

**UCGE Reports
Number 20347**

Department of Geomatics Engineering

**Analysis of Radarsat-2 Full Polarimetric Data for
Forest Mapping**

(URL: <http://www.geomatics.ucalgary.ca/graduatetheses>)

by

Yasser Maghsoudi

December 2011



THE UNIVERSITY OF CALGARY

Analysis of Radarsat-2 Full Polarimetric Data for Forest Mapping

by

Yasser Maghsoudi

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF PhD

DEPARTMENT OF GEOMATICS

CALGARY, ALBERTA

December, 2011

© Yasser Maghsoudi 2011

Abstract

Forests are a major natural resource of the Earth and control a wide range of environmental processes. Forests comprise a major part of the planet's plant biodiversity and have an important role in the global hydrological and biochemical cycles. Among the numerous potential applications of remote sensing in forestry, forest mapping plays a vital role for characterization of the forest in terms of species. Particularly, in Canada where forests occupy 45% of the territory, representing more than 400 million hectares of the total Canadian continental area. In this thesis, the potential of polarimetric SAR (PolSAR) Radarsat-2 data for forest mapping is investigated.

This thesis has two principle objectives. First is to propose algorithms for analyzing the PolSAR image data for forest mapping. There are a wide range of SAR parameters that can be derived from PolSAR data. In order to make full use of the discriminative power offered by all these parameters, two categories of methods are proposed. The methods are based on the concept of feature selection and classifier ensemble. First, a nonparametric definition of the evaluation function is proposed and hence the methods NFS and CBFS. Second, a fast wrapper algorithm is proposed for the evaluation function in feature selection and hence the methods FWFS and FWCBFS. Finally, to incorporate the neighboring pixels information in classification an extension of the FWCBFS method i.e. CCBFS is proposed. The second objective of this thesis is to provide a comparison between leaf-on (summer) and leaf-off (fall) season images for forest mapping. Two Radarsat-2 images acquired in fine quad-polarized mode were chosen for this study. The images were collected in leaf-on and leaf-off seasons. We also test the hypothesis whether combining the SAR parameters obtained from both images can provide better results than either individual datasets. The rationale for this combination is that every dataset has some parameters which

may be useful for forest mapping.

To assess the potential of the proposed methods their performance have been compared with each other and with the baseline classifiers. The baseline methods include the Wishart classifier, which is a commonly used classification method in PolSAR community, as well as an SVM classifier with the full set of parameters. Experimental results showed a better performance of the leaf-off image compared to that of leaf-on image for forest mapping. It is also shown that combining leaf-off parameters with leaf-on parameters can significantly improve the classification accuracy. Also, the classification results (in terms of the overall accuracy) compared to the baseline classifiers demonstrate the effectiveness of the proposed nonparametric scheme for forest mapping.

Acknowledgements

First of all, I would like to take this opportunity to express my deep gratitude to my supervisor Dr. Michael Collins for his guidance, advice and invaluable assistance during my studies at the university of Calgary. Without his kind direction and constant encouragements, I would have never been able to develop my project.

I would like to extend my appreciation to Dr. Donald Leckie for his many valuable comments and constructive suggestions. His expert knowledge of the Petawawa Research Forest was very helpful in this research.

I also wish to acknowledge the Canadian Forest Service, Natural Resource Canada in Victoria BC and Canadian Space Agency. They provided me valuable images and field data as well as advice on forest conditions in Petawawa. I would like to thank Dave Hill for helping me on the preparation of the ground truth data and pre-processing of Radarsat-2 images.

I also thank the fellow graduate students from the department of Geomatics Engineering, particularly the fellow students in Trailer H.

Finally, I would like to thanks my parents for their strong support of my studies, for their courage, for standing behind me from thousands of miles away. Mom and Dad you have been and always will be a source of inspiration. You stood over my shoulder whenever I needed you.

Contents

Approval Page	ii
Abstract	iii
Acknowledgments	v
Contents	vi
List of Tables	viii
List of Figures	ix
Epigraph	xi
1 Introduction	1
1.1 Background	1
1.2 Motivations and Innovations	2
1.3 Organization of the proposal	6
2 Polarimetry and Polarimetric Decomposition	8
2.1 Introduction	8
2.2 SAR Polarimetry	9
2.2.1 Basics	9
2.2.2 Stokes vs. Jones Formalism	12
2.2.3 Scattering Matrix vs. Muller Matrix	16
2.2.4 Polarization Basis Transformation	21
2.2.5 Deterministic vs. Non-Deterministic Scatterers	23
2.3 Polarimetric Decomposition	25
2.3.1 Coherent Decomposition	26
2.3.2 Incoherent Decomposition	31
2.4 SAR discriminators	42
2.5 summary	44
3 PolSAR Classification	47
3.1 Supervised vs. unsupervised	48
3.1.1 Supervised	48
3.1.2 Unsupervised	50
3.2 Full covariance matrix vs. selected features	54
3.3 Single- vs. multi-classifiers	57
3.4 Incorporating spatial context	59
3.5 Parametric vs. nonparametric classifiers	61
3.6 Summary	63

4	Methods	64
4.1	Nonparametric methods	65
4.1.1	Single Classifier: NFS	65
4.1.2	Multiple Classifier System: CBFS	72
4.2	Wrapper methods	75
4.2.1	Single Classifier: FWFS	77
4.2.2	Multiple Classifier System: FWCBFS	79
4.3	Contextual Multiple Classifier System: CCBFS	79
5	Study area and dataset description	82
5.1	Study area description	82
5.2	Dataset description	83
5.2.1	Ground truth data	83
5.2.2	SAR data	87
5.3	Data preprocessing	89
5.3.1	Speckle filtering	91
5.3.2	SAR parameters	93
5.3.3	Georeferencing	93
6	Results and Discussion	96
6.1	Nonparametric Methods: Single (NFS) vs. Multiple(CBFS) Classifiers	97
6.2	Wrapper Methods: Single (FWFS) vs. Multiple (FWCBFS) Classifiers	105
6.3	Context-based Method: CCBFS	110
6.4	Summary	116
7	Conclusions and Further Works	120
7.1	Conclusions	120
7.2	Further Works	121
	Bibliography	123

List of Tables

2.2	polarimetric parameters used in this research	46
5.2	List of classes and training and testing sample sizes	87
5.4	Overview of the principal characteristics of the RS-2 data	88
5.7	Acquired image details	89
5.8	PolSAR parameters used in this thesis	94
6.1	Cross tabulation of number of correct and wrongly classified pixels for two classifications	97
6.2	the calculated p-values for the leaf-on vs. leaf-off	99
6.3	the calculated p-values for the leaf-off vs. leaf-on-off	99
6.4	classification accuracies in different classes using the combined leaf-on-off dataset	102
6.5	selected features (the superscript 1 and 2 in the third row refer to leaf-on and leaf-off features respectively)	103
6.6	the calculated p-values for the leaf-on vs. leaf-off	106
6.7	the calculated p-values for the leaf-off vs. leaf-on-off	106
6.8	selected features (the superscript 1 and 2 in the third row refer to leaf-on and leaf-off features respectively)	109
6.9	selected features for each class (the superscript 1 and 2 in the third row refer to leaf-on and leaf-off features respectively)	111
6.10	classification accuracies in different classes using the combined leaf-on-off dataset	114
6.11	classification accuracies in different classes using the combined leaf-on-off dataset	117

List of Figures

1.1	Research outline	7
2.1	The propagation of an EM wave. The Electric field vector (red) comprises horizontal (green) and vertical (blue) components (adapted from [83])	10
2.2	Polarization ellipse (adapted from [83])	11
2.3	The Poincare Sphere	14
2.4	Coherent response of a resolution cell: a) without a dominant scatterer b) with a dominant scatterer	23
2.5	Decomposition taxonomy	26
3.1	A review of PolSAR classification methods	47
3.2	Entropy-alpha feature space (adapted from [79])	51
4.1	Steps taken to analyze RS-2 images for forest mapping	65
4.2	Separating hyperplanes in SVM: left (linear case) and right (non-linear case)	69
4.3	The proposed multi-classifier schema	76
4.4	simple (<i>a</i>) vs. spatially consistent (<i>b</i>) combination rules	81
5.1	Petawawa Research Forest (adapted from Google maps)	83
5.2	The forest inventory stands overlaid on high resolution images obtained from Google earth	85
5.3	Undesirable polygons:(<i>a</i>) mixture of red pine and poplar (<i>b</i>) edgy polygon(<i>c</i>) clear-cut area(<i>d</i>) partial cut area (<i>e</i>) low density area	86
5.4	leaf-off (<i>a</i>) and leaf-on (<i>b</i>) images in Pauli basis R: $ HH - VV $, G: $ HV $ and B: $ HH + VV $	90
5.5	The T_{11} image before (left) and after (right) applying speckle filter	93
6.1	classification accuracy vs. number of features in three datasets	100
6.2	Quantitative comparison of nonparametric methods with baseline methods using the three datasets	100
6.3	classification accuracy vs. number of features in the leaf-on, leaf-off and combined leaf-on-off datasets using the FWFS method	107
6.4	Quantitative comparison of wrapper methods with baseline classifiers using the three datasets	108
6.5	FWCBFS vs. FWFS using the combined leaf-on-off dataset	108
6.6	the leaf-on features selected by FWCBFS and their frequency	112
6.7	the leaf-off features selected by FWCBFS and their frequency	113
6.8	increasing the window size in CCBFS method	115
6.9	the window size effect over different classes: solid lines represent the classes with fairly large number of samples and dotted lines indicate the classes with small number of samples	116

6.10	summary of the results: the classification accuracies in nonparametric, wrapper and context-based methods in different classes. Wishart classification results in different classes served as the baseline and the differences in classification accuracies with respect to the baseline are plotted.	119
------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----

Glory be to Almighty God who is the source of all knowledge, wisdom
and understanding.

Chapter 1

Introduction

1.1 Background

Forest mapping is one of the core applications in remote sensing. Many studies are based on multispectral optical images. Also, the use of hyperspectral data are increasing due to their increased information content. Unfortunately, the weather conditions limit the use of those optical images. On the other hand the Synthetic Aperture Radar (SAR) data are not only independent of the weather conditions, they are also sensitive to target geometry. For these reasons, these properties make the SAR data a useful tool for applications such as forest mapping.

The extraction of information from SAR data has been an active area of research for many years and they are becoming more and more important in remote sensing applications. Despite all their advantages, the use of SAR data are subject to geometric limitations which affects the SAR data by shadows, layover, foreshortening and variation in pixel resolution over the image. In addition, because of the coherent nature of the SAR sensors, the images suffers from speckle noise [94]. Because of these reasons, it is very difficult to obtain satisfactory results from classification of single channel SAR data even if advanced classification techniques are used.

In order to achieve reliable results, multichannel measurements are generally necessary. The multichannel datasets can be achieved by using multi-temporal data [104, 135], multi-frequency data [70], multi-polarization data [38], fusion of different SAR sensors [117] and fusion of SAR data and optical data [21]. A combination of these approaches are also considered in several studies [21, 16, 32, 41]. While multi-temporal, multi-frequency and multi sensor approaches are widely used and fairly well documented, SAR polarimetry is a relatively new approach which yields some

advantages over the conventional methods. In this context, the remote sensing community has become increasingly interested in the use of polarimetric SAR data for the land cover mapping and forest mapping.

Polarimetric data allow to identify different classes by analyzing the multi-polarization of the backscattering coefficient; on the other hand, there are a lot of SAR parameters which can be extracted from the polarimetric data. These parameters are either derived using the well-known decomposition methods, or obtained directly from original data, or can be the SAR discriminators. The use of these parameters can improve the separability of the classes in the feature space. For these reasons, polarimetric SAR data have become a relatively operational tool for classification problems.

Extraction of SAR parameters is a primary step prior to classification. Polarimetric target decomposition (TD) is developed to separate polarimetric radar measurements into basic scattering mechanisms. TD methods are often categorized as coherent and incoherent methods. The coherent methods are based on the scattering matrix that possess 5 independent parameters while incoherent methods are based on the incoherently averaged covariance or coherency matrices that have 9 independent parameters. These target decomposition parameters along with the original data features and SAR discriminators are fully explained in section 2.3.

1.2 Motivations and Innovations

There are a wide range of SAR parameters (features) which can be extracted from polarimetric SAR data. Target decomposition theory laid down the basis for the classification of polarimetric SAR images and almost all classification algorithms are based on them. Particularly, the formalism worked out by Cloude [30] led to the introduction of an unsupervised classification scheme [29], further augmented and improved by subsequent contributions [99, 78, 80].

Despite the significant number of works carried out for terrain and land-use clas-

sification [108, 32, 16, 29, 10], very few researches have been performed to investigate the potential of C-band polarimetric data for forest classification [128] [58]. Proisy et al. [102] in 2000 employed Radarsat-1 HH polarization and ERS-1 VV polarization for forest mapping and concluded that the C-band SAR data can not provide a good discrimination of forest species. Touzi et. al. [128] in 2006 used the airborne C-band polarimetric Convair-580 SAR data for forest classification. They showed that using radiometric intensity of the conventional C-band SAR polarizations HH, HV, and VV can only perform a limited discrimination of various tree species. However, the polarization information provided by fully polarimetric SAR clearly improved forest type discrimination under leafy and no-leaves conditions and permits the demonstration of the significance of SAR illumination angle on forest scattering mechanisms.

Although a large variety of works have already taken place for the classification of polarimetric SAR images, most of them have concentrated on the use of a very limited number of features. For instance, in the method proposed by Cloude and Potter [33], which is one of the most used approaches, the polarimetric information is converted into three parameters (entropy H , α -angle and anisotropy A) each of which have been associated to an elegant physical interpretation. Then, they subdivided the feature space formed by the three parameters into regions that corresponds to distinct scattering behaviors (See chapter 3 for a review of various SAR polarimetric data classification approaches). In very complex scenes like forests it is very useful to make full use of the discriminative power offered by all these features. However, due to the small training sample size problem, using all these features for the classification is not feasible. Furthermore, some of these features might carry redundant information. Therefore, a key stage in a classifier design is the selection of most discriminative and informative features. Most of the SAR parameters are of complex and sometimes unknown statistical properties. For this reason, the conventional fea-

ture selection algorithms cannot be applied. To account for this, a new classification approach, which is based on a nonparametric feature selection (NFS) and support vector machine classifier is proposed. The feature selection process generally involves a search strategy and an evaluation function. In this research, the sequential forward floating selection (SFFS) [64] method will be used as the search algorithm to generate subsets of features from the original features. For the evaluation function, a non-parametric separability measure will be adopted to evaluate the generated feature subsets. To formulate the criteria of class separability, the between-class and within class scattering matrices have to be calculated. The employed separability measure is the ratio of the determinant of the between-class scatter matrix to the determinant of the sum of within-class scatter matrices. Upon the selection of the most appropriate features, they are transferred to the classification step. Because of its ability to take numerous and heterogeneous features into account, as well as its ability handle linearly non separable cases, the support vector machine (SVM) algorithm is proposed as the classifier.

Most of the feature selection algorithms seek only one set of features that distinguish among all the classes simultaneously and hence a limited amount of classification accuracy. Recently, there has been a great interest for using an ensemble of classifiers for solving problems in pattern recognition community. Thus, in order to take advantage of heterogeneous features provided by the polarimetric target decomposition and hence to improve the classification accuracy, a multi classifier schema is used for the next part of this research. In doing so, a class-based feature selection (CBFS), which is based on the theory of multiple classifiers, is proposed. In this schema, instead of using feature selection for the whole classes, the features are selected for each class separately. The selection is based on the calculation of the determinant of the between-class scatter matrix to the determinant of within-class scatter matrices for a

specific class. It should be noted that unlike the previous method, the between class scatter matrix is defined as the distance between that specific class and the rest of classes (and not the distance between all classes). Also, instead of the determinant of the sum of within-class scatter matrices, the determinant is calculated for the class of interest. Afterwards, an SVM classifier is trained on each of the selected feature subsets. Finally, the outputs of the classifiers are combined through a combination mechanism. The proposed schema was already successfully tested in our previous works for the classification of hyperspectral images [86] and multitemporal Radarsat-1 images [87]. In this study, we will investigate the potential of the CBFS method for the classification of polarimetric SAR images.

The inter-class distance measures as the evaluation function of the feature selection although a reasonable method of the similarity and dissimilarity, they are not directly related to the ultimate classification accuracy. A question arises as to whether it is possible to use a more direct criterion as the evaluation function. According to the evaluation function, the feature selection approaches can be broadly grouped into filter and wrapper methods [69]. Wrappers utilize the classification accuracy as the evaluation function whereas filters uses the inter-class distance measures as the evaluation function. The optimized problem in filters is different from the real problem. Nonetheless, filters are faster because the problem they solve is in general simpler. Alternatively, wrappers try to solve the real problem and the considered criterion is really optimized which means that the ultimate problem has to be solved numerous times. For this reason wrappers are potentially very time consuming. This time complexity in our work is mainly due to the SVM training time. In this thesis, in order to alleviate this problem, a fast wrapper feature selection (FWFS) is proposed. We tried to reduce the training time by reducing the number of training samples. The number of support vectors for each class was taken as the degree of training

reduction for that class. Like the NFS method, the method was used in the context of a class-based feature selection schema and hence the name FWCBFS.

One of the disadvantages of the classification methods described so far is that each pixel is classified independently of its neighbors. When a majority of pixels in a certain region are assigned to the a single class, it becomes highly unlikely that a pixel in this region belongs to another class. This misassignment could likely be due to the speckle noise. Therefore, in the last part of this research we will take the spatial consistency into account. In particular, in the combination mechanism of the multi classifier schema we will incorporate the spatial context. A more detailed explanation of these methodologies can be found in chapter 4. In summary, all the proposed methodologies are illustrated in figure 1.1.

1.3 Organization of the proposal

This thesis is organized in seven chapters. In the previous sections, the background and motivation of the research being conducted are explained. In Chapter 2, first some of the basic concepts on polarimetry (section 2.2) are explained. Next, an overview on target decomposition methods is given based on a categorization scheme (section 2.3). Some of the well-known approaches for the classification of polarimetric SAR images are then explained in chapter 3. Chapter 4 focuses on the description of the proposed methodologies. Then, a description of the dataset, study area and the preprocessing results are given in chapter 5. Discussion of the experimental results and their analysis are given in chapter 6. Finally, conclusions and further works are described in chapter 7.

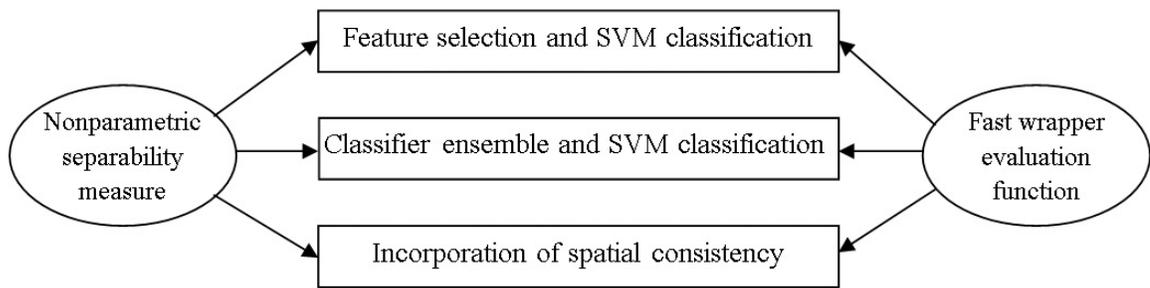


Figure 1.1: Research outline

Chapter 2

Polarimetry and Polarimetric Decomposition

2.1 Introduction

Radar polarimetry is the science and technology of acquiring, processing and analyzing the polarization state of an electromagnetic field [9]. The complex structure of forests manifests itself in polarimetric SAR (PolSAR) data by a challenging diversity of scattering mechanisms. This extensive source of polarization information contained in fully polarimetric SAR data shows great potential for measuring forest scattering characteristics and producing separation between different forest species. In order to understand the information content of polarimetry and the physical relationship of polarimetric parameters with natural media, such as forests, a theoretical grounding is necessary.

The main goal of this chapter is to explain the SAR parameters that can be extracted from PolSAR data. These are the parameters that will be used as the input features in the forest classification step next. Therefore, a detailed understanding of these parameters plays a vital role in interpretation and evaluation of the final results. The polarimetric features can be divided into three categories: the features obtained directly from original data, the features which are derived using the well-known decomposition methods, and the SAR discriminators.

The first part of this chapter (section 2.2) provides a mathematical and physical background to polarimetry which is essential for understanding the PolSAR parameters. Also, the explanation of the original data features i.e. scattering matrix, covariance matrix and coherency matrix can be found in this section. Section 2.3 deals with different target decomposition methods. Within this framework, different coherent and incoherent decompositions are explained. There are several quantities

derived from PolSAR data to be used as indicators to discriminate among surface types or land covers. These SAR discriminators are explained in section 2.4.

The main reason to consider different SAR parameters is to take advantage of the complementary information provided by the different parameters in the field of forest classification. For instance, incoherent algorithms performs an averaging of the returned signals. This provides a statistically smoother description of the behaviour of the scatterers. These parameters are mainly believed to be better for randomly distributed targets. For instance, within forests they can distinguish between volumes and surface scatterers. It has also been shown that some of the incoherent parameters are very promising for forest structure characterization and detection of forest changes between leaf-on and leaf-off conditions [126]. On the contrary, the coherent methods can be used to maintain the full resolution. Indeed, by avoiding the averaging at the first stage, the coherent methods are able to extract features that may be lost in the early averaging step. By combining these parameters, various forest features are emphasized. Coherent methods focuses on the more detailed characteristics of the trees such as the leaves, branches and twigs whereas the incoherent parameters reveals the high scale components e.g. the canopy structure.

Because of these reasons, it is expected that the concurrent application of both coherent and incoherent methods would provide complementary information and hence improve the forest classification results.

2.2 SAR Polarimetry

2.2.1 Basics

Polarization is an important property of a plane electromagnetic (EM) wave. It refers to the alignment and regularity of the electric field component of the wave, in a plane perpendicular to the direction of propagation (as the magnetic field is directly related to electric field and can always be obtained from it, we direct our attention to the

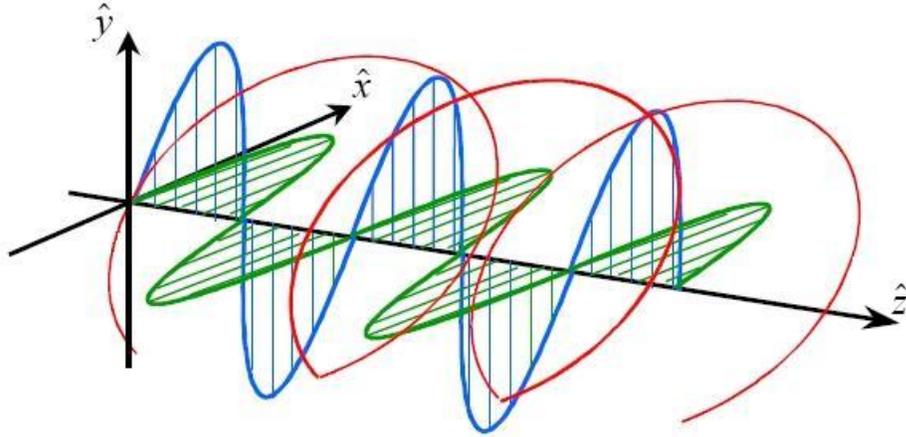


Figure 2.1: The propagation of an EM wave. The Electric field vector (red) comprises horizontal (green) and vertical (blue) components (adapted from [83])

electric field component). Figure 2.1 illustrates the propagation of an EM wave.

