SAR measurements into components that describes the scattering behavior of the target. This thesis deals with evaluating the suitability of the various decomposition theorems to describe and classify areas of tropical forest. High resolution images provided by an optical spaceborne instrument is used as ground truth information. These images are used to determine classes for segmenting the polarimetric image, and picking training and testing data for the classification procedure. The thesis focuses on multivariate Gaussian classifiers engaging the parameters associated with the components of the polarimetric decomposition theorems. There are two main goals of the project, the first is to provide a ranking to which polarimetric decomposition theorem is the best fit to describe this kind of landscape and the second is to find an optimal subset of the polarimetric features. It is shown that using compositions of polarimetric features from the decomposition theorems increases accuracy significantly compared to a classification based on intensities. Methods are first used on one test site to find an optimal composition of features, then the same features are used on another test site to prove that the composition will be effective on another site as well.

Acknowledgements

I would like to thank my supervisors through the pilot study and the main project, Camilla Brekke and Anthony Doulgeris. Thank you both for all the meetings, for reading through my work so many times and for giving me help and advise. Also, thanks to Stian Anfinsen for the work with finding proper data for the project.

List of Tables

List of Figures

Contents

Chapter 1

Introduction

1.1 Motivation

There is a big interest for radar in surveillance of areas of tropical forest. In the Amazon forest, deforestation is caused by human settlement and development of the land. The authorities wish to prevent and surveillance the deforestation in order to protect biodiversity and prevent climate change. In 2008 Brazil created the Amazon fund. The Norwegian government has committed to contributing with 6 billion NOK in the period until 2015 [27], [28]. Forest inventory is neither cost effective or frequently available. Remote sensing of these areas provide a more frequent source of information about the deforestation. A radar transmits microwaves towards the target and measures the reflected radiation. The polarimetric decomposition theorems extract physical information from the radar measurements. As opposed to optical instruments the radar is not dependant on daylight and it also has the advantage that it can penetrate cloud cover. Hence it is more frequently available than the measurements from the optical instruments. Because the SAR data is available to a bigger extent than the optical instruments it is desirable to be able to extract as much and precise information as possible from the SAR measurements. In this study high resolution optical images are used as groundtruth and are compared with SAR data from the same region.

1.2 Aim of project

There are two main goals of this project. The first is to determine which decomposition theorem is best fit to use on tropical forest. The other goal is to find the most favourable subset of polarimetric features from the decom-

position theorems. The features are evaluated on the basis of an accuracy measure of a classification. The main contribution in this study is an evaluation of how well the polarimetric decomposition theorems perform in a classification with the components of each theorem as features, and then to recommend a composition of polarimetric features provided by the different decomposition theorems to apply on tropical forest.

1.3 Structure of thesis

This thesis starts with a theoretical part. Chapter 2 discusses the SAR theory, imaging, speckle and polarimetry. Chapter 3 explains and introduces the polarimetric SAR decomposition theorems. Chapter 4 discusses the methods used in the thesis, i.e. choice of ground truth areas and classes, probability theory, classifiers, accuracy measures and feature selection methods. Chapter 5 presents the data used in the thesis. Chapter 6 is the result chapter. Chapter 7 concludes the thesis.

1.4 Earlier work

There has been done some work on land cover classification based on SAR data for various purposes [31–42]. Some of that work has been done on areas covered by tropical forest [33, 42].

In [33] it is performed classification of a site in the Amazon forest in order to map deforestation. This study focused on distinguishing forest from non-forested areas. The report concluded that when corrected for shadowing caused by height differences and applying data obtained under representative conditions, SAR measurements could be used for distinguishing forest and non-forested areas in the Amazon. In [42] polarimetric decomposition theorems are engaged in land cover classification in a site of tropical rain forest in the Brazilian Amazon. The purpose of this study was to investigate the potential for P-band for these applications. The study concluded that the data had better separability for older forest than other classes.

Land cover classification has been performed on urban areas with some surrounding land in [31] to evaluate the potential of data from an instrument. There has been performed land cover classification on mixtures of cultivated land, not cultivated land, water and urban areas in regions that are not tropical forest in [32, 35] for the same purpose.

This study will engage more polarimetric features and more combinations of the polarimetric features on another site of tropical forest. And as opposed to [33, 42] the aim is to study the performance of the polarimetric features.

Chapter 2

SAR theory

A synthetic aperture radar is a radar that transmits microwaves towards a target and measures the phase, polarization and intensity of the backscattered waves. The SAR may transmit and receive waves with vertical and horizontal linear polarization. A fully polarimetric SAR system has all 4 channels HH, HV, VH and VV. Information about the transmitted and backscattered wave can be used to retrieve information about the target. The main data format for describing the backscatter is the scattering matrix given in equation 2.1, the covariance matrix given in equation 2.2 and the coherency matrix given in equation 3.3. S_i is the scattering matrix for pixel i reshaped as a 3×1 vector where it is assumed that $S_{hv} = S_{vh}$, $S_i = [S_{hh} \quad \sqrt{2}S_{hv} \quad S_{vv}]^T$. This is called the reciprocity assumption. The elements given in equation 2.1 is the scatter where the polarization has gone from horizontal to horizontal, horizontal to vertical, vertical to horizontal and vertical to vertical.

$$\mathbf{S}_{2\times 2} = \begin{pmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{pmatrix} \tag{2.1}$$

$\mathbf{C}_{3\times 3,j}$ in equation 2.2 is the single look covariance matrix. Note that $^*T$ denotes the complex conjugate transpose. $\mathbf{C}_{3\times 3,j}$ is preferably multilooked, see equation 2.5.

$$\mathbf{C}_{3\times 3,j} = S_i S_i^{*T} = \begin{pmatrix} S_{hh}S_{hh}^* & \sqrt{2}S_{hh}S_{hv}^* & S_{hh}S_{vv}^* \\ \sqrt{2}S_{hv}S_{hh}^* & 2S_{hv}S_{hv}^* & \sqrt{2}S_{hv}S_{vv}^* \\ S_{vv}S_{hh}^* & \sqrt{2}S_{vv}S_{hv}^* & S_{vv}S_{vv}^* \end{pmatrix} \tag{2.2}$$

2.1 Imaging geometry

The geometry of the SAR system is given in figure 2.1. The flight direction of the satellite is called *azimuth* and the direction perpendicular to the flight direction when projected on to the ground is called *range*. The point directly below the SAR is called the *nadir*. The SAR system looks in the range direction. The radar travels in the azimuth direction and has objects inside the beam for a time ΔT. By having objects inside the beam for a longer time the system synthezises a larger antenna which gives a higher resolution in the azimuth direction. The resolution of the system in the range direction is given by equation 2.3 and in the azimuth direction is given by equation 2.4 [8]. In equation 2.3, c denotes the speed of light, τ is the pulse length and β is the bandwidth. In equation 2.4, λ denotes the wavelength of the transmitted wave, R is the slant range distance to the target and the D_a is the length of the antenna. And by advanced signal processing techniques the system synthezise a larger dimension of the antenna in the azimuth direction [17]. The incidence angle is denoted as θ_i in figure 2.1.

$$\Delta R = \frac{c\tau}{2} = \frac{c}{2\beta} \tag{2.3}$$

$$\Delta A = \frac{\lambda R}{D_a} \tag{2.4}$$

2.2 Speckle statistics

Inside each resolution cell there are many scatterers that each contribute to the signal received by the sensor. The signal received at the sensor is a sum of the contribution from all scatterers inside the resolution cell. Thus, each scatterer is a vector in the real-imaginary plane. This is illustrated in figure 2.2. Constructive and destructive interference between the scattering mechanisms inside the resolution cell can give speckle in an image. Speckle appears as bright and dark dots in an image. Reducing speckle can be done by performing multilooking.

2.3 Multilooking

Multilooking is performed to reduce speckle in an image. It is done by averaging over a number of pixels. The cost of reducing the speckle is loss of resolution. In equation 2.5, L is the number of looks.

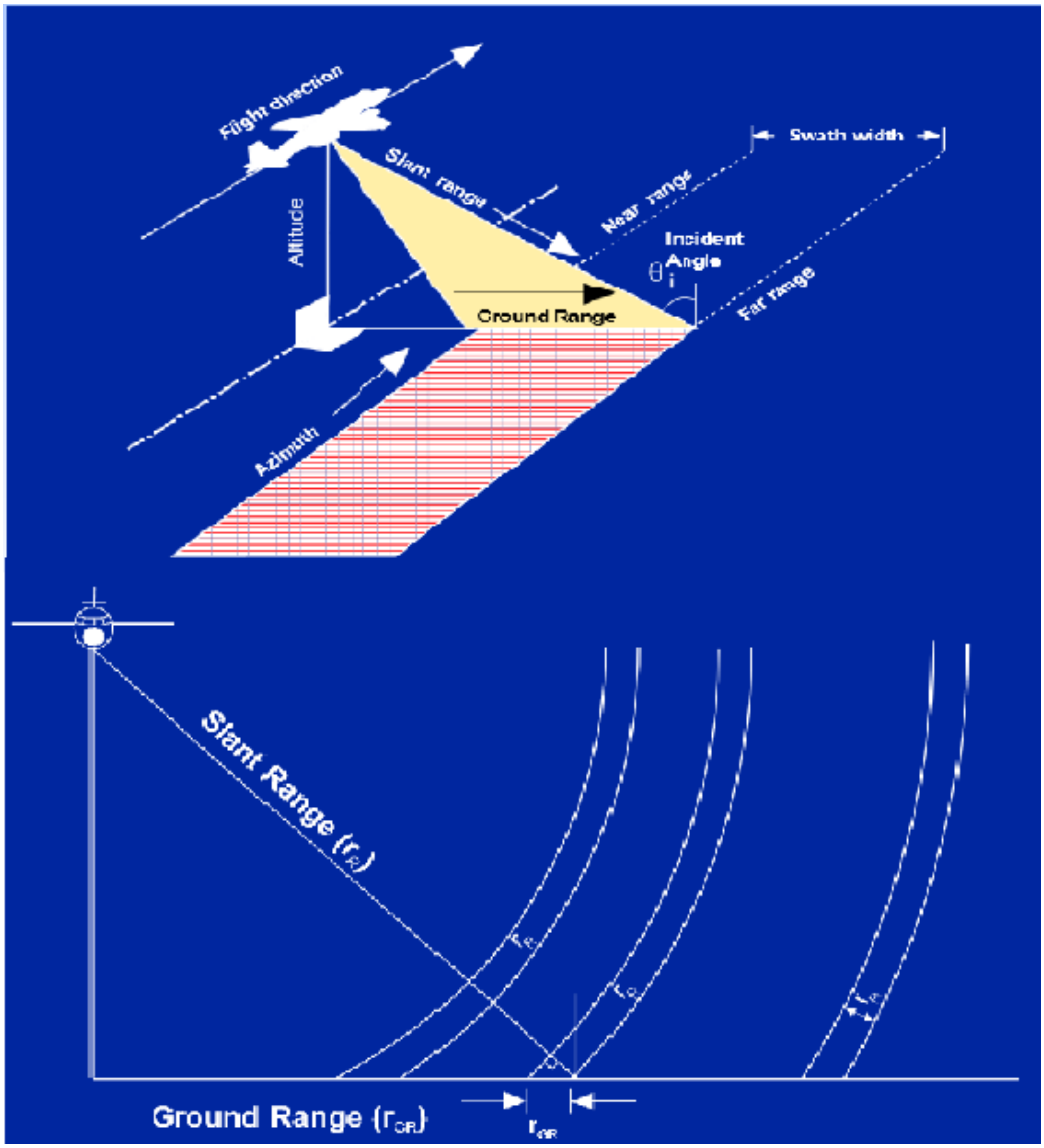


Figure 2.1: SAR imaging geometry, figures from [6]

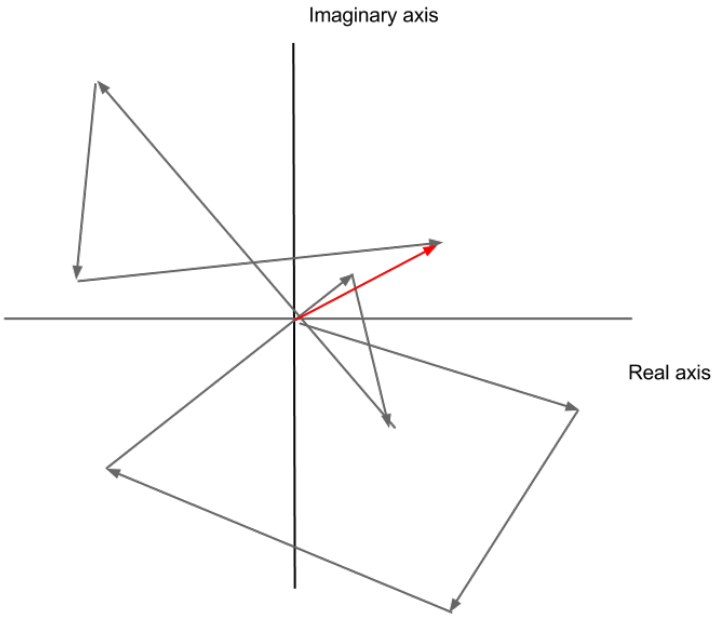


Figure 2.2: Figure illustrating destructive interference.

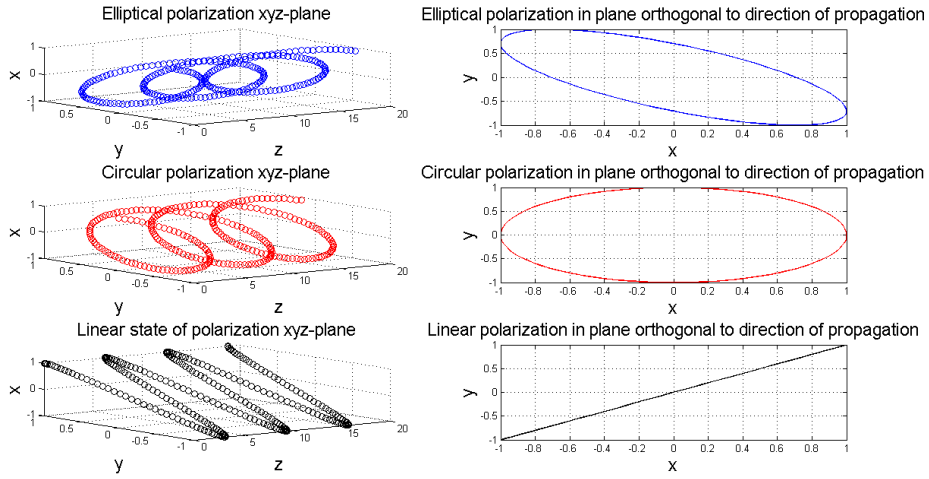


Figure 2.3: Figure describing the state of polarization of an electromagnetic wave. Note that $E_{0x} = E_{0y}$ in this plot.

$$\mathbf{C}_{3\times3} = \frac{1}{L}\sum_{j=1}^{L}\mathbf{C}_{3\times3,j} \tag{2.5}$$

2.4 Polarimetry

There are three possible states of polarization that an electromagnetic wave can have. Polarization can be either linear, circular or elliptical. The state of polarization describes the oscillation in the plane orthogonal to the direction the wave is propagating in. The propagation of an electromagnetic wave can be fully described by equation 2.6. The z denotes the position, t denotes time, E_{0x}, E_{0y} denotes the amplitudes of the wave in the x and y direction, k is the wavenumber, δ_x, δ_y is the phase shift in the respective direction, $\hat{x}$ and $\hat{y}$ is the unit vector in the x and y direction and ω is the angular frequency. Figure 2.3 illustrates waves with the three different types of polarization. The direction of propagation is the z-direction. The plots on the right side is the waves motion in the plane orthogonal to the direction of propagation.

$$\mathbf{E}(z,t) = \Re\{E_{0x}e^{j(kz-\omega t+\delta_x)}\}\hat{x} + \Re\{E_{0y}e^{j(kz-\omega t+\delta_y)}\}\hat{y} \tag{2.6}$$

The SAR system transmits waves that have a linear state of polarization. It can have one, two or four of the transmit-receive linear polarization combinations given in equation 2.1. The case discussed in this study has all four

combinations. Thus all resolution elements in the image has a value for all four elements in equation 2.1.

Chapter 3

Polarimetric decomposition theorems

The polarimetric decomposition theorems project the matrices that describe the backscattering, as given in equations 2.1,2.2 and 3.3 on to a set of basis matrices and express the backscatter as a linear sum of the basis matrices multiplied with corresponding coefficients. Thus, the scattering of each resolution cell is expressed as a linear sum of scattering mechanisms [5, 16]. In order to understand the decompositions it is necessary to discuss the interaction between the transmitted waves and the targets.

3.1 Backscattered waves

A transmitted wave interacting with a medium can be absorbed, transmitted or reflected [7]. The SAR system measures reflected radiation.
When a transmitted wave interacts with a bare surface the reflected wave will depend on the the roughness of the surface in comparison to the wavelength. Rayleigh and Fraunhofer criterions describe what is regarded as smooth or rough surfaces [29]. A smooth surface will typically give specular reflection. Specular reflection is strong reflection in one direction away from the surface. Rough surfaces will tend to give diffuse backscatter. Diffuse backscatter is scattering with uniform strength in all directions. Rough surfaces generally give stronger backscatter than smooth surfaces [8]. For bare soil the signal to the S_{hh} and the S_{vv} channels would be approximately equal [9]. The strength can also be affected by a slope in the terrain and dielectric properties of the surface. When a transmitted wave is reflected by the canopy of a forest the scatter will be depolarized such that it gives contribution to the S_{hv} and S_{vh} channels [9]. This means that scattering has bounced more

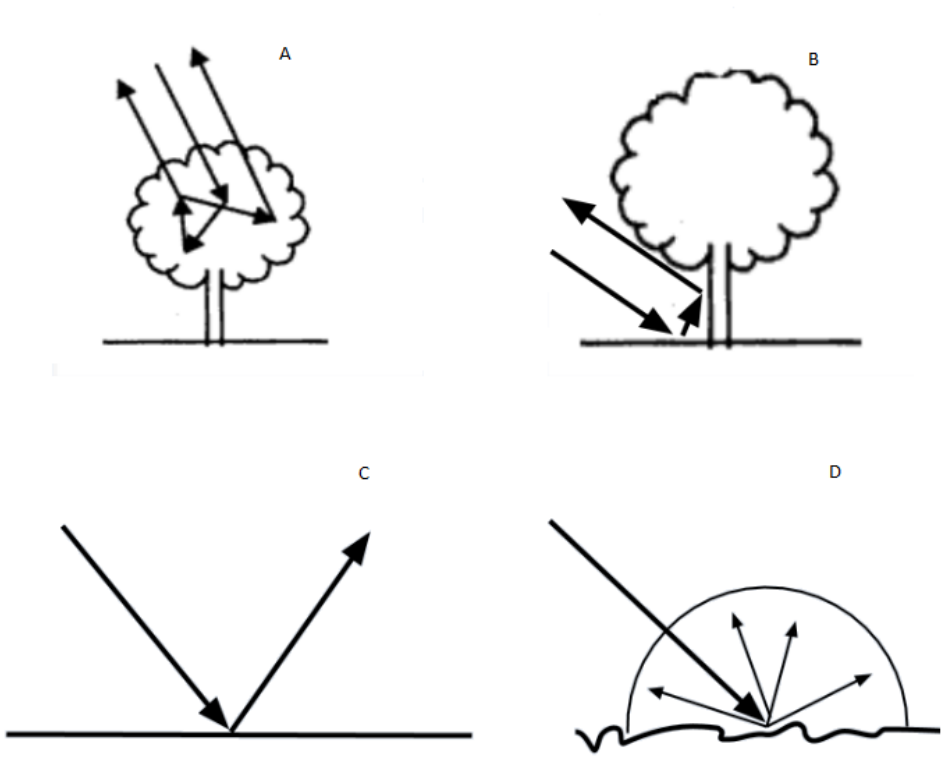


Figure 3.1: A describes volume scattering from forest canopy. B describes double bounce scattering from the ground and a tree trunk. C describes specular reflection from a smooth surface. D describes diffuse scattering from a rough surface. Figure based on [30].

than twice. Presence of dihedral corner reflectors will give a stronger double bounce component. This could be man made buildings, tree trunks or any other geometric formation that could cause the wave to reflect two times and back towards the sensor. In a forested environment this type of scatter is typically caused by tree trunks. Double bounce will give strong response in the S_{hh} and weaker response in the S_{vv} [9]. Figure 3.1 illustrates possible scattering mechanisms in a forest.

3.2 Coherent decomposition theorems

The coherent decomposition theorems aim to describe the scattering matrix in equation 2.1 as a sum of scattering mechanisms inside a resolution cell [10]. These theorems are better fit to describe areas dominated by man made

structures. The Pauli decomposition will be the only one of the coherent decomposition theorems to be discussed in this study as the other theorems are regarded to be less fit to describe the type of terrain described in the data section.

3.2.1 Pauli's decomposition theorem

Pauli's decomposition theorem aim to describe the $\mathbf{S}_{2\times2}$ matrix as a sum of three scattering mechanisms. The decomposition is described by equation 3.1.

$$\mathbf{S}_{2\times2} = \alpha \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \beta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \gamma \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.1}$$

Note that each pixel in an image has a value for each of the coefficients. The α parameter describes the odd or single bounce as discussed in section 3.1. Bare soil gives approximately equal contribution to S_{hh} and S_{vv}. Single or odd bounce typically origin from an open surface. The β parameter describes double or even bounce. This type of scatter gives a stronger contribution to the S_{hh} channel. This kind of scatter could origin from for instance interaction between the ground and walls on buildings or tree trunks. The γ parameter in the Pauli decomposition would be interpreted as volume scatter, i.e scatter from canopy. An RGB-composite image is made by putting the squared values of the β parameter in the red channel, the γ parameter in the green channel and the α parameter in the blue channel. As an example there are RGB composite images of the 6990 and 7000 site shown in figure 5.3. Coefficients α, β and γ are given in equation 3.2.

$$\alpha = \frac{S_{hh} + S_{vv}}{\sqrt{2}}, \beta = \frac{S_{hh} - S_{vv}}{\sqrt{2}}, \gamma = \sqrt{2}S_{hv} \tag{3.2}$$

3.3 Incoherent decomposition theorems

The incoherent decomposition theorems aim to describe the scattering by the $\mathbf{C}_{3\times3}$ covariance matrix or the $\mathbf{T}_{3\times3}$ coherency matrix [10], [16]. The single look coherency matrix $\mathbf{T}_{3\times3,j}$ is defined by equation 3.3. $\mathbf{T}_{3\times3,j}$ is also multilooked.

$$\mathbf{T}_{3\times3,j} = \mathbf{k}\mathbf{k}^{*T}, \quad \mathbf{k} = \begin{bmatrix} S_{hh} + S_{vv} & S_{hh} - S_{vv} & S_{hv} + S_{vh} \end{bmatrix}^T \tag{3.3}$$

So the incoherent decomposition theorems can be expressed as a sum of components of the $\mathbf{C}_{3\times3}$ or the $\mathbf{T}_{3\times3}$ matrices, see equation 3.4.

$$\mathbf{C}_{3\times3} = \sum_{i=1}^{N} q_i C_i, \quad \mathbf{T}_{3\times3} = \sum_{j=1}^{M} q_j T_j \tag{3.4}$$

C_i, T_j are the basis matrices and q_i, q_j are the coefficients of the respective basis matrices.

3.3.1 Model based decomposition theorems

A target decomposition theorem being model based means that the components are justified by a physical interpretation. The drawback is that the parameters may be harder to compute. The physical models for describing backscatter can be complicated.

Freeman Durden's 3 component decomposition theorem

Freeman Durden's 3 component decomposition theorem expresses the $\mathbf{C}_{3\times3}$ matrix as a sum of three components. The components are volume, double bounce and surface scatter [11]. Volume scatter is the multiple bounce depolarized backscatter described in section 3.1. The volume scatter is modeled as the backscatter from a cloud of very thin, randomly oriented, cylinder-like dipoles. The scattering matrix for a dipole is given by equation 3.5.

$$\mathbf{S}_{2\times2,dipole} = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix}, \alpha_1 >> \alpha_2 \tag{3.5}$$

α_1, α_2 are complex scattering coefficients. If the dipole is under a rotation of an angle θ around the radar line of sight the scattering matrix becomes as in equation 3.6.

$$\mathbf{S}_{2\times2,cyl}(\theta) = \begin{pmatrix} \alpha_1 cos^2(\theta) + \alpha_2 sin^2(\theta) & (\alpha_2 - \alpha_1)cos(\theta)sin(\theta) \\ (\alpha_2 - \alpha_1)cos(\theta)sin(\theta) & \alpha_1 sin^2(\theta) + \alpha_2 cos^2(\theta) \end{pmatrix} \tag{3.6}$$

The assumption that the mechanism is a very thin horizontal scatterer means that $\alpha_2 \to 0$. Freeman-Durden assumed the probability distribution for the orientation to be uniform, i.e. $\Theta \sim \mathcal{U}(-\pi, \pi)$. Given these assumptions the mean-angle average becomes the matrix given in equation 3.7 [11].

$$\mathbf{C}_{3\times3,VOL} = \frac{f_{VOL}}{8} \begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix} \tag{3.7}$$

The double bounce component is modeled as the contribution from two orthogonal surfaces that may have different dielectric properties, for instance the ground and a tree trunk. In this study it is assumed that the surfaces are in fact the ground and a tree trunk. The tree trunks horizontal and vertical reflection coefficients are given the symbols R_{TH} and R_{TV}. The grounds reflection coefficients are denoted as R_{GH} and R_{GV}. The factors $e^{2j\gamma_h}$ and $e^{2j\gamma_v}$ are propagation attenuation factors. The scattering matrix for this component will be given as in equation 3.8.

$$\mathbf{S}_{2\times 2,DB} = \begin{pmatrix} e^{2j\gamma_h} R_{TH} R_{GH} & 0 \\ 0 & e^{2j\gamma_v} R_{TV} R_{GV} \end{pmatrix} \tag{3.8}$$

This gives the basis matrix given in equations 3.9,3.10, with $\alpha = e^{2j(\gamma_h - \gamma_v)} \frac{R_{TH} R_{GH}}{R_{TV} R_{GV}}$ and $f_{DB} = |R_{TV} R_{GV}|^2$.

$$\mathbf{C}_{3\times 3,DB} = \begin{pmatrix} |R_{TH} R_{GH}|^2 & 0 & e^{2j(\gamma_h - \gamma_v)} R_{TH} R_{GH} R_{TV}^* R_{GV}^* \\ 0 & 0 & 0 \\ e^{2j(\gamma_v - \gamma_h)} R_{TV} R_{GV} R_{TH}^* R_{GH}^* & 0 & |R_{TV} R_{GV}|^2 \end{pmatrix} \tag{3.9}$$

$$= f_{DB} \begin{pmatrix} |\alpha|^2 & 0 & \alpha \\ 0 & 0 & 0 \\ \alpha^* & 0 & 1 \end{pmatrix} \tag{3.10}$$

The surface component is modeled as odd bounce scattering from bare soil. The $\mathbf{S}_{2\times 2}$ matrix for this component is given by equation 3.11. R_H, R_V are the dielectric constants for the ground.

$$\mathbf{S}_{2\times 2,S} = \begin{pmatrix} |R_H| & 0 \\ 0 & |R_V| \end{pmatrix} \tag{3.11}$$

The contribution to the $\mathbf{C}_{3\times 3}$ matrix becomes the expression in equation 3.12. $\beta = \frac{R_H}{R_V}$, $f_S = |R_V|^2$.

$$\mathbf{C}_{3\times 3,S} = \begin{pmatrix} |R_H|^2 & 0 & R_H R_V^* \\ 0 & 0 & 0 \\ R_V R_H^* & 0 & |R_V|^2 \end{pmatrix} = f_S \begin{pmatrix} |\beta|^2 & 0 & \beta \\ 0 & 0 & 0 \\ \beta^* & 0 & 1 \end{pmatrix} \tag{3.12}$$

The $\mathbf{C}_{3\times 3}$ matrix is expressed as a sum over the three discussed components, as in equation 3.13.

$$\mathbf{C}_{3\times 3} = \frac{f_{VOL}}{8} \begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix} + f_{DB} \begin{pmatrix} |\alpha|^2 & 0 & \alpha \\ 0 & 0 & 0 \\ \alpha^* & 0 & 1 \end{pmatrix} + f_S \begin{pmatrix} |\beta|^2 & 0 & \beta \\ 0 & 0 & 0 \\ \beta^* & 0 & 1 \end{pmatrix} \tag{3.13}$$

An RGB composite image would be made by putting the double bounce component in the red channel, the volume component in the green channel and the surface component in the blue channel.

Yamaguchi's 4 component decomposition theorem

Yamaguchi's 4 component decomposition theorem expresses the $\mathbf{C}_{3\times3}$ matrix as a sum of volume scattering, helix scattering, double bounce scattering and odd bounce scattering [12]. The surface and double bounce components are modeled in the same way as in Freeman-Durdens theorem [11] [12].
Yamaguchi assumes a different probability distribution function than Freeman-Durden for the orientation of the cylinders representing the volume scatter. The probability distribution function proposed in Yamaguchi's theorem is given in equation 3.14. Yamaguchi argue that because the vertical structure is more dominant in a forest, equation 3.14 describes the forest case better [12].

$$p_\theta(\theta) = \begin{cases} \frac{1}{2}cos(\theta), & |\theta| < \frac{\pi}{2} \\ 0, \text{otherwise} \end{cases} \tag{3.14}$$

Assuming this distribution and equation 3.5 it is necessary to discuss two cases to which basis matrix should be used to describe the volume scatter [10]. The basis matrix representing the volume scatter could be either of the two given in equation 3.15. $C_{3\times3,VOL1}$ assumes a very thin horizontal cylinder-like scatterer, i.e. $\alpha_2 \to 0$. $C_{3\times3,VOL2}$ assumes a very thin vertical cylinder-like scatterer, i.e. $\alpha_1 \to 0$.

$$\mathbf{C}_{3\times3,VOL1} = \frac{f_{VOL}}{15}\begin{pmatrix} 3 & 0 & 2 \\ 0 & 4 & 0 \\ 2 & 0 & 8 \end{pmatrix}, \text{or } \mathbf{C}_{3\times3,VOL2} = \frac{f_{VOL}}{15}\begin{pmatrix} 8 & 0 & 2 \\ 0 & 4 & 0 \\ 2 & 0 & 8 \end{pmatrix} \tag{3.15}$$

Yamaguchi recommends that if the ratio $10log(|S_{vv}|^2/|S_{hh}|^2)$ is close to zero the volume component should be a weighted mean of the two components, if the ratio is >0 Yamaguchi claims that $\mathbf{C}_{3\times3,VOL1}$ should be applied and if the ratio is <0 it should be $\mathbf{C}_{3\times3,VOL2}$. The limit is $\pm2dB$.
The helix component was first introduced by Krogager [18]. It is meant to describe areas with sharp corners and edges, usually man made structures. The helix component is typically multiple scattering bouncing back and forth between tall buildings that eventually escape towards the satellite. The scattering will usually be specular, thus a relatively large proportion of

the scattering power remains despite the multiple bounces. The scattering matrices for a left sense and right sense helix are given in equation 3.16.

$$\mathbf{S}_{2\times2,LH} = \frac{1}{2}\begin{pmatrix} 1 & j \\ j & -1 \end{pmatrix}, \text{or } \mathbf{S}_{2\times2,RH} = \frac{1}{2}\begin{pmatrix} 1 & -j \\ -j & -1 \end{pmatrix} \tag{3.16}$$

The helix component is modeled by the basis matrices given in equation 3.17 [12], [18].

$$\mathbf{C}_{3\times3,LH} = \frac{f_C}{4}\begin{pmatrix} 1 & -j\sqrt{2} & -1 \\ j\sqrt{2} & 2 & -j\sqrt{2} \\ -1 & j\sqrt{2} & 1 \end{pmatrix}, \mathbf{C}_{3\times3,RH} = \frac{f_C}{4}\begin{pmatrix} 1 & j\sqrt{2} & -1 \\ -j\sqrt{2} & 2 & j\sqrt{2} \\ -1 & -j\sqrt{2} & 1 \end{pmatrix} \tag{3.17}$$

The covariance matrix is given by the sum over all these components, as in equation 3.18. Yamaguchi's theorem has two of the same components as the Freeman-Durden theorem, it has added the helix component to interpret scattering in areas of tall buildings and sharp corners and it has a modified probability density function for the orientation of the cylinders that represent the volume scattering. So it is fair to say that Yamaguchi's theorem is an extension to the Freeman-Durden theorem.

$$\mathbf{C}_{3\times3} = \mathbf{C}_{3\times3,S} + \mathbf{C}_{3\times3,DB} + \mathbf{C}_{3\times3,VOL} + \mathbf{C}_{3\times3,LH/RH} \tag{3.18}$$

Freeman's 2 component decomposition theorem

Freeman's 2 component decomposition theorem was introduced in 2007, posterior to the three component model(1998). This theorem assumes that the $\mathbf{C}_{3\times3}$ matrix is a sum of a surface component and a volume component [13]. The basis matrices are given in equation 3.19.

$$\mathbf{C}_{3\times3,VOL} = f_{VOL}\begin{pmatrix} 1 & 0 & \rho \\ 0 & 1-\rho & 0 \\ \rho^* & 0 & 1 \end{pmatrix}, \quad \mathbf{C}_{3\times3,SURFACE} = f_{SURFACE}\begin{pmatrix} 1 & 0 & \alpha \\ 0 & 0 & 0 \\ \alpha^* & 0 & |\alpha|^2 \end{pmatrix} \tag{3.19}$$

$\mathbf{C}_{3\times3,VOL}$ is the covariance matrix for the volume scattering from a reciprocal medium. The reciprocity assumption is discussed in chapter 2. The parameters f_{VOL} and ρ describes the contribution from the volume scattering. $\mathbf{C}_{3\times3,SURFACE}$ is the covariance matrix describing either double bounce scattering or scattering from bare soil. The parameters $f_{SURFACE}$ and α describe the contribution from the surface component. In [13] Freeman argues that for $\mathbf{C}_{3\times3,SURFACE}$ to be considered as scattering from bare soil, the model requires that $|\alpha| < 1$ and that the argument of the complex coefficient

α is approximately equal to two times the phase difference(HH-VV) that models the propagation delay effects for the waves to travel from the radar to the scattering mechanism and back again. Freeman suggests that this delay could be caused by for instance the wave travelling through a canopy layer [13]. For $\mathbf{C}_{3\times3,SURFACE}$ to be considered as double bounce scattering it requires that $|\alpha| > 1$ and that $arg(\alpha) = \pm\pi$ [5] [13]. The $\mathbf{C}_{3\times3}$ matrix is expressed as the sum of these components, see equation 3.20. Each of them could be used to produce a grayscale image. Freeman argues that one of the biggest advantages with this decomposition theorem is that it requires less assumptions, as it only needs to solve four unknown parameters with four equations.

$$\mathbf{C}_{3\times3,TOT} = \mathbf{C}_{3\times3,VOL} + \mathbf{C}_{3\times3,SURFACE} = \begin{pmatrix} f_{SURFACE} + f_{VOL} & 0 & f_{SURFACE}\alpha + f_{VOL}\rho \\ 0 & f_{VOL}(1-\rho) & 0 \\ f_{SURFACE}\alpha^* + f_{VOL}\rho^* & 0 & f_{SURFACE}|\alpha|^2 + f_{VOL} \end{pmatrix} \tag{3.20}$$

The contribution provided from each of the mechanisms can be estimated by equation 3.21.

$$\mathbf{P}_{SURFACE} = f_{SURFACE}(1 + |\alpha|^2), \quad \mathbf{P}_{VOL} = f_{VOL}(3 - \rho) \tag{3.21}$$

Van Zyl's Nonnegative eigenvalue decomposition theorem(NNED)

Van Zyl et.al. argues that the Freeman-Durden and Yamaguchi's theorem overestimate the volume component and produce negative eigenvalues for the other components, which is a nonphysical result as it is not possible to have a negative presence of scattering mechanisms [14] [15]. The nonnegative eigenvalue decomposition guarantees that no eigenvalues are negative. The reason why the volume component is overestimated in Freeman-Durdens 3 component decomposition theorem is the assumption that neither the double bounce or the odd bounce components add to the cross polarized term $|S_{hv}|^2$, and that therefore the cross polarized term can be used to find the value of f_{VOL} [14]. The cross polarized term can be significantly affected by terrain slopes in the along track direction. The cross polarized power can also get contributions from a rough surface. Furthermore Freeman-Durden's theorem subtracts the volume component from the measured covariance matrix, and calculate the two remaining components. But the power in the remaining matrix after having subtracted the volume scatter from the measured covariance matrix, may be insufficient to guarantee that all eigenvalues are nonnegative.
Yamaguchi's procedure is to first subtract the helix component that has the

copol-crosspol correlation terms and a crosspol term, see equation 3.17. Afterwards f_{VOL} is calculated based on the remaining power in $|S_{hv}|^2$. Furthermore Yamaguchi follows the same procedure as Freeman-Durden calculating the two remaining components. Since Yamaguchi's theorem subtracts the helix component before calculating f_{VOL} it will not overestimate the volume component to the same extent as the Freeman-Durden decomposition theorem. However, it may still produce negative eigenvalues. In [14] and [15] the authors present results where the volume scatter is overestimated by 10-20% by Freeman-Durden's decomposition theorem.
The matrix $\mathbf{C}_{3\times3,VOL}$ has a predicted form based on a physical model of volume scattering. The restriction put on the value a is that it has to allow the eigenvalues of the $\mathbf{C}_{3\times3,remainder}$ matrix to be equal to or greater than zero [14] [15].

$$\mathbf{C}_{3\times3,remainder} = \mathbf{C}_{3\times3} - a\mathbf{C}_{3\times3,VOL} \tag{3.22}$$

The two remaining components are calculated from the matrix $\mathbf{C}_{3\times3,remainder}$. Hence the decomposition theorem can be written as a sum of four components, see equation 3.23. The last component, $\mathbf{C}_{DIFF}$ represents the remaining part of the matrix after the volume scattering, odd bounce scattering and double bounce scattering has been subtracted.

$$\mathbf{C}_{3\times3} = a\mathbf{C}_{3\times3,VOL} + \lambda_{ODD}\mathbf{C}_{3\times3,ODD} + \lambda_{DOUBLE}\mathbf{C}_{3\times3,DOUBLE} + \lambda_{DIFF}\mathbf{C}_{3\times3,DIFF} \tag{3.23}$$

Note that a is not necessarily given the biggest value that results in all eigenvectors being nonnegative. In [15] the authors propose that a should be given the value that gives the least power in the difference matrix. In both [14] and [15] the authors have pointed out errors and false assumptions with the Freeman-Durden and Yamaguchi's decomposition theorems, and based on that argued how their model is stronger. So it is fair to say that the nonnegative eigenvalue decomposition theorem is an extension to both Freeman-Durden's decomposition theorem and Yamaguchi's decomposition theorem.