The Electric field of a plane wave can be described by the sum of two orthogonal components, i.e horizontal and vertical components [15]. The path of the end point of the Electric wave vector traces out an ellipse in its general form as shown in figure 2.2. The ellipse has a semi-major axis of length E_o^x and a semi-minor axis of length E_o^y . A is the wave amplitude, $\phi(0^\circ \leq \phi \leq 180^\circ)$ is the orientation angle which is the angle of the semi-major axis, $\tau(0 \leq \tau \leq 45^\circ)$ is the ellipticity angle, defined as $\tau = \arctan(E_o^y/E_o^x)$. τ describes the degree to which the ellipse is oval (figure 2.2). The magnitudes and relative phase between the horizontal and vertical components of the Electric field vector governs the shape of the ellipse. For instance, when the components are in phase ($\tau = 0$), the polarization ion is linear . As the relative phase angle increases to 90° , the ellipticity increases to 45° , representing circular polarization.

The propagation of a plane EM wave is described in a three dimensional space with the coordinates given by the three axes x , y and z . The z axis is in the direction of propagation, while the x and y axes lie in a plane perpendicular to the direction

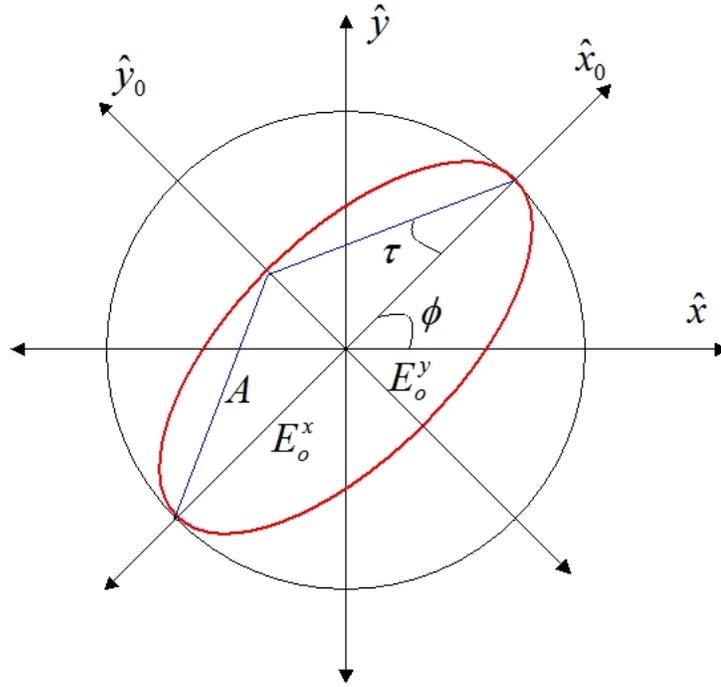


Figure 2.2: Polarization ellipse (adapted from [83])

of propagation, with (x, y, z) forming a right hand orthogonal set. In scattering situations, the coordinate space has to be defined for both the incident wave and the scattered wave. There are two common conventions: the forward scatter alignment (FSA) and the backscatter alignment (BSA). The FSA takes the positive z -axis in the same direction as the travel of the wave for both the incident and scattered wave, whereas, in BSA, the positive z -axis is towards the target for both the incident and scattered wave. Therefore, in both conventions, the z -axis points to the same direction for the incident wave, but in opposite directions for the scattered wave. For the monostatic radar, the coordinate systems are the same for the incident and scattered wave in the BSA convention, so it is more commonly used for imaging radars.

2.2.2 Stokes vs. Jones Formalism

In order to characterize the polarization state of a plane wave, George Stokes in 1852 introduced a 4-element vector, $\begin{bmatrix} S_0 & S_1 & S_2 & S_3 \end{bmatrix}^T$, known as the Stokes vector.

$$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} |E_y|^2 + |E_x|^2 \\ |E_y|^2 - |E_x|^2 \\ 2Re\{E_y E_x^*\} \\ 2Im\{E_y E_x^*\} \end{bmatrix} = \begin{bmatrix} |E_y|^2 + |E_x|^2 \\ |E_y|^2 - |E_x|^2 \\ 2E_y E_x \cos \delta \\ 2E_y E_x \sin \delta \end{bmatrix} = \begin{bmatrix} S_0 \\ S_0 \cos 2\phi \cos 2\tau \\ S_0 \sin 2\phi \cos 2\tau \\ S_0 \sin 2\tau \end{bmatrix} \quad (2.1)$$

where E_y and E_x are the vertical and horizontal components of Electric field, $\delta = \delta_x - \delta_y$ is the phase difference between E_y and E_x , ϕ and τ are the orientation and ellipticity angles respectively, $|\cdot|$ is the absolute value and $*$ is the complex conjugate. This formalism indicates that the polarization state of a plane wave can be described by orientation and ellipticity, plus a parameter S_0 . S_0 is proportional to the total intensity of the wave, S_1 is the difference between the density powers related to the horizontal and vertical polarizations. Parameters S_2 and S_3 are related to the phase difference between the horizontal and vertical components of the electric field.

An EM plane wave can be completely polarized, partially polarized and completely unpolarized. In the completely polarized case, only 3 of the Stokes parameters are independent because

$$S_0^2 = S_1^2 + S_2^2 + S_3^2 \quad (2.2)$$

In this case, a geometrical interpretation of the Stokes parameters is used to map the polarization state on a sphere with radius S_0 and Cartesian coordinates (S_1, S_2, S_3) . This visualized representation of the polarization state of the wave, which is called the Poincare sphere, is shown in figure 2.3. The latitude and longitude of a point on the sphere corresponds to 2τ and 2ϕ . Based on this notation, the linear polarizations

lie on the equator, with horizontal and vertical polarizations opposite each other. Left hand and right hand circular polarizations lie on the north and south poles respectively. All other points on the sphere refer to the elliptical polarizations with different τ and ϕ . Points on the sphere which are opposite to each other are referred to as cross polarizations and they represent polarizations that are orthogonal to each other.

For partially polarized waves, not all the superficial density powers is contained in the polarized components and thus the total intensity of the wave is greater than the polarized components

$$S_0^2 > S_1^2 + S_2^2 + S_3^2 \quad (2.3)$$

The degree of polarization is the ratio of the polarized power to the total power

$$p = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \quad (2.4)$$

If the EM wave is partially polarized, it can be expressed as the sum of a completely polarized wave and a completely unpolarized or noise-like wave as follow

$$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} 1-p \\ 0 \\ 0 \\ 0 \end{bmatrix} + S_0 p \begin{bmatrix} 1 \\ \cos 2\phi \cos 2\tau \\ \sin 2\phi \cos 2\tau \\ \sin 2\tau \end{bmatrix} \quad (2.5)$$

in which the first term and the second term in the right side represent the completely unpolarized and completely polarized components.

The Jones vector [65] is another formalism for characterizing the polarization state of a plane wave. In this formalism, instead of a 3D real space, as of the Stokes vector, a 2D complex space is used. The Electric field of a wave propagating in the z can be

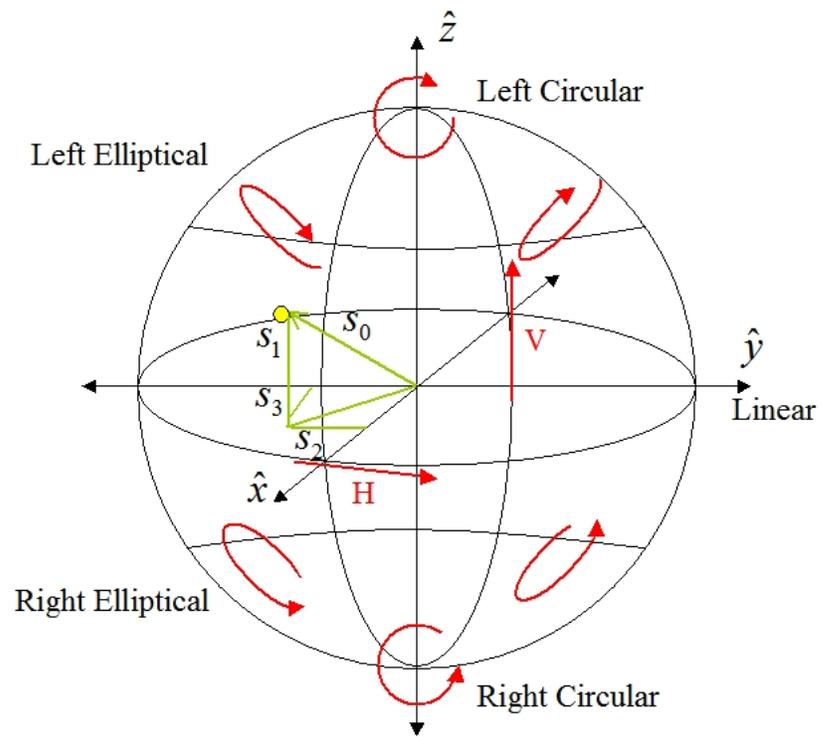


Figure 2.3: The Poincare Sphere

written as

$$\vec{E}(z, t) = \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \begin{pmatrix} E_{ox} \cos(\omega t - kz - \delta_x) \\ E_{oy} \cos(\omega t - kz - \delta_y) \\ 0 \end{pmatrix} = \begin{pmatrix} E_{ox} \exp(-jkz) \exp(j\delta_x) \\ E_{oy} \exp(-jkz) \exp(j\delta_y) \\ 0 \end{pmatrix} \quad (2.6)$$

where E_x, E_y, E_z are the real electric field vector components, E_{ox}, E_{oy} are the Cartesian components of the real electric field vector $\vec{E}(z, t)$, ω is the angular frequency, t is the time, k is the wave number, and δ_x and δ_y are the x and y phases of the electric field components when projected onto the $x - y$ plane.

The Jones vector is the phasor of the real electric field vector given by 2.6

$$J = \begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} E_{ox} \exp(j\delta_x) \\ E_{oy} \exp(j\delta_y) \end{pmatrix} \quad (2.7)$$

The Jones vector can be written as

$$J = E_{ox} \begin{pmatrix} 1 \\ \rho \end{pmatrix} \quad (2.8)$$

in which ρ is called complex polarization ratio [1]

$$\rho = \frac{E_{oy}}{E_{ox}} \cdot e^{j(\delta_y - \delta_x)} = \frac{\cos(2\tau) + j \sin(2\tau)}{1 - \cos(2\phi) \cos(2\tau)} \quad (2.9)$$

Assuming $|J| = 1$, the following equation can be obtained as the normalized Jones vector \hat{J} [119]

$$\hat{J}(\tau, \phi) = \begin{pmatrix} \cos(\phi) \cos(\tau) - j \sin(\phi) \sin(\tau) \\ \sin(\phi) \cos(\tau) + j \cos(\phi) \sin(\tau) \end{pmatrix} \quad (2.10)$$

Unlike the Stokes formalism, Jones vector does not depend on the intensity power of

the Electric field. For linear polarization ($\tau = 0$), the Jones vector is only dependent on the orientation angle

$$\hat{J}_{lin}(\phi) = \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} \quad (2.11)$$

Alternatively, the circular ($\tau = 45$) right-hand and left-hand polarization can be expressed as

$$\hat{J}_R = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -j \end{pmatrix} \quad \hat{J}_L = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ j \end{pmatrix} \quad (2.12)$$

2.2.3 Scattering Matrix vs. Muller Matrix Scattering Matrix

Fully polarimetric radar antennas are able to transmit and then receive waves in both x and y polarizations. Horizontal (H) and Vertical (V) polarizations are often chosen as x and y . The H polarization is first transmitted and both backscattered H and V polarizations are simultaneously received. Also, the V polarization is transmitted and both H and V backscattered polarizations are received simultaneously. Thus, for a given ground resolution cell, all four transmitting and receiving configurations are recorded at the same time.

There are two ways for representing the scattering behavior of the target: Scattering matrix and the Muller matrix. These matrices can be used to relate the backscattered wave to the incident wave. The scattering matrix is based on the Jones formalism whereas Muller matrix is based on Stokes formalism.

Given the Jones vectors of the incident and the scattered waves, \vec{E}^i and \vec{E}^s respectively, the scattering process occurring at the target of interest is [116]

$$\vec{E}^s = \frac{e^{jkr}}{r} [S] \vec{E}^i = \frac{e^{jkr}}{r} \begin{pmatrix} S_{HH} & S_{VH} \\ S_{HV} & S_{VV} \end{pmatrix} \vec{E}^i \quad (2.13)$$

where r is the distance between target and antenna, $[S]$ is a complex scattering matrix, called Sinclair matrix. The elements of $[S]$ are complex as $S_{ij} = |S_{ij}| e^{j\phi_{ij}}$ where $i, j \in \{H, V\}$. The diagonal elements of the scattering matrix receive the name of co-polar terms, since they relate the same polarization for the incident and the scattered fields. The off-diagonal elements are known as cross-polar terms as they relate orthogonal polarization states. Finally, the term $\frac{e^{jkr}}{r}$ takes into account the propagation effects both, in amplitude and phase.

It can be deduced from equation 2.13 that the characterization of a given target by means of the scattering matrix allows the possibility to explore the phase information provided by the phase of complex scattering coefficients, and not only the intensity or amplitude. It can be concluded that polarimetry opens the door to consider phase measurements to characterize the targets. There are objects which cannot be differentiated in terms of the radar cross section coefficients, whereas they are seen as different objects if they are analyzed by means of the corresponding scattering matrices (an example for this are trihedral and dihedral objects).

Since the scattering matrix $[S]$ is employed to characterize a given target, it can be parametrized as follows

$$[S] = \frac{e^{-jkr}}{r} \begin{pmatrix} |S_{HH}| e^{j\varphi_{HH}} & |S_{VH}| e^{j\varphi_{HV}} \\ |S_{HV}| e^{j\varphi_{VH}} & |S_{VV}| e^{j\varphi_{VV}} \end{pmatrix} = \quad (2.14)$$

$$\frac{e^{-jkr} e^{j\varphi_{HH}}}{r} \begin{pmatrix} |S_{HH}| & |S_{VH}| e^{(j\varphi_{HV} - \varphi_{HH})} \\ |S_{HV}| e^{j(\varphi_{VH} - \varphi_{HH})} & |S_{VV}| e^{(j\varphi_{VV} - \varphi_{HH})} \end{pmatrix}$$

The absolute phase term in 2.14 is not considered as an independent parameter since it presents an arbitrary value due to its dependence on the distance between the radar and the target. Consequently, it is assumed that the scattering matrix can be parametrized by 7 independent parameters: four amplitudes $|S_{HH}|, |S_{HV}|, |S_{VH}|, |S_{VV}|$

and three relative phases $(\varphi_{HV} - \varphi_{HH}), (\varphi_{VH} - \varphi_{HH}), (\varphi_{VV} - \varphi_{HH})$. As a conclusion, a given target of interest is determined by 7 independent parameters in the most general case and an absolute value. Usually, in SAR applications a monostatic configuration is used in which the transmitter and receiver use the same antenna which for reciprocal targets states that

$$S_{HV} = S_{VH} \quad (2.15)$$

One important property of this configuration is that a given target is characterized by 5 independent parameters namely three amplitudes $\{|S_{HH}|, |S_{HV}|, |S_{VV}|\}$ and two relative phases $\{(\varphi_{HV} - \varphi_{HH}), (\varphi_{VV} - \varphi_{HH})\}$ and one additional absolute phase.

The scattering matrix can be represented by the following four component complex vector [28].

$$\vec{k} = \frac{1}{2} \text{tr}([S]\Psi) = [k_0, k_1, k_2, k_3] \quad (2.16)$$

in which $\text{tr}[S]$ is the sum of the diagonal elements of $[S]$, Ψ is a complete set of 2×2 complex basis matrices. Two bases are commonly used:

- Lexicographic basis

Also referred as Borgeaud basis [10]. It is formed by the following matrices

$$\Psi_B = \left\{ 2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, 2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, 2 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, 2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \quad (2.17)$$

This corresponds to the following complex vector k_B called lexicographic vector

$$\vec{k}_B = \left[S_{HH}, S_{HV}, S_{VH}, S_{VV} \right]^T \quad (2.18)$$

This vector is directly related to the system measurables.

- Pauli basis

This basis is more related to the physics of wave scattering. It is formed by the following matrices [10]

$$\Psi_P = \left\{ \sqrt{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sqrt{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \sqrt{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sqrt{2} \begin{bmatrix} 0 & -i \\ i & 1 \end{bmatrix} \right\} \quad (2.19)$$

The complex vector corresponding to Pauli basis is

$$\vec{k}_P = \frac{1}{\sqrt{2}} \left[S_{HH} + S_{VV}, S_{HH} - S_{VV}, S_{HV} + S_{VH}, j(S_{VH} - S_{HV}) \right]^T \quad (2.20)$$

The norm of the scattering vector \vec{k} is equal to the total scattered power

$$\left| \vec{k} \right|^2 = \vec{k}_P^{*T} * \vec{k}_P = \vec{k}_B^{*T} * \vec{k}_B = (|S_{HH}|^2 + |S_{HV}|^2 + |S_{VH}|^2 + |S_{VV}|^2) \quad (2.21)$$

This justifies the use of factor 2 in 2.17 and the factor $\sqrt{2}$ in 2.19.

Muller Matrix

The Muller matrix is another way of transforming the incident EM wave into the backscattered wave [131]. If \vec{S}_i is the Stokes vector of the incident wave and \vec{S}_s is the Stokes vector of the backscattered wave, these two are related through the Mueller matrix $[M]$ as follows

$$\vec{S}^s = \frac{1}{r^2} \cdot [M] \cdot \vec{S}^i \quad (2.22)$$

The Matrix $[M]$ is a 4×4 real matrix, which is completely defined as [52]

$$[M] = \begin{bmatrix} |S_{HH}|^2 & |S_{HV}|^2 \\ |S_{VH}|^2 & |S_{VV}|^2 \\ 2\text{Re}(S_{HH}S_{VH}^*) & 2\text{Re}(S_{HV}S_{VV}^*) \\ -2\text{Im}(S_{HH}S_{VH}^*) & -2\text{Im}(S_{HV}S_{VV}^*) \\ \text{Re}(S_{HH}S_{HV}^*) & \text{Im}(S_{HH}S_{HV}^*) \\ \text{Re}(S_{VH}S_{VV}^*) & \text{Im}(S_{VH}S_{VV}^*) \\ 2\text{Re}(S_{HH}S_{VV}^* + S_{HV}S_{VH}^*) & \text{Im}(S_{HH}S_{VV}^* + S_{VH}S_{HV}^*) \\ -\text{Im}(S_{HH}S_{VV}^* + S_{HV}S_{VH}^*) & \text{Re}(S_{HH}S_{VV}^* - S_{HV}S_{VH}^*) \end{bmatrix} \quad (2.23)$$

The Mueller matrix is used in the FSA convention. The Kennaugh matrix $[K]$ is the version of the Mueller matrix used in BSA convention. They are related by $[M] = \text{diag}[1, 1, 1, -1][K]$. The trace of the matrix $[K]$ equals the total power, while the trace of the matrix $[M]$ does not. The $[K]$ matrix can be written under the following form

$$[K] = \begin{bmatrix} A_0 + B_0 & C & H & F \\ C & A_0 + B & E & G \\ H & E & A_0 - B & D \\ F & G & D & -A_0 + B \end{bmatrix} \quad (2.24)$$

where the parameters are called Huynen parameters and are given by

$$\begin{aligned}
A_0 &= \frac{1}{4} |S_{HH} + S_{VV}|^2 \\
B_0 &= \frac{1}{4} |S_{HH} - S_{VV}|^2 + |S_{HV}|^2 & B &= \frac{1}{4} |S_{HH} - S_{VV}|^2 - |S_{HV}|^2 \\
C &= \frac{1}{2} |S_{HH} + S_{VV}|^2 & D &= \text{Im}(S_{HH}S_{VV}^*) \\
E &= \text{Re}(S_{HV}^*(S_{HH} - S_{VV})) & F &= \text{Im}(S_{HV}^*(S_{HH} - S_{VV})) \\
G &= \text{Im}(S_{HV}^*(S_{HH} + S_{VV})) & H &= \text{Re}(S_{HV}^*(S_{HH} + S_{VV}))
\end{aligned} \tag{2.25}$$

It should be noted that the $[K]$ matrix is symmetric like the scattering $[S]$ matrix. As the monostatic polarimetric dimension of the target is equal to five, it is concluded that the nine Huynen parameters are related to each other by $(9-5)=4$ equations which are called the monostatic target structure equations and are given by

$$\begin{aligned}
2A_0(B_0 + B) &= C^2 + D^2 \\
2A_0(B_0 - B) &= G^2 + H^2 \\
2A_0E &= CH - DG \\
2A_0F &= CG - DH
\end{aligned} \tag{2.26}$$

Another dependency relationship which will be important in Huynen decomposition (section 2.3.2) is

$$B_0 = B^2 + E^2 + F^2 \tag{2.27}$$

2.2.4 Polarization Basis Transformation

One of the main advantages of radar polarimetry is that once a target response is measured in one polarization basis (transmitting/receiving transformation), the response in any basis can be obtained from by a simple mathematical transformation

without any additional measurements. This can be performed by applying a special unitary transformation as follow

$$S_{(B,B_\perp)} = U^T S_{(A,A_\perp)} U \quad (2.28)$$

where $S_{(B,B_\perp)}$ is the desired Sinclair matrix, $S_{(A,A_\perp)}$ is the current Sinclair matrix and U is the unitary basis transformation matrix defined as

$$U = \begin{pmatrix} \cos(\tau) & -j \sin(\tau) \\ -j \sin(\tau) & \cos(\tau) \end{pmatrix} \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix} = \frac{1}{\sqrt{1 + \rho\rho^*}} \begin{pmatrix} 1 & \rho^* \\ \rho & 1 \end{pmatrix} \quad (2.29)$$

in which ϕ and τ are the relative orientation and ellipticity angles between the two polarization bases. To transform the scattering matrix $[S]$ in the linear (H, V) polarization basis to an arbitrary basis (x, y) using equations 2.28 and 2.29 and assuming 2.15 we have

$$\begin{aligned} S_{xx} &= \frac{1}{\sqrt{1 + \rho\rho^*}} [S_{HH} + 2\rho S_{HV} + \rho^2 S_{VV}] \\ S_{xy} &= \frac{1}{\sqrt{1 + \rho\rho^*}} [\rho S_{HH} + (1 - \rho\rho^*) S_{HV} - \rho^* S_{VV}] \\ S_{yx} &= \frac{1}{\sqrt{1 + \rho\rho^*}} [\rho S_{HH} + (\rho\rho^* - 1) S_{HV} - \rho^* S_{VV}] \\ S_{yy} &= \frac{1}{\sqrt{1 + \rho\rho^*}} [\rho^2 S_{HH} + 2\rho S_{HV} + S_{VV}] \end{aligned} \quad (2.30)$$

It is a common practice to measure the scattering matrix in a linear polarization basis (H, V) and transform to a circular polarization (R, L) , where R and L are the right and left circular polarization respectively. The elements of $[S]_{RL}$ can be calculated as

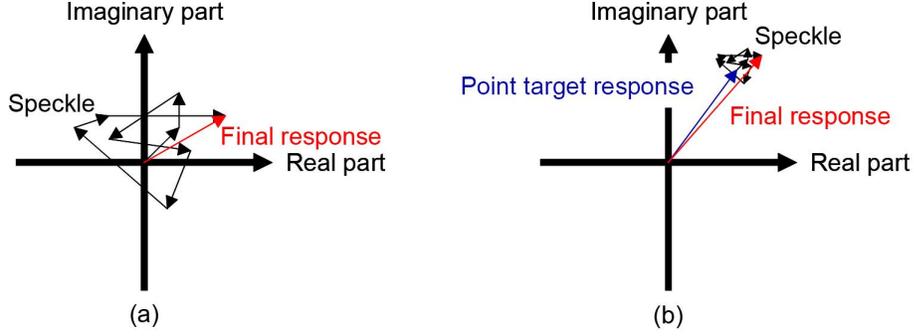


Figure 2.4: Coherent response of a resolution cell: a) without a dominant scatterer b) with a dominant scatterer

$$\begin{aligned}
 S_{RR} &= jS_{HV} + \frac{1}{2}(S_{HH} - S_{VV}) \\
 S_{LL} &= jS_{HV} - \frac{1}{2}(S_{HH} - S_{VV}) \\
 S_{RL} &= \frac{j}{2}(S_{HH} - S_{VV})
 \end{aligned} \tag{2.31}$$

It should be noted that the following properties of the scattering matrix are transformation invariant: *SPAN* of the matrix ($SPAN = |S_{HH}|^2 + |S_{VV}|^2 + 2|S_{HV}|^2$), the determinant of the scattering matrix and the symmetry of the scattering matrix.