3.3.2 Eigenvector based decomposition theorems

Some of the target decomposition theorems are based on eigenvector decomposition of the $\mathbf{C}_{3\times3}$ or the $\mathbf{T}_{3\times3}$ matrix. This ensures that the basis matrices are orthogonal. The $\mathbf{T}_{3\times3}$ can be expressed with the diagonal matrix $\mathbf{D}_{3\times3}$. This matrix contains the eigenvalues on the diagonal and $\mathbf{U}_{3\times3}$ contains the

eigenvectors. The $\mathbf{T}_{3\times3}$ can then be expressed as the product, as in equation 3.24 [10].

$$\mathbf{T}_{3\times3} = \mathbf{U}_{3\times3}\mathbf{D}_{3\times3}\mathbf{U}_{3\times3}^{-1} \tag{3.24}$$

This means that the $\mathbf{T}_{3\times3}$ can be expressed as a sum of the eigenvalues multiplied with the corresponding eigenvectors as in equation 3.25.

$$\mathbf{T}_{3\times3} = \sum_{i=1}^{3} \lambda_i \mathbf{u}_i {\mathbf{u}_i}^{*T} \tag{3.25}$$

H/A/$\overline{\alpha}$ decomposition theorem

The H/A/$\overline{\alpha}$ theorem is based on eigenvalue decomposition of the $\mathbf{T}_{3\times3}$ matrix. The theorem was introduced in 1997 [23]. An entropy based classification scheme was first introduced in 1995 [24]. The parameters in the theorem can be suitable for classification [19]. The $\mathbf{U}_{3\times3}$ unitary matrix from equation 3.24 is parametrized with 5 parameters in equation 3.26.

$$\mathbf{U}_{3\times3} = \begin{pmatrix} cos(\alpha_1)e^{j\phi_1} & cos(\alpha_2)e^{j\phi_2} & cos(\alpha_3)e^{j\phi_3} \\ sin(\alpha_1)cos(\beta_1)e^{j(\delta_1+\phi_1)} & sin(\alpha_2)cos(\beta_2)e^{j(\delta_2+\phi_2)} & sin(\alpha_3)cos(\beta_3)e^{j(\delta_3+\phi_3)} \\ sin(\alpha_1)sin(\beta_1)e^{j(\gamma_1+\phi_1)} & sin(\alpha_2)sin(\beta_2)e^{j(\gamma_2+\phi_2)} & sin(\alpha_3)sin(\beta_3)e^{j(\gamma_3+\phi_3)} \end{pmatrix} \tag{3.26}$$

The theorem defines pseudo probabilities based on the eigenvalues of $\mathbf{T}_{3\times3}$. The pseudo probability for class i is given in equation 3.27.

$$P_i = \frac{\lambda_i}{\sum_{k=1}^{3}\lambda_k} \tag{3.27}$$

The mean of the parameters in the unitary matrix are defined as in equation 3.28.

$$\overline{\alpha} = \sum_{k=1}^{3} P_k\alpha_k, \quad \overline{\beta} = \sum_{k=1}^{3} P_k\beta_k, \quad \overline{\delta} = \sum_{k=1}^{3} P_k\delta_k, \quad \overline{\gamma} = \sum_{k=1}^{3} P_k\gamma_k \tag{3.28}$$

The interpretation of the $\overline{\alpha}$ parameter is that it is depending on the average dominant scattering mechanism. Low values of $\overline{\alpha}$ is interpreted as single bounce scattering. For instance it takes the lowest values over oceans. Values of $\overline{\alpha}$ around 45^o indicates that there is some depolarization, could be for instance forested areas. When $\overline{\alpha}$ values approach 90^o it is interpreted as double bounce.
$\overline{\beta}$ is interpreted as the correlation between $|S_{hv}|^2$ and $|S_{vv}|^2$, $\overline{\delta}$ is the phase of

$(HH-VV)/(HH+VV)$, $\overline{\gamma}$ is interpreted as the phase of $HV/(HH+VV)$ [24] and ϕ is physically equivalent to the absolute target phase [5].
The polarimetric scattering entropy is defined by equation 3.29.

$$H = -\sum_{k=1}^{3} P_k log_3(P_k) \tag{3.29}$$

The interpretation of the entropy is that it is a measure to what extent the scattering is a random process. In the case when $H = 1$ it is a uniform distribution and the scattering is completely random. Lower values for H means that it is easier to extract information from the scattering. A higher value for H indicates that there are more than one scattering mechanisms and that they are equal in strength [5]. The last parameter in this decomposition theorem is the polarimetric scattering anisotropy given in equation 3.30. The interpretation of the anisotropy parameter is that it is the relative importance between the second and third eigenvector. When the anisotropy takes the highest possible value $A = 1$, it means that the the third scattering mechanism disappears completely. When it takes the other extreme $A = 0$, it means that the second and third scattering mechanism has equal power. For higher values of H the anisotropy becomes noisy as there are few scattering mechanisms contributing to the signal. For lower values of H the anisotropy could be a useful feature.

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \tag{3.30}$$

To summarize, the entropy and anisotropy parameters are interpreted as a level of dominance between the scattering mechanisms. The entropy is the level of dominance of the first scattering mechanism versus the two others, hence it is a useful parameter to determine whether there is one or more scattering mechanisms present. The anisotropy is the level of dominance between the second and third scattering mechanism, hence it is a useful parameter to determine whether there is two or three scattering mechanisms present.

3.3.3 Obtaining parameters of the decomposition theorems

In this project the polarimetric features in the decomposition theorems were computed from Polarimetric SAR Data Processing and Educational tool(PolSARpro). This tool has built-in functions to calculate the parameters of the various decomposition theorems [22].

Chapter 4

Methods

4.1 Choice of ground truth areas

The choice of ground truth data is based on visual inspection of the optical images, and recognizing areas in the optical data by shape and size in a Pauli image of the radar data. It was found that the radar data should be segmented into 6 different classes. The classes are labeled dense forest, medium dense forest, sparse forest, no forest, burned land and grassland.
The dense forest appears in the optical images with a more intense green color than the surrounding forest. In optical images this type is very distinguishable from other classes. There are visible areas of dense forest in the northwestern area of the 411 image, the northwestern area of the 511 image, in the middle of the 610 image and in the southwestern quarter of the 710 image, see figure 5.4.
The medium dense forest is visible in the optical images as green areas, but without the intense green color that stands out in the areas of dense forest. Among the forest types, the medium dense forest would be the most frequently occuring within the area covered by radar measurements. Most of the green areas in image 411 and the green areas in the center of image 511 should be regarded as medium dense forest.
The sparse forest are areas with a less intense green than the medium dense and dense forest. In image 411 the sparse forest is visible in bordering areas with the medium dense forest. In general it appears as bright green openings within or next to medium or dense forest.
The no forest areas appear in the optical images as areas with a desert sand color. These areas are most common in image 611 and also to a lesser extent in image 511.
The burned land appears in the optical images as blue or purple areas. In

image 710 to the top right there is a fire taking place. In several of the images it is possible to find areas of blue or purple with white smoke drifting away from it. This indicates that fires occur frequently in the area. The 611 image contains most of the burned land in these optical images.
The grassland class is distinguished with a bright green color and a different texture than other classes. In optical images this class is easily disinguishable from the other classes. In particular this class is seen in the top center of image 511 and in the center bottom of image 611.

4.2 Probability theory

In order to segment images and attribute a pixel to a given land cover class, like the ones describes in section 4.1, it is necessary to have some kind of measurement of the probability that it belongs to that particular class. It is common to assume that the data has a Gaussian distribution, or if there is more than one variable, a multivariate Gaussian distribution. SAR measurements does not necessarily produce data that follows the Gaussian distribution and in that case it is necessary to use some kind of transformation to make it fit the Gaussian distribution better.

4.2.1 Multivariate Gaussian distribution

The probability density function(pdf) of a multivariate Gaussian distribution with k features is given in equation 4.1.

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \tag{4.1}$$

The vector x with k values, is the datapoint for which the pdf is evaluated. μ is a $k \times 1$ vector containing the mean values for each of the features. Σ is a $k \times k$ matrix, and is the covariance matrix of the features. $|\Sigma|^{\frac{1}{2}}$ denotes the square root of the determinant of the covariance matrix.

4.2.2 Classifiers

A Bayesian classifier will take into account the value of the pdf for the given datapoint and the *a priori* probabilities. A higher value of a class' probability density function indicates a higher probability that the datapoint belongs to that class. The *a priori* probabilities is the probability that a point belongs

to a given class without any knowledge of the datapoint. Bayes' formula is given in equation 4.2.

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}, \quad \text{with} \quad p(x) = \sum_{j=1}^{N} p(x|\omega_j)P(\omega_j) \tag{4.2}$$

The symbol ω_j denotes class j, N is the number of classes and and $P(\omega_j)$ is the *a priori* probability of class j. A Bayesian classification will classify the pixel with the value x to the class j that maximize the value for $P(\omega_j|x)$ in equation 4.2. If the classifier is to make a choice between two classes m and n, the classification rule can be given as below.

$$\text{If } P(\omega_m|x) > P(\omega_n|x) \;\; x \text{ is classified to class } m \tag{4.3}$$
$$\text{If } P(\omega_m|x) < P(\omega_n|x) \;\; x \text{ is classified to class } n \tag{4.4}$$

This only describes the case when there are two classes, but the principle is the same when there are more than two classes. Values of $p(x)$ in equation 4.2 is equal for all classes and can be ignored. Assuming that the data follows a Gaussian distribution, the choice of class will be determined by the value in equation 4.5. This is called a *maximum likelihood classifier*.

$$p(\omega_j|x) = P(\omega_j)\frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma_j|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)} \tag{4.5}$$

4.2.3 Accuracy assessment

Confusion matrix

The confusion matrix is a table that gives the agreement between the classifier and the ground truth data. In this paper the columns will indicate the reference class and the rows indicate the choice of the classifier. The matrix has $d \times d$ elements where d is the number of classes, which in this study is 6. The elements on the diagonal are the amount of pixels that have been classified to the reference class. The form of the confusion matrix is given in table 4.1. From the confusion matrix it is possible to calculate various accuracy measurements. Producer accuracy of a given class is defined as the number of pixels correctly classified to that class class divided by the total number of pixels belonging to that class. The complementary error is called the omission error. User accuracy of a given class is defined as the number of pixels correctly classified to that class divided by the total number

Table 4.1: Confusion matrix

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total
Dense	$c_{1,1}$	$c_{1,2}$	$c_{1,3}$	$c_{1,4}$	$c_{1,5}$	$c_{1,6}$	$\sum_{k=1}^{6} c_{1,k}$
Medium	$c_{2,1}$	$c_{2,2}$	$c_{2,3}$	$c_{2,4}$	$c_{2,5}$	$c_{2,6}$	$\sum_{k=1}^{6} c_{2,k}$
Sparse	$c_{3,1}$	$c_{3,2}$	$c_{3,3}$	$c_{3,4}$	$c_{3,5}$	$c_{3,6}$	$\sum_{k=1}^{6} c_{3,k}$
Grass	$c_{4,1}$	$c_{4,2}$	$c_{4,3}$	$c_{4,4}$	$c_{4,5}$	$c_{4,6}$	$\sum_{k=1}^{6} c_{4,k}$
Burned	$c_{5,1}$	$c_{5,2}$	$c_{5,3}$	$c_{5,4}$	$c_{5,5}$	$c_{5,6}$	$\sum_{k=1}^{6} c_{5,k}$
No forest	$c_{6,1}$	$c_{6,2}$	$c_{6,3}$	$c_{6,4}$	$c_{6,5}$	$c_{6,6}$	$\sum_{k=1}^{6} c_{6,k}$
Total	$\sum_{k=1}^{6} c_{k,1}$	$\sum_{k=1}^{6} c_{k,2}$	$\sum_{k=1}^{6} c_{k,3}$	$\sum_{k=1}^{6} c_{k,4}$	$\sum_{k=1}^{6} c_{k,5}$	$\sum_{k=1}^{6} c_{k,6}$	

of pixels classified to the given class. The complementary error is called the comission error. Global accuracy is defined as sum of pixels correctly defined to their class divided by the total number of pixels in the testing data, see equation 4.6. A_P, A_U and A_G respectively denote producer accuracy, user accuracy and global accuracy. The symbol q here denotes the class label.

$$A_P = \frac{c_{q,q}}{\sum_{n=1}^{d} c_{q,n}}, A_U = \frac{c_{q,q}}{\sum_{n=1}^{d} c_{n,q}}, A_G = \frac{\sum_{i=1}^{d} c_{i,i}}{\sum_{i=1}^{d}\sum_{j=1}^{d} c_{i,j}} \tag{4.6}$$

4.2.4 Transformation of data

The expressions in equations 4.1 and 4.5 assume a Gaussian distribution. Radar backscattering in general does not follow a Gaussian distribution. For instance a Gaussian distribution allows infinitely negative values, and the backscatter can not take negative values. Figure 6.1 shows the histograms of the intensities from the 6990 site and their log transform. To fit a multivariate Gaussian pdf to the data it is necessary to perform a transformation to ensure that the distribution of the features look more like a Gaussian distribution. Log transformation of the data will make the histograms more symmetric around the peak. Figure 6.1 shows that the histogram of $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$ have a relatively steep slope for the low values and a long tail for the higher values. The log transformed values looks more like a mixture of Gaussian distribution, without the long tail and the steep slope. The log transform may also be necessary for some polarimetric features produced by the decomposition theorems.

4.3 Feature selection

In order to determine which features are the most useful for classification, it is interesting to see how the features perform together in terms of classi-

fication. One of the objectives of this study is to pick the most beneficial combination of the features in terms of classification. The selection is based on a seperability measure, i.e. a criterion to rank the performance of the various features. In this study it is the global accuracy defined in equation 4.6. Backward and forward sequential feature selection will be performed to find the optimal composition of features.

4.3.1 Backward sequential method

Given n features, the first step of the backward sequential method is to perform classification with all the available features. The second step is to perform classification with all subsets containing *n-1* features, i.e. all possible subsets where one of the features has been left out. The classification that achieves the best separability result of the subsets containing *n-1* features determines which combination of the features that is considered to be the best. In the following steps the feature that was not considered to be included in the best *n-1* subset is disregarded. The next step is to test all possible combinations of *n-2* features, only considering the features included the subset that was considered to be the best *n-1* combination. For each step, one feature is dropped and it is possible to obtain a combination of k features, with $k \leq n$. This is not an optimal method of finding the best combination of k features as there is no guarantee that the first features being dropped can not be included in the best k features. However it requires less computational efforts than evaluating all possible permutations [25, 26].

4.3.2 Forward sequential method

The forward sequential method starts with first evaluating classification with all features individually. The feature with the best separability result is considered to be the best feature. In the second step, classification with each of the remaining features paired with the one regarded as the best in the first step, is evaluated. The pair with the best separability measure is regarded as the best pair. In the next step, classification with each of the remaining features together with the two features that was regarded as the best in step 2, are evaluated. Step by step one feature is added to the combination regarded as the best, until it has reached a set of k features, with $k \leq n$. This is also a suboptimal selection of features as there is no guarantee that the subset of k features considered as the best set by this feature selection method is in fact the subset that gives the overall best separability result. However this also requires less computational efforts than evaluating the full set [25, 26].

Choice of separability measure

In this study, the global accuracy defined in equation 4.6 is used as separability measure in the feature selection. The main weakness with the global accuracy as separability measure is that it will favor the classes with many pixels as ground truth. However it is certainly the most popular measure of performance of classification, and because of that it is used in this study.

Chapter 5

Study area and data

5.1 Location

The polarimetric SAR data used in this study covers two sites that each form a rectangular area that has length 63,5 km and width of 26 km in Lindi county, Tanzania. The location of the sites is shown in figure 5.1. The sites have an overlap of 7,5 km. Optical data consists of 8 adjacent images. The area covered by radar measurements and the area covered by optical data have an overlap as shown in figure 5.2. All 8 boxes are square and have a side length of 25 km. The boxes have an overlap of 1 km.

5.2 Instruments

Polarimetric SAR data from both sites are obtained by the Advanced land observing satellites(ALOS) phased array type L-band synthetic aperture radar (PALSAR). The satellite operates at L-band with center frequency 1270 MHz [2], [3]. The instrument applies incident angle from 8^o to 30^o. The dataset consists of four polarimetric channels. The raw data of each site has a size of 18432 pixels in the azimuth direction and 1248 in the range direction. Both images are obtained at 07.12.2010. Figure 5.3 shows Pauli images of the test sites.
Optical images are captured by the RapidEye sensor. Image 410 is captured 20.06.2010, image 411 is captured 09.07.2010, image 510 is captured 20.06.2010, image 511 is captured 09.07.2010, image 610 is captured 10.07.2010, image 611 is captured 09.07.2010, image 710 is captured 10.07.2010 and image 711 is captured 20.06.2010. Each of the RapidEye's satellites carries the Jena Spaceborne Scanner JSS 56 which is a pushbroom sensor [4]. It does measurements in the 5 bands given in table 5.1. The sensor provides images

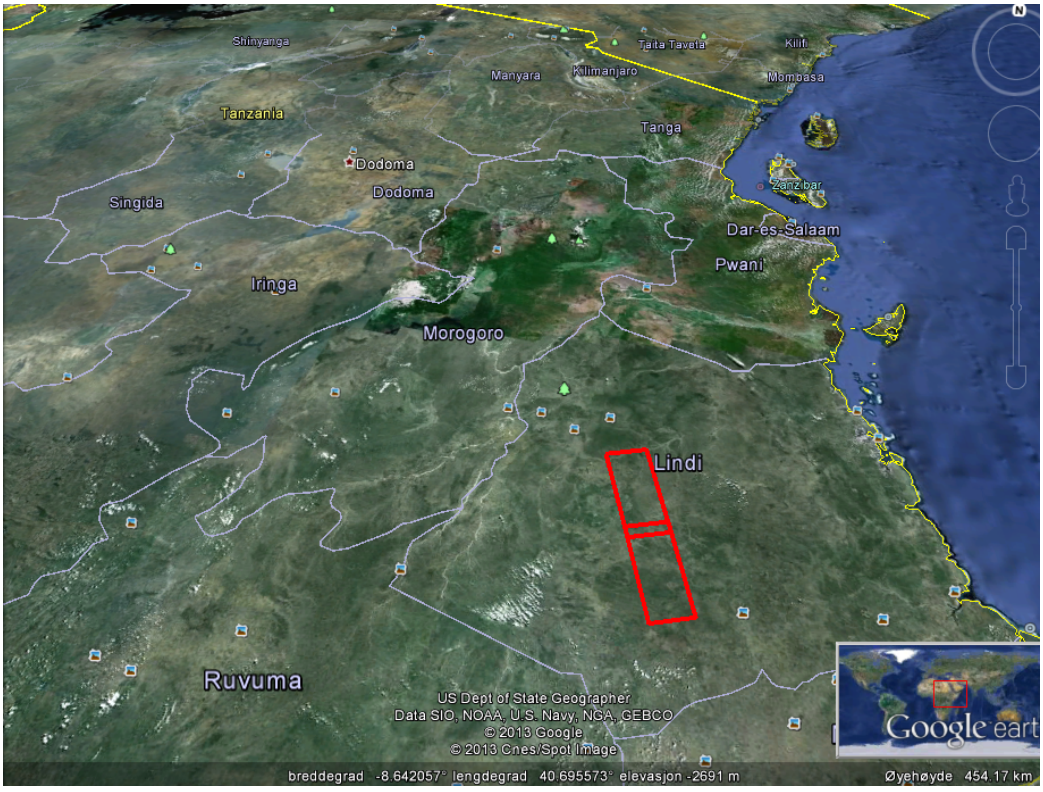


Figure 5.1: Geographical view of area covered by SAR data, Google earth [1]

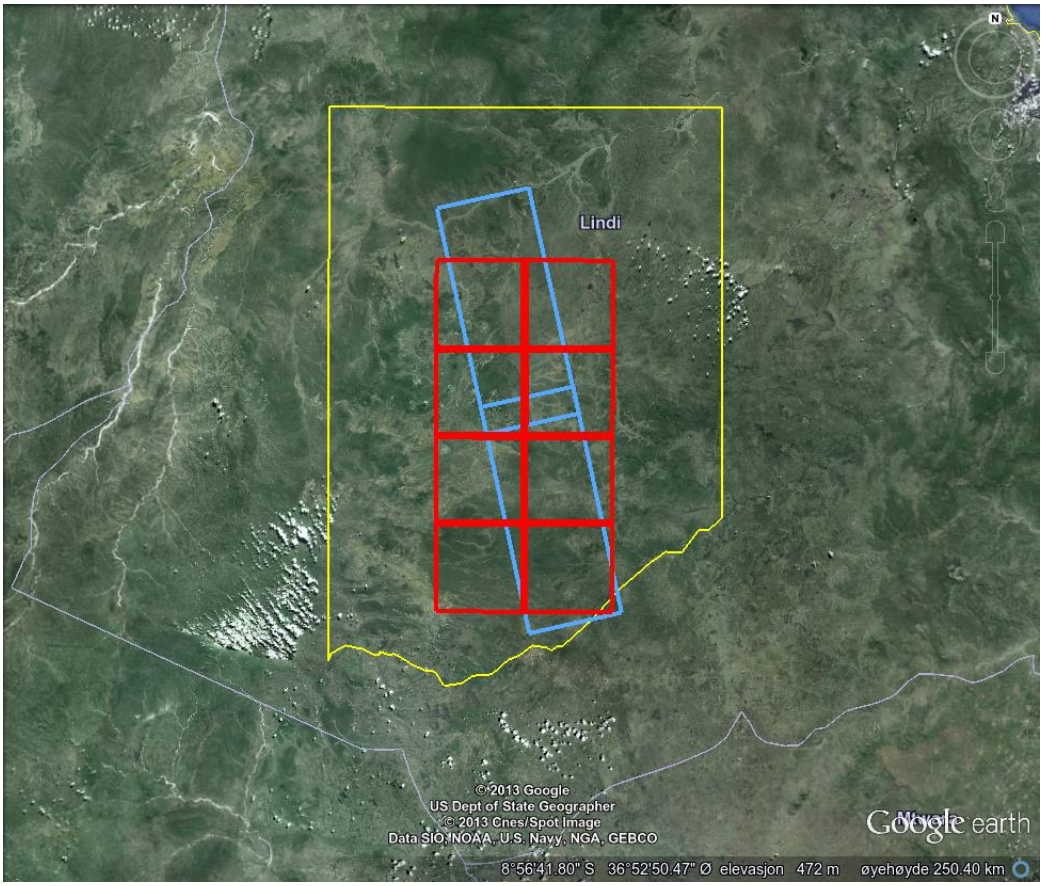


Figure 5.2: Area covered by optical data marked in red, area covered by SAR data marked in blue and Lindi county is marked in yellow, Google earth [1].

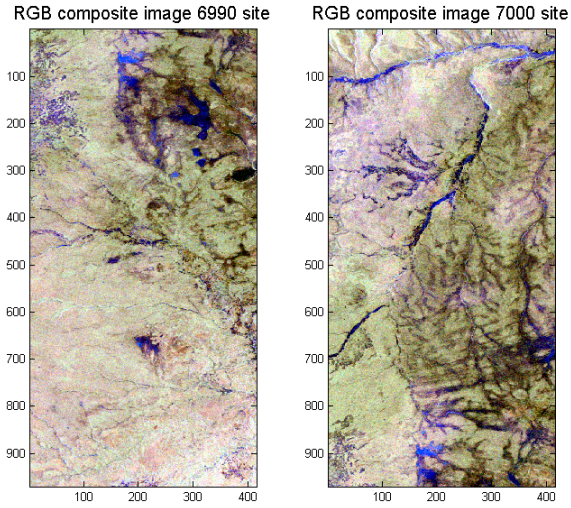


Figure 5.3: Pauli images of the polarimetric data. 6990 is south and 7000 is north.

Table 5.1: Wavelengths and bands of the Jena spaceborne scanner, [4]

Wavelength	Band
440 - 510 nm	Blue
520 - 590 nm	Green
630 - 690 nm	Red
690 - 730 nm	Red edge
760 - 880 nm	Near infrared

with pixel size 5 m [4]. RGB composite images of the optical data are shown in figure 5.4. Black areas in the optical images are areas of no information.

5.3 Processing of data

The polarimetric data has been multilooked with 19 pixels in azimuth direction and 3 pixels in range direction in order to reduce speckle and to make the axes of the image proportional to the actual length on the ground. This makes images on the format 970 times 416 which makes each pixel equivalent to a square box with side length 62,5 meters. Canopy of the vegetation in the area varies from 10-20 meters in diameter, so this multilooking should be sufficient. If a pixel would cover less area than what is covered by the canopy of a single tree, it would be hard to distinguish different forest classes.

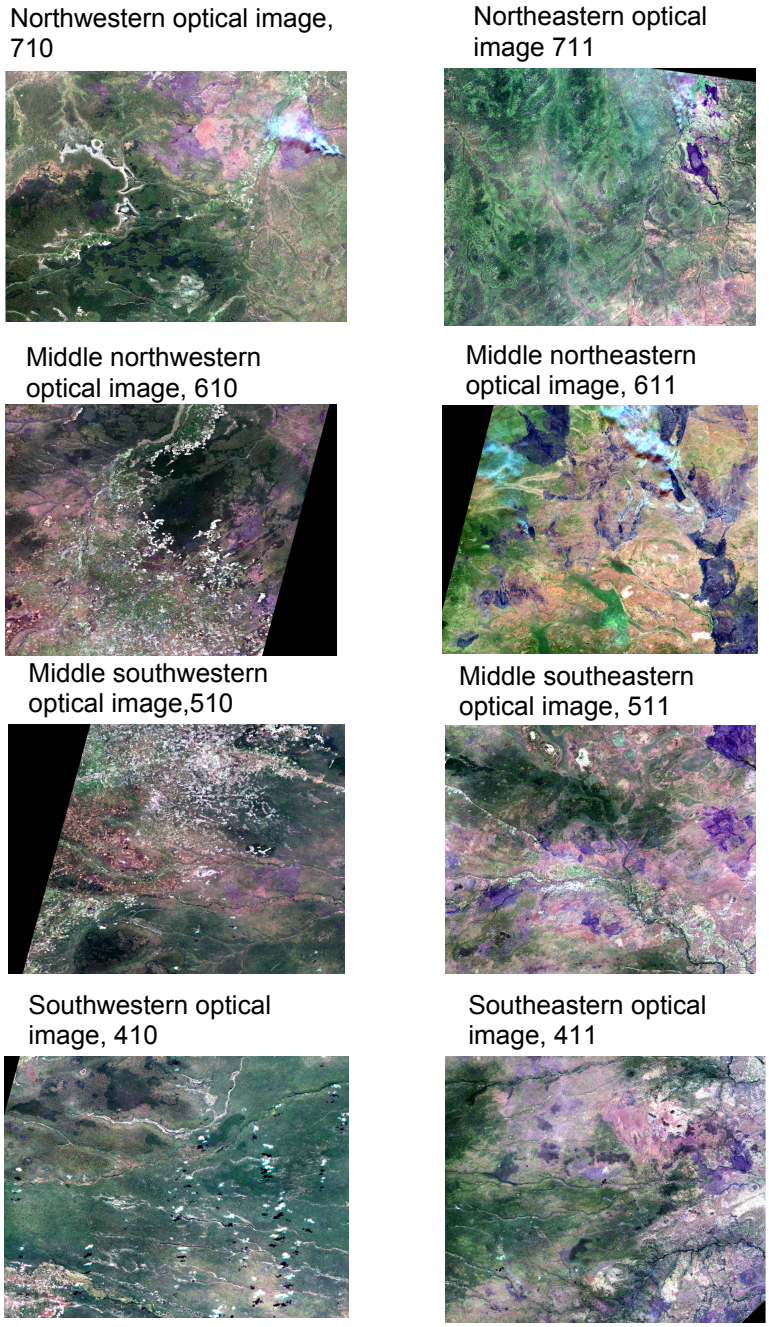


Figure 5.4: RGB composite images of the optical data.

Chapter 6

Results

6.1 Classification using theorems separately on the 6990 site

6.1.1 Intensities

The following results were obtained by using the log transformed intensities $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$ as features for classfication of the radar image. Classification based on intensities was performed because it was initially thought to produce similar results to the parameters from the Pauli decomposition theorem, see section 6.9. The histogram of the intensities and their log transform is shown in figure 6.1. Based on the shape of the histograms the log transformed intensities were considered to be the most appropriate data to use in the classifier, as it looked more like a mixture of Gaussians than the original data. For each of the ground truth classes, an algorithm was used to pick half of the pixels as training data and the other half as testing data. Random numbers determined which of the pixels were used as training and testing data. In the case where the ground truth class has an odd number of pixels, either the training or the testing data will have one pixel more. Classification was performed as described in section 4.2.2. Since the selection of ground truth pixels do not necessarily reflect the presence of the classes in the image, *a priori* probabilities were assumed to be equal, i.e. $P(\omega_k) = \frac{1}{6}, k = 1, \ldots, 6$ where k is the class label. Mean values and covariance matrices are calculated from the training data. Throughout the thesis, the same sets of pixels that constitute the training and testing data are applied. The classification map is given in figure 6.2 and the classification result is shown in table 6.1. The classfier achieved a global accuracy of 77.53%.

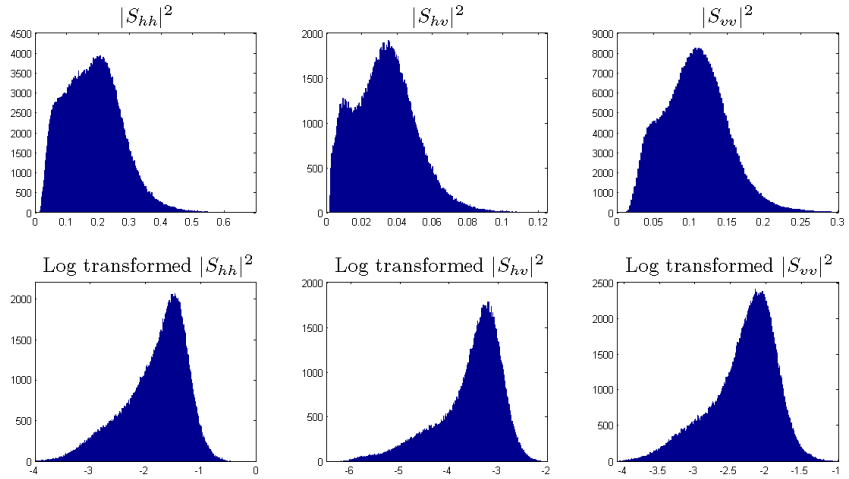


Figure 6.1: Histogram intensities with and without the log transformation, 6990 site.

Table 6.1: Classification results using a multivariate Gaussian classifier with the log transformed intensities $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$ as features.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	82	48	0	0	0	0	130	63.08%	36.92%
Medium	24	139	12	0	0	0	175	79.43%	20.57%
Sparse	0	20	51	0	0	6	77	66.23%	33.77%
Grass	0	0	0	146	18	0	164	89.02%	10.98%
Burned	0	0	0	23	53	7	83	63.86%	36.14%
No forest	0	0	2	0	4	95	101	94.06%	5.94%
Total	106	207	65	169	75	108			
Producer's accuracy	77.36%	67.15%	78.46%	86.39%	70.67%	87.96%			
Omission error	22.64%	32.85%	21.54%	13.61%	29.33%	12.04%			

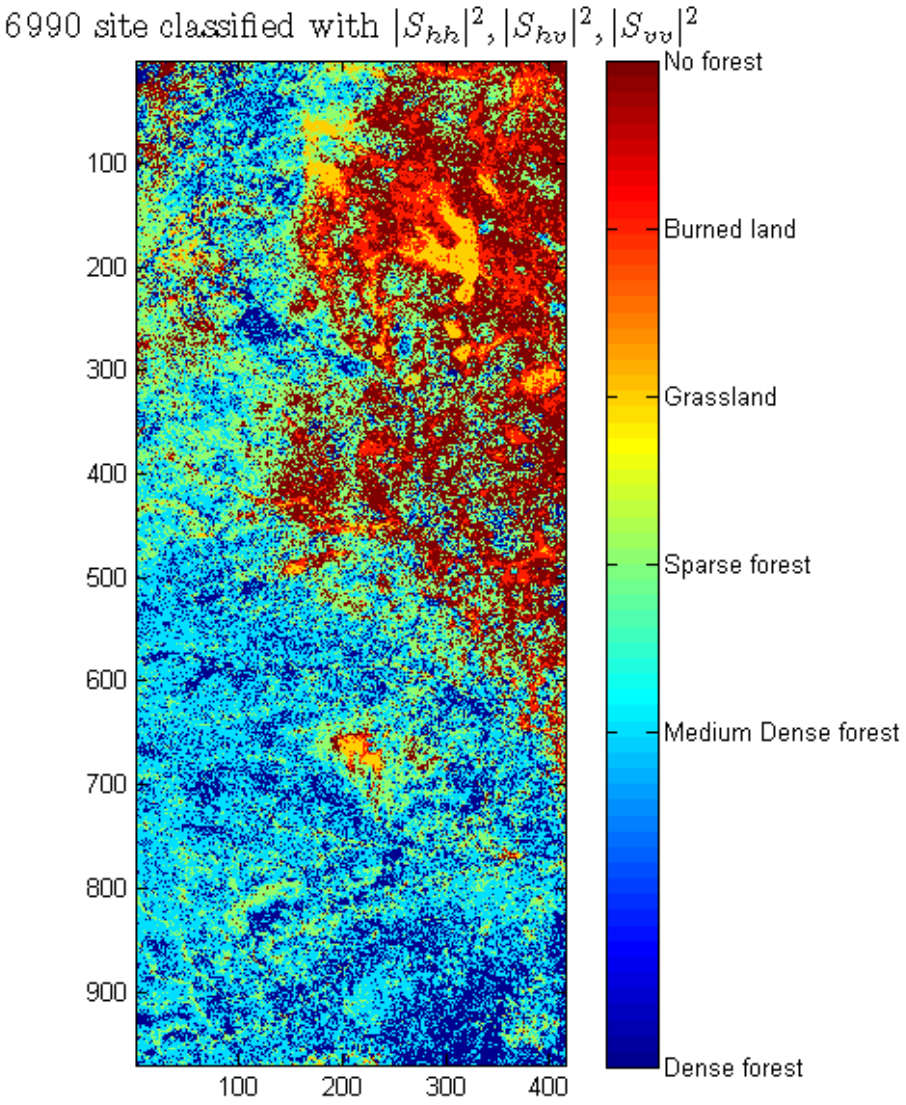


Figure 6.2: Classification map with log transformed intensities as features.

Table 6.2: Classification result using a multivariate Gaussian classifier with the log transformed parameters of Yamaguchi's theorem as features.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	77	42	0	0	0	0	119	64.71%	35.29%
Medium	28	150	12	0	0	0	190	78.95%	21.05%
Sparse	0	15	50	0	0	5	70	71.43%	28.57%
Grass	0	0	0	150	16	0	166	90.36%	9.64%
Burned	0	0	0	19	55	4	78	70.51%	29.49%
No forest	1	0	3	0	4	99	107	92.52%	7.48%
Total	106	207	65	169	75	108			
Producer's accuracy	72.64%	72.46%	76.92%	88.76%	73.33%	91.67%			
Omission error	27.36%	27.54%	23.08%	11.24%	26.67%	8.33%			

6.1.2 Yamaguchi's 4 component decomposition theorem

In this classification the log transformed values of the parameters of Yamaguchi's decomposition theorem were used as features. The log transform was applied because the histograms looked more like a mixture of Gaussians than the original data. The classification gave the result presented in table 6.2. This classfication achieved a global accuracy of 79.59% which is slightly better than the result the classifier based on the intensities achieved. Classification map is shown in figure 6.3.