2.2.5 Deterministic vs. Non-Deterministic Scatterers

A resolution cell in a SAR image is formed by the coherent addition of the responses of the elementary scatterers. In those cases where there is not a dominant scatterer, the statistic of the response is given by the complex Gaussian scattering model, given rise to the so-called speckle [94]. On the other hand, the resolution cell may contain a point target, which dominates the response of the resolution cell. In this case, the scattering response is due to the coherent combination of two components: the dominant scatterer and the coherent combination due to the clutter. Figure 2.4 compares the resolution cell response with and without the presence of a point scatterer.

Therefore, the SAR response of some resolution cells may be dominated by a

strong contribution of a deterministic scatterer. Indeed, deterministic scatterers can be completely defined by the scattering matrix [S]. Nonetheless, the concept of the deterministic scatterer is not appropriate for radar remote sensing, because the resolution cell is larger than the transmitted wavelength and thus contains many deterministic scatterers, each of which is completely represented by an individual $[S]_i$ matrix. As a result, the measured [S] matrix is the coherent superposition of the individual $[S]_i$ matrices within the resolution cell. Therefore, for a complete analysis of the effects connected to non-deterministic scatterers, the target covariance and coherency matrices can be used [31, 129]. The covariance matrix is the ensemble averaged complex outer-product of the lexicographic scattering vector as follows [31]

$$[C] = \left\langle \vec{k}_B \cdot \vec{k}_B^{*T} \right\rangle = \begin{pmatrix} \langle |S_{HH}|^2 \rangle & \langle S_{HH}S_{HV}^* \rangle & \langle S_{HH}S_{VH}^* \rangle & \langle S_{HH}S_{VV}^* \rangle \\ \langle S_{HV}S_{HH}^* \rangle & \langle |S_{HV}|^2 \rangle & \langle S_{HV}S_{VH}^* \rangle & \langle S_{HV}S_{VV}^* \rangle \\ \langle S_{VH}S_{HH}^* \rangle & \langle S_{VH}S_{HV}^* \rangle & \langle |S_{VH}|^2 \rangle & \langle S_{VH}S_{VV}^* \rangle \\ \langle S_{VV}S_{HH}^* \rangle & \langle S_{VV}S_{HV}^* \rangle & \langle S_{VV}S_{VH}^* \rangle & \langle |S_{VV}|^2 \rangle \end{pmatrix} \quad (2.32)$$

where $\langle \dots \rangle$ represents ensemble averaging assuming that the spatial scattering medium to be averaged is homogeneous. The diagonal elements correspond to the backscattered intensities. The off-diagonal elements represents the complex covariance of the respective polarization configurations. Alternatively, the Coherency matrix is calculated using the Pauli scattering vector [31]

$$[T]_{4 \times 4} = \left\langle \vec{k}_P * \vec{k}_P^{*T} \right\rangle \quad (2.33)$$

Some properties of these two matrices are

- Both matrices have the same real positive eigenvalues but different eigenvectors.
- The matrices are full rank.
- Without the ensemble averaging, their rank is 1 indicating a deterministic scattering process [96].

- As we shall see later in section 2.3, the interpretation of physical scattering mechanism is easier using the coherency matrix.

It should be noted that assuming reciprocity (equation 2.15), the Pauli scattering vector would take the following form

$$\vec{k}_{P3} = \frac{1}{\sqrt{2}} \left[S_{HH} + S_{VV}, \quad S_{HH} - S_{VV}, \quad 2S_{HV} \right]^T \quad (2.34)$$

Applying \vec{k}_{P3} , the following 3×3 coherency matrix can be defined

$$[T]_{3 \times 3} = \left\langle \vec{k}_{P3} \cdot \vec{k}_{P3}^{*T} \right\rangle = \frac{1}{2} \begin{bmatrix} \langle |M|^2 \rangle & \langle MN^* \rangle & \langle MP^* \rangle \\ \langle M^*N \rangle & \langle |N|^2 \rangle & \langle NP^* \rangle \\ \langle M^*P \rangle & \langle N^*P \rangle & \langle |P|^2 \rangle \end{bmatrix} \quad (2.35)$$

where $M = S_{HH} + S_{VV}$, $N = S_{HH} - S_{VV}$ and $P = 2S_{HV}$.

It should be noted that in a pure target case, there exist a one-to-one correspondence between the Kennaugh matrix and the $[T]_{3 \times 3}$ matrix, given by [125]

$$[T]_{3 \times 3} = \begin{bmatrix} 2A_0 & C - jD & H + jG \\ C + jD & B_0 + B & E + jF \\ H - jG & E - jF & B_0 - B \end{bmatrix} \quad (2.36)$$

2.3 Polarimetric Decomposition

The main goal of target decomposition (TD) methods is to decompose or express the average matrix into a sum of independent matrices representing independent elements and to associate a physical mechanism with each element. This decomposition facilitates the interpretation of the scattering process.

Based on the type of matrix that is used for decomposition, Cloude [33] categorized the TD methods into three groups: those employing the coherent decomposition of the scattering matrix, those employing Muller matrix and Stokes vector and those

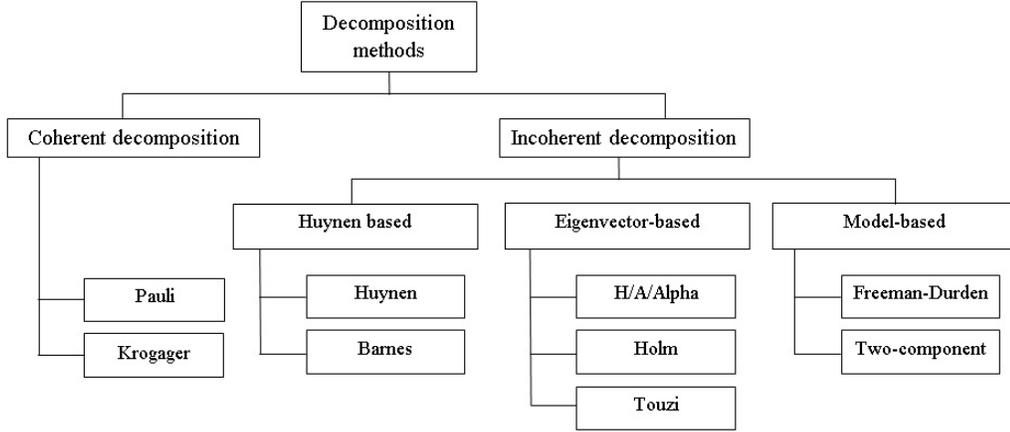


Figure 2.5: Decomposition taxonomy

using an eigenvector analysis of the covariance or coherency matrix.

In this research we divide the TD methods into coherent and incoherent approaches. The incoherent approaches are divided into Huynen based, eigenvector based and model based methods. Figure 2.5 shows a taxonomy of the methods we investigated in this research.

The main reason for using this categorization is that the SAR parameters provided by coherent and incoherent parameters can complement each other for our goal of forest classification. Incoherent algorithms performs an averaging of the returned signals. This provides a statistically smoother description of the behaviour of the scatterers. Alternatively, the coherent methods, by avoiding the averaging at the first stage, can be used to maintain the full resolution. By combining these parameters, various forest features are emphasized. Coherent methods focuses on the more detailed characteristics of the trees such as the leaves, branches and twigs whereas the incoherent parameters reveals the high scale components e.g. the canopy structure.

2.3.1 Coherent Decomposition

A first class of TD theorems are coherent decomposition methods. The objective of the coherent decompositions is to express the measured scattering matrix by the

radar, i.e. $[S]$, as a combination of the scattering responses of simpler objects [33]

$$[S] = \sum_{i=1}^k c_i [S]_i \quad (2.37)$$

In 2.37, the symbol $[S]_i$ stands for the response of every simpler objects, also known as canonical objects, whereas c_i indicates the weight of $[S]_i$ in the combination leading to the measured $[S]$.

In general, a direct analysis of the matrix $[S]$, with the objective to infer the physical properties of the scatterers under study, is shown very difficult. Thus, the physical properties of the target under study are extracted and interpreted through the analysis of the simpler responses $[S]_i$ and the corresponding coefficients c_i in 2.37.

Pauli and Krogager are some of the important examples of such a decomposition which are explained in the following sections.

The Pauli Decomposition

The Pauli decomposition expresses the measured scattering matrix $[S]$ in the Pauli basis. Recalling the four matrices in equation 2.19, the Pauli basis $[S]_a$, $[S]_b$, $[S]_c$, $[S]_d$ is given by the following four 2×2 matrices

$$\begin{aligned} [S]_a &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ [S]_b &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ [S]_c &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ [S]_d &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \end{aligned} \quad (2.38)$$

Since reciprocity applies in a monostatic system configuration, the Pauli basis can be reduced to ,

$$[S]_a, [S]_b, [S]_c \quad (2.39)$$

Consequently, given a measured scattering matrix $[S]$, it can be expressed as follows

$$[S] = \begin{pmatrix} S_{HH} & S_{HV} \\ S_{HV} & S_{VH} \end{pmatrix} = \alpha [S]_a + \beta [S]_b + \gamma [S]_c \quad (2.40)$$

where

$$\begin{aligned} \alpha &= \frac{S_{HH} + S_{VV}}{\sqrt{2}} \\ \beta &= \frac{S_{HH} - S_{VV}}{\sqrt{2}} \\ \gamma &= \sqrt{2}S_{HV} \end{aligned} \quad (2.41)$$

In general, $[S]_a$ is referred to single-bounce or odd-bounce scattering which corresponds to the scattering matrix of a sphere, a plate or a trihedral. Thus, the complex coefficient α , represents the contribution of $[S]_a$ to the final measured scattering matrix. Generally, $[S]_b$ indicates a scattering mechanism characterized by double-bounce or even-bounce scattering which represents the scattering mechanism of a dihedral oriented at 0 degrees. Consequently, β stands for the complex coefficient of this scattering mechanism. Finally, the third matrix $[S]_c$ corresponds to the scattering mechanism of a diplane oriented at 45 degrees. As it can be observed from $[S]_c$ of 2.38, the target returns a wave with a polarization orthogonal to the one of the incident wave. Thus, the scattering mechanism represented by $[S]_c$ is referred to those scatterers which are able to return the orthogonal polarization. One of the best examples is the volume scattering produced by the forest canopy. The complex coefficient

γ is thus the contribution of $[S]_c$ to $[S]$.

The Krogager Decomposition

Krogager [72] has proposed an alternative to factorize the scattering matrix as the combination of the responses of a sphere, a diplane and a helix. The last two components present an orientation angle θ . The Krogager decomposition presents the following formulation

$$\begin{aligned}
[S_{HV}] &= e^{j\varphi} \{ e^{j\varphi_s} k_s [S]_s + k_d [S]_d + k_h [S]_h \} \\
&= e^{j\varphi} \left\{ e^{j\varphi_s} k_s \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + k_d \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix} \right. \\
&\quad \left. + k_h e^{\pm j2\theta} \begin{pmatrix} 1 & \pm j \\ \pm j & 1 \end{pmatrix} \right\} \quad (2.42)
\end{aligned}$$

The number of independent parameters in both Krogager and Pauli is the same, i.e. six. In the case of the Pauli decomposition, the complex coefficients α , β and γ , whereas in the case of the Krogager decomposition, the three angles φ , φ_s and θ and the three real coefficients k_s , k_d and k_h : $\{\varphi, \varphi_s, \theta, k_s, k_d, k_h\}$.

The absolute phase φ can contain information about the scatterer under study. But, since its value depends also on the distance between the radar and the target, it is considered as a irrelevant parameter and it is often considered that the Krogager decomposition presents 5 independent parameters. The parameters φ_s and k_s characterize the sphere component of the Krogager decomposition. The phase φ_s represents a displacement of the sphere with respect to the diplane and the helix components. The real parameter k_s represents the contribution of the sphere component to the final scattering matrix $[S]$. Consequently, $|k_s|^2$ is interpreted as the power scattered by the sphere-like component of the matrix $[S]$. The phase parameter θ stands for the orientation angle of the diplane and the helix components of the Krogager decompo-

sition. Finally, the coefficients k_d and k_h correspond to the weights of the diplane and the helix components. Thus, $|kd|^2$ and $|kh|^2$ are interpreted as the power scattered by the diplane and the helix components of the Krogager decomposition.

The value of these 5 parameters can be easier derived in a circular basis (r, l) . Reformulation of 2.42 in (r, l) gives

$$\begin{aligned}
[S_{rl}] &= \begin{pmatrix} S_{rr} & S_{rl} \\ S_{lr} & S_{ll} \end{pmatrix} = \begin{pmatrix} |S_{rr}| e^{j\phi_{rr}} & |S_{rl}| e^{j\phi_{rl}} \\ |S_{rl}| e^{j\phi_{rl}} & -|S_{ll}| e^{j(\phi_{rr}+\pi)} \end{pmatrix} \\
&= e^{j\varphi} \left\{ e^{j\varphi_s} k_s \begin{pmatrix} 0 & j \\ j & 0 \end{pmatrix} + k_d \begin{pmatrix} e^{j2\theta} & 0 \\ 0 & -e^{-j2\theta} \end{pmatrix} + k_h \begin{pmatrix} e^{j2\theta} & 0 \\ 0 & 0 \end{pmatrix} \right\}
\end{aligned} \tag{2.43}$$

According to this formulation:

$$\begin{aligned}
k_s &= |S_{rl}| \\
\phi &= \frac{1}{2}(\phi_{rr} + \phi_{ll} - \pi) \\
\theta &= \frac{1}{4}(\phi_{rr} - \phi_{ll} + \pi) \\
\phi_s &= \phi_{rl} - \frac{1}{2}(\phi_{rr} + \phi_{ll})
\end{aligned} \tag{2.44}$$

Based on the difference in absolute value of S_{rr} and S_{ll} , Krogager considered two cases:

$$\begin{aligned}
|S_{rr}| \geq |S_{ll}| &\Rightarrow \begin{cases} k_d^+ = |S_{ll}| \\ k_h^+ = |S_{rr}| - |S_{ll}| \end{cases} \Leftarrow \text{Left sense helix} \\
|S_{rr}| \leq |S_{ll}| &\Rightarrow \begin{cases} k_d^- = |S_{rr}| \\ k_h^- = |S_{ll}| - |S_{rr}| \end{cases} \Leftarrow \text{Right sense helix}
\end{aligned} \tag{2.45}$$

The formulations in 2.42 and 2.43 can be related using

$$\begin{aligned}
S_{rr} &= jS_{HV} + \frac{1}{2}(S_{HH} - S_{VV}) \\
S_{ll} &= jS_{HV} - \frac{1}{2}(S_{HH} - S_{VV}) \\
S_{rl} &= \frac{j}{2}(S_{HH} + S_{VV})
\end{aligned} \tag{2.46}$$

In general, coherent decomposition methods, are exposed to problems due to speckle, so they are not appropriate when applied to radar remote sensing of natural random targets such as vegetation. Nevertheless, they are still suitable when the scene is dominated by a single scattering element or a few of them, and a radar with high resolution is applied [71].

2.3.2 Incoherent Decomposition

The scattering matrix $[S]$ can not be employed to characterize the distributed scatterers. It is only able to characterize the coherent or pure scatterers [33, 7]. The distributed scatterers can be only characterized statistically, due to the presence of speckle noise. Since speckle noise must be reduced, only second order polarimetric representations can be employed to analyze distributed scatterers.

The complexity of the scattering process makes extremely difficult the physical study of a given scatterer through the direct analysis of $[C]$ or $[T]$. Hence, the objective of the incoherent decompositions is to separate the $[C]$ or $[T]$ matrices as the combination of second order descriptors corresponding to simpler or canonical objects, presenting an easier physical interpretation. These decomposition theorems can be expressed as

$$[C] = \sum_{i=1}^k p_i [C]_i \tag{2.47}$$

$$[T] = \sum_{i=1}^k q_i [T]_i \tag{2.48}$$

where p_i and q_i denote the coefficients of these components in $[C]$ or $[T]$, respectively. Different compositions can be presented based on this formulation.

Huynen-based decomposition Huynen decomposition

The main idea of the Huynen target decomposition is to separate from the incoming data stream, a single average target and a residue component called 'N-target' [62]. To represent the averaged distributed target, the ensemble average value of the $[K]$ matrix (equation 2.24) or $[T]_{3 \times 3}$ (equation 2.36) can be taken.

$$[T]_{3 \times 3} = \begin{bmatrix} 2 \langle A_0 \rangle & \langle C \rangle - j \langle D \rangle & \langle H \rangle + j \langle G \rangle \\ \langle C \rangle + j \langle D \rangle & \langle B_0 \rangle + \langle B \rangle & \langle E \rangle + j \langle F \rangle \\ \langle H \rangle - j \langle G \rangle & \langle E \rangle - j \langle F \rangle & \langle B_0 \rangle - \langle B \rangle \end{bmatrix} \quad (2.49)$$

This averaged $[T]_{3 \times 3}$ is described by nine parameters which lose their dependency relations and become independent, whereas as we know a fixed single object is described by 5 parameters. Thus, the main goal of the Huynen decomposition is to represent the averaged target as an effective single target T_0 (given 5 parameters) and a residue target T_N which contain the 4 remaining degrees of freedom. Recalling from section 2.2.3 one of the dependency relationships was $B_0 = B^2 + E^2 + F^2$. The Huynen approach decomposes the vector (B_0, B, E, F) into two vectors of equivalent single target and the residue target as follows

$$\begin{aligned} B_0 &= B_{0T} + B_{0N} & B &= B_T + B_N \\ E &= E_T + E_N & F &= F_T + F_N \end{aligned} \quad (2.50)$$

in which T and N denote the equivalent single target (T) and N-target (N). Based on this, Huynen decomposed the $[K]$ matrix or $[T]_{3 \times 3}$ matrix according to

$$[T]_{3 \times 3} = \begin{bmatrix} \langle 2A_0 \rangle & \langle C \rangle - j \langle D \rangle & \langle H \rangle + j \langle G \rangle \\ \langle C \rangle + j \langle D \rangle & \langle B_0 \rangle + \langle B \rangle & \langle E \rangle + j \langle F \rangle \\ \langle H \rangle - j \langle G \rangle & \langle E \rangle - j \langle F \rangle & \langle B_0 \rangle - \langle B \rangle \end{bmatrix} = T_0 + T_N \quad (2.51)$$

where

$$T_0 = \begin{bmatrix} \langle 2A_0 \rangle & \langle C \rangle - j \langle D \rangle & \langle H \rangle + j \langle G \rangle \\ \langle C \rangle + j \langle D \rangle & B_{0T} + B_T & E_T + jF_T \\ \langle H \rangle - j \langle G \rangle & E_T - jF_T & B_{0T} - B_T \end{bmatrix} \text{ and} \quad (2.52)$$

$$T_N = \begin{bmatrix} 0 & 0 & 0 \\ 0 & B_{0N} + B_N & E_N + jF_N \\ 0 & E_N - jF_N & B_{0N} - B_N \end{bmatrix} \quad (2.53)$$

It is important to note that the N-target corresponds to a perfectly non-symmetric target (because it is defined with only the parameters (B_{0N}, B_N, E_N, F_N)). Due to this fact, the N-target does not change with target tilt angle (roll-invariant). In other words, the N-target is independent of rotation along the line of sight between radar and target. These 4 parameters are determined using 2.51. Alternatively, the parameters (B_{0T}, B_T, E_T, F_T) , which corresponds to the single target, can be reconstructed uniquely using the target structure equations (equation 2.26) which can be rewritten as [98]

$$\begin{aligned} 2A_0(B_{0T} + B_T) &= C^2 + D^2 \\ 2A_0(B_{0T} - B_T) &= G^2 + H^2 \\ 2A_0E_T &= CH - DG \\ 2A_0F_T &= CG - DH \end{aligned} \quad (2.54)$$

In summary, the Huynen target decomposition method factorizes the measured coherency matrix $[T]_{3 \times 3}$ into a rank one pure target T_0 and into a distributed and roll-invariant N-target T_N .

Barnes decomposition

Barnes [5] proposed the general form of the Huynen decomposition. He proved that the structure proposed by Huynen is not unique and other decompositions can be realized with the same structure. As was mentioned in Huynen decomposition, T_N is roll-invariant. This can be interpreted as the fact that the vector space generated by T_N is orthogonal to the vector space generated by the pure target T_0 . Assume that an arbitrary vector q belongs to the space of N-target: $T_N q = 0$. For T_N to be roll-invariant, it requires that

$$T_N(\theta) q = 0 \Rightarrow U(\theta) T_N U(\theta)^{-1} q = 0 \quad (2.55)$$

where $U(\theta)$ is a 3×3 rotation matrix. The condition in 2.55 is met for any q only if

$$U(\theta)^{-1} q = \lambda q \quad (2.56)$$

which means that q is the eigenvector of the matrix $U(\theta)^{-1}$ which are as follow

$$q_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad q_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ j \end{bmatrix} \quad q_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ j \\ 1 \end{bmatrix} \quad (2.57)$$

This means that there are three ways to decompose the measured coherency matrix $[T]_{3 \times 3}$ into a pure target T_0 and a distributed N-target T_N . The eigenvector q_1 corresponds to the Huynen decomposition, while q_1 and q_2 corresponds to the Barnes decomposition theorem. The normalized target vector k_0 associated with each of these eigenvector can be determined using

$$\left. \begin{aligned} Tq &= T_0q + T_Nq = T_0q = k_0k_0^{T*}q \\ q^{T*}qT &= q^{T*}k_0k_0^{T*}q = |k_0^{T*}q|^2 \end{aligned} \right\} \Rightarrow K_0 = \frac{Tq}{k_0^{T*}q} = \frac{Tq}{\sqrt{q^{T*}qT}} \quad (2.58)$$

Plugging the 2.49 and 2.57 into the equation 2.58 the three normalized target vectors can be obtained as

$$\begin{aligned} k_{01} &= \frac{1}{\sqrt{\langle 2A_0 \rangle}} \begin{bmatrix} 2\langle A_0 \rangle \\ \langle C \rangle + j\langle D \rangle \\ \langle H \rangle - j\langle G \rangle \end{bmatrix} \\ k_{02} &= \frac{1}{\sqrt{2(\langle B_0 \rangle - \langle F \rangle)}} \begin{bmatrix} \langle C \rangle - \langle G \rangle + j\langle H \rangle - j\langle D \rangle \\ \langle B_0 \rangle + \langle B \rangle - \langle F \rangle + j\langle E \rangle \\ \langle E \rangle + j\langle B_0 \rangle - j\langle B \rangle - j\langle F \rangle \end{bmatrix} \\ k_{03} &= \frac{1}{\sqrt{2(\langle B_0 \rangle + \langle F \rangle)}} \begin{bmatrix} \langle H \rangle + \langle D \rangle + j\langle C \rangle + j\langle G \rangle \\ \langle E \rangle + j\langle B_0 \rangle + j\langle B \rangle + j\langle F \rangle \\ \langle B_0 \rangle - \langle B \rangle + \langle F \rangle + j\langle E \rangle \end{bmatrix} \end{aligned} \quad (2.59)$$

Eigenvector based decomposition H/A/ α decomposition

The eigenvector-eigenvalue based decomposition is based on the eigen decomposition of the coherency matrix $[T]$. According to the eigen decomposition theorem, the 3×3 Hermitian matrix $[T]$ can be decomposed as follows [29]

$$[T] = [U_3] \left[\sum \right] [U_3]^{-1} \quad (2.60)$$

The 3×3 , real, diagonal matrix $[\sum]$ contains the eigenvalues of $[T]$

$$[\Sigma] = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad (2.61)$$

where $\lambda_1 > \lambda_2 > \lambda_3 > 0$. The 3×3 unitary matrix $[U_3]$ contains the eigenvectors u_i for $i = 1, 2, 3$ of $[T]$

$$[U_3] = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \quad (2.62)$$

The eigenvectors u_i of $[T]$ can be formulated as follows

$$u_i = \begin{bmatrix} \cos \alpha_i & \sin \alpha_i \cos \beta_i e^{j\delta_i} & \sin \alpha_i \cos \beta_i e^{j\gamma_i} \end{bmatrix}^T \quad (2.63)$$

Considering the expressions 2.61 and 2.62, the eigen decomposition of $[T]$, i.e. 2.60, can be written as follows

$$[T] = \sum_{j=1}^3 \lambda_j u_j u_j^{*T} \quad (2.64)$$

As 2.64 shows, the rank 3 matrix $[T]$ can be decomposed as the combination of three rank 1 coherency matrices formed as

$$[T] = T_{01} + T_{02} + T_{03} \quad (2.65)$$

The obtained eigenvalues and eigenvectors are considered as the primary parameters of the eigen decomposition of $[T]$. In order to simplify the analysis of the physical information provided by this eigen decomposition, three secondary parameters are defined as a function of the eigenvalues and the eigenvectors of $[T]$ [31, 29]:

- Scattering Entropy H

The degree of randomness of target scattering is represented by H . It is computed from the eigenvalues of the target coherency matrix according to:

$$H = \sum_{i=1}^3 -P_i \log_n (P_i) \quad (2.66)$$

where p_i is defined as:

$$P_i = \frac{\lambda_i}{\sum_{k=1}^3 \lambda_k} \quad (2.67)$$

The entropy is a scalar between 0 and 1. It can be also interpreted as the degree of statistical disorder. In this way [100]:

$$\begin{aligned} H \rightarrow 0 &\Rightarrow \lambda_1 = SPAN \quad \lambda_2 = 0 \quad \lambda_3 = 0 && \Rightarrow \text{one pure target} \\ H \rightarrow 1 &\Rightarrow \lambda_1 = \frac{SPAN}{3} \quad \lambda_2 = \frac{SPAN}{3} \quad \lambda_3 = \frac{SPAN}{3} && \Rightarrow 3 \text{ pure targets} \\ 0 < H < 1 &\Rightarrow \text{three different eigen value } \lambda_i && \Rightarrow 3 \text{ weighted pure targets} \end{aligned} \quad (2.68)$$

In the first case, the scattering matrix $[T]$ presents rank 1 and the scattering process corresponds to a pure target. In the second case, the scattering matrix $[T]$ presents rank 3, that is the scattering process is due to the combination of three pure targets (distributed targets). In the third case, the final scattering mechanism given by $[T]$ results from the combination of the three pure targets given by u_i , but weighted by the corresponding eigenvalues.

- Anisotropy A

The anisotropy can be derived from eigenvalues of the target coherency matrix according to

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \quad (2.69)$$

The anisotropy A , is a parameter complementary to the entropy. The anisotropy measures the relative importance of the second and the third eigenvalues of the eigen decomposition. From a practical point of view, the anisotropy can be employed as a source of discrimination only when $H > 0.7$. The reason is that for lower entropies, the second and third eigenvalues are highly affected by noise. Consequently, the anisotropy is also very noisy.

- Alpha angle

The parameter alpha provides information on the dominant scattering mechanism. It is computed as the weighted average of the value

$$\alpha = \sum_{i=1}^3 p_i \alpha_i \quad (2.70)$$

The next list reports the interpretation of α :

$\alpha \rightarrow 0$: The scattering corresponds to single-bounce scattering produced by a rough surface.

$\alpha \rightarrow \frac{\pi}{4}$: The scattering mechanism corresponds to volume scattering.

$\alpha \rightarrow \frac{\pi}{2}$: The scattering mechanism is due to double-bounce scattering. The eigen decomposition of the coherency matrix is also referred as the $H/A/\alpha$ decomposition [99].

Holm decomposition

Holm decomposition [59] improves the Huynen approach by combining the concept of the single target plus noise model of the Huynen method with eigenvalue analysis. Based upon this, the measured coherency matrix (T) is decomposed into a pure target matrix (T_1) plus a mixed target state (T_2) and an unpolarized mixed state equivalent to a noise term (T_3)

$$T = T_1 + T_2 + T_3 \quad (2.71)$$

where

$$\begin{aligned} T_1 &\Rightarrow \text{Pure target state} \\ T_2 &\Rightarrow \text{Mixed target state} \\ T_3 &\Rightarrow \text{Unpolarized mixed state (noise)} \end{aligned} \quad (2.72)$$

For a detailed explanation of the Holm decomposition method the reader is referred to [83].

Model-based decomposition **Freeman-Durden decomposition**

The Freeman and Durden [47] presented a method which is based on the physics of radar scattering and less bound to pure mathematical models. The Freeman decomposition describes the scattering as due to three physical mechanisms, i.e. first order Bragg surface scatterer from a moderately rough surface (s), even- or double-bounce scattering mechanism (d) and canopy (or volume) scattering from randomly oriented dipoles (v). According to this model, the measured power P can be finally expressed as

$$P = SPAN = \langle |S_{HH}|^2 \rangle + \langle |S_{VV}|^2 \rangle + 2 \langle |S_{HV}|^2 \rangle = P_s + P_d + P_v \quad (2.73)$$

These three components can be calculated using the elements of the covariance matrix. In this process, a series of intermediate parameters i.e. f_s , f_d , f_v , α and β are first introduced as follows

$$\begin{aligned}
P_s &= f_s(1 + |\beta|^2) \\
P_d &= f_d(1 + |\alpha|^2) \\
P_v &= \frac{8}{3}f_v
\end{aligned} \tag{2.74}$$

These parameters are then related to elements of covariance matrix using

$$\begin{aligned}
\langle |S_{HH}|^2 \rangle &= f_s |\beta|^2 + f_d |\alpha|^2 + f_v \\
\langle |S_{VV}|^2 \rangle &= f_s + f_d + f_v \\
\langle S_{HH}S_{VV}^* \rangle &= f_s \beta + f_d \alpha + \frac{f_v}{3} \\
\langle |S_{HV}|^2 \rangle &= \frac{f_v}{3}
\end{aligned} \tag{2.75}$$

It should be noted that due to reflection symmetry i.e. $\langle S_{HH}S_{HV}^* \rangle = \langle S_{HV}S_{VV}^* \rangle = 0$, the remaining covariance matrix element were omitted. There are 4 equations with 5 unknowns. One of the unknowns can be fixed using the method of van Zyl by deciding whether double-bounce or surface scatter is the dominant contribution based on the sign of the real part of $S_{HH}S_{VV}^*$. If $Re \{ \langle S_{HH}S_{VV}^* \rangle \} \geq 0$, then surface scatter is dominant and α is fixed with $\alpha = -1$. If $Re \{ \langle S_{HH}S_{VV}^* \rangle \} \leq 0$, then double-bounce scatter is dominant and β is fixed with $\beta = 1$.

While this decomposition is useful in providing features for distinguishing between different surface cover types, it has two limiting assumptions, namely the three component scattering model which is not always applicable and the reflection symmetry assumption [33].

Generally, forests have strong volume scattering. However, due to different canopy structure as well as different shape of the leaves, this volume scattering varies among different trees. These can make the canopy scattering as a useful parameter for our

case of forest mapping.

Freeman two component decomposition

Unlike the first Freeman model which had 3 parameters, Freeman proposed a two-component model to polarimetric SAR observations of forests [46]. The considered mechanisms are: 1) a canopy scatter from a reciprocal medium with reflection symmetry and 2) a ground scatter term representing either a double bounce scatter (ground-trunk interaction) or a Bragg scatter from a moderately rough surface. In this new decomposition method there are 4 equations and 4 unknowns

$$\begin{aligned}
 \langle |S_{HH}|^2 \rangle &= f_G + \rho f_V \\
 \langle |S_{VV}|^2 \rangle &= f_G |\alpha|^2 + \rho f_V \\
 \langle S_{HH} S_{VV}^* \rangle &= f_G \alpha + (\rho - 1) f_V \\
 \langle |S_{HV}|^2 \rangle &= \rho f_V
 \end{aligned} \tag{2.76}$$

where f_V and ρ correspond to the volume scattering component contribution, while f_G and α corresponds to the double- or single-bounce scattering component contribution. The double- and single-bounce can be distinguished using the amplitude and phase of the parameter α

$$\begin{cases} |\alpha| \leq 1 \text{ } arg(\alpha) = \pm\pi & \Leftarrow \text{double bounce scatter} \\ |\alpha| \leq 1 \text{ } arg(\alpha) \approx 2\phi & \Leftarrow \text{single surface scatter} \end{cases} \tag{2.77}$$

This decomposition provides useful features for forest application as the model is sensitive to forest canopy structure and to the ratio of the canopy to the ground returns.

Touzi decomposition

In contrast to the $H/A/\alpha$ decomposition, which uses the α angle to describe target scattering type, the Touzi decomposition characterizes uniquely the scattering type with the following parameters: the symmetric scattering type magnitude (α_s) and phase (ϕ_s), the target helicity (τ_s), the orientation angle (ψ_s) and the dominant eigenvalue (λ_s).

Touzi et al. have shown that the use of α_s, τ_s and λ_s parameters can lead to efficient wetland classification [126]. In this study, we investigate the usefulness of these parameters for forest mapping.

2.4 SAR discriminators

Several quantities have been derived from PolSAR data to be used as indicators to discriminate among surface types or land covers. These include:

1. SPAN: the SPAN of the scattering matrix is defined as the sum of the squares of all the original scattering matrix elements

$$SPAN = |S_{HH}|^2 + |S_{VV}|^2 + 2|S_{HV}|^2 \quad (2.78)$$

1. Extrema of the received power: Evans et al. [40] used the maximum and minimum of the received power at different polarizations to discriminate different land cover types. The procedure for calculating these extrema comprises varying the polarization angles (ϕ, τ) of the transmitted wave and computation of the corresponding received powers for each transmitted polarization angle. Although, the calculation of these extrema is computationally expensive, they can be useful for separating different areas in polarimetric image.
2. Fractional polarization: it is defined as

$$F = \frac{(P_{max} - P_{min})}{(P_{max} + P_{min})} \quad (2.79)$$

in which P_{max} and P_{min} are the maximum and minimum of the received power. It can be used as a measure of the polarization purity of the return signal.

3. Extrema of the degree of polarization (polarized and unpolarized intensity extrema): The degree of polarization is the ratio between the intensity of the polarized part and the total scattered intensity. Touzi et al. [127] proposed a systematic and analytic computation method for the calculation of the maximum and minimum degree of polarization

$$p = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \quad (2.80)$$

S_0 is proportional to the total intensity of the wave, S_1 is the difference between the density powers related to the horizontal and vertical polarizations. Parameters S_2 and S_3 are related to the phase difference between the horizontal and vertical components of the electric field. In this thesis, we used the maximum and minimum degree of polarization.

4. Extrema of the total scattered intensity: Touzi et al. [127] divided the total scattered intensity into the completely polarized and completely unpolarized components, and for each one the extrema was calculated. He showed that these indicators along with the extrema of the degree of the polarization can be combined with other indices for target discrimination. Note that for the case of extrema of the completely unpolarized scattered intensity, the minimum component was only considered because the maximum of the completely unpolarized scattered intensity and the coefficient of fractional polarization are correlated.
5. Pedestal height: The pedestal height is defined as the minimum value of the

co-polarization response and is based on the polarization synthesis at each pixel. It indicates the depolarization within the image. The pedestal height is higher for a higher degree of depolarization.

6. Complex correlation coefficients (coherence): the four correlation coefficients between polarizations are important parameters in PolSAR data, and were used in this study

$$\begin{aligned} \rho_{12} &= \frac{\langle S_{HH}S_{HV}^* \rangle}{\sqrt{|S_{HH}|^2 |S_{HV}|^2}} & \rho_{13} &= \frac{\langle S_{HH}S_{VV}^* \rangle}{\sqrt{|S_{HH}|^2 |S_{VV}|^2}} \\ \rho_{23} &= \frac{\langle S_{HH}S_{VV}^* \rangle}{\sqrt{|S_{HH}|^2 |S_{VV}|^2}} & \rho_{RL} &= \frac{\langle S_{RR}S_{LL}^* \rangle}{\sqrt{|S_{RR}|^2 |S_{LL}|^2}} \end{aligned} \quad (2.81)$$

7. depolarization ratio: The depolarization ratio describes how completely a target depolarizes incident polarized signal. It can be calculated as

$$d = \frac{\langle S_{hv}S_{hv}^* \rangle}{\langle S_{hh}S_{hh}^* \rangle + \langle S_{vv}S_{vv}^* \rangle} \quad (2.82)$$

Roughness causes a depolarization of the incident wave and the cross-polarized response is very sensitive to the roughness variations. Therefore, it may be used to separate forest types with different roughness types.

2.5 summary

The chapter aimed at providing various PolSAR features that might be useful for forest classification. At the beginning, a brief introduction of SAR polarimetry including polarimetry basics, different formalism (Stokes vs. Jones Formalism, Scattering Matrix vs. Muller Matrix) and deterministic vs. non-deterministic scatterers was given. This theoretical grounding is necessary for a better understanding of the different SAR parameters.

Afterwards, the chapter reviewed different target decomposition approaches i.e. coherent and incoherent methods. Then, the SAR discriminators that are used in this thesis were listed. These PolSAR parameters provide the necessary features for forest classification. Table 2.2 lists the parameters used in this thesis. Some remarks on these parameters are given below

- The amplitude of the 3 upper triangle elements of the covariance matrix are used in this research. It should be noted that the diagonal elements of the covariance matrix were already employed as the scattering matrix elements
- The amplitude of the 5 diagonal and upper triangle coherency matrix elements were used. Note that the third diagonal element of the coherency matrix ($[T]_{33}$) was not considered because it was exactly twice as much as the second element of the covariance matrix ($[C]_{22}$), and hence redundant.
- In the case of Huynen , Barnes and Holm decompositions the 6 amplitudes of the upper triangle elements of the corresponding matrices were considered in this thesis

#	name	description
1	hh	scattering matrix elements
2	vv	
3	hv	
4	SPAN	
5	T11_coh	coherency matrix elements
6	T12_coh	
7	T13_coh	
8	T22_coh	
9	T23_coh	
10	C12_cov	covariance matrix elements
11	C13_cov	
12	C23_cov	
13	T11_huy	Huynen matrix elements
14	T12_huy	
15	T13_huy	
16	T22_huy	
17	T23_huy	
18	T33_huy	
19	Entropy	Cloud and Pottier decomposition
20	Alpha	
21	Anisotropy	
22	H(1-A)	
23	(1-H)A	
24	Ks_Krog	Krogager decomposition elements
25	Kd_Krog	
26	kh_Krog	
27	Free_odd	Freeman decomposition elements
28	Free_dbl	
29	Free_vol	
30	T11_bar	Barnes decomposition
31	T12_bar	
32	T13_bar	
33	T22_bar	
34	T23_bar	
35	T33_bar	Holm decomposition
36	T11_holm	
37	T12_holm	
38	T13_holm	
39	T22_holm	
40	T23_holm	Touzi decomposition
41	T33_holm	
42	psi_angle	
43	dom_landa	
44	tau_angle	
45	max_deg	max degree of pol.
46	min_deg	min degree of pol.
47	max_pol	completely pol. max
48	min_pol	completely pol. min
49	max_unpo	completely unpol. max
50	max_pow	max received pow
51	min_pow	min received pow
52	fract_pow	fractional pow
53	ped_hgh	pedestal height
54	ro12	correlation coefficients
55	ro13	
56	ro23	
57	CCC	depolarization ratio
58	depol_ind	

Table 2.2: polarimetric parameters used in this research

Chapter 3

PolSAR Classification

PolSAR image classification is one of the most important applications in remote sensing. Many methods have been proposed. In this chapter we will review some of the well-known algorithms for the classification of polarimetric SAR data. Generally, there are 3 fundamental issues to be addressed in the classification problem.

- The first issue is the selection of the ground truth data. There are two approaches for the classification whether to use the training samples or not i.e. supervised or unsupervised respectively. Section 3.1 of this chapter reviews some of the methods belonging to each category. Earlier polarimetric classification methods were based on the supervised method of statistical characteristics of PolSAR images. One of the advantages of polarimetry for constructing classification schemes is that the observables have an intrinsic physical meaning. In this framework, several algorithms were developed based on the scattering mechanism of electromagnetic waves. All these methods are described in this very first section.
- The second issue is to identify the features which allow different classes to be distinguished from each other. There are several approaches in the literature which used the whole covariance matrix (or coherency matrix) for the classifi-

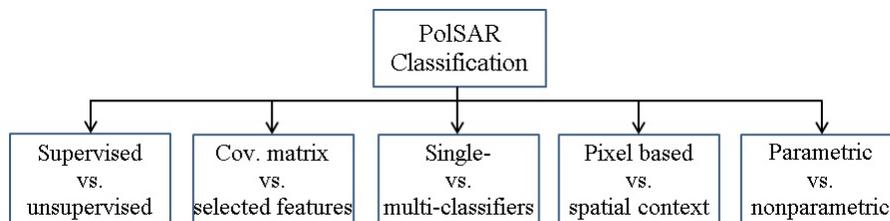


Figure 3.1: A review of PolSAR classification methods

cation. On the other hand, there are approaches which employ a set of selected features for that specified problem. These methods are discussed in section 3.2.

- The third issue is to devise a classification technique which uses these features in the classification problem. Depending on whether the classifier assumes the data fit a specific distribution, the classification techniques are divided into parametric and nonparametric classifiers (section 3.5). Recently, there has been a great interest for using an ensemble of classifiers for solving problems in pattern recognition community. There are different methods for creating such an ensemble, which are discussed in section 3.3. There are another category of methods which tend to incorporate spatial context into classification. The main idea is that two neighboring pixels are not entirely statistically independent. Some of these methods are reviewed in section 3.4. Figure 3.1 shows an overview of different categories for PolSAR classification.

3.1 Supervised vs. unsupervised

There are basically two approaches to classification: unsupervised or supervised. The supervised approach (section 3.1.1) requires ground truth. However it may lead to ambiguities because the scene characteristics required by the analyst may not necessarily be supported by the properties of the data. Unsupervised classification (section 3.1.2) leads to an understanding of the class separability in the scene that is supported by the polarimetric signatures of the data. A difficulty with unsupervised classification is that convergence depends on the initial seeding of candidate classes [108].

3.1.1 Supervised

As one of the earlier algorithms, Kong et al. derived a distance measure based on the complex Gaussian distribution and used it for maximum likelihood (ML) classification of single look complex PolSAR data:

$$d(k, \omega_i) = k^{*T} C_i^{-1} k + \ln |\hat{C}_i| \quad (3.1)$$

in which k is the SAR observable $k = [S_{HH}, \sqrt{2}S_{HV}, S_{VV}]$ which has a multivariate complex circular Gaussian probability density function $N_c(0, C_i)$, where C_i is the complex covariance matrix for class ω_i . $|C_i|$ is the determinant of C_i . During the classification process a pixel is assigned to the minimum distance cluster.