6.1.3 Freeman's 2 component decomposition theorem

The Polsarpro software does a histogram squeezing when it calculates the parameters of the decomposition theorems in order to enhance the contrast in images. Values that are significantly lower or higher than the marjority of the pixels can cause images to become either very dark or very bright. The lowest values are replaced by a mean value, such that it creates a spike in the histograms. Unfortunately the replaced values appear in predictable regions of the image. They almost exclusively appear in the regions covered by grass, burned land or no forest. So the surface component is severely contaminated by squeezed values. The volume component does not have the same problem. So the following results were obtained by using only the log transformed values of the volume component from Freeman's 2 component decomposition theorem. The log transform was applied because the histogram looked more like a mixture of Gaussians after being transformed. The classification map is shown in figure 6.4 and the result is shown in table 6.3. The classification achieved a global accuracy of 50.96%.

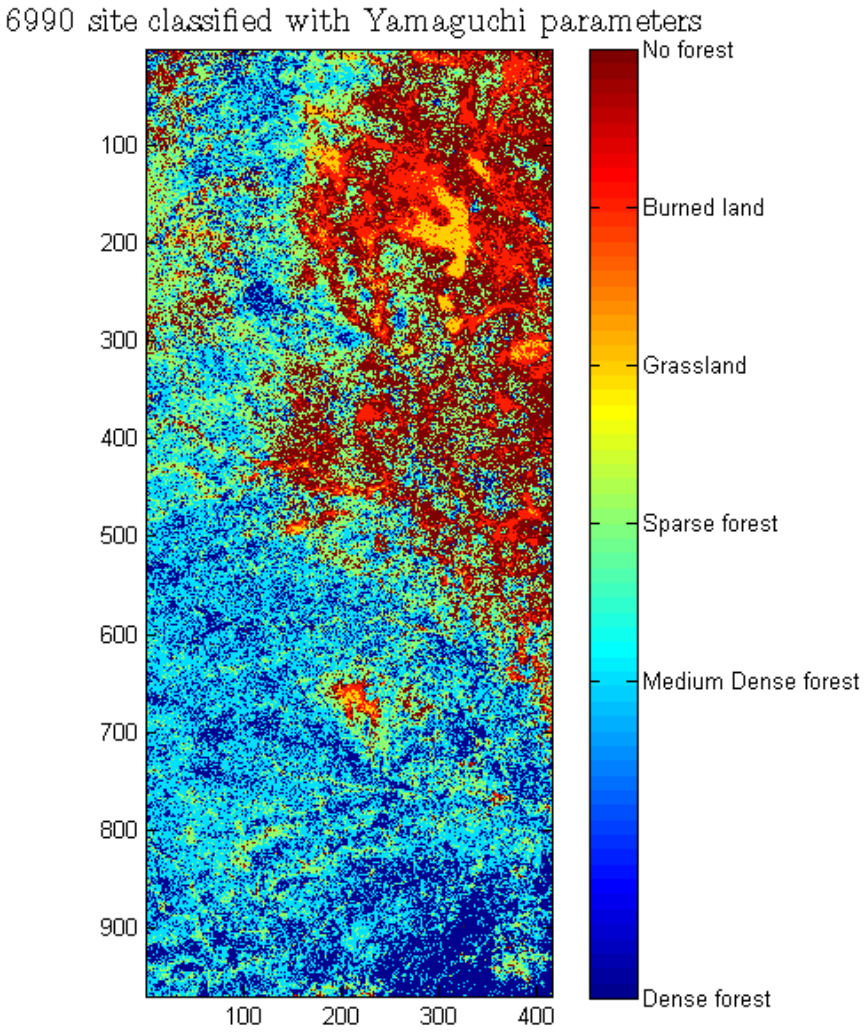


Figure 6.3: Classification map using the log transformed parameters from Yamaguchi's theorem as features.

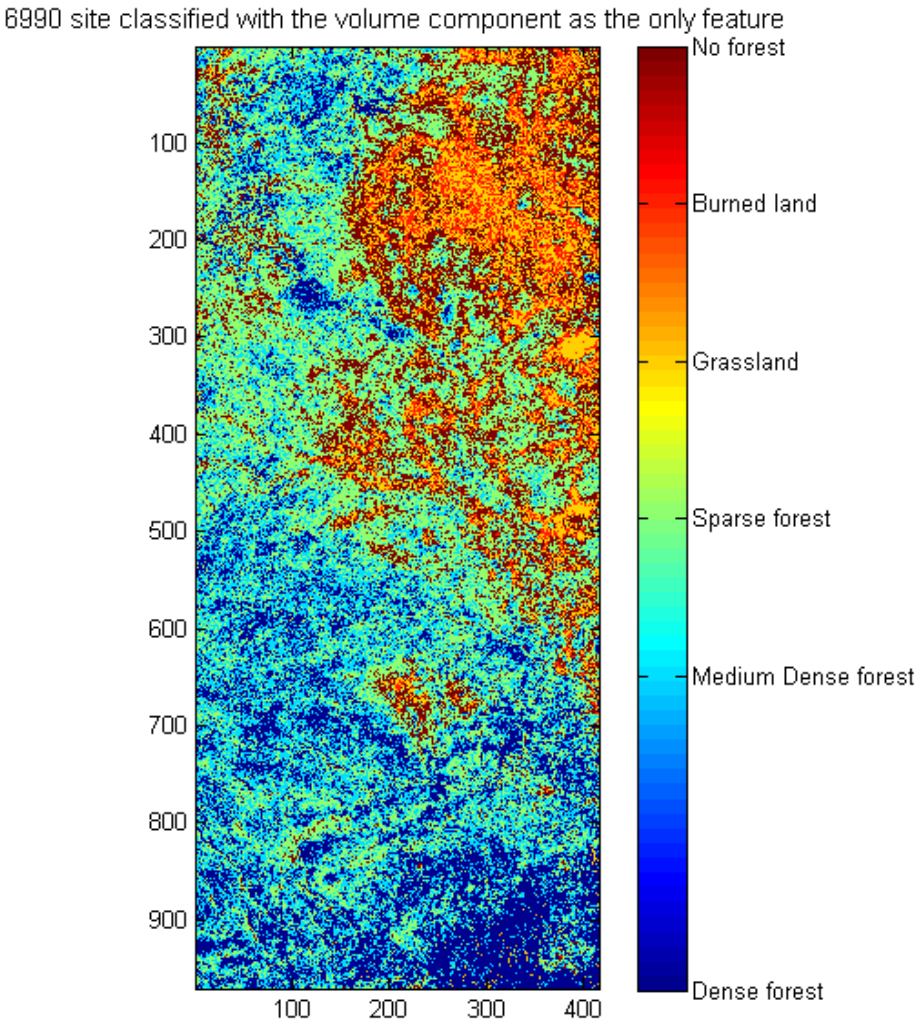


Figure 6.4: Classification map where the log transformed volume component of Freeman's 2 component decomposition theorem was used as the only feature for classification.

Table 6.3: Classification result using a Gaussian classifier with the log transform of the volume component of Freeman's 2 component decomposition theorem as the only feature.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	73	56	3	0	0	0	132	55.30%	44.70%
Medium	33	123	13	0	0	0	169	72.78%	27.22%
Sparse	0	28	41	9	0	9	87	47.13%	52.87%
Grass	0	0	0	22	8	11	41	53.66%	46.34%
Burned	0	0	0	77	48	23	148	32.43%	67.57%
No forest	0	0	8	61	19	65	153	42.48%	57.52%
Total	106	207	65	169	75	108			
Producer's accuracy	68.87%	59.42%	63.08%	13.02%	64%	60.19%			
Omission error	31.13%	40.58%	36.92%	86.98%	36%	39.81%			

Table 6.4: Classification result using a multivariate Gaussian classifier with the H/A/$\overline{\alpha}$ parameters as features.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	59	37	3	0	0	28	127	46.46%	53.54%
Medium	19	102	20	0	0	22	163	62.58%	37.42%
Sparse	3	23	24	0	2	14	66	36.36%	63.64%
Grass	0	0	1	153	11	0	165	92.73%	7.27%
Burned	3	1	6	16	57	5	88	64.77%	35.23%
No forest	22	44	11	0	5	39	121	32.23%	67.77%
Total	106	207	65	169	75	108			
Producer's accuracy	55.66%	49.28%	36.92%	90.53%	76%	36.11%			
Omission error	44.34%	50.72%	63.08%	9.47%	24%	63.89%			

6.1.4 H/A/$\overline{\alpha}$ decomposition theorem

The following results were obtained by using the H/A/$\overline{\alpha}$-parameters as features in a multivariate Gaussian classifier. The classification map is given in figure 6.5 and the classification results are given in table 6.4. The classification map seems to contain a lot of variation, so this classification may benefit from extensive multilooking or use of a smoothing filter. But to avoid giving this classifier an unfair advantage towards the others, no further multilooking or filtering is performed. This classifier achieved a global accuracy of 59.45%. Although it achieves a modest result, these parameters can be useful in fusion with the model based decomposition theorems.

6.1.5 Other decomposition theorems

The output from the Freeman-Durden and the nonnegative eigenvalue decomposition theorem in Polsarpro also suffer from histogram squeezing that pollute the burned land, grassland and no forest classes with replaced values. As mentioned earlier the surface component of the Freeman 2 component decomposition theorem has the same problem. It would have been preferable to have the components from at least one of the two mentioned decomposition

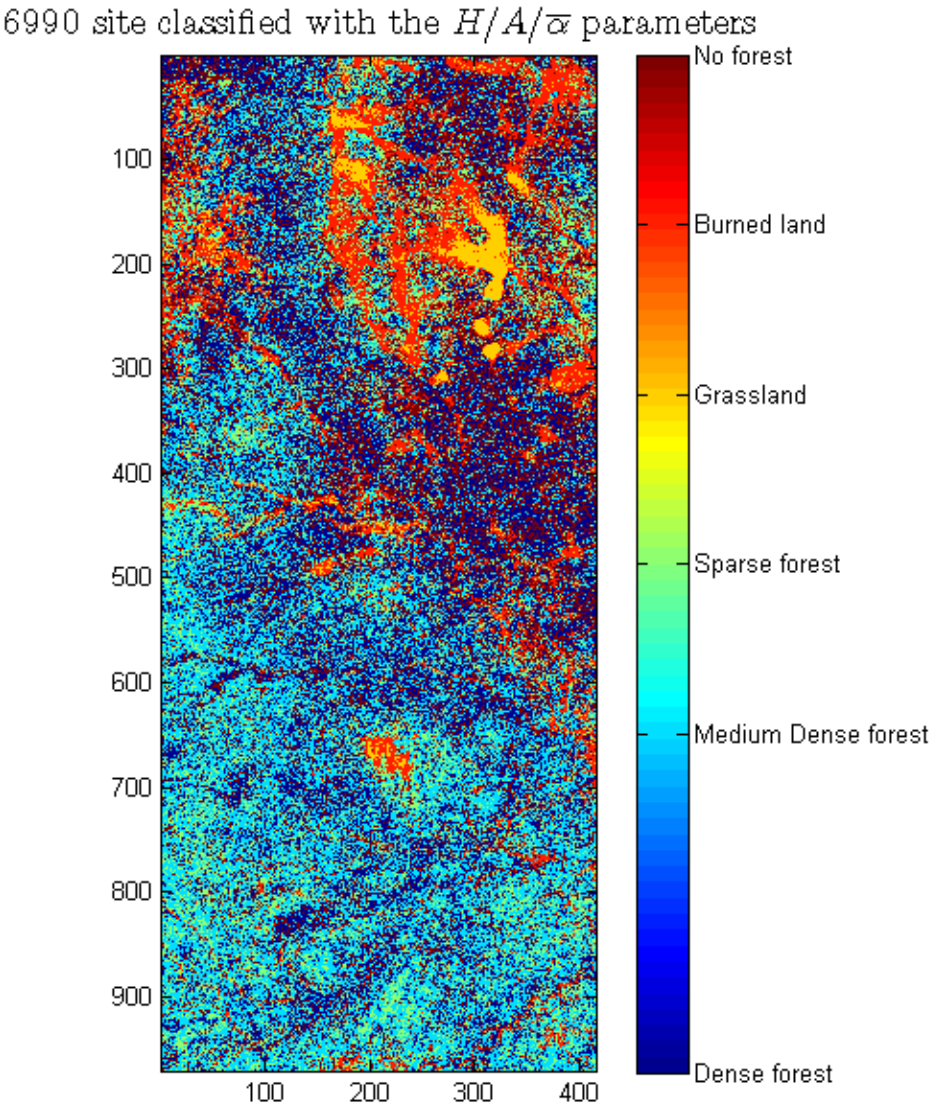


Figure 6.5: Classification map where H/A/$\overline{\alpha}$ parameters were used as features.

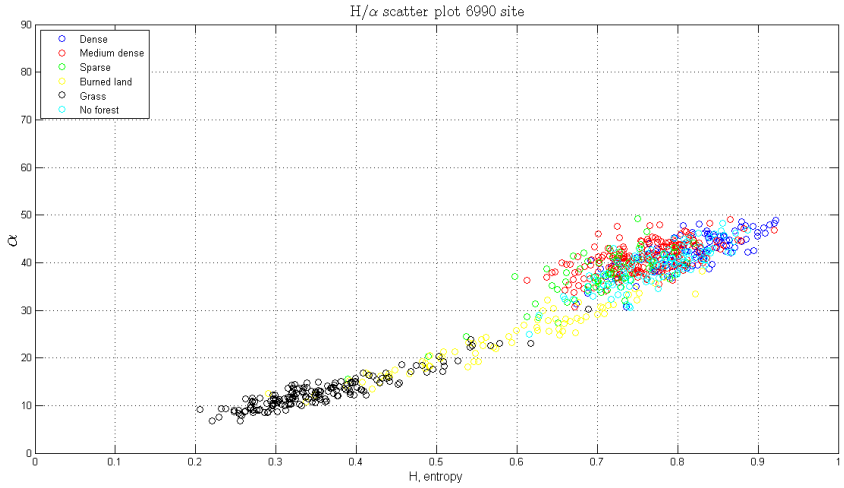


Figure 6.6: Scatter plot for all classes with the entropy and $\overline{\alpha}$ parameter.

theorems available. The most favourable would be the nonnegative eigenvalue decomposition theorem as it is an extension of the Freeman-Durden theorem [14].

6.2 Classification with 10 polarimetric features 6990 site

The following results were obtained by classification with the 10 polarimetric features given in table 6.5. The log transformed volume component of Freeman's 2 component decomposition theorem was omitted, as it was highly correlated with other features for some of the classes. Highly correlated features in the training data could generate a singular covariance matrix. It is impossible to calculate the pdf value given in equation 4.1 when the covariance matrix is singular, because a singular matrix is not invertible and the determinant is equal to zero. Even if it was possible to calculate the pdf value with a singular matrix, a highly correlated feature would not add much information because of the correlation. Yamaguchi's log transformed volume component is included in table 6.5, so the log transformed volume component from Freeman's 2 component decomposition theorem was considered to be obsolete. The correlation coefficient is defined in equation 6.1 [26].

$$\rho_{X,Y} = \frac{E\{(X - \mu_X)(Y - \mu_Y)\}}{\sigma_X \sigma_Y} \tag{6.1}$$

Table 6.5: Individual accuracy and feature label of polarimetric features.

Feature label	Feature	Individual accuracy
1	$log(\|S_{hh}\|^2)$	52.21%
2	$log(\|S_{hv}\|^2)$	66.98%
3	$log(\|S_{vv}\|^2)$	49.73%
4	Entropy	52.47%
5	$\overline{\alpha}$	47.95%
6	Anisotropy	24.66%
7	log transformed double bounce component of Yamaguchi's theorem	62.19%
8	log transformed odd bounce component of Yamaguchi's theorem	43.69%
9	log transformed volume component of Yamaguchi's theorem	56.99%
10	log transformed helix component of Yamaguchi's theorem	25.89%

Table 6.6: Classification result using a multivariate Gaussian classifier with 10 polarimetric features.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	78	46	1	0	0	0	125	62.40%	37.60%
Medium	25	135	10	0	0	0	170	79.41%	20.59%
Sparse	0	25	52	0	0	10	87	59.77%	40.23%
Grass	0	0	0	159	14	0	173	91.91%	8.09%
Burned	0	0	1	10	59	5	75	78.67%	21.33%
No forest	3	1	1	0	2	93	100	93%	7%
Total	106	207	65	169	75	108			
Producer's accuracy	73.58%	65.22%	80%	94.08%	78.67%	86.11%			
Omission error	26.41%	34.78%	20%	5.92%	21.33%	13.89%			

X, Y are the correlated features, μ_X, μ_Y are their mean values and σ_X, σ_Y are their variances. Among others, the log transformed volume component of Freeman's 2 component theorem was highly correlated with the log transformed odd bounce component of Yamaguchi's theorem in the grass class($\rho = 0.896$), $log(|S_{hh}|^2)$ in the grass class($\rho = 0.9497$), $log(|S_{vv}|^2)$ in the grass class($\rho = 0.8392$) and $log(|S_{vv}|^2)$ in the no forest class($\rho = 0.9231$). The component is modeled with a factor 1 multiplied with the intensities $|S_{hh}|^2$ and $|S_{vv}|^2$, see equation 3.19. The component is correlated with these intensities in the non-forested classes. Assuming smooth surfaces in these areas, the contribution to the S_{hh} and the S_{vv} channels would be approximately equal. So given the physical properties of the land cover class and the modeling of the component, it is reasonable that those features are correlated. Figure 6.7 illustrates the correlation between the log transformed odd bounce component of Yamaguchi's theorem and the log transformed volume component from Freeman's 2 component decomposition theorem.

The classification map is given in figure 6.8 and the classification result is given in table 6.6. The classifier achieved a global accuracy of 78.90%. The classifier based on Yamaguchi's parameters performed slightly better. This indicates that some of the features are confusing the classifier. Data fusion in remote sensing is expected to result in higher accuracy [9, 20, 21].