For multilook data represented in covariance or coherency matrices, Lee et al. [82, 81] proposed a method based on the maximum likelihood classifier using the complex Wishart classifier. They show that assuming that target vectors have a $N_c(0, C_i)$ distribution, a sample n -look covariance matrix $Z = \frac{1}{n} \sum_n k k^{*T}$ follows a complex Wishart distribution with n degrees of freedom, given by

$$p(Z) = \frac{n^{qn} |Z|^{n-q} \exp[-\text{tr}(nC^{-1}Z)]}{K(n, q) |C|^n} \quad (3.2)$$

$$\text{with } K(n, q) = \pi^{\frac{q(q-1)}{2}} \prod_{i=1}^q \Gamma(n - i + 1)$$

where q is the dimension of the vector k which is 3 in our case, $\Gamma()$ represents the Gamma function. Based on equation 3.2, the Wishart distance is defined as

$$d(Z, \omega_i) = n \ln |\hat{C}_i| + n \text{tr}(\hat{C}_i Z) + qn \ln n - (n - q) \ln |Z| + \ln K(n, q) \quad (3.3)$$

and removing the unnecessary terms from the above distance, the final decision rule will be

$$\text{Decide } k \text{ belong to the class } \omega_i, \text{ if } \arg \min(d(Z, \omega_i)) \text{ with} \quad (3.4)$$

$$d(Z, \omega_i) = n \ln |\hat{C}_i| + n \text{tr}(\hat{C}_i Z)$$

For the supervised classification, the covariance matrix for class i i.e. \hat{C}_i is estimated. Then, for each pixel the Wishart distance measure $d(Z, \omega_i)$ is calculated for each class and the pixel is assigned to the class with the minimum distance. Lee et al. [81] showed that by the use of supervised classification and a Wishart classifier, high classification accuracies for various land cover types could be obtained.

3.1.2 Unsupervised

One of the advantages of PolSAR data for classification is that no prior knowledge is required about the scene because the observables have an intrinsic physical meaning. This makes the PolSAR data ideal for unsupervised classification.

Cloude and Pottier [33] found that the target entropy and alpha angle can be effective at classifying the SAR image according to scattering mechanism using P,L or C-band SAR data. This is done by forming a two dimensional feature space and subdividing it into eight possible categories. The categories are found by defining the ranges of alpha angle that gives rise to surface scattering ($0^\circ - 42^\circ$), volume scattering ($42^\circ - 48^\circ$) and double bounce scattering ($48^\circ - 90^\circ$) and defining low, medium and high levels of entropy [29]. Figure 2 shows the subdivided entropy-alpha feature space and an example of how the pixels in a SAR image may be distributed among the partitioned zones. Here is a summary of the physical scattering characteristics of each of the nine zones [29]:

- Zone 9: Low entropy surface scattering (e.g. Bragg surface scattering, specular scattering phenomena which do not involve 180 phase inversions between HH and VV and very smooth land surfaces such as water at L and P-Bands).

- Zone 8: Low entropy dipole scattering (e.g. strongly correlated mechanisms which have a large imbalance between HH and VV in amplitude and scattering from vegetation with strongly correlated orientation of anisotropic scattering elements).

- Zone 7: Low entropy multiple scattering events(double or 'even' bounce scat-

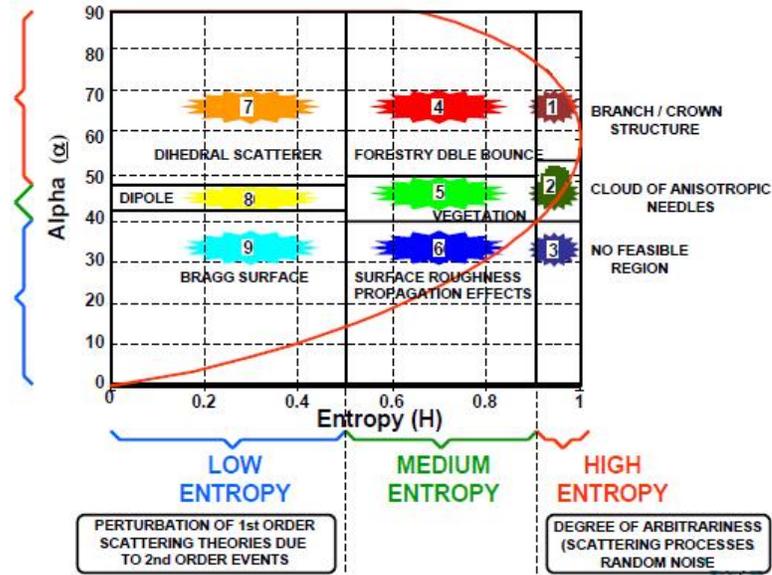


Figure 3.2: Entropy-alpha feature space (adapted from [79])

tering events).

- Zone 6: Medium entropy surface scatter (surface cover comprising oblate spheroidal scatterers e.g. leaves or discs for example)
- Zone 5: Medium entropy vegetation (dipole type) scattering (scattering from vegetated surfaces with anisotropic scatterers and moderate correlation of scatterer orientations).
- Zone 4: Medium entropy multiple (dihedral) scattering (e.g. in forestry applications, where double bounce mechanisms occur at P and L bands following propagation through a canopy or in urban areas, where dense packing of localized scattering centers can generate moderate entropy with low order multiple scattering dominant).
- Zone 3: High entropy surface scattering (not a feasible region in H/α plane since with entropy $H > 0.9$, surface scattering cannot be distinguished).
- Zone 2: High entropy vegetation (multiple) scattering (e.g. forest canopy or the scattering from some types of vegetated surfaces with random highly anisotropic scattering elements).

- Zone 1: High entropy double bounce scattering (e.g. in forestry applications or in scattering from vegetation which has a well developed branch and crown structure).

The basic scattering mechanism of each pixel of a polarimetric SAR image can then be identified by comparing its entropy and α parameters to fixed thresholds. The segmentation in the H/α plane permits to identify in a macroscopic way the type of scattering mechanism. Agricultural fields and bare soils are characterized by surface scattering. Scattering over forested areas is dominated by volume diffusion while urban areas are mainly characterized by double bounce scattering.

Lee et al. [78] used the H/α decomposition as a basis for the unsupervised classification of the polarimetric data. A segmentation based on the H/α feature space [33] is first applied. The result of this segmentation was then used as the initial training set. Using this obtained training set, the Wishart distribution parameters and then the Wishart distance measure are obtained. In the next step the classified results are used as new training sets based on which the new distance measures are calculated. This process will continue until the number of pixels migrating between classes becomes less than a predefined threshold. One of the limitations in using such H/α initialization is that different objects, like water and road might be assigned to the same class due to the same surface scattering behavior. In order to overcome this problem some other information has to be used. The next two approaches are examples of this.

Pottier et al. [99] showed that by explicitly including the anisotropy information during the segmentation procedure the H/α method can be further improved. As mentioned in section 2.3, the parameter A indicates the relative importance of secondary mechanisms. Thus, using this parameter could be useful to discriminate scattering mechanisms with different eigenvalue distributions but with similar intermediate entropy values. In such cases, a high value of A represents two dominant

scattering mechanisms and a less significant third mechanism. A low value of A shows that there is a dominant first scattering mechanism and two non-negligible secondary mechanisms with equal importance. The A parameter can be used to provide further class refinement [41]. The method is implemented in two successive classification procedures. At first a H/α classification approach is employed. Once the convergence is reached, the 8 output classes of the first step is decomposed into 16 classes by comparing the parameter A of each pixel to a threshold fixed to 0.5. The final Wishart ML is applied on these new sets of clusters.

Since the H and α parameters alone are not sufficient for a good interclass resolution, additional information is needed. Hellmann [54] showed by using the backscatter intensity information contained in the first eigenvalue λ_1 , a better interclass resolution can be achieved. This is because the pure H/α approach has a limited interclass resolution. Therefore, a combination of H , α with λ_1 can improve the interclass resolution. This improvement is mainly for the areas where surface scattering occurs. For example, for the classes of low vegetation, road like structures and water the first eigenvalue λ_1 is dominant. Applying this approach to the E-SAR data, it provided the best interclass resolution and a good classification accuracy was also reported. But the interpretation of the data is more difficult [54].

One of the shortcomings of the previous approaches was that the classification result lacked details, because of the preset zone boundaries in the H and α plane. This could cause the clusters to fall on the boundaries. Also, more than one cluster may be enclosed in a zone. To solve this problem, Lee et al. [80] proposed a new method which not only uses a statistical classifier, but also preserves the dominant polarimetric scattering properties. The algorithm first applies the Freeman and Durden decomposition to divide pixels into three scattering categories: surface scattering, volume scattering, and double-bounce scattering. A merge criterion was also devel-

oped using the Wishart distance measure to merge the clusters in each scattering category. Finally, an iterative Wishart classifier was applied. In comparison with $H/A/\alpha$ Wishart classifier, this algorithm was more efficient in terms of the stability in convergence.

3.2 Full covariance matrix vs. selected features

There are basically two approaches for the use of input features in PolSAR classification. One approach is the use of full covariance (coherency) matrix as the input features. One of the common approaches in this category is the method proposed by Lee et al. [81]. The classification method is a supervised method, using a distance measure based on Wishart statistics and this measure could be incorporated in several classification algorithms [78, 41]. A possible drawback to these approaches is that only second order representations were considered when operating these classifiers [80] and it does not allow combining the covariance data with other types of features such as those obtained from other decomposition methods or textural features or backscattered intensity. Alberga [3] showed that the methods based on the covariance matrix provide important information about the scatterers on the ground, but extra features would be needed in order to distinguish complex classes.

The second approach try to employ the best features in the classification problem. The solution of applying feature selection before classification was suggested by different authors [50, 67, 64, 111, 66, 113]. The feature selection problem can be stated as follows: Given a set of N features find the best subset of m features to be used for classification. This process generally involves a search strategy and an evaluation function [50]. The aim of the search algorithm is to generate subsets of features from the original feature space and the evaluation function compares these feature subsets in terms of discrimination. The output of the feature selection algorithm is the best feature subset found for this purpose. Optimal search algorithms determine

the best feature subset in terms of an evaluation function, whereas suboptimal search algorithms determine a good feature subset. When the number of features increased, using an optimal search algorithm is computationally expensive and thus not feasible.

According to the evaluation function, the feature selection approaches can be broadly grouped into filter and wrapper methods[69]. Wrapper methods utilize the classification accuracy as the evaluation function whereas filter methods uses the inter-class distance measures as the evaluation function. The most widely used inter-class measures are Bhattacharyya distance, divergence and Jeffries-Matusita (JM) distance.

Zhang et al.[138] employed Multiple-Component Scattering Model (MCSM) as the decomposition method to generate features. These features along with the texture features formed the input features for classification. The Sequential Backward Selection (SBS) was chosen in the feature selection procedure, in which the features were sequentially removed from a full candidate set until the removal of further features increases the criterion. The number of support vectors was considered as the measurement of feature selection. They found that with the decreasing number of support vectors, the precision of classification increases. Finally a SVM was finally used for the classification. This schema, though simple and straightforward, has a weakness: the authors limited themselves to the MCSM decomposition method. The other decomposition methods could have been used to expand the input feature vector.

The Logistic Regression (LR), which is a statistical tool for distinguishing the target class from background, has been adopted for the classification of polarimetric SAR data [14]. LR's task is optimize the β_i 's coefficients in the following nonlinear function called logistic function:

$$p(\textit{target} | x) = \frac{\exp[\beta_0 + \sum_i x_i \beta_i]}{1 + \exp[\beta_0 + \sum_i x_i \beta_i]} \quad (3.5)$$

$p(\textit{target} | x)$ is the conditional probability that a pixel belongs to the target class given the vector of input features x for that class. For a dichotomous problem, the probability of background will be:

$$p(\textit{background} | x) = 1 - p(\textit{target} | x) \quad (3.6)$$

LR finds a combination of the features that optimizes the following log-likelihood:

$$L(\beta) = \sum_{i=1}^n \{y_i \ln [p(\textit{target} | x)] + (1 - y_i) \ln [1 - p(\textit{target} | x)]\} \quad (3.7)$$

in which y_i is 1 for targets and 0 for background. Therefore, LR implicitly performs a feature selection: the features that contribute significantly to the discrimination between the background and the target are added to the model.

Shimoni et al. [114] derived a large feature set from multi-frequency multi-polarization PolSAR and PolInSAR data. They developed two level fusion method i.e. LR as feature-level fusion and neural network (NN) method for higher level fusion. According to their work, LR selects the most discriminative features automatically and combines them in order to distinguish the investigated class from all others. Good results were obtained and LR was proved to be as a powerful tool for fusion.

In order to improve the discrimination power and classification accuracy, the texture measures of the SAR image were also in the feature set [139, 27, 130]. For example in [139], gray-level co-occurrence matrix (GLCM) based texture features were added to the span image and the H, Alpha and A parameters. A principle component analysis (PCA) was then used to reduce the feature set. A 3-layer neural network is finally constructed for the classification. In some other studies, fusion of physical and textural information that are derived from various SAR polarizations

has improved the classification results [35, 11].

In [55] a supervised classification method based on neural networks and fuzzy logic is used to learn the class borders from the available learning samples. The advantage of this approach is that other input features can be easily added in order to increase the discrimination ability of the classification.

In addition, a method based on the manual selection of features was evaluated on a subset of the Flevoland dataset [57].

3.3 Single- vs. multi-classifiers

Multiple classifier systems (MCS) or classifier ensembles are another machine learning concept; there has been a great interest for using such an ensemble for solving problems in pattern recognition community recently[6]. By combining different independent classifiers, MCS can improve the classification accuracy in comparison to a single classifier.

There are different methods for creating such an ensemble. These methods include modifying the training samples (e.g. bagging [18] and boosting [42]), manipulating the input features (the input feature space is divided into multiple subspaces [56]), manipulating the output classes (multi-class problem is decomposed into multiple two-class problems, e.g. the error correcting output code (ECOC) [36]). After creation of an ensemble of classifiers, a decision fusion is used to combine the outputs of the classifiers. Several fusion algorithms have been developed and employed, for example, majority voting, D-S evidence theory, fuzzy integral, weighted summation, consensus, mixed neural network and hierarchical classifier system [120, 2, 20].

Although classifier ensembles have given promising results, only a few applications are known for SAR data, in particular for PolSAR data. She et al. [112] introduced Adaboost for PolSAR image classification. Comparing with traditional methods such as Wishart distance classifier, it was found to be more flexible and robust. Chen

et al. [25] proposed a supervised classification scheme based on Adaboost . Each independent element of the Mueller matrix as well as the extracted parameters formed a weak classifier. By the Adaboost procedure, each classifier was endowed with a weight. The weight presented the effectiveness of the corresponding feature. All the weak classifiers were combined together to form a strong classifier using majority weighted voting. They found that this scheme was more robust and more accurate than the traditional maximum likelihood classifier. Min et al. [93] also employed polarimetric decomposition and the Adaboost algorithm to solve a PolSAR image classification problem. Their simulation results validated their method compared with H/α classification algorithm.

In 2010, Zou et al. [141] took advantage of Random Forests [19] for the generation of multiple classifiers. They proposed two feature combination strategies and in the classification stage they employed a robust classifier named Extremely randomized Clustering Forests (ERCFs), which are ensembles of randomly created clustering trees, for terrain classification using Polarimetric data. Comparing with other classifiers, ERCFs got slightly better classification accuracy than SVM with less computational time.

The Logistic Regression (LR), which is a statistical tool for distinguishing the target class from background, has been adopted for the classification of polarimetric SAR data: Shimoni et al. [114] derived a large feature set from multi-frequency multi-polarization PolSAR and PolInSAR data. They developed two level fusion method i.e. LR as feature-level fusion and neural network (NN) method for higher level fusion. According to their work, LR selects the most discriminative features automatically and combines them in order to distinguish the investigated class from all others. Good results were obtained and LR was proved to be as a powerful tool for fusion.

3.4 Incorporating spatial context

One disadvantage of most of the methods reviewed so far is that each pixel is treated to be independent of its neighbors. The local neighborhood have a significant influence on a pixel's class membership. Spatial context as an additional classification knowledge was taken into account in the literature. The main idea behind these methods is that when a certain region has already been classified, with high confidence, as belonging to a single class, it becomes comparatively unlikely that a pixel in this region belongs to another class. This means that two neighboring pixels are not entirely statistically independent.

Basically, there are three ways to incorporate spatial context into a classification i.e. preprocessing, post-processing and processing.

In preprocessing methods, texture channels are often generated and are added to the input features for classification. These spatial features can be extracted using the texture features.

The post-processing approaches are applied after the classification of the original image. The essence of these approaches is to filter the labels in the thematic map, e.g. a mode filter. Another approach in this category for incorporating the spatial context into classification is the probabilistic label relaxation [109, 51, 53]. Label relaxation is based on the the modification of the probability values of each pixel according to the probability compatibility among pixels in a local neighborhood. The goal of label relaxation is to reduce the uncertainty and improve the consistency in the assignment of one of the labels to each pixel in a set of related pixels [106]. In its simplest form, the probabilistic label relaxation involves defining a neighborhood function that expresses support from the neighbors for the labeling of the g^{th} pixel at the k^{th} iteration of the form

$$Q_g^k(w_i) = \sum_n d_n \sum_j p_{gn}(w_i|w_j) p_n^k(w_j) \quad (3.8)$$

in which d_n weights the influence of the n^{th} neighbor, $p_n^k(w_j)$ is the posterior probability that the neighbor n has the label w_j , $\sum_j p_{gn}(w_i|w_j)$ is the conditional probability that pixel g is labeled w_i if the neighbor n is labeled w_j .

Finally, the methods on the processing level incorporate spatial reasoning into the classification. These classifiers are termed as contextual classifiers. For classifying a certain pixel, a contextual classifier directly make use of the contextual information from a pixel window by means of a spectral-spatial distance.

In the context of SAR data, there have been several works taken place in recent years. Reigber et al. [105] employed probabilistic label relaxation in the framework of an unsupervised classification scheme of PolSAR data based on the statistical partitioning of the covariance matrix feature space. They took advantage of this spatial context by taking into account the local neighborhood by altering the probabilities of class membership using a neighborhood function. In this way, robust and homogeneous classification results were obtained even in the presence of strong speckle noise.

Markov Random Fields (MRFs) is a popular model for incorporating spatial context into image classification. They are frequently employed to model neighborhood and class label structure for the classification of remotely sensed data. In particular, [107] and [118] have used MRF models in the classification of SAR data and obtained better results than ordinary classification methods without use of spatial consistency. Some newer works in the context of SAR data can be found in [121] and [43]. One of the weaknesses of the MRFs is that they are computationally complex due to the complicated structure of their likelihood function.

In [12, 13, 14] spatial consistency was also taken into account. They used majority

voting for spatial regularization (MVR): in a neighborhood of each pixel the sum of the conditional probability for each class is determined and the pixel is assigned to the class corresponding to the highest sum. In [14], in order to avoid blurring of linear objects such as roads by the MVR method, they improved this method by replacing the MVR by the second best class regularization (SBR) i.e. in miss-classified pixels, the correct class often corresponds to the second highest value. This method significantly reduced isolated misclassified pixels.

3.5 Parametric vs. nonparametric classifiers

The main objective in this section is to review the works according to the type of classifiers that they are employing. During the past years, different classifier types and different classifier architecture were used for the classification of PolSAR data [82, 26, 63, 8, 49, 24, 37]. These methods can be broadly categorized into parametric and nonparametric methods. Parametric methods, for example the maximum likelihood (ML) [82], is based on the assumption of the complex Wishart distribution of the covariance matrix. If the distribution is not good enough, the classification result is not good either. However, the assumption on the distribution is not necessary when nonparametric classifiers are employed for a classification task.

In 1991, Pottier et al. [101] firstly introduced the Neural Networks (NNs) to PolSAR image classification. In 1999, Hellmann [92] further introduced fuzzy logic with neural networks classifier. More recently, neural network based approaches [139, 137] for classification of polarimetric synthetic aperture radar data have been shown to outperform other aforementioned well-known techniques. Compared with other approaches, neural network classifiers have the advantage of adaptability to the data without making a priori assumption of a particular probability model or distribution. However, their performance depends on the network structure, training data, initialization and parameters. Also, they usually converge slowly and tends to converge to

a local optimization.

Support Vector Machine (SVM) has become an increasingly popular tool for machine learning tasks involving classification, recognition, or detection. SVM is based on statistical learning theorem; however, SVM method does not need the statistical features of the training samples and it can deal with high dimension data and nonlinear problem easily and can also achieve global optimization [132, 133]. It has been mostly applied to hyperspectral remote sensed data and few studies have also been conducted with SAR data [49, 91].

PolSAR image classification using SVM has shown exceptional growth in recent years. Fukuda et al. [48] introduced Support Vector Machine (SVM) to land cover classification with higher accuracy. Additionally, Lardeux et al. [76] showed that the SVM classification algorithm is well suited for full polarimetric data with similar accuracy as Wishart classification. Also, the possibility to add different polarimetric or textural indices makes the SVM algorithm very interesting as the overall performance increased.

Another approach that is applied on remote sensing imagery is Self-learning decision tree classifiers (DT) [95, 115]. The handling of DTs is rather simple and their training time is relatively low compared to computationally complex approaches such as neural networks [95]. In contrast to other classifier algorithms, which use the whole feature space at once and make a single membership decision, a decision tree is based on a hierarchical concept. It is composed of an initial root node, several split nodes and the final leaf nodes. At each node the most relevant feature is selected and used for the construction of a binary test (i.e., decision boundary). In the context of PolSAR data Zhixin et al. [140] integrated polarimetric decomposition, object-oriented image analysis, and decision tree algorithms for RADARSAT-2 PolSAR data. The data were then combined with the parameters of the backscattering and coherency

matrix to form a multichannel image. The decision tree algorithm was used to select features and create a decision tree for the classification. Their obtained results showed that the method outperforms the Wishart supervised classification.

3.6 Summary

The chapter reviewed some of the common algorithms for the classification of PolSAR data. The algorithms were grouped into 5 categories. In the first category, several supervised and unsupervised methods were reviewed. In this research, the reference data have already been collected from the study area. Thus, the proposed methods are supervised. In the second category, several methods were discussed in terms of the input features for classification. In this thesis, a feature selection is adopted to select the most informative features for the classification. In the third category, it was shown that using a classifier ensemble might provide better results than single classifier scheme. Therefore, a multiple classifier scheme is proposed in this thesis. Next, the methods were categorized into parametric and nonparametric methods. In this work, a support vector machine, which is a nonparametric method, is used as the core classification algorithm. Finally, as was shown in the last category, using the contextual information can improve the classification results. Therefore, the neighboring pixels information are incorporated in the proposed schema.

Chapter 4

Methods

A flowchart (figure 4.1) was used to describe the steps taken in order to analyze Radarsat-2 images for forest mapping. Dataset preparation along with the ground truth preparation will be explained in chapter 5. After generating SAR parameters which was explained in chapter 2, this chapter focuses on the methodologies and experimental design used for the classification of polarimetric data. As can be seen in figure 4.1, the classification section is categorized into single classifier, multi classifier and one that incorporates context. For the single classifier, two methods are proposed. The first method is a nonparametric feature selection (NFS) algorithm (section 4.1.1) which is based on a nonparametric definition of the evaluation function in feature selection. The second is a fast wrapper feature selection (FWFS) method (section 4.2.1). FWFS uses the classification accuracy as the evaluation function. In this regard, a random sampling strategy is proposed to speed up the training process. The single classifier methods are extended to multiple classifiers schemes, by proposing a class-based definition of the evaluation function to create an ensemble of classifiers. In this framework, two methods are proposed: class-based feature selection (CBFS) method (section 4.1.2) and fast wrapper class-based feature selection (FWCBFS) method (section 4.2.2) which are the ensemble-based versions of the NFS and FWFS respectively. Finally, in order to incorporate the neighboring pixels information in the classification, a contextual class-based feature selection (CCBFS) method (section 4.3) is proposed. In a certain point of view, CCBFS can be considered as the evolution of the FWCBFS.

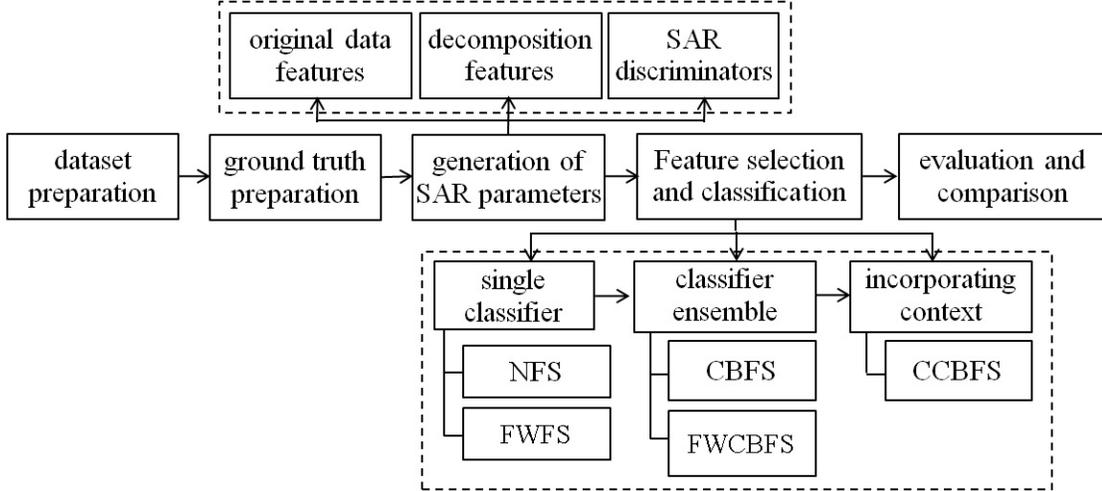


Figure 4.1: Steps taken to analyze RS-2 images for forest mapping

4.1 Nonparametric methods

4.1.1 Single Classifier: NFS

Feature selection

The feature selection process generally entails a search strategy and an evaluation function [50]. The former aims to generate subsets of features from the original feature space while the latter compares these feature subsets using a specific criterion. The output of the feature selection algorithm is the best feature subset found for this purpose.

For the evaluation function, inter-class distance measures are often chosen. The most widely used inter-class measures, such as Bhattacharyya distance, divergence and Jeffries-Matusita (JM) distance, are all parametric methods and assume that the data are following a known distribution. But, most of the SAR parameters are of complex with unknown statistical properties. For this reason, the above-mentioned inter-class measures are not appropriate. To account for this, a non-parametric separability measure is adopted to evaluate the generated feature subsets. This measure is based on the calculation of the between-class M_b and within class M_w scattering matrices. The within-class scatter matrix is a measure of how compact a class is. It

is defined as

$$M_w = \sum_{i=1}^N M_i \quad \text{where} \quad M_i = \sum_{j=1}^{n_i} (x_j - \mu_i)(x_j - \mu_i)^T \quad (4.1)$$

where n_i is the number of samples in class i , x_j is the j^{th} sample in class i , μ_i is the mean vector for class i , M_i is the within class scattering matrix for class i , M_w is the sum of the within class scattering matrices for all classes and N is the number of classes. The between-class scatter matrix measures the separation between classes and is defined as

$$M_b = \sum_{j=1}^N n_j (\mu_j - \mu)(\mu_j - \mu)^T \quad (4.2)$$

where M_b is the between class scattering matrix.

An objective function F can be obtained using the Fisher class separability measure which is the ratio of the determinant of the between-class scatter matrix to the determinant of the sum of within-class scatter matrices[39]

$$F = \frac{|M_b|}{|M_w|} \quad (4.3)$$

The measure described above, although simple and straightforward, has several disadvantages. One is that it only works well when the distribution of the classes are normal. But, in SAR parameters the data distributions are often complicated and not normal. Second, the rank of the matrix M_b equals $N - 1$, so the maximum number of features that can be selected is $N - 1$. But in applications where the classes are very close, a large number of features would be required to solve the classification problem. So, this limits the performance of the final classification.

In order to solve these limitations Kuo and Landgrebe [74] proposed a nonparametric definition for the calculation of M_b and M_w . Although, the method was originally proposed for the feature extraction in hyperspectral data, we are using the method

for the calculation of Fisher ratio as the evaluation function for feature selection. The method is based on the discriminant analysis feature extraction by focusing on samples near the decision boundaries. The main idea is to put different weights on each sample to compute the “weighted means” and defining new nonparametric between-class and within-class scatter matrices to obtain more than $N-1$ features [75]. In this method, the nonparametric between-class scatter matrix for N classes is defined as:

$$M_b = \sum_{i=1}^N \frac{p_i}{N_i} \sum_{j=1}^N \sum_{k=1}^{N_i} \lambda_k^{(i,j)} (x_k^{(i)} - M_j(x_k^{(i)}))(x_k^{(i)} - M_j(x_k^{(i)}))^T \quad (4.4)$$

where $x_k^{(i)}$ refers to the k^{th} sample from class i , N_i is training sample size of class i , P_i denotes the prior probability of class i . The scatter matrix weight $\lambda_k^{(i,j)}$ is a function of $x_k^{(i)}$ and $M_j(x_k^{(i)})$, and defined as:

$$\lambda_k^{(i,j)} = \frac{dist(x_k^{(i)}, M_j(x_k^{(i)}))^{-1}}{\sum_{l=1}^{N_i} dist(x_l^{(i)}, M_j(x_l^{(i)}))^{-1}} \quad (4.5)$$

where $dist(a, b)$ denotes the Euclidean distance from a to b . $M_j(x_k^{(i)})$ denotes the weighted mean $x_k^{(i)}$ in class j and defined as:

$$M_j(x_k^{(i)}) = \sum_{l=1}^{N_i} W_{lk}^{(i,j)} x_l^{(i)} \quad (4.6)$$

where,

$$W_{lk}^{(i,j)} = \frac{dist(x_k^{(i)}, x_l^{(j)})^{-1}}{\sum_{l=1}^{N_i} dist(x_k^{(i)}, x_l^{(j)})^{-1}} \quad (4.7)$$

The nonparametric within-class scatter matrix is defined as:

$$M_w = \sum_{i=1}^N \frac{p_i}{N_i} \sum_{k=1}^{N_i} \lambda_k^{(i,j)} (x_k^{(i)} - M_i(x_k^{(i)}))(x_k^{(i)} - M_i(x_k^{(i)}))^T \quad (4.8)$$

Once the between-class and within-class scatter matrices are defined, they are plugged into equation 4.3 to calculate the fisher ratio for the evaluation function of

the feature selection.

For the search strategy, the sequential forward floating selection (SFFS) method [103] was used in this thesis. According to Jain and Zongker [64], sequential floating search methods are probably the most effective search algorithms for feature selection. SFFS starts with a null feature set and, for each step, the best feature that satisfies the evaluation function is included with the current feature set. The algorithm also verifies the possibility of improvement of the evaluation function if a feature is excluded. In this case, the worst feature (in terms of the evaluation function) is eliminated from the set. Therefore, the SFFS proceeds dynamically increasing and decreasing the number of features until the desired number of features is reached.

Classification

Upon the selection of the most appropriate features, they are transferred to the classification step. Because of its ability to take numerous and heterogeneous features into account, such as the intensity channels and different SAR polarimetric parameters, as well as its ability handle linearly non separable cases, a support vector machine (SVM) algorithm is proposed for the classification step. The LIBSVM package [23] available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm> was used for the SVM classification in our work.

The Support Vector Machine (SVM) is a relatively recent approach introduced by Vapnik [132, 133], for solving supervised classification and regression problems. SVM based approaches have received much attention as a promising tool for the classification of remotely sensed images. It has been shown in several studies that classification by a SVM can be more accurate than other nonparametric classifiers such as neural networks and decision trees as well as parametric classifiers such as maximum likelihood classification [44, 61, 90]. In the following a brief description of the SVM is given, and more detail can be found in [22].

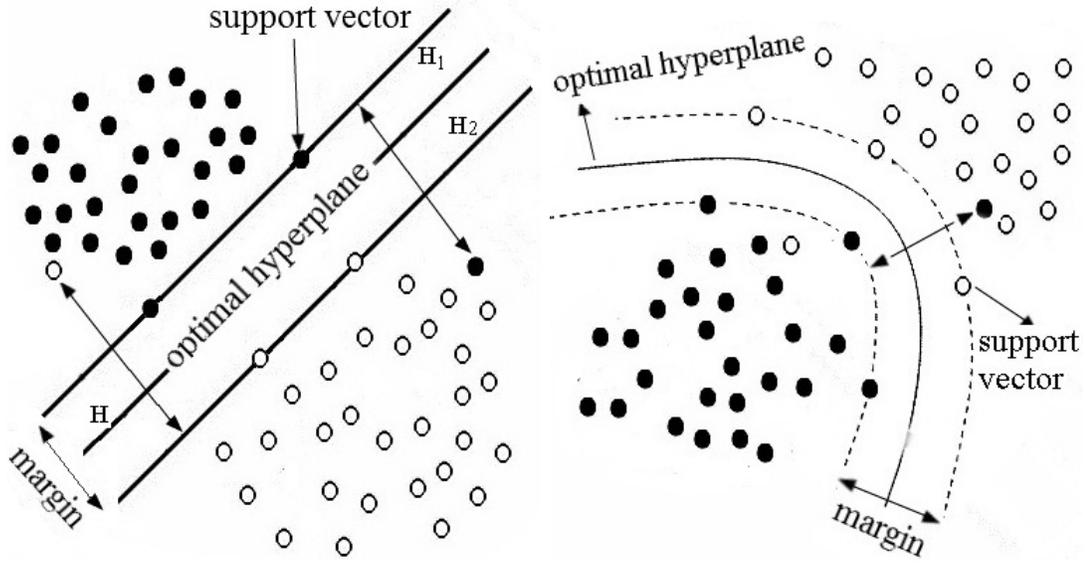


Figure 4.2: Separating hyperplanes in SVM: left (linear case) and right (non-linear case)

The basic approach to SVM is first explained in a linear case, then the extension to non-linear classification is described.

1. Linear classification: The simplest starting point to outline the SVM based classification is to consider situation in which there are two linearly separable classes. Let's assume that there are r training samples each of which is described by an N dimensional vector x_j ($j \leq r$). The N vector, in our case, are the SAR polarimetric parameters. y_j is the label assigned to the j th sample x_j . For the present two-class problem $y_j \in \{-1; +1\}$. The aim of the SVM classification is to separate the two classes by finding an optimal separating hyperplane (OSH). This hyperplane in feature space is defined by the equation $H : w \cdot x + b = 0$, where x is a point lying on the hyperplane, w is the normal vector to the hyperplane and b is the offset (figure 4.2). The classification function f assigning the label y_j to the sample x_j is represented by

$$f(x_j) = \text{Sign} \{ (w \cdot x_j) + b \} \quad (4.9)$$

where $w \cdot x_j$ is the dot product of the vectors w and x_j and $Sign$ is the sign function.

A separating hyperplane can be defined as $w \cdot x_j + b \geq 1$ for the class $y_j = +1$ and $w \cdot x + b \leq -1$ for the class $y_j = -1$. The training samples on the two hyperplanes $H_1 : w \cdot x + b = +1$ and $H_2 : w \cdot x_j + b = -1$ are called the support vectors. These samples are found to be nearest the hyperplane H . Combining the above two inequality equations we will have

$$y_j(w \cdot x_j + b) - 1 \geq 0 \quad (4.10)$$

The distance between these planes H_1 and H_2 is called the margin and it is $\frac{2}{\|w\|}$, where $\|w\|$ refers to the norm of w vector. The goal of SVM is to compute the OSH or equivalently, to maximize the margin by finding $\min \left\{ \frac{1}{2} \|w\|^2 \right\}$ under the inequality condition of 4.10.

Using the Lagrangian formalism, the optimal decision function can be obtained

$$f(x_j) = Sign \left\{ \sum_{i=1}^r y_i \alpha_i (w \cdot x_j) + b \right\} \quad (4.11)$$

in which α_i are the Lagrange multipliers.

2. Non-linear classification: Due to speckle, the SAR classification problem may not be linearly separable. In this case, no hyperplane can be found in the input feature space between the two classes. To solve this, SVM uses the kernel method to map the data with a non-linear transformation to a space of higher dimension $M (M > N)$, in which a linear separating plane between the two classes can be found. This transformation, which is realized through a transformation function $\Phi (\Phi : R^N \rightarrow R^M)$, tends to spread the data out in a way that facilitates finding of a linear separating plane. In this case the classification

function f becomes

$$f(x_j) = \text{Sign} \left\{ \sum_{i=1}^r y_i \alpha_i (\Phi(x_j) \cdot \Phi(x_i)) + b \right\} \quad (4.12)$$

where the $\Phi(x_j) \cdot \Phi(x_i)$ is called the kernel function K

$$K(x_j, x_i) = \Phi(x_j) \cdot \Phi(x_i) \quad (4.13)$$

Combining 4.12 and 4.13 we have

$$f(x_j) = \text{Sign} \left\{ \sum_{i=1}^r y_i \alpha_i K(x_j, x_i) + b \right\} \quad (4.14)$$

Three kernels are commonly used

- (a) The polynomial kernel $K(x, x_i) = (x \cdot x_i + 1)^r$
- (b) The sigmoid kernel $K(x, x_i) = \tanh(x \cdot x_i + 1)$
- (c) The radial basis function (RBF) kernel $K(x, x_i) = \exp \left\{ -\frac{|x-x_i|^2}{2\sigma^2} \right\}$

The basic SVM approach for a two-class problem may be extended for multi-class classification. This can be achieved using the One Against All (OAA) or One Against One (OAO) strategies [60]. Considering the classification problem with c classes, the OAA algorithm constructs c hyperplanes that separate respectively one class and $c - 1$ other classes. Alternatively, the OAO algorithm tries to separate each pair of classes and thus needs $\frac{c(c-1)}{2}$ hyperplanes. In the two cases the final label is the one which has been chosen most often.

The total training time seems to be longer in OAO, but actually the individual problems are smaller and hence its training time is shorter than the OAA method with comparable performance. Because of this, the LIBSVM package uses the OAO training method.

4.1.2 Multiple Classifier System: CBFS

Most of the feature selection algorithms seek only one set of features that distinguish among all the classes simultaneously. This, in one hand, can increase the complexity of the decision boundary between classes in the feature space [73] and limit the classification accuracy. On the other hand, considering one set of features for all the classes requires a large number of features.

To overcome these problems, a multi classifier schema is used in this research. In doing so, a class-based feature selection (CBFS), is proposed. The main idea of the CBFS is that from the large number of SAR parameters extracted from PolSAR data there are some parameters which can discriminate each class better than the others. The method comprises two parts: the feature selection, and classification.

Feature selection

Instead of using feature selection for all classes, CBFS selects the features for each class separately. First of all, the feature selection process is applied for the first class; hence, the most appropriate features for discriminating the first class from the others is selected. Next, the most discriminative features for the second class are selected by using the same procedure for the second class. This process is repeated until all the feature subsets for all classes are selected (see figure 4.3).

The evaluation function used for NFS can not be applied in this method. Because, NFS considers all the classes in each round of feature selection whereas CBFS focuses on one class in each round of feature selection. Thus, here we are proposing a class-based definition of the Fisher ratio

$$F^{CB}(i) = \frac{|M_b^{CB}(i)|}{|M_w^{CB}(i)|} \quad (4.15)$$

where $M_b^{CB}(i)$ and $M_w^{CB}(i)$ are the class-based definition of the between-class and within-class scatter matrices for class i . For each class m the M_b^{CB} is defined as

$$M_b^{CB}(m) = \frac{p_m}{N_m} \sum_{j=1}^N \sum_{k=1}^{N_m} \lambda_k^{(m,j)} (x_k^{(m)} - M_j(x_k^{(m)}))(x_k^{(m)} - M_j(x_k^{(m)}))^T \quad (4.16)$$

As can be seen, the between-class scatter matrix is defined as the distance between the specific class m and the rest of classes (and not the distance between all classes).

The M_w^{CB} is also defined as

$$M_w^{CB}(m) = \frac{p_m}{N_m} \sum_{k=1}^{N_m} \lambda_k^{(m,j)} (x_k^{(m)} - M_m(x_k^{(m)}))(x_k^{(m)} - M_m(x_k^{(m)}))^T \quad (4.17)$$

It is obvious that instead of sum of within-class scatter matrices, the class-based within class scattering matrix is only calculated for the class of interest m at each time.

Classification

After creating the feature subsets, an SVM classifier is trained on each of the selected feature subsets and the outputs of the classifiers are combined through a combination mechanism.

Based on the classifiers outputs there are several consensus rules for the combination process. When the outputs of classifiers are labels, voting techniques can be used for combination. Apparently, the most informative outputs are when the classifiers provide probabilities. These probabilities states the confidence of the classifier on different classes. Thus, it is highly desirable to use a special SVM classifier with multi-class probability estimates.

Basically, SVM is a discriminative classifier that predicts only the class label, without probability information. However, it can be extended to give the probability estimates. Methods for translating the output of SVM into probability estimates are discussed and compared in [136]. In our implementation we use the LIBSVM package

which supports probability estimates for SVM classifier based on the methods of [136]. The objective is to estimate, for each pixel x , the probabilities to belong to each class of interest:

$$p_k = p(y = k | x), k = 1, \dots, M$$

For this purpose, the outputs of each SVM are mapped into the pairwise class probabilities r_{ij} for all classes i and j using a simple logistic sigmoid model [97]

$$r_{mn} = \frac{1}{1 + \exp(Af(x) + B)} \quad (4.18)$$

where $f(x)$ represents the output of the SVM trained to separate the class m from class n for the input value x . A and B are estimated by minimizing the negative log likelihood of training data using their labels and decision values. Once the values of r_{mn} are estimated for all classes, the posterior probabilities for all classes can be estimated as follows

$$\begin{aligned} & \arg \min_{p_m} \sum_{m=1}^N \sum_{n:n \neq m} (r_{nm} p_m - r_{mn} p_n)^2 \\ & \text{subject to } \sum_{m=1}^N p_m = 1, p_m \leq 0, \forall m. \end{aligned}$$

When each classifier produce N measurements (one for each of N classes) for each classifying pixel, the simplest means of combining them to obtain a decision are the the measurement-level methods [68]. The most commonly used measurement level methods are mean and product combination rules. Mean rule simply adds the probabilities provided by each classifier for every class, and assigns the class label with the maximum score to the given input pattern. Analogously, product-rule multiplies the probabilities for every class and then outputs the class with the maximum score.

These rules perform the same classification in most cases. Kittler et al. [68] showed that the mean rule is less sensitive to noise than other rules. Therefore, in this research, we applied the mean rule as the combination method. According to mean combination rule the pixel x is assigned to the class ω_i if

$$\sum_{j=1}^M p(x_j|\omega_i) = \max_{k=1}^N \sum_{j=1}^M p(x_j|\omega_k) \quad (4.19)$$

in which M is the number of classifiers and N is the number of classes. In our case $N = M$.

4.2 Wrapper methods

The inter-class distance measures as the evaluation function of the feature selection although a reasonable method of the similarity and dissimilarity, they are not directly related to the ultimate classification accuracy. A question arises as to whether it is possible to use a more direct criterion as the evaluation function. According to the evaluation function, the feature selection approaches can be broadly grouped into filter and wrapper methods [69]. Wrappers utilize the classification accuracy as the evaluation function whereas filters uses the inter-class distance measures as the evaluation function. The optimized problem in filters is different from the real problem. Nonetheless, filters are faster because the problem they solve is in general simpler. Alternatively, wrappers try to solve the real problem and the considered criterion is really optimized which means that the ultimate problem has to be solved numerous times. For this reason wrappers are potentially very time consuming. In particular, as SVM is used as the ultimate classifier in this research, it involves time consuming optimization. Furthermore, as the number of data points increases, the limitation of SVM as the evaluation function becomes more significant.

This time complexity is mainly due to the SVM training time. This is because

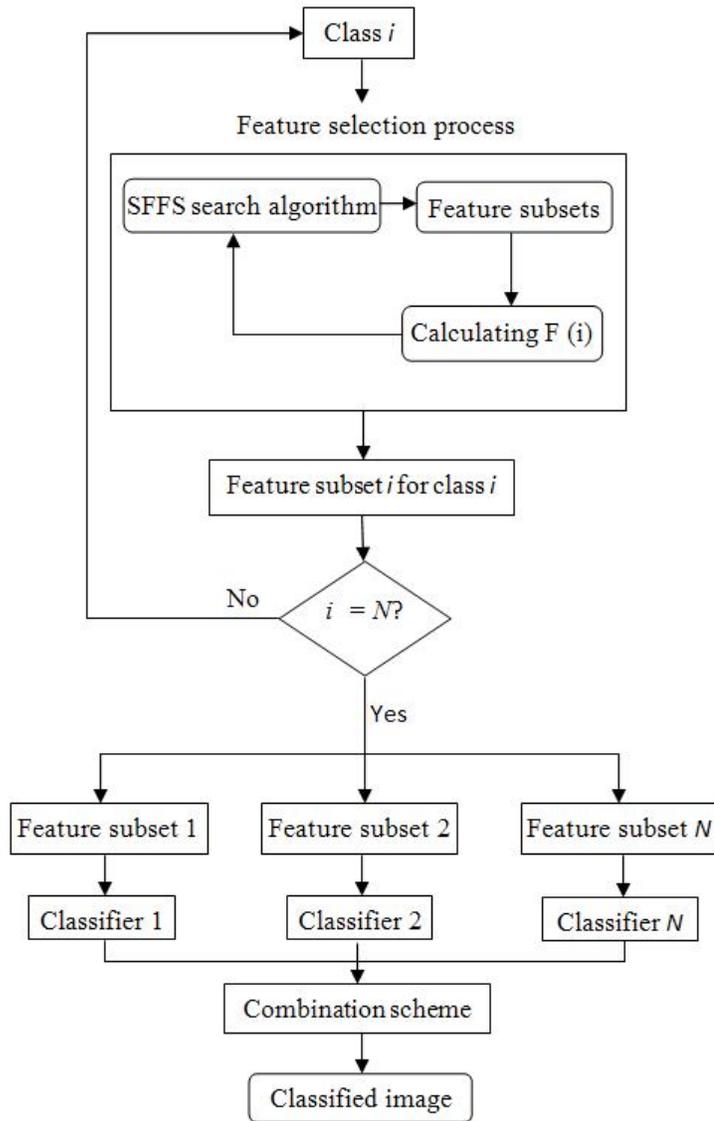


Figure 4.3: The proposed multi-classifier schema

quadratic programming (QP) problems need to be solved for training SVMs. There have been several methods in the literature to accelerate the SVM training process. Some methods, called subset selection, try to decompose the original QP problem into a number of sub-problems which decrease the training time [65, 34]. Despite reducing the time complexity, it still cannot meet the requirements needed in optimization. Because, in the optimization process, the SVM classifier will repeat over and over again until a pre-specified condition (i.e. the desired number of features) is achieved. Therefore, these algorithms are not efficient for optimization. Sampling a small number of training data out of the original dataset has also been used to reduce the time complexity [122]. Specifically, some random sampling methods e.g. [134] has been used to speed up the SVM training process [84]. Random sampling is a simple and efficient method for reducing time complexity. However, it has some shortcomings: 1) some of the effective samples could be sacrificed during the sampling process; and 2) the randomly selected points may not be a true representative of the whole dataset.