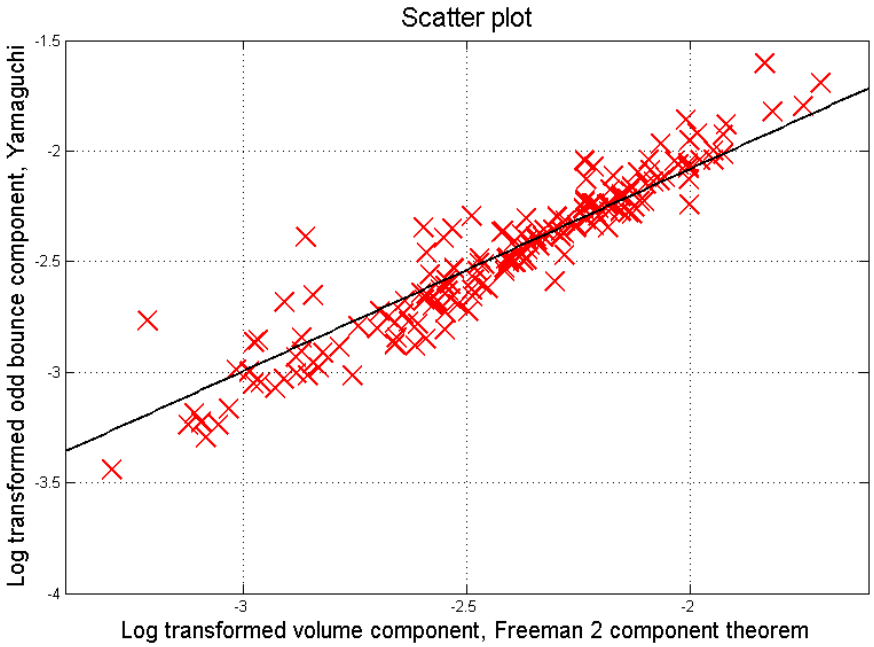


Figure 6.7: Scatter plot illustrating the correlation between components in the training data for the grass class.

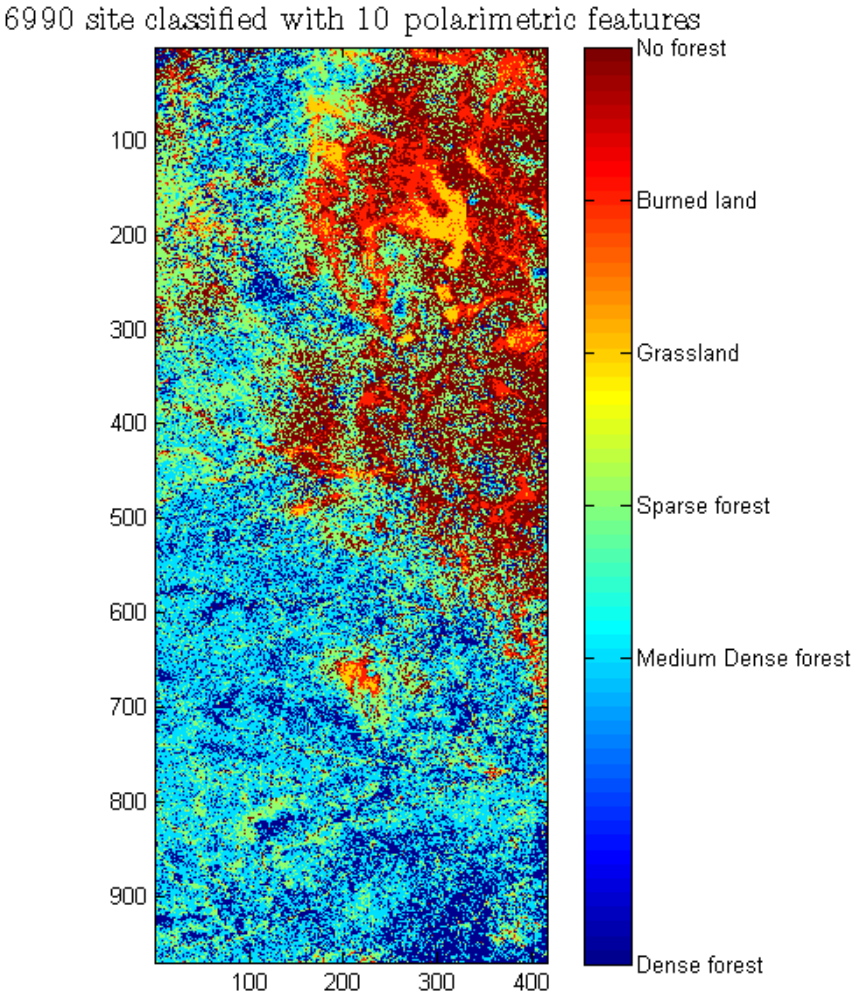


Figure 6.8: Classification map using all 10 polarimetric features given in table 6.5.

Table 6.7: Classification results from combined datasets obtained by backward sequential feature selection on the 6990 site.

	Polarimetric features used in classifier	Global accuracy	Features left out
1	All	78.90%	None
2	1,2,3,4,5,7,8,9 and 10	79.45%	6
3	1,2,3,4,5,7,9 and 10	79.86%	6 and 8
4	1,2,3,4,5,7 and 10	81.78%	6,8 and 9
5	1,2,3,4,7 and 10	82.60%	5,6,8 and 9
6	1,2,3,4 and 7	83.01%	5,6,8,9 and 10
7	1,2,3 and 4	82.60%	5,6,7,8,9 and 10
8	1,3 and 4	82.19%	2,5,6,7,8,9 and 10
9	1 and 4	80.27%	2,3,5,6,7,8,9 and 10
10	4	52.47%	1,2,3,5,6,7,8,9 and 10

6.3 Backward sequential feature selection method 6990 site

The results from the backwards sequential feature selection is shown in table 6.7 and in figure 6.9. The backward feature selection achieved the highest accuracy when using five of the polarimetric features in table 6.5. The optimum found by backward feature selection consisted of the log tranformed intensities $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$, the polarimetric entropy from the H/A/$\overline{\alpha}$ theorem and the log transformed double bounce component from Yamaguchi's theorem. With this composition of features the classifier achieved a global accuracy of 83.01%. Notice that the global accuracy increases when the first 5 features are removed, and decreases when removing additional features. It seems the first 5 features rather confuse the classifier than add information.

6.4 Forward sequential feature selection method 6990 site

The results from the forward sequential feature selection is shown in table 6.8 and figure 6.9. The global accuracy optimum is found when using 5 polarimetric features. The optimal composition consisted of the log transformed values of the intensities $|S_{hh}|^2$ and $|S_{hv}|^2$, the polarimetric entropy and anisotropy from the H/A/$\overline{\alpha}$ decomposition theorem and the log transformed values of the helix component of Yamaguchi's decomposition theorem. The classification with this composition of features achieved a global accuracy of 82.19%. As with the backward feature selection, accuracy decreased for both

Table 6.8: Classification results from combined datasets obtained by forward sequential feature selection on the 6990 site.

	Polarimetric features used in classifier	Global accuracy	Features left out
1	2	66.98%	1,3,4,5,6,7,8,9 and 10
2	2 and 6	76.71%	1,3,4,5,7,8,9 and 10
3	1,2, and 6	80.82%	3,4,5,7,8,9, and 10
4	1,2,4 and 6	81.78%	3,5,7,8,9 and 10
5	1,2,4,6 and 10	82.19%	3,5,7,8 and 9
6	1,2,3,4,6 and 10	81.64%	5,7,8 and 9
7	1,2,3,4,5,6 and 10	81.10%	7,8 and 9
8	1,2,3,4,5,6,9 and 10	80%	7 and 8
9	1,2,3,4,5,6,7,9 and 10	79.45%	8
10	All	78.90%	None

more and less than five features.

6.5 Classification with an educated choice of 5 polarimetric features 6990 site

It is interesting to see the performance of a classifier using features that are chosen based on their performance in both the backward and forward sequential feature selection. Perhaps a composition that is a compromise between the composition optima found by the two feature selection methods give a higher accuracy. The accuracy peaked at 5 features for both the forward and backward selection method, therefore it was decided that the ideal composition should consist of 5 features. Selection of features is based on the average finishing position in the backward and forward selection. The 5 features with the best average performance is the log transformed values of the intensities $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$, the entropy from the H/A/$\overline{\alpha}$ theorem and the log transformed values of the helix component from Yamaguchi's decomposition theorem. This composition is the same as the optimum for the backward selection except that the log transformed value of the double bounce component of Yamaguchi's decomposition theorem is replaced with the log transformed helix component. The classification map is shown in figure 6.10 and the classification result is shown in table 6.9. This classification achieves a global accuracy of 83.01% which is equal to what the composition of accuracy optimum in the backward feature selection.

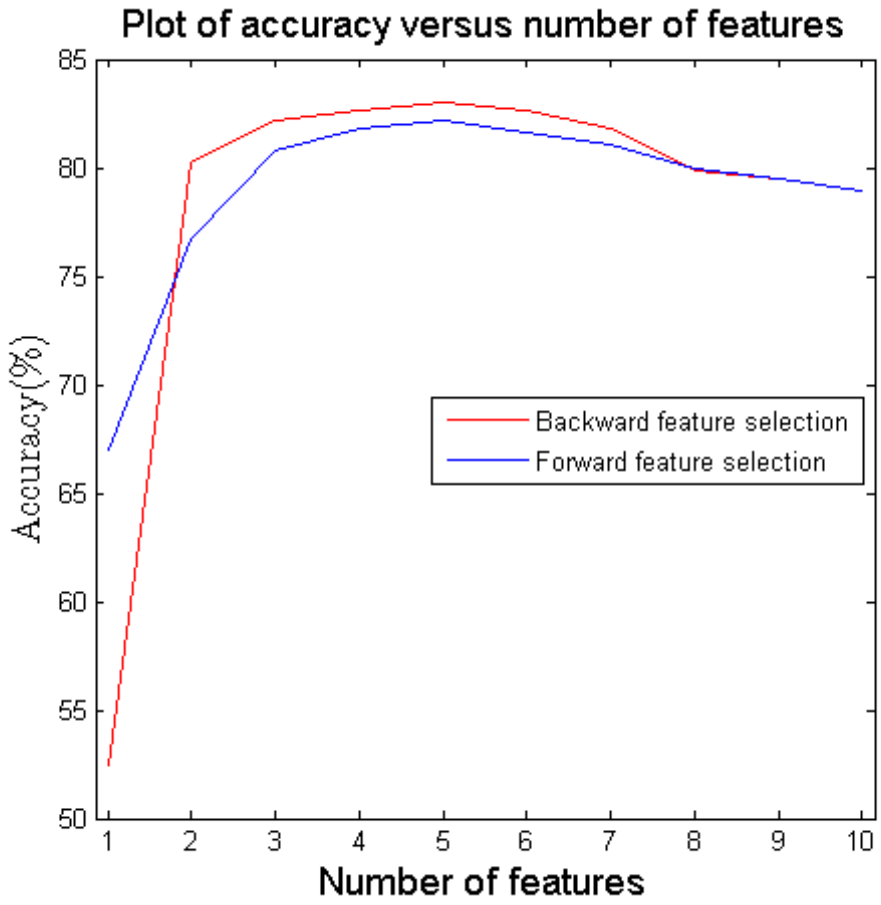


Figure 6.9: Plot of accuracy versus the number of features used in classifier using the feature compositions found by backward and forward sequential feature selection on the 6990 site.

Table 6.9: Classification result using an educated choice of 5 polarimetric features on the 6990 site.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	86	38	0	0	0	0	124	69.35%	30.65%
Medium	20	157	10	0	0	0	187	83.96%	16.04%
Sparse	0	12	50	0	0	4	66	75.76%	24.24%
Grass	0	0	0	150	10	0	160	93.75%	6.25%
Burned	0	0	0	19	62	3	84	73.81%	26.19%
No forest	0	0	5	0	3	101	109	92.66%	7.34%
Total	106	207	65	169	75	108			
Producer's accuracy	81.13%	75.85%	76.92%	88.76%	82.67%	93.52%			
Omission error	18.87%	24.15%	23.08%	11.24%	17.33%	6.48%			

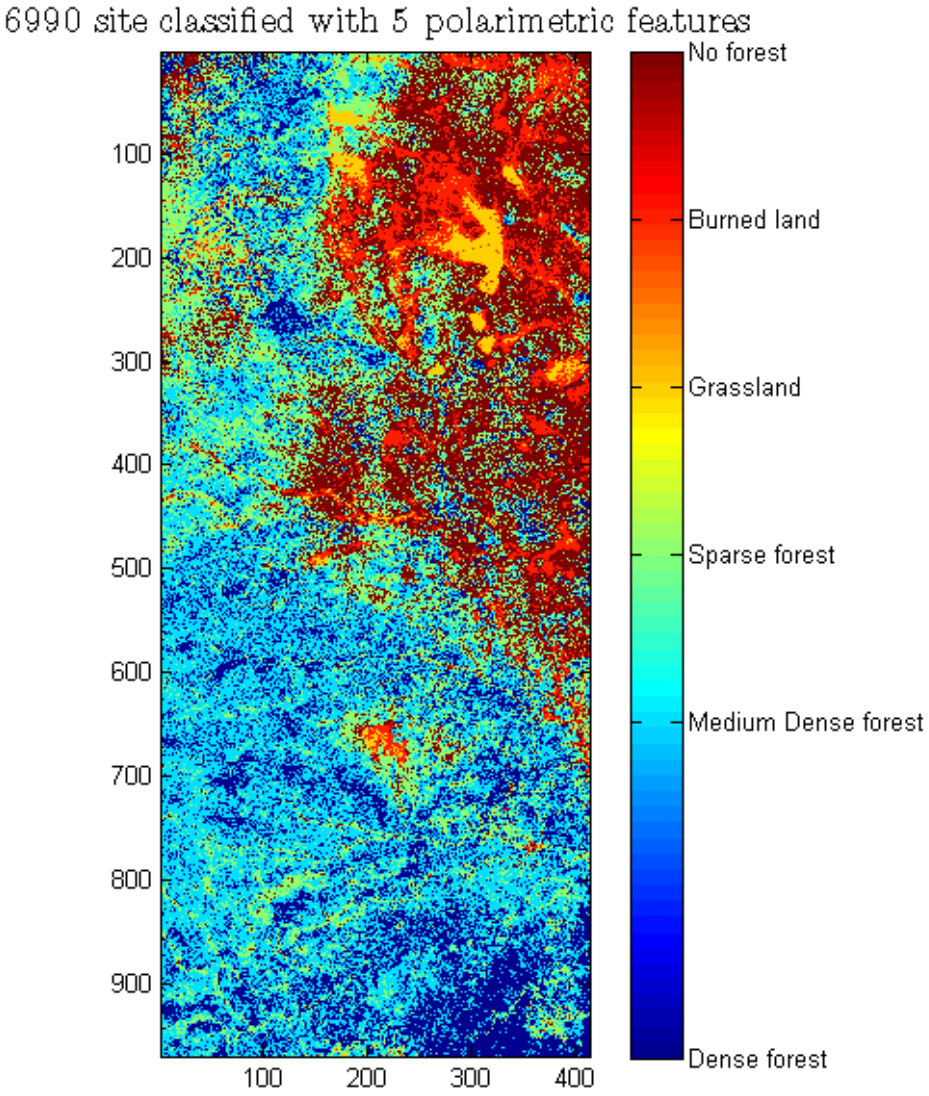


Figure 6.10: Classification map using an educated choice of 5 polarimetric features, 6990 site.

Table 6.10: Classification result using a composition of 5 polarimetric features found by sequential backward feature selection on the 6990 site.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	91	15	14	0	0	0	120	75.83%	24.17%
Medium	28	93	8	0	0	0	129	72.09%	17.81%
Sparse	3	3	108	0	0	5	119	90.76%	9.24%
Grass	0	0	0	67	1	0	68	98.53%	1.47%
Burned	0	0	2	4	45	1	52	86.54%	13.46%
No forest	2	0	18	0	0	10	30	33.33%	66.67%
Total	124	111	150	71	46	16			
Producer's accuracy	73.39%	83.78%	72%	94.37%	97.83%	62.50%			
Omission error	26.61%	16.22%	28%	5.63%	2.17%	37.50%			

6.6 Classification with 5 polarimetric features 7000 site

In this section, classification with the optimum feature compositions found by sequential backward and forward feature selection on the 6990 site are applied to the 7000 site.

6.6.1 Classification with optimum feature composition found by sequential backward selection on 6990 site

The following results were obtained by using the log transformed intensities $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$, the entropy from the H/A/$\overline{\alpha}$ theorem and the log transformed double bounce component from Yamaguchi's decomposition theorem as features. These features were the optimal combination found by the sequential backward feature selection. The classification map is shown in figure 6.11 and the classification result is shown in table 6.10. This classification achieved a global accuracy of 79.92%, which is just below the optimum found by sequential backward feature selection of the 6990 site.

6.6.2 Classification with optimum feature composition found by sequential forward selection on 6990 site

The following results were obtained by using the log transformed intensities $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$ and the entropy and anisotropy from the H/A/$\overline{\alpha}$ decomposition theorem. The choice of features is based on the result from forward feature selection. The log transformed $|S_{vv}|^2$ has replaced the log transformed helix component. Unfortunately the helix component of the 7000 site contained a large number of pixels that took the value 0. This creates problems when applying the log transform. The log transformed

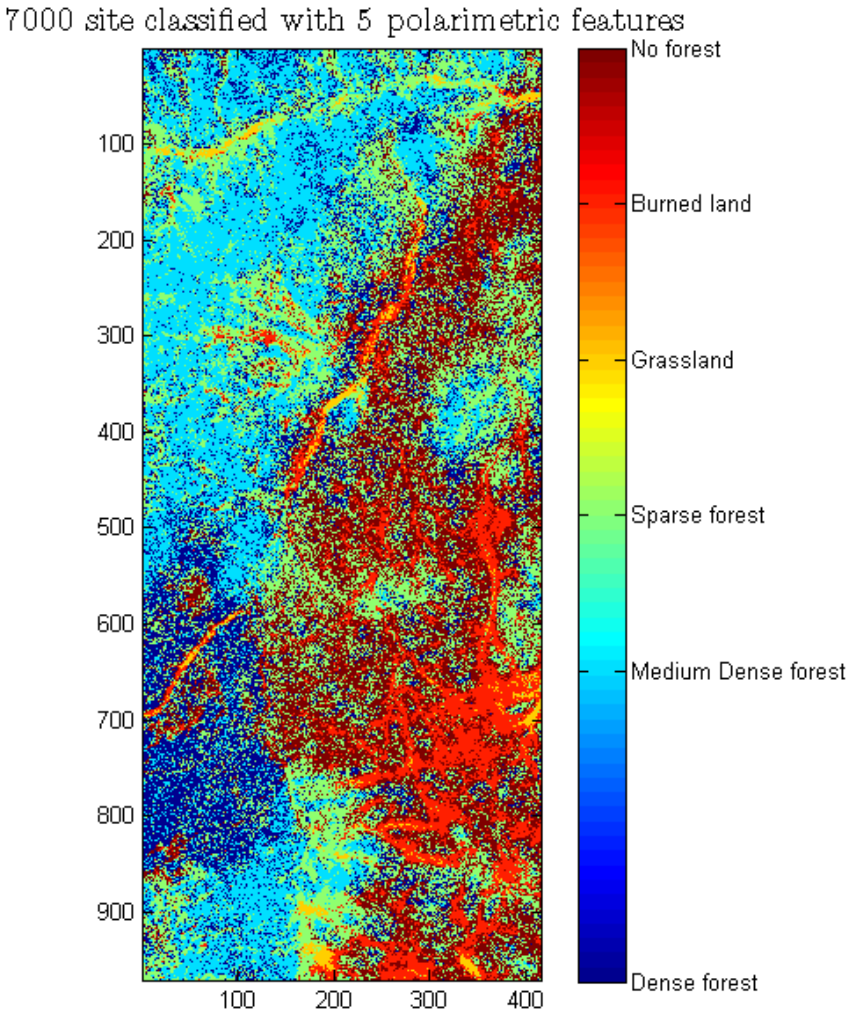


Figure 6.11: Classification map 7000 site using a composition of 5 features found by sequential backward feature selection on the 6990 site.

Table 6.11: Classification result using a composition of 5 polarimetric features found by sequential forward feature selection on the 6990 site.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	94	15	21	0	0	0	130	72.31%	27.69%
Medium	24	94	7	0	0	0	125	75.20%	24.80%
Sparse	4	2	103	0	0	5	114	90.35%	9.65%
Grass	0	0	0	66	2	0	68	97.06%	2.94%
Burned	0	0	1	5	42	0	48	87.50%	12.50%
No forest	2	0	18	0	2	11	33	33.33%	66.67%
Total	124	111	150	71	46	16			
Producer's accuracy	75.81%	84.68%	68.67%	92.96%	91.30%	68.75%			
Omission error	24.19%	15.32%	31.33%	7.04%	8.70%	31.25%			

$|S_{vv}|^2$ component was added to the composition of features the step after the log transformed helix component was added. The classification map is shown in figure 6.12 and the classification result is shown in table 6.11. The classifier achieved a global accuracy of 79.15%, which is slightly lower than what the classifier with the same composition of features achieved at the 6990 site.

6.6.3 Classification of 7000 site with training data from 6990 site

All previous experiments have been conducted by first picking training and testing data from one site, and then performing classification on the same site. It is interesting to see the performance of a classification based on using the training data from one site to classify a different site. Information from this kind of classification can be valuable to evaluate the performance of the features, because in this case they are applied to an independent dataset.

Classification with optimum feature composition found by sequential backward selection

The following results were obtained by using the feature composition found by sequential backward selection on the 6990 site in a multivariate Gaussian classifier. This optimum composition consisted of the log transformed intensities $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$, the entropy and the log transformed double bounce component from Yamaguchi's theorem. Mean values and covariance matrices for the various classes were calculated from training data from the 6990 site, and classification was performed on the 7000 site. The classification map is given in figure 6.13 and the classification result is given in table 6.12. This classification achieved a global accuracy of 69.50%.

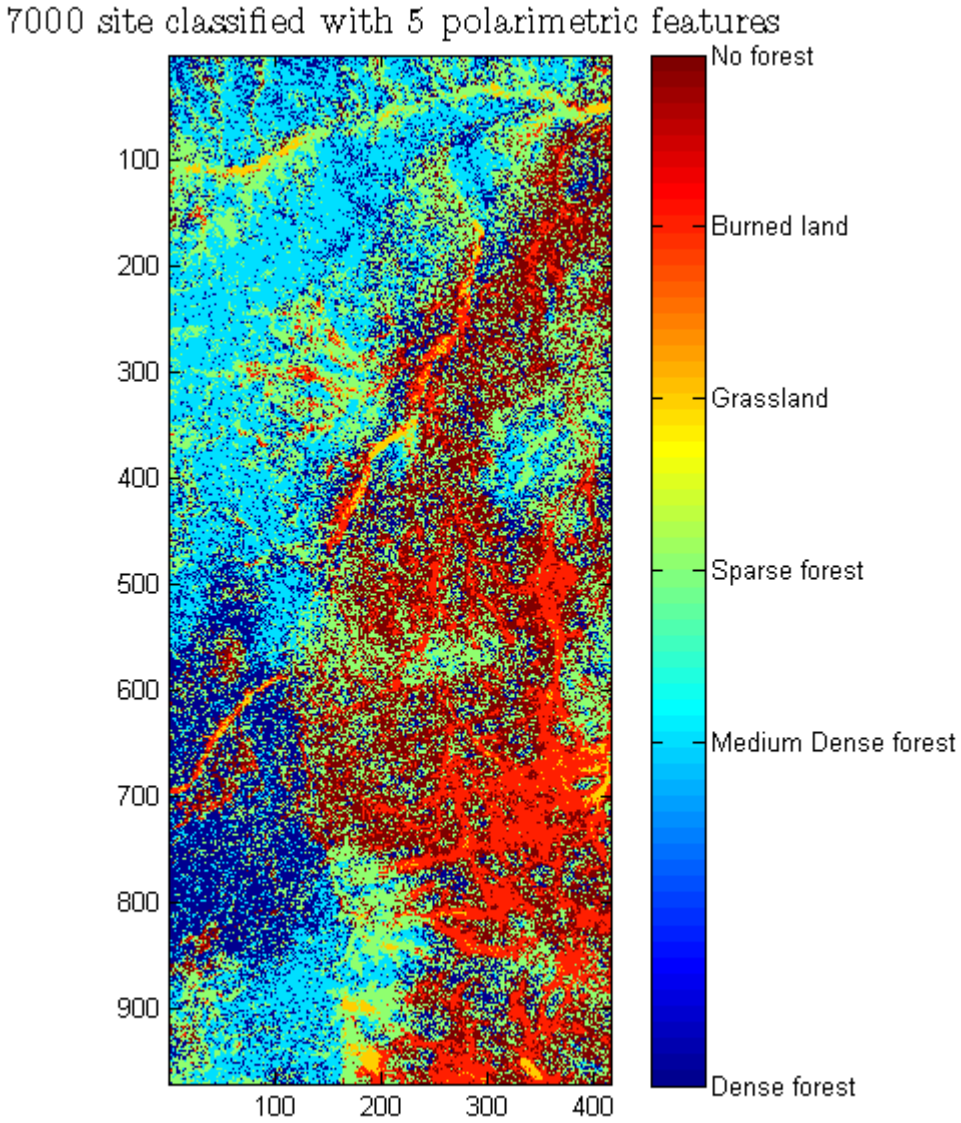


Figure 6.12: Classification map 7000 site using a composition of 5 features found by sequential forward feature selection on the 6990 site.

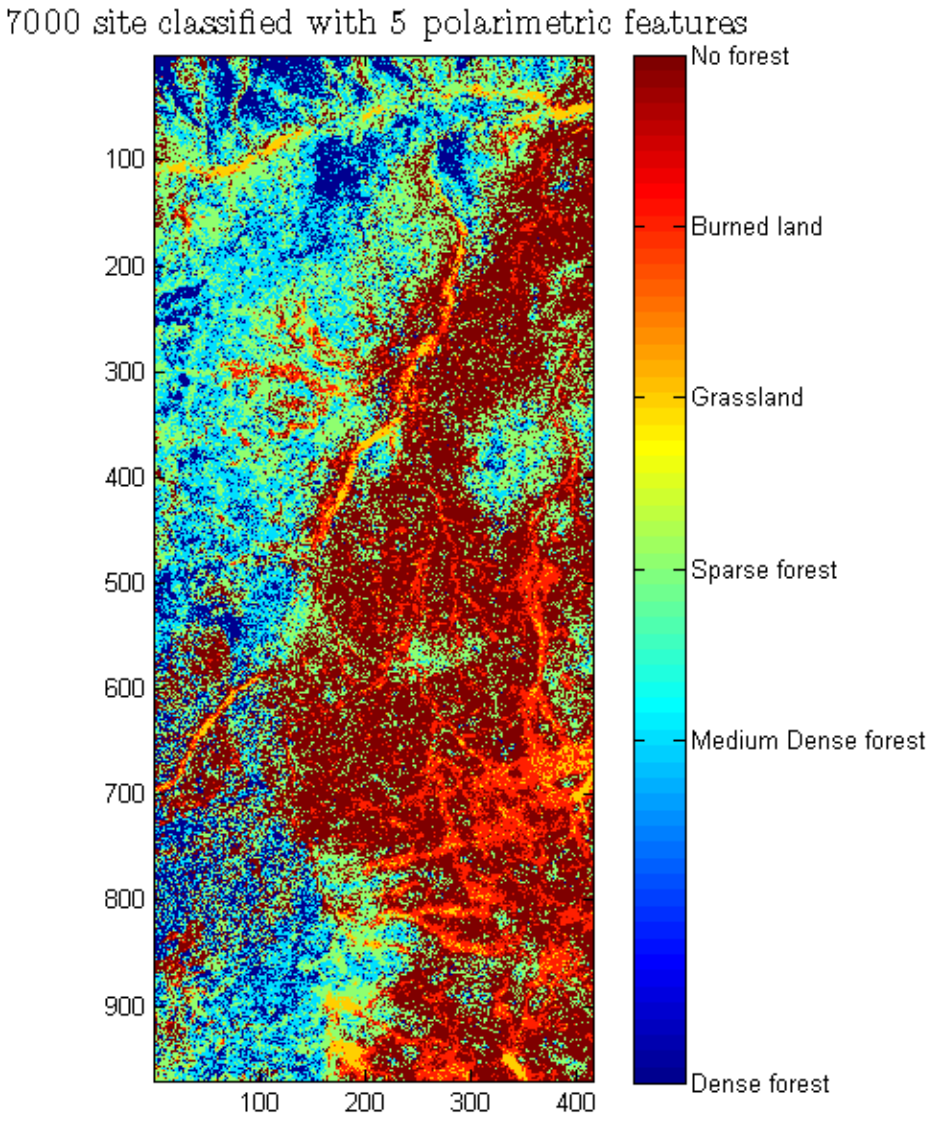


Figure 6.13: Classification map having used training data from 6990 site to classify 7000 site. The feature composition found by backward feature selection is engaged in this classification.

Table 6.12: Classification result having used training data from 6990 site and testing data from 7000 site. The feature composition found by backward feature selection is applied here.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	80	36	1	0	0	0	117	68.38%	31.62%
Medium	30	68	2	0	0	0	100	68%	32%
Sparse	8	7	95	0	0	1	111	85.59%	14.41%
Grass	0	0	2	68	11	0	81	83.85%	16.15%
Burned	2	0	2	3	34	0	41	82.93%	17.07%
No forest	4	0	48	0	1	15	68	22.06%	77.94%
Total	124	111	150	71	46	16			
Producer's accuracy	64.52%	61.26%	63.33%	95.77%	73.91%	93.75%			
Omission error	35.48%	38.74%	36.67%	4.23%	26.09%	6.25%			

Table 6.13: Classification result having used training data from 6990 site and testing data from 7000 site. The feature composition found by forward feature selection is applied here.

Classifier/Ref	Dense	Medium	Sparse	Grass	Burned	No forest	Total	User's accuracy	Comission error
Dense	74	30	0	0	0	0	104	71.15%	28.85%
Medium	27	66	2	0	0	0	95	69.47%	30.53%
Sparse	8	8	83	0	0	1	100	83%	17%
Grass	0	0	0	43	3	0	46	93.48%	6.52%
Burned	4	4	15	28	43	2	96	44.79%	55.21%
No forest	11	3	50	0	0	13	77	16.88%	83.12%
Total	124	111	150	71	46	16			
Producer's accuracy	59.68%	59.46%	55.33%	60.56%	93.48%	81.25%			
Omission error	40.32%	40.54%	44.67%	39.44%	6.52%	18.75%			

Classification with optimum feature composition found by sequential forward selection

The following results were obtained by using the feature composition found by sequential forward feature selection on the 6990 site. Mean values and covariance matrices were calculated based on training data from the 6990 site, and were applied to classification on the 7000 site. The feature composition consisted of the log transformed intensities $|S_{hh}|^2$ and $|S_{hv}|^2$, entropy and anisotropy from the H/A/$\overline{\alpha}$ theorem and the log transformed helix component from Yamaguchi's theorem. Classification map is given in figure 6.14 and classification results are given in table 6.13. This classification achieved a global accuracy of 62.16%.

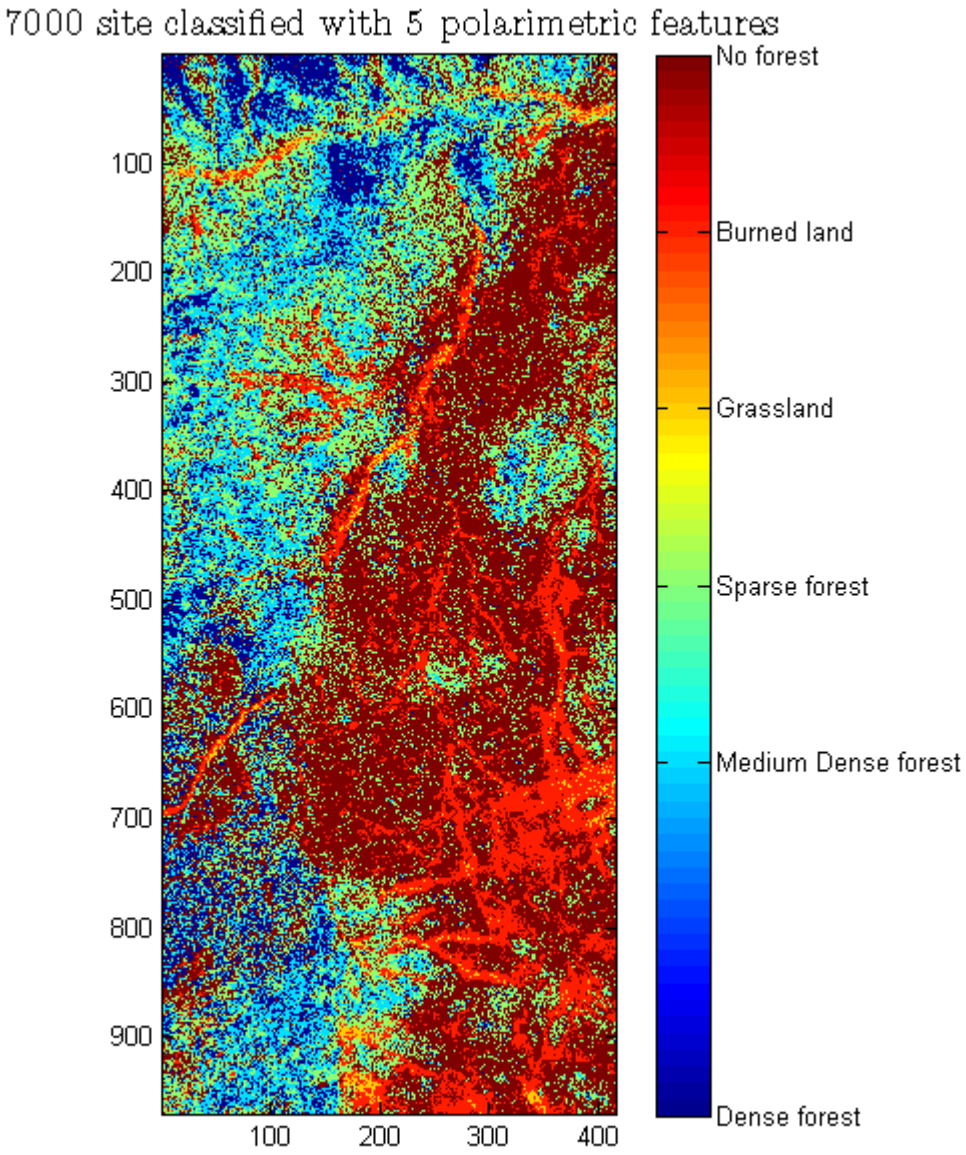


Figure 6.14: Classification map having used training data from 6990 site to classify 7000 site. The feature composition found by forward feature selection is engaged in this classification.

6.7 Discussion

6.7.1 Classification using theorems separately on the 6990 site

Classification using the log transformed intensities as features

The classifier that engaged the log transformed intensities as features achieved the second highest global accuracy among the classifiers that only engaged the components from each theorem separately. The classifier achieved higher producer accuracy than user accuracy for the dense and sparse forest which suggests that these classes are overestimated. The result is the opposite for the medium dense forest. It seems that the mean values of the medium dense forest is squeezed between the dense and sparse forest classes. For the other three ground truth classes the producer accuracy and user accuracy are approximately equal. The burned land achieved lower accuracy than grassland and no forest. The grass and no forest classes achieved significantly higher accuracy than the other classes. The polarimetric signature of grass, burned land and no forest should be easier to distinguish from each other than the forest classes. Burned land have some of the testing data confused with both grass and no forest. None of the testing data for any of the grass, burned land or no forest classes are attributed to any of the forest classes.

Classification using the log transformed values of Yamaguchi's parameters as features

The classifier that engaged the log transformed parameters of Yamaguchi's theorem achieved the highest global accuracy of the classifiers that only engaged the parameters from each theorem separately. The classification map from this classifier is almost identical to the classification map from the classification based on the intensities. Given the relatively complex model of the volume component described in section 3.3.1, one could expect that this classifier to a greater extent would distinguish between the forest classes. But it does not get significantly stronger results for these classes. The grass and no forest classes achieve a higher accuracy than the other classes. The no forest class class has the highest accuracy. This classifier also confuse some of the burned land with grass and no forest.

Freeman's 2 component decomposition theorem

Because of the histogram squeezing performed in Polsarpro, the surface component of Freeman's 2 component theorem could not be engaged as a feature.

This is a severe disadvantage for the classifier, and makes it harder to evaluate and compare the results. The dense forest and sparse forest classes seems to be overestimated based on the producer and user accuracy. The medium dense forest is equivalently underestimated. Most of the grass class is attributed to the burned land and no forest. Hence, the burned land and no forest are overestimated. An accuracy of 50.96% using a single feature is not a weak result. 64.02% of the accuracy achieved by the classification that engaged Yamaguchi's 4 parameters is retained by classification with the log transformed volume component of Freeman's 2 component theorem.

Classification using the H/A/$\overline{\alpha}$ parameters

The classifier that engaged the H/A/$\overline{\alpha}$ parameters as features achieved a lower accuracy score than the other classifiers that only engaged the parameters from each theorem separately. The grassland and burned land classes achieved higher accuracy scores than the other classes. Notice that this classifier attributes a relatively big proportion of the pixels in the center-right region of the image to the dense forest class. This region is not covered by any of the groundtruth pixels, but by inspecting images 411 and 511 one can tell that it is certainly not dense forest, see figure 5.4. The other classifiers have classified most of these areas as no forest or burned land. Figure 6.6 illustrates the separability of the classes in the H/$\overline{\alpha}$-plane, and it shows that the separability between the no forest, dense forest, medium dense forest and sparse forest classes is poor. Parameters used in this classification does not contain any information about the intensity, it only contains information abut the scattering mechanisms and the dominance relationship between them. That could be the reason that it scored lower than the other decomposition theorems. This classification also applied the mean entropy and the mean angle($\overline{\alpha}$). It is possible possible to calculate the angle and entropy for each of the scattering mechanisms, see section 3.3.2. But it would make it even harder to interpret, and is beyond the scope of this study.
In [23] the authors present a scheme for interpreting various regions in the H/$\overline{\alpha}$ plane. The scheme is divided into 9 zones. The grass class lies in zone 9 which is called *low entropy surface scatter*. This kind of scattering occur over surfaces such as bare soil and open waters. The burned land lies in zone 6. This zone is called *medium entropy surface scatter*. This is interpreted as surface scatter where the entropy has increased due to roughness of the surface. The four remaining classes lies mainly in zone 5 which is called *medium entropy vegetation scattering*. The author argues that the increase in entropy from zone 6 is caused by increased variability due to canopy. The no forest class has little separability with the forest classes. The reason for the high $\overline{\alpha}$

angle values in the no forest class could be an increased contribution to the S_{hv} channel due to surface roughness or depolarizing effects from branches and leaves on bushes.