4.2.1 Single Classifier: FWFS

In order to alleviate the above-mentioned shortcomings, a fast wrapper feature selection (FWFS) method is proposed in this thesis. The method adopts a random selection type algorithm based on the notion of support vectors. The support vectors are the essential or critical training samples which lie closest to the decision boundary. If all other training samples were removed and training was repeated, the same separating hyperplanes would be obtained. The number of support vectors (nSV) also characterizes the complexity of the classification task: if nSV is small, then that suggests that only a few examples are important, and that many samples can be disregarded without any loss in accuracy. Alternatively, if nSV is large, then nearly every sample is important. This concept is used in our study for training data reduction. This reduction procedure is performed in two steps:

1. Random sampling: Instead of sampling the same number of data points from all the classes, the number of the support vectors for each class was taken as a measure for the amount of training points reduction for that class. An SVM classification with full feature set is run before the feature selection process begins. Depending on the number of support vectors in each class, the random sampling of the training data is followed next. The amount of reduction in each class i is obtained using

$$R(i) = 1 - \frac{nSV(i)}{Nr(i)} \quad i = 1, 2, \dots, M \quad (4.20)$$

in which $Nr(i)$ and $nSV(i)$ are the number of samples and the number of support vectors in each class i respectively. $R(i)$ is the amount of reduction for the class i .

2. Using support vectors: Although, the above step reduces the number of training samples, this reduced training sample set is still not enough for our goal of using SVM for feature selection. To further reduce the number of training samples, using the above reduced samples, an SVM classifier is trained on the full feature set. Finally, the support vectors of each class are used as the training samples.

After reducing the training samples, we go to the feature selection step. Like the NFS method, we again used the sequential forward floating selection (SFFS) method [103] as the search strategy in this thesis to generate subsets of features from the original feature space. However, unlike the NFS method which used a nonparametric definition of separability measure, here we use the overall accuracy of the SVM classifier as the evaluation function for feature selection. For each feature subset s this evaluation function F is defined as:

$$F = overall(s) \quad (4.21)$$

The searching process proceeds until the desired number of features is reached.

4.2.2 Multiple Classifier System: FWCBFS

The same as the NFS method , we propose the FWFS in the context of a class-based feature selection schema and hence the name FWCBFS. Trying to optimize the classification accuracy for each class, the FWCBFS method finds the best SAR parameters for each class separately. For each class i , the producer accuracy for that class i is used as the evaluation function.

$$F(i) = \text{producer}(s(i)) \quad (4.22)$$

This process is repeated until all the feature subsets for all classes are selected. The general schema of the CBFS method was already described in section 4.1.2.

4.3 Contextual Multiple Classifier System: CCBFS

One of the disadvantages of the classification methods described above is that each pixel is classified independently of its neighbors. When a majority of pixels in a certain region are assigned to a single class, it becomes highly unlikely that a pixel in this region belongs to another class. This misassignment could likely be due to the speckle noise. Therefore, an improvement was proposed to the FWCBFS method to take the spatial context into account.

In this research, we propose to incorporate the spatial consistency in the combination mechanism of the class-based schema and hence the name contextual class-based feature selection (CCBFS). The method can be categorized as a post-processing technique. It does not only consider the generated probabilities of the classifying pixel in different classifiers, the neighboring probabilities of the that pixel are also taken into account. Recalling the simple combination rule in equation 4.19, we see that the combination is performed pixel by pixel regardless of their neighborhood. Instead, the

proposed spatial consistent decision rule can be performed for each pixel by considering the likelihoods of its neighboring pixel in different classifiers output. A weighting function $w(l)$, centered on the pixel, was employed. Based on their distance, the pixels which are closer to the central pixel are given more weights. For each pixel, the weighted sum of the neighboring likelihoods can be calculated for each class in all classifiers and the pixel will be assigned to the class with the highest sum. According to this combination rule, the pixel x is assigned to the class ω_i if

$$\sum_{j=1}^N \sum_{l=1}^n w(l) p(x_{lj} | \omega_i) = \max_{k=1}^M \sum_{j=1}^N \sum_{l=1}^n w(l) p(x_{lj} | \omega_k) \quad (4.23)$$

where $p(x_{lj} | \omega_k)$ is the probability that a pixel x_{lj} belongs to the class ω_k , N is the number of classifiers, M is the number of classes and n is the number of neighboring pixels. For instance, $n = 9$ if a 3×3 neighborhood is taken into consideration. This could provide a very robust and homogeneous classification results even in the presence of strong speckle noise. Figure 4.4 shows the combination rules in simple and contextual class-based schemes.

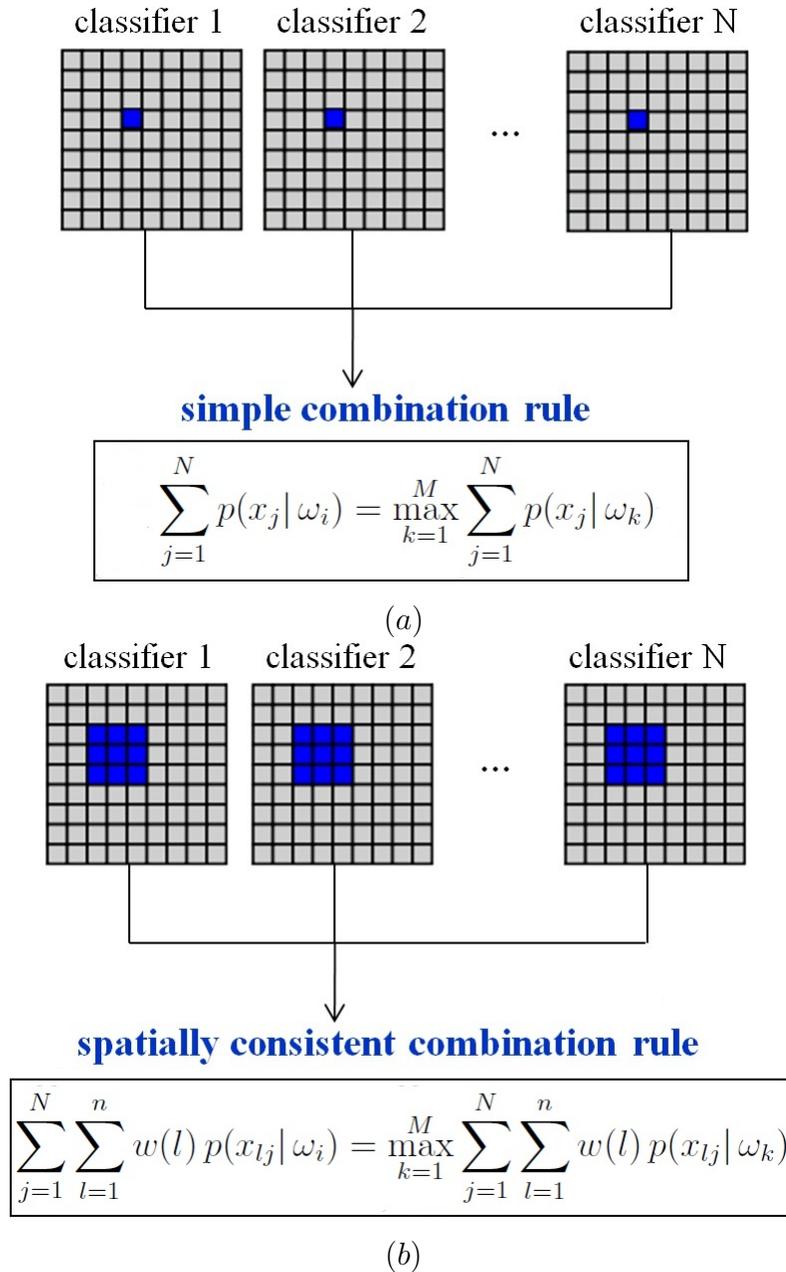


Figure 4.4: simple (a) vs. spatially consistent (b) combination rules

Chapter 5

Study area and dataset description

5.1 Study area description

The study site selected is the Petawawa Research Forest (PRF) located near Chalk River, Ontario (45°57' N, 77°34' W). It is approximately 200 km west of Ottawa and 180 km East of North Bay, Ontario, Canada. Figure 5.1 shows the map of the study area.

The PRF is the oldest continuously operated forest research center in Canada. It maintains more than 2000 experimental plots and sites making it an excellent resource for advanced remote sensing technologies. This experimental forest is larger than 100 km² in size and is characterized by white, red and jack pine, white and black spruce, poplar and red oak. About 85% of the PRF is productive forest land with growing stock estimated to be 1.5 million m². Harvesting schedules vary depending on research program needs but the typical volume of harvested wood is 2400 - 7000 m² per year.

The forest lies on the southern edge of the Canadian Shield with bedrock consisting largely of granites and gneisses. Thin soils are a result of significant glacial impact. This also makes the terrain relatively flat consisting of sandy plains, gently rolling hills and bedrock outcrops. The elevation ranges from 140 - 280 m above sea level. Climate is continental with winter temperatures cold enough to support a fixed period of snow cover. The precipitation is evenly distributed through the year with 600 mm in the form of rain and 200 mm in the form of snow. January is the coldest month with a mean temperature of -12.4°C and July is the warmest with a mean temperature of 19.1°C. The site is in the Great Lakes St. Lawrence forest region [110] and contains both boreal and temperate forest species. Common species are white pine (*Pinus*



Figure 5.1: Petawawa Research Forest (adapted from Google maps)

strobus L.) (Pw), red pine (*Pinus resinosa* Ait.) (Pr), jack pine (*Pinus banksiana* Lamb.) (Pj), white spruce (*Picea glauca* (Moench Voss) (Sw), black spruce (*Picea mariana* (Mill) B.S.P.) (Sb), poplar (*Populus* L.) (Po) and red oak (*Quercus rubra* L.) (Or). More information about the study area including map of the site and other radar results from the site can be found in [77].

5.2 Dataset description

5.2.1 Ground truth data

Ground truth data were collected from a circa 2002 forest inventory map, aerial photographic interpretation, Landsat ETM + images and field visits. The forest inventory map contained information on large forest stands representing 18 different species of trees. Figure 5.2 shows some of these stands overlaid on the Google earth images. The necessary reference data for our analysis were collected from these polygons. For this, the following preparation steps were taken prior to this:

1. The polygons with small number of samples were deleted
2. The polygons in which there were a mixture of two or more species were deleted (figure 5.3 a)

3. Edgy stands i.e. those polygons existing on the steep slope were not considered (5.3 b)
4. The low density stands or stands with two different densities were ignored (figure 5.3 e)
5. Depleted areas including clearcut (figure 5.3 c) or partial cut (figure 5.3 d) after the inventory were deleted

As some changes could have occurred in the study area before or after the collection of image data, a consultation was held with Don Leckie of the Pacific Forestry Center who has expert knowledge of the study area as well as remote sensing principles. He provided us with their most recent Petawawa site evaluation carried out in March 2008. In this documentation, the sites were examined for oddities such as: forest composition, terrain in terms of topography, species, damage, holes, non-uniformity, two storied stands, residual trees that might affect radar, inclusions of wetlands, roads and streams that might affect radar, etc. The inventory data were also checked for changes from the date of the inventory in 2002 including clear-cuts, partial-cuts and sites which have now been split because of a road or other features. All these notes were taken into account in the preparation step. Figure 5.3 shows some examples of this pruning process.

In order to evaluate the classification algorithms with different types of ground truth data, a variety of forest classes were considered in this study including highly similar species (e.g. red pine, white pine and jack pine) and highly different classes (e.g. ground vegetation and poplar), classes with a fairly large number of samples (e.g. red oak) and finally classes with very small sample size (e.g. white spruce and mixed species). A total of nine classes were finally considered in this study including red oak, poplar, mixed hardwood species (mostly maple (*Acer* L.), birch (*Betula* L.) and beech (*Fagus grandifolia* Ehrh.)), white pine, red pine, jack pine, black spruce,



Figure 5.2: The forest inventory stands overlaid on high resolution images obtained from Google earth



(a)



(b)



(c)



(d)



(e)

Figure 5.3: Undesirable polygons:(a) mixture of red pine and poplar (b) edge polygon(c) clear-cut area(d) partial cut area (e) low density area

type	name	species	# of training	# of testings
hardwood	Or	Red oak	1109	1239
	Po	Poplar	867	769
	Hx	Mixed species	276	233
softwood	Pr	Red pine	882	802
	Pj	Jack pine	272	232
	Pw	White pine	1257	1265
	Sb	Black spruce	876	875
	Sw	White spruce	393	375
vegetation	GV	Ground vegetation	1531	1737
Total			7463	7527

Table 5.2: List of classes and training and testing sample sizes

white spruce, and ground vegetation. The sites for the forest classes were dense, mature and of pure species. The ground vegetation sites consisted of grass, herbs, sweet fern, bracken and low shrubs of moderate or high density. Although, classes with a mixture of species were removed in the preparation step, Hx was included in the study because its constituent species do not appear in other classes. Also, this mixture is a common occurrence in forests in Canada. Table 5.2 shows the final set of classes with their corresponding number of training and test samples. Almost half of the samples were used for training and the rest for testing.

As can be seen in table 5.2, the forest species used in this study were a combination of both softwood and hardwood. The hardwood classes (e.g. Or, Po and Hx) are deciduous trees that drop their leaves every year, whereas softwoods (e.g. Pr, Pw, Sb and Sw) have needles that persist throughout the year. The ground vegetation in the leaf-off image will have senesced and for the low shrub the leaves were dropped.

5.2.2 SAR data

A set of images acquired in fine quad-polarized (FQ) mode by Radarsat-2 (RS-2) on the PRF test site was provided through Pacific Forestry Center. RS-2 was successfully launched on 14 December, 2007 and commissioned on 25 April, 2008. Radarsat-2

band	c-band
frequency	5.405 GHz
polarization	HH, HV, VV, VH
Resolution [range×azimuth](m)	11×9
Nominal swath width (km)	25
Near incidence angle (°)	28.0
Far incidence angle (°)	29.8
Altitude (average)	798 km
Inclination	98.6 degrees
Period	100.7 minutes
Ascending node	18:00 hrs
Sun-synchronous	14 orbits per day
Repeat cycle	24 days

Table 5.4: Overview of the principal characteristics of the RS-2 data

offers fully polarimetric (HH, HV,VH, VV) radar imagery as well as multiple other beam modes, resolutions and incidence angles. Radarsat-2 operates in the C-band (5.405 GHz) with a wave length of 5.55 cm. The satellite is in a polar sun synchronous orbit at an altitude of approximately 798 km. The orbit has a period of 101 minutes and allows for a 24 day ground repeat. Due to its ability to acquire images at different incidence angles, it is possible to image the same area more than once in a 24 day repeat period. In order to obtain the maximum amount of data from each image at the highest resolution possible, the fine quad polarization beam mode was chosen. RS-2 uses horizontal and vertical polarizations sending alternating pulses of each. Radarsat-2 has two receivers and so can receive both horizontal and vertical polarizations simultaneously. Table 5.4 presents principal characteristics of the Radarsat-2 data in the fine quad beam mode.

Two images of the site, which were collected in the leaf-on and leaf-off season, were chosen for this study. All data were delivered as Single-Look Complex (SLC) data. The images were acquired in the same imaging mode (FQ9) and the same orbit (ascending) as can be seen in table 5.7. Monitoring of environmental conditions

#	leaf on/off	mode	orbit	season	date
308	leaf on	fq9	ascend	summer	2009/8/4
347	leaf off	fq9	ascend	winter	2009/11/8

#	local time	surface moisture	snow on tree	soil moisture	wetness	air temp
308	5:55 pm	low	no	low	N/A	22
347	5:55 pm	variable	no*	mod	N/A	7

*no snow on the trees but some patchy wet snow on the ground

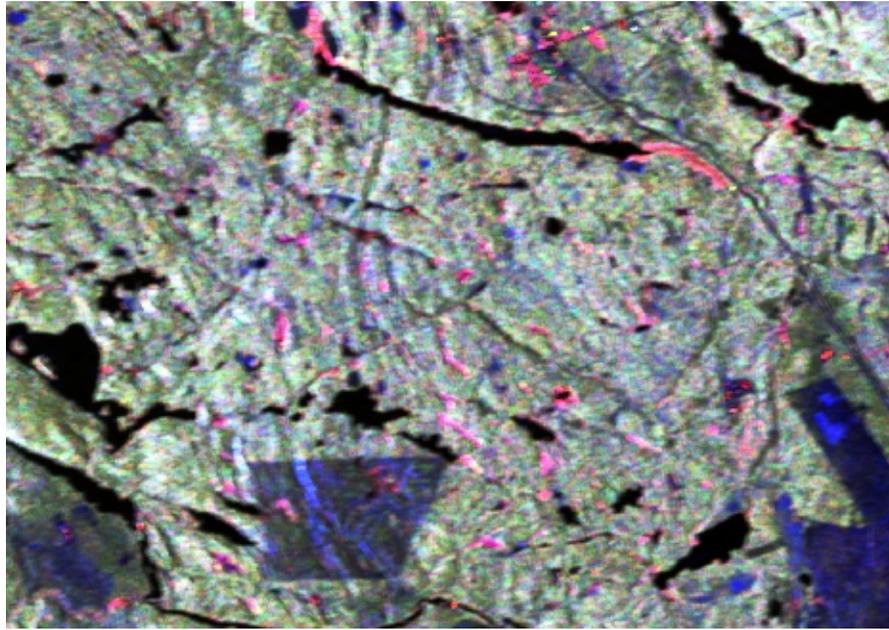
Table 5.7: Acquired image details

indicated that both images had some (low) moisture on the canopy and open ground vegetation at the time of acquisition. An image subset, that covers a forest area of approximately $15 \times 16 \text{ km}^2$, was selected for testing the developed classification algorithms.

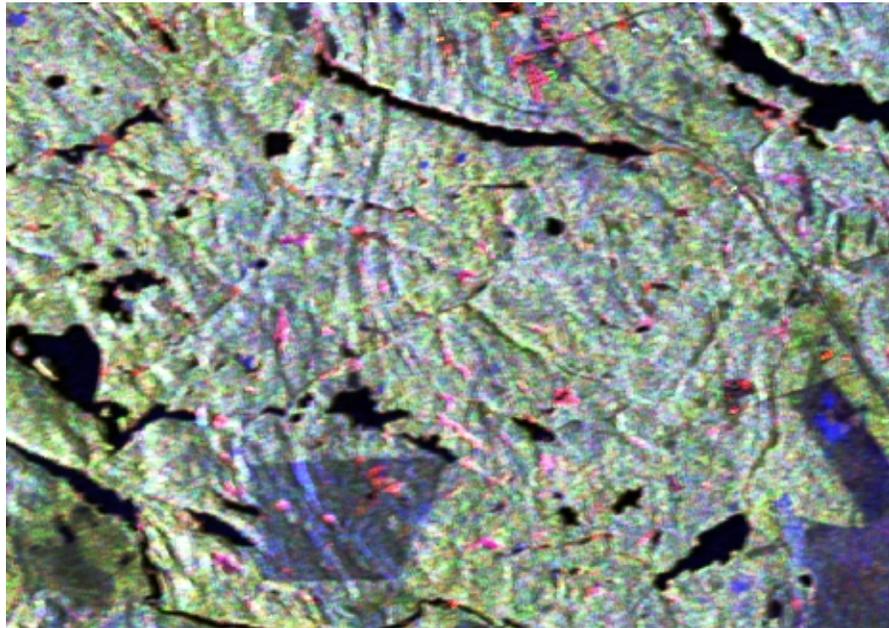
The polarization color composite of the subset of the leaf-on and leaf-off images are displayed in figure 5.4. The color scheme is based on the Pauli vector by assigning $|HH - VV|$, $|HV|$ and $|HH + VV|$ as red, green and blue. In the Pauli basis, $|HH + VV|$, $|HH - VV|$ and $|HV|$ represent single bounce, double bounce and volume scattering respectively. As can be seen in figure the leaf-on image looks greener than the leaf-off season. This can be interpreted as the larger contribution of the volume scattering in leaf-on season. On the other hand, the leaf-off season looks more reddish which is explained as the higher contribution of the double bounce scattering (trunk-ground scattering) due to the penetration of the signals in the absence of the leaves.

5.3 Data preprocessing

Three preprocessing steps were taken before using the SAR data for classification: speckle filtering, SAR parameter extraction and georeferencing.



(a)



(b)

Figure 5.4: leaf-off (a) and leaf-on (b) images in Pauli basis $R:|HH - VV|$, $G:|HV|$ and $B:|HH + VV|$

5.3.1 Speckle filtering

Radarsat-2 transmits pulses to illuminate the earth's surface. When that surface is rough on the scale of the radar wavelength (5 cm), the return signal is a combination of waves reflected from a number of elementary scatterers within the IFOV. If all the individual reflected waves add in a relatively constructive manner, a strong signal is received whereas if they are relatively out of phase, the returned signal is weak. The resulting effect, namely, speckle, complicates image interpretation and analysis as well as decreases the effectiveness of feature selection and image classification. Because of this, speckle reduction in PolSAR data is an important step for the extraction of useful information from polarimetric images.

On the other hand, speckle significantly affects the accuracy of the extracted target decomposition parameters. Several studies have been reported to investigate the speckle effect on target decomposition [123, 124, 85]. Lopez et al. [85] showed that the PolSAR speckle noise has a big impact on quantitative physical parameter estimation, especially in high entropy environments. They showed that the sample eigenvalues are biased, with bias that decreases with an increase in the number of independent samples.

Touzi [124] showed that the processing window size significantly influences the accuracy of the estimates derived from the incoherent decomposition parameters. The results indicated that for unbiased estimation of the incoherent target decomposition parameters, the coherency matrix has to be estimated within a moving window that includes a minimum of 60 independent samples. However, coherent decomposition should be limited to coherent targets with sufficiently high signal to clutter ratios. It was also shown that the averaged parameters derived from coherent decomposition in application that involve extended natural targets might be significantly biased.

As a result, we followed two scenarios in our study. For the incoherent parameters,

the coherency matrix elements were first speckle filtered. The incoherent observables were then extracted from the resulting coherency matrix. Alternatively, for the coherent parameters, the averaging was avoided at the first processing stage and coherent observables were extracted from unfiltered coherency matrix. Thereby, the coherent techniques maintain the full resolution during the feature extraction process, while filtering is applied once the features are obtained. We used the method proposed by [4] to calculate the ENL. Using this method, a processing window size of 13 was found sufficient to maintain the minimum ENL of 60 as proposed by Touzi [124].