6.7.2 Classfication with 10 polarimetric features 6990 site

The classifier that engaged 10 polarimetric features achieved a lower accuracy than the classifier based on the parameters of Yamaguchi's theorem and only slightly higher than what the classifier based on the intensities achieved. This indicates that some of the features are confusing the classifier, rather than adding information. This classifier appears to underestimate the medium dense forest, as the producer accuracy for this class is significantly lower than the user accuracy. The dense forest and sparse forest classes are equivalently overestimated. It seems that the medium dense forest is squeezed between the two other forest classes, and thus medium dense forest pixels are attributed to the other forest classes. The grass class has the highest accuracy with 94.08%. The class with the lowest accuracy in this classifier, is also the class with the highest number of pixels in the testing data. This is a disadvantage for the classifier.
The feature with the highest individual accuracy was the log transformed intensities $|S_{hv}|^2$. In total there are 378 pixels of testing data in the forest classes. Forest classes give contribution to the S_{hv} channel due to multiple bounce of the leaves and branches which leads to depolarization. So it is not surprising that this feature achieved the highest individual accuracy.

6.7.3 Backward sequential feature selection on 6990 site

The optimum composition of features found by the sequential backward feature selection consisted of 5 features. The optimum composition was the log transformed intensities $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$, the entropy from the H/A/$\overline{\alpha}$ theorem and the log transformed double bounce component from Yamaguchi's theorem. This composition achieved a global accuracy of 83.01%. When the first 5 features were removed, the accuracy increased. When the number of features decreased below 5, the accuracy decreased as well. Despite the high accuracy achieved by the classifier based on the parameters of Yamaguchi's theorem, they were all left out early in the backward feature selection. Surprisingly the helix component was left out as the last one of the Yamaguchi parameters. The helix component is expected to almost disappear in natural terrain, and was not expected to be a particularly valuable component in this classification. In the individual classification it achieved

the second lowest accuracy. The helix scattering is modeled as strong depolarized backscattering. This means that it has strong contribution in the S_{hv} channel which is expected to be correlated with the presence of canopy, so perhaps the helix component can help distinguishing between forest classes. When applying 3 of the 5 features in the optimum composition, 99.01% of the accuracy achieved by the optimum composition is retained.

6.7.4 Forward sequential feature selection on 6990 site

The optimum composition of features found by the sequential forward feature selection consisted of 5 features. The optimum composition was the log transformed intensitites $|S_{hh}|^2$ and $|S_{hv}|^2$, the entropy and anisotropy from the H/A/$\overline{\alpha}$ decomposition theorem and the log transformed helix component from Yamaguchi's decomposition theorem. It is surprising that the polarimetric anisotropy from the H/A/$\overline{\alpha}$ theorem is in the best pair of features, as it gave the worst accuracy result of all the features individually. The accuracy of the classifier engaging the optimum composition of features found by sequential forward feature selection was slightly lower than the one found by sequential backward feature selection. When applying 3 of the 5 features in the optimum composition, 98.33% of the accuracy achieved by the optimum composition is retained.

6.7.5 Classification with an educated choice of 5 polarimetric features 6990 site

The classifier that engaged 5 manually chosen polarimetric features on the 6990 site achieved the same accuracy as the optimum found in the backward sequential feature selection. The only difference between this classifier and the optimum composition of features found by the sequential forward feature selection, was that the log transformed double bounce component from Yamaguchi's theorem replaced the log transformed helix component. This classifier achieved high accuracy results for the forest classes. Based on producer and user accuracy, it seems as if the dense forest and burned land is overestimated, and that the medium dense forest is underestimated.

6.7.6 Classification with 5 polarimetric features 7000 site

Classification with optimum feature composition found by sequential backward selection on 6990 site

The classifier that used the 5 features found by backwards selection on the 6990 site in a classification on the 7000 site achieved a global accuracy of 79.92%. It seems like it underestimates the sparse forest, as it has significantly higher user accuracy than producer accuracy. The grass and no forest classes have higher accuracy than the other classes.

Classification with optimum feature composition found by sequential forward selection on 6990 site

The classifier that used the 5 features found by forward selection on the 6990 site in a classification on the 7000 site achieved a global accuracy of 79.15%. This is also slightly lower than the accuracy achieved when using the same feature composition on the 6990 site. This classifier also seems to underestimate the sparse forest class.

Both of these two classifications achieved a lower global accuracy than classification with the same feature compositions on the 6990 site achieved. Choices made by feature selection methods when applied to the 6990 site could make it biased towards that site, in the sense that the applied feature composition is specialized towards maximizing accuracy score on that particular site. There is also a relatively bigger proportion of forest classes in the testing data from the 7000 site than in the testing data from the 6990 site. In other classifications these classes have lower accuracy scores than the non-forest classes.

Classification of 7000 site with training data from 6990 site

Both of the two classifications that used training data from the 6990 site and testing data from the 7000 achieved significantly lower accuracy scores than when the same feature compositions were used in a classification with training and testing data from the same dataset. This is mainly due to different scores in the non-forest classes. A difference in accuracy score is expected. It will be an advantage for a classifier to have training data obtained close to the testing data. A decreased accuracy could be caused by local variations in vegetation and land cover. Global accuracy for the classification that used

the features found by forward feature selection achieved a lower accuracy than the classification that used the features found by backward feature selection.

6.8 Summary

The parameters of Yamaguchi's decomposition theorem achieved a higher accuracy score than any of the other decomposition theorems when the theorems were tested separately. The classifier that engaged the log transformed intensities as features achieved a slightly lower global accuracy. The classification based on the $H/A/\overline{\alpha}$ parameters and the classification based on the log transformed volume component from Freeman's 2 component decomposition theorem achieved lower accuracy scores. When applying the feature selection methods, the log transformed intensities performed better than Yamaguchi's parameters. It is important to remember that the accuracy obtained by engaging Yamaguchi's parameters as features used 4 different features, and the one based on the log transformed intensities only engaged 3 features. The entropy was included in the feature composition optima found by both the forward and backward feature selection and the anisotropy was included in the optimum composition found by the forward feature selection.
When using the feature composition optima found by sequential backward and forward feature selection on the 6990 site in a classifier on the 7000 site, the classifiers achieved nearly the same accuracy as the same feature composition achieved on the 6990 site. Classification achieved a lower accuracy score when applying training data from the 6990 site in a classifier on the 7000 site.

6.9 Comments

Initially it was assumed that classification based on the log transformed intensities would be equivalent to classification based on the Pauli parameters, since the α, β and γ parameters are linear combinations of the S_{hh}, S_{hv} and S_{vv} channels. But the classification of intensities involves squaring the values in these channels since they are complex, which would also be necessary with the Pauli parameters. Squaring the Pauli parameters α and β results in crossterms given in equation 6.2. So the assumption that classification based on the intensities would be equivalent to classification based on the Pauli parameters requires that there is no significant contribution from crossterms, which may not be true. Furthermore the data is log transformed which is not a linear transformation. Unfortunately this was not noticed until a late

stage in the project, and time did not allow rerunning all experiments.

$$\begin{aligned} \alpha = \frac{S_{hh}+S_{vv}}{\sqrt{2}} \Rightarrow |\alpha|^2 = \frac{|S_{hh}|^2+|S_{vv}|^2+S_{hh}S_{vv}^*+S_{hh}^*S_{vv}}{2} \\ \beta = \frac{S_{hh}-S_{vv}}{\sqrt{2}} \Rightarrow |\beta|^2 = \frac{|S_{hh}|^2+|S_{vv}|^2-S_{hh}S_{vv}^*-S_{hh}^*S_{vv}}{2} \end{aligned} \tag{6.2}$$

The Polsarpro software performs a histogram squeezing when it calculates some of the components from the decomposition theorems. Low values are replaced with a fixed mean value, probably to enhance the display. This creates a spike in the histograms of the components. Unfortunately these replaced values appear in ground truth regions of the image. The grassland, burned land and no forest classes are severely contaminated with these replaced mean values. In order to get around this problem it would be necessary to make a program that calculated components from the raw data, which time did not allow.

Chapter 7

Conclusion

In this thesis, the performance of polarimetric decomposition theorems and their individual components have been evaluated on an area of tropical forest in Lindi county, Tanzania.

The first stage of the work was to determine which classes are present in the images. This was done by visual inspection of high resolution optical images that covered parts of the SAR measurements. A decision was made that images should be segmented into 6 different classes.

In the preprocessing stage, the polarimetric data was multilooked. Ground truth areas were picked and divided in to testing and training data.

All parameters of all theorems used in this thesis were calculated through the educational tool Polsarpro. Features and theorems are tested and evaulated through supervised classification. The theorems were evaluated separately and an optimal composition of polarimetric features was found by feature selection methods. When applying the theorems separately, Yamaguchi's 4 component decomposition theorem gave the highest accuracy. When searching with feature selection methods for an optimal composition of individual polarimetric features, the intensities on the diagonal of the multilooked covariance matrix achieved higher average finishing position in the ranking than the Yamaguchi parameters. Two different feature composition were found by applying feature selection methods. One of the feature compositions consisted of the log transformed intensities $|S_{hh}|^2$, $|S_{hv}|^2$ and $|S_{vv}|^2$, the polarimetric entropy from the H/A/$\overline{\alpha}$ theorem and the log transformed double bounce component of Yamaguchi's decomposition theorem. The other feature composition consisted of the log transformed intensities $|S_{hh}|^2$ and $|S_{hv}|^2$, the entropy and anisotropy from the H/A/$\overline{\alpha}$ theorem and the log

transformed helix component from Yamaguchi's theorem. Note that parameters from the H/A/$\overline{\alpha}$ decomposition theorem were included in the feature composition optima. When applying the optimum combinations found by the feature selection methods on another site, accuracies were close to equal. Applying training data from one site on testing data from another site gave lower accuracy scores. It is shown that when applying compositions of features found by selection methods, accuracy scores increased compared to the results obtained by an intensity based classification.

The features used in this thesis have shown potential for land cover classification of this type of landscape. This study has given recommendation to combinations of features to use for these purposes based on accuracy scores. Note that these scores depend on the size of each respective class in the testing data. Which features should be used for classification of tropical forest depends on the application. Some features are more useful for distinguishing forest classes and some features can distinguish other types of land cover, and this should be taken in to account when choosing features.

Bibliography

[1] Google Earth,
earth.google.com/

[2] ALOS research and application project, Earth Observation Research Center.
http://www.eorc.jaxa.jp/ALOS/en/about/palsar.htm

[3] Japanese Aerospace Exploration Agency,
http://www.jaxa.jp/projects/sat/alos/index_e.html

[4] RapidEye's homepages,
http://www.rapideye.com/about/satellites.htm

[5] Jong-Sen Lee and Eric Pottier, *Lee & Pottier - Polarimetric Radar imaging - From basics to imaging*, CRC Press 2009.

[6] Canadian natural resources homepage,
www.cnrl.com

[7] Emilio Chuvieco and Alfredo Huete, *Fundamentals of Satellite Remote Sensing*, CRC Press 2009

[8] Lecture notes, FYS-3001 given spring 2012, University of Tromsø.

[9] Temesgen Gebrie Yitayew, *Multi-sensor Data Fusion and Feature Extraction for Forest Applications*, May 2012.

[10] Stein Erik Sakshaug, *Polarimetric SAR decompositions - pilot project*, 2012.

[11] Anthony Freeman and Stephen L. Durden, *A Three-Component Scattering Model for Polarimetric SAR Data*, IEEE Transactions on geoscience and remote sensing, vol. 36, May 1998.

[12] Yoshio Yamaguchi, Toshifumi Moriyama, Motoi Ishido and Hiroyoshi Yamada, *Four-Component Scattering Model for Polarimetric SAR Image Decomposition*, IEEE Transactions on geoscience and remote sensing, vol. 43, NO. 8, August 2005.

[13] Anthony Freeman, *Fitting a Two-Component Scattering Model to Polarimetric SAR Data From Forests*, IEEE Transactions on geoscience and remote sensing, vol. 45, NO. 8, August 2007.

[14] Jakob J. van Zyl, Motofumi Arii and Yunjin Kim, *Model-Based Decomposition of Polarimetric SAR Covariance Matrices Constrained for Nonnegative Eigenvalues*, IEEE transactions on geoscience and remote sensing, vol. 49, NO. 9, September 2011

[15] Jakob J. van Zyl, Motofumi Arii and Yunjin Kim, *Requirements for model-based polarimetric decompositions*, Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, vol. 5. Pages: V - 417 - V - 420.

[16] SR Cloude and Eric Pottier, *A Review of Target Decomposition Theorems in Radar Polarimetry*, IEEE transactions on geoscience and remote sensing, VOL 34, NO 2, March 1996

[17] Charles Elachi and Jakob Van Zyl, *Introduction to the physics and techniques of remote sensing, 2e*, Wiley 2006

[18] Ernst Krogager and Anthony Freeman, *Three component break-downs of scattering matrices for radar target identification and classification*, Piers 94, Noordwijk, The Netherlands, July 1994, p. 391.

[19] T. Macrì Pellizzeri and P. Lombardo, P. Ferriero, *Polarimetric SAR image processing: Wishart vs "H/A/alpha" segmentation and classification*

schemes, Geoscience and Remote Sensing Symposium, 2003.

[20] M. Shimoni, D. Borghys, R. Heremans, C. Perneel and M. Acheroy. *Fusion of PolSAR and PolInSAR data for land cover classification*, International Journal of Applied Earth Observation and Geoinformation, 2009.
`http://earth.esa.int/workshops/polinsar2007/papers/173_shimoni.pdf`

[21] A.H.S. Solberg, A.K. Jain and T. Taxt. *Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images*, IEEE Transactions on geoscience and remote sensing, VOL. 32, July 1994.

[22] European space agency,
`http://earth.eo.esa.int`

[23] Shane Robert Cloude and Eric Pottier, *An Entropy Based Classification Scheme for Land Applications of Polarimetric SAR*, IEEE transactions on geoscience and remote sensing, vol. 35, NO. 1, January 1997.

[24] Cloude, S.R, *An entropy based classification scheme for polarimetric SAR data*, Geoscience and Remote Sensing Symposium, 1995.

[25] A.L. Blum and P. Langley, *Selection of relevant features and examples in machine learning.* Artificial intelligence, 1997.

[26] Segios Theodoris and Konstantinos Koutrombas, *Pattern recognition*, Academic Press, 2008.

[27] *Norwegian government home pages on the Amazon fund*, `http://www.regjeringen.no/nb/dep/md/tema/klima/klimaogskogprosjektet/norge-brasil-amazonasfond.html?id=593978`

[28] *Home pages of the Amazon fund* `http://www.amazonfund.gov.br/FundoAmazonia/fam/site_en/Esquerdo/doacoes/`

[29] W.G. Rees, *Physical principles of remote sensing*, Press syndicate of the University of Cambridge, 1990.

[30] Yumiko Wada and Ryosuke Shibasaki, *Estimation of forest biomass density in Kalimantan Island using JERS-1 SAR data*, ACRS 1999, `http://www.a-a-r-s.org/acrs/proceeding/ACRS1999/Papers/PS299-1.htm`

[31] Hasi Bagan, Member, IEEE, Tsuguki Kinoshita, and Yoshiki Yamagata, *Combination of AVNIR-2, PALSAR, and Polarimetric Parameters for Land Cover Classification*, IEEE transactions on geoscience and remote sensing, vol. 50, NO. 4, April 2012.

[32] Anne Lönnqvist, Yrjö Rauste, Matthieu Molinier, and Tuomas Häme, *Polarimetric SAR Data in Land Cover Mapping in Boreal Zone*, IEEE transactions of geoscience and remote sensing, vol. 48, NO. 10, October 2010.

[33] Emerson Luiz Servello,Tatiana Mora Kuplich and Yosio Edemir Shimabukuro, *Tropical Land Cover Change Detection with Polarimetric SAR data*, Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International.

[34] Henning Skriver, *Signatures of Polarimetric Parameters and their Implications on Land Cover Classification*, Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007.

[35] Alexandra Rusanova and Irina Smirnova, *Evalutation of the potential of th ALOS PALSAR polarimetric data for land cover classification in the northern part of Timan-Pechorian petroleum province*, Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International .

[36] Tatiana Mora Kuplich, Yosio Edemir Shimabukuro, Emerson Servello and Edson Sano, *Polarimetric signatures and classification of tropical land covers*, Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009.

[37] Ken Yoong LEE, Soo Chin LIEW, Leong Keong KWOH, *Land Cover Classification of Polarimetric Synthetic Aperture Radar (POLSAR) Data Based on Scattering Mechanisms and Complex Wishart Distribution*, Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International.

[38] P. Lumsdon, S.R. Cloude and G. Wright, *Polarimetric classification of land cover for Glen Affric radar project*, Radar, Sonar and Navigation, IEE Proceedings - Volume:152, Issue: 6, 9 December 2005.

[39] B. Ripka and O. K. Atwood, *Accuracy Assessment of Polarimetric SAR Land Cover Classification for Boreal Regions with Moderate Topography*, Synthetic Aperture Radar (EUSAR), 2010 8th European Conference, 7-10 June 2010.

[40] Zhongsheng Xia, Maosong Xu, Chou Xie, Ridha Touzi, Fengli Zhang, Huazhe Gong and Wei Tian, *Karst forest type discrimination in southwest China using spaceborne polarimetric SAR data*, Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009, 12-17 July 2009.

[41] Gulab Singh and G. Venkataraman, *LOS Palsar Data Analysis of Snow Cover Area in Himalayan Region Using Four Component Scattering Decomposition Technique*, Conference on Recent Advances in Microwave Theory and Applications, 21-24 Nov. 2008.

[42] Corina da Costa Freitas, Luciana de Souza Soler, Sidnei João Siqueira Sant'Anna, Luciano Vieira Dutra, João Roberto dos Santos, José Claudio Mura and Antônio Henrique Correia, *Land Use and Land Cover Mapping in the Brazilian Amazon Using Polarimetric Airborne P-Band SAR Data*, IEEE transcation on geoscience and remote sensing, vol. 46, NO. 10, October 2008.


UNIVERSITY
OF TROMSØ