The next issue that should be addressed is the type of the filter and filter size. Nowadays, there are various approaches to filter speckle noise in case of PolSAR data. An evaluation of some of the PolSAR speckle filters can be found in [45].

In this study, a simple boxcar filter was used. The boxcar filter has three advantages: first, it is simple to apply, second, it is an effective method for speckle reduction in homogeneous areas and finally it preserves the mean value. Therefore, over homogeneous areas, the boxcar filter presents the best filtering performance. However, in the presence of inhomogeneous areas the performance significantly decreases due to indiscriminately averaging pixels with inhomogeneous objects such as roads and buildings.

As was mentioned in the dataset description, the original ground truth data were pruned of any non-uniformity in terms of holes, roads and streams, residual trees, inclusion of wetlands and any damages. On the other hand, the selected ground truth regions are within large fairly homogeneous forest regions, thus there are no edge effect within the sample areas. Because of these reasons, we used a boxcar filter in this study for speckle suppression. The speckle filter cannot be directly applied onto scattering matrix elements. The scattering matrix $[S]$ has to be converted to an incoherent second order representation, e.g. [T3] matrices. Figure 5.5 shows the a

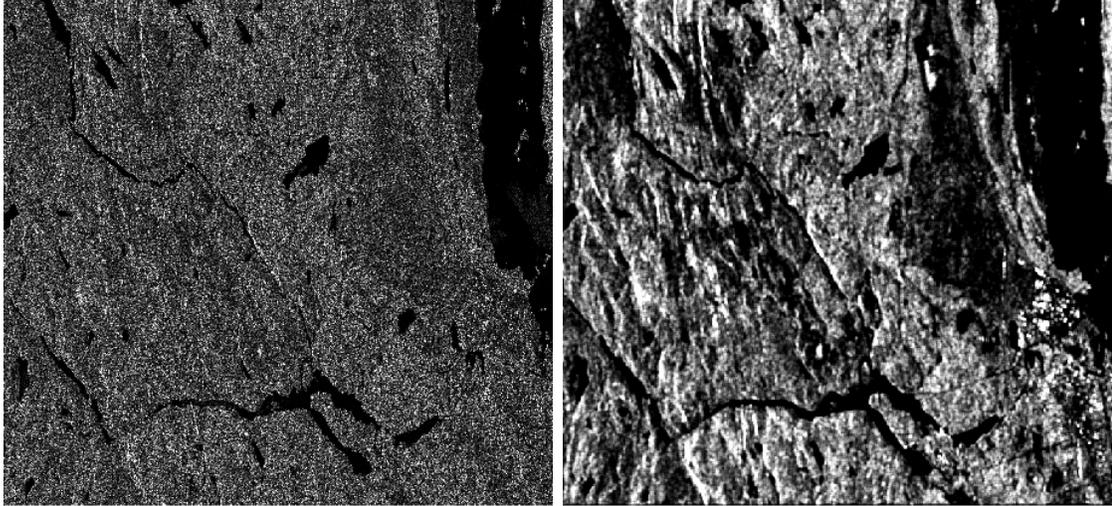


Figure 5.5: The T_{11} image before (left) and after (right) applying speckle filter subset of T_{11} image before and after applying speckle filter.

5.3.2 SAR parameters

The SAR parameters that we used in this research were detailed in chapter 2 and in table 2.2 and are summarized here in table 5.8.

5.3.3 Georeferencing

As the ground truth data are in the ground coordinate system, the SAR image data must be geocoded. The phase information is included in the scattering matrix elements. Therefore, performing the geocoding on the scattering matrix elements changes the phase information in the resampling step. To avoid this, all the SAR parameters are extracted from scattering matrix in the first step. In this way, we have made use of all the information included in the scattering matrix to obtain the PolSAR parameters. Once the PolSAR parameters are obtained, they were georeferenced to the ground coordinate system.

We already had a series of georeferenced Radarsat-1 data used in an earlier project over the same area [88]. The Radarsat-1 image was taken as the reference image and image to image registration was used to register the PolSAR parameters. For

feature	description	symbol	# of pars
original features	scattering matrix elements	[S]	3
	coherency matrix elements	[T]	5
	covariance matrix elements	[C]	3
decomposition features	Krogager [72]	[Krog]	3
	Huynen [62]	[T_huy]	6
	$H/A/\alpha$ [29]	H, A, α	5
	Freeman [47]	[Free]	3
	Touzi [126]	[Touzi]	3
	Barnes [5]	[T_bar]	6
	Holm [59]	[T_holm]	6
SAR discriminators	SPAN	SPAN	1
	received power extrema [40]	[rec_pow]	2
	fractional polarization	F	1
	degree of polarization extrema [127]	[pol_deg]	2
	polarized intensity extrema [127]	[pol]	2
	unpolarized intensity extrema [127]	[unpol]	1
	pedestal height	ped	1
	complex correlation coefficients	$[\rho]$	4
	depolarization ratio	d	1

Table 5.8: PolSAR parameters used in this thesis

this goal, a set of 45 ground control points (GCPs) were collected over the image. A digital elevation model (DEM) of the area was obtained from Geobase and also used to correct for relief displacement. Geobase (www.geobase.ca) is a portal for Canadian geospatial information operated by the Canadian Council on Geomatics. The total root mean squared (RMS) of the registration was 1.1 pixels and the x and y RMS were both less than 1.0 pixel.

Chapter 6

Results and Discussion

In this thesis, the proposed algorithms were compared with two baseline methods. The Wishart classifier proposed by Lee [81] was used as the first baseline method in this thesis. This is one of the most common methods for the classification of PolSAR data. The main idea of the Wishart classifier is that the polarimetric covariance matrix Z may be described by a complex multivariate Wishart distribution. Let $C_m = E(Z|\omega_m)$ be the mean covariance matrix for class ω_m , based on the maximum likelihood classifier and the complex Wishart distribution, the Wishart distance measure was derived as

$$d_m = \ln |C_m| + Tr(C_m^{-1}Z) \quad (6.1)$$

The supervised classification rule becomes as follows: assign the pixel to class ω_i $i \in \{1, 2, \dots, k\}$, if

$$d_m^{(i)} \leq d_m^{(j)} \quad \forall \omega_j \neq \omega_i \quad (6.2)$$

The common framework of all the proposed algorithms is a feature selection scheme. Therefore, in order to investigate the effectiveness of the proposed methods, an SVM with the full set of features was used as the second baseline method. Throughout this thesis this baseline method is called SVM.

The comparison of classification accuracies are undertaken using a statistical method. For this goal, a statistical test is used to determine whether the differences between the classification results are statistically significant or not. Various tests have been proposed to evaluate such significance [17]. In this thesis, a McNemar test [89] was used to assess the statistical significance of differences in the accuracy of classi-

		C2	
		incorrect	correct
C1	incorrect	f_{11}	f_{12}
	correct	f_{21}	f_{22}

Table 6.1: Cross tabulation of number of correct and wrongly classified pixels for two classifications

fications obtained using different methods. The McNemar test summarizes the joint performance of the two classification methods in a 2×2 confusion matrix as per table 6.1. In this table, f_{11} denotes the number of samples which are correctly classified in both C1 and C2 classification algorithms, f_{22} represents the number of samples which are incorrectly classified in both methods, f_{21} is the number of correctly classified samples using C1, which are falsely classified when using C2 and f_{12} denote the number of samples, correctly classified by C2 but wrongly classified when using C1. Then, the McNemar’s test statistic T which is approximately χ^2 distributed with 1 degree of freedom is computed as

$$T = \frac{(f_{12} - f_{21})^2}{(f_{12} + f_{21})} \quad (6.3)$$

The null-hypothesis H_0 is that both classifications lead to equal accuracies. At a given significance level (we used $\alpha = 0.025$), H_0 is rejected if the test statistic T is greater than $\chi^2_{(1,1-\alpha)}$. McNemar’s test calculates a p-value. When the two-sided p-value is less than the 0.025, the conclusion is that there is a significant difference between the two classification algorithms. In other words, with 97.5% confidence we can state that the two classification algorithms are significantly different.

6.1 Nonparametric Methods: Single (NFS) vs. Multiple(CBFS) Classifiers

In this section NFS and CBFS results are presented. In the first experiment NFS method was compared to CBFS method. NFS method uses a nonparametric defini-

tion of evaluation function (equation 4.3) for selecting the best parameters from the available feature set. An SVM classifier is then used for the classification step. CBFS is based on the theory of multiple classifiers. It provides a class-based definition of the evaluation function (equation 4.15) to generate multiple features sets. Afterwards, for each feature set an SVM classifier is trained. Finally, a combination scheme is used to combine the outputs of the classifiers.

The first question that arises here is: what is the optimal number of parameters for classification? To answer this question, given the initial set of 58 features, the NFS method was used for selecting different number of features for leaf-on and leaf-off datasets. The results of classification for different number of features are shown in figure 6.1. As can be seen, increasing the number of features the classification accuracy at first increases and then saturates. For both leaf-on and leaf-off cases the best classification accuracy were obtained when using 12 features.

Figure 6.1 also shows that the leaf-off dataset provided better results than leaf-on image. This can be interpreted as the higher penetration of the SAR signal in leaf-off season. In leaf-on season the canopy surface is mainly responsible for the backscattering whereas in leaf-off season the backscattering is substantially from inside canopy branches and twigs. These results indicate that leaf-off conditions were a better data source for our case of forest mapping. Also, it should be noted that in the leaf-off season the hardwood classes receive double bounce scattering while the softwood classes do not. This difference in backscattering mechanisms increases the discrimination between hardwood and softwood classes and thus improves the hardwood/softwood classification. Further analysis is required to fully elucidate different species behaviour in leaf-off and leaf-on seasons.

A McNemar test was performed to analyze the statistical significance of the differences between three datasets. The p-values are given in table 6.2. As can be seen,

#	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
p	0.091	0	0	0	0	0	0	0	0	0	0	0	0	0	0.003	0.002

Table 6.2: the calculated p-values for the leaf-on vs. leaf-off

#	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
p	0	0	0	0.071	0.004	0	0	0	0	0	0	0	0	0	0	0

Table 6.3: the calculated p-values for the leaf-off vs. leaf-on-off

except for the cases of the 3, 17 and 18 features the p-values are on the order of 10^{-5} for the other cases. Using the 3 features the calculated p-value is larger than 0.025. This shows that in this case the difference is not significant. For other cases, the leaf-off dataset was significantly better than the leaf-on dataset. The same comparison was made using the leaf-on dataset and the combined leaf-on-off dataset. Table 6.3 shows that except for the case of 6 features, the combined leaf-on-off dataset significantly provided better results than leaf-off dataset.

The results are compared with baseline classifiers. Here, the Wishart classifier employs 9 elements of the covariance matrix. The classification accuracies for the leaf-off and leaf-on seasons are summarized in figure 6.2. A look at figure 6.2 reveals that the NFS method gives much better results than those obtained with Wishart and SVM classifiers. Compared to the Wishart classifier, this improvement is 9% and 10% for the leaf-off and leaf-on datasets respectively. Also, in comparison with the SVM method, an improvement of 6% and 4% were obtained for the leaf-off and leaf-on datasets respectively.

In the next experiment, the leaf-on and leaf-off features are combined. The rationale for this combination is that every dataset has some features which may be useful for forest mapping. The features related to leaf-on season give information relating to the characteristics of leaves and twigs whereas there are some leaf-off features concentrate mainly on the main branches, trunk and even ground information. This difference in backscattering mechanisms between leaf-on and leaf-off seasons especially

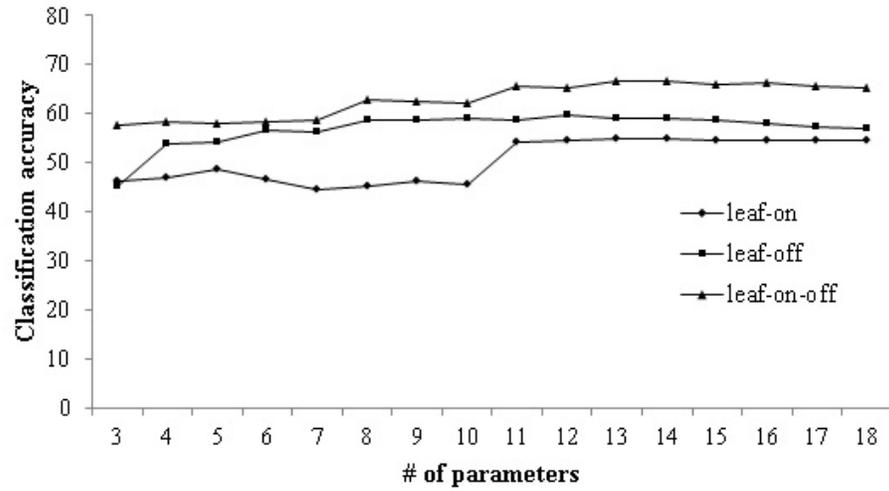


Figure 6.1: classification accuracy vs. number of features in three datasets

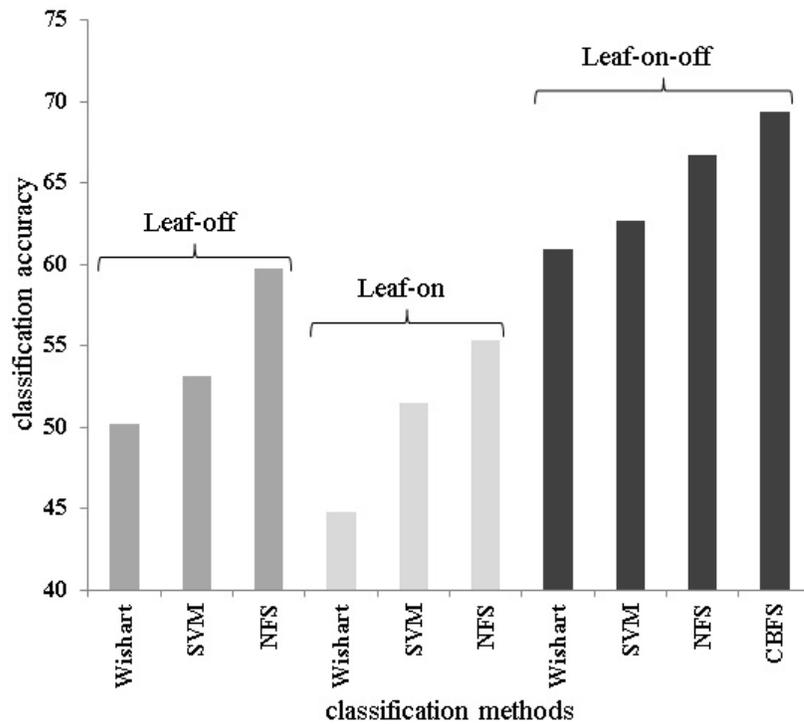


Figure 6.2: Quantitative comparison of nonparametric methods with baseline methods using the three datasets

for the hardwood classes may add to the effectiveness in distinguishing different forest species. For this reason, the combined dataset may have complementary information for forest mapping.

Similarly, in order to find the optimal number of features for classification, the NFS method was performed on this new dataset. The results are shown in figure 6.1. As expected, increasing the number of selected features, the classification accuracy first increases and then a saturating behavior is obtained. In this dataset the best classification accuracy was obtained when using 14 features.

Next, the performance of the the proposed NFS and CBFS methods were compared to the baseline methods on this combined dataset. It should be noted that the Wishart classifier, here, uses 18 elements, i.e. the 2×9 elements of the covariance matrix of the leaf-on and leaf-off datasets. Also, it should be noted that as the concept of multiple classifiers works well for the large feature sets [86], the CBFS method was applied only on the combined dataset.

Figure 6.2 shows that combining leaf-off features with leaf-on features can significantly improve the classification accuracy. Using this combined dataset increased the classification accuracy by 16% for Wishart classifier, 11% for the NFS classifier and 11% for the SVM method than the leaf-on dataset. The NFS method again outperformed the Wishart and SVM classifiers by 6% and 4% using the combined dataset.

The highest accuracy was obtained using the CBFS method. It improved the classification accuracy by 9%, 7% and 3% comparing to Wishart, SVM and NFS methods respectively. One of the benefits of the class-based method comparing to the NFS is that all the classes have their own features participating in the final classification. When there is only one set of features for all the classes, the classes with small number of training samples may be sacrificed, whereas in the class-based schema

name	Wishart	SVM	NFS	CBFS
Or	43.2	60.5	65.9	62.1
Po	51.5	56.2	58.4	58.4
Hx	45.5	30.4	32.2	42.1
Pr	52.2	37.7	40.2	46.4
Pj	34.5	37.7	25.9	31.5
Pw	62.8	78.1	82.8	84.7
Sb	53.3	57.6	64.0	70.9
Sw	57.1	40.0	29.6	33.6
GV	90.7	81.7	91.0	93.6
overall	60.9	62.62	66.7	69.2

Table 6.4: classification accuracies in different classes using the combined leaf-on-off dataset

even the classes with small number of samples can survive by having their own specific features. Table 6.10 shows the Wishart, NFS, SVM and CBFS classification accuracies in different classes. Three classes with small number of training samples are Hx, Pj and Sw (table 5.2). For these classes, CBFS has improved the classification accuracy by 10%, 6% and 4% comparing to the single classifier NFS method respectively (table 6.10). However, a closer look at table 6.4 reveals that for these classes the Wishart classifiers outperformed the NFS and the CBFS methods. This is likely related to the fact that the proposed nonparametric methods are more sensitive to the sample size than the Wishart classifier. Although, the CBFS method has significantly improved the classification accuracy, the Wishart classifier has better performance for these small classes. Table 6.4 also shows an improvement in classification accuracy in most of the classes compared to SVM method.

Next, we applied the McNemar test to two different cases: 1) CBFS and NFS and 2) NFS and Wishart classification results. The results of the test showed that with a 97.5% confidence interval, in both cases the null hypothesis was rejected and thus the differences were statistically significant. The obtained p-values were both very small and less than 0.025.

In the next experiment the leaf-on image was compared to leaf-off image in terms

	selected parameters		
leaf-off	$SPAN$ $[Krog]_{ks}$ $[T_bar]_{11}$ $[pol]_{max}$	$[T_huy]_{11}$ $[Krog]_{kd}$ $[T_bar]_{23}$ $[pol]_{min}$	$[T_huy]_{12}$ $[Free]_{Pd}$ $[T_holm]_{22}$ $[unpol]_{max}$
leaf-on	$[S]_{vv}$ H $[Free]_{Pv}$ $[T_holm]_{22}$	$[T]_{22}$ $[Krog]_{kd}$ $[T_bar]_{11}$ F	$[T_huy]_{22}$ $[Free]_{Pd}$ $[T_bar]_{22}$ ped
leaf-on-off	$[T_huy]_{22}^1$ $[Free]_{Pv}^1$ $[T]_{22}^2$ $[Free]_{Pd}^2$ $[T_holm]_{22}^2$	$[Krog]_{kd}^1$ $[T_holm]_{22}^1$ $[T_huy]_{22}^2$ $[T_bar]_{11}^2$ $[pol]_{max}^2$	$[Free]_{Pd}^1$ $SPAN^2$ $[Krog]_{kd}^2$ $[T_bar]_{23}^2$

Table 6.5: selected features (the superscript 1 and 2 in the third row refer to leaf-on and leaf-off features respectively)

of selected parameters. The 12 selected parameters for leaf-off and leaf-on datasets as well as the 14 selected parameters for the combined dataset are shown in table 6.5. For a detailed description of the symbols used in this table refer to table 5.8. It should be noted that the subscripts used in this table represent the related element in that category. For instance, $[T_huy]_{11}$ represents the parameter in the first row and first column of the Huynen matrix, or $[pol]_{max}$ shows the maximum of the polarized scattered intensity, or $[Free]_{Pv}$ refers to the volume scattering element in the Freeman-Durden decomposition, and so on.

Given hereunder is a description of some of the selected features. Pedestal height (ped), being a measure of the amount of unpolarized backscattered energy, is selected as a discriminative parameter for the leaf-on case. This is mainly because components within the canopy such as leaves and twigs play a significant role in the amount of volume scattering, suggesting a higher pedestal height. Therefore, canopies with different amount of volume scattering represent different amount of pedestal height. On the other hand, the leaf-off condition results in a smaller amount of volume scattering and hence a significantly lower pedestal height.

The Freeman-Durden volume scattering element ($[Free]_{Pv}$) was also selected for the leaf-on case. The volume scattering coming from the leaf canopy at the tops of the trees is mainly responsible for this parameter. This volume scattering varies among trees with different shape of the leaves and different canopy structure. This makes the the freeman canopy scattering parameter a useful feature for the leaf-on case. As expected, for the case of leaf-off the Freeman-Durden double-bounce scattering parameter ($[Free]_{Pd}$) is selected. This can be interpreted as the higher interaction of trunk-ground scattering mechanism.

In the leaf-on season there is an increase in entropy (H) due to the higher volume scattering and due to canopy propagation effects. Thus, different trees with different types of leaves and canopy structure result in different amount of entropy. For this, the entropy (H) is selected as an informative parameter for the leaf-on season.

However, it should be noted that the above discussion is only valid for the hardwood classes as they drop their leaves in the leaf-off season. For the softwood classes, on the other hand, the green foliage is persistent throughout the year, and thus the signal penetration does not vary from season to season.

Table 6.5 also shows that there are some PolSAR parameters that play role in both leaf-on and leaf-on cases. For example, the Krogager decomposition parameters were selected in both cases. This shows that the coherent decomposition, which operates on the individual pixels on a coherent basis, can also provide useful information for the forest mapping. Other instances of this type of parameters are $[T_bar]_{11}$ and $[T_holm]_{22}$.

The third row in table 6.5 shows the selected features for the combined dataset. The superscript 1 and 2 in the third row refer to leaf-on and leaf-off features respectively. In this case, 5 features were selected from the leaf-on image and 9 features were from leaf-off image. This shows the higher significance of the leaf-off image as

expected.

6.2 Wrapper Methods: Single (FWFS) vs. Multiple (FWCBFS) Classifiers

One of the main disadvantages of the nonparametric methods is that the evaluation function it uses for feature selection is not directly related to the final goal of classification. Wrappers on the other hand optimizes the real problem by utilizing the classification accuracy as the evaluation function of the feature selection. However, they are time consuming. In order to overcome this, the notion of the support vectors was used to reduce the training samples and a fast wrapper feature selection (FWFS) was proposed. The method was also proposed in a class-based schema (FWCBFS). For a detailed explanation, the reader is referred to section 4.2. Regarding the random nature of the proposed method for the training sample reduction, the reduction procedure was performed once and the reduced training samples were used for all the experiments in this section. In this way, the differences in the obtained classification results can be thoroughly related to the performance of the methods rather than the employed training samples.

Similar to the nonparametric methods, the leaf-on, leaf-off and the the combined leaf-on-off images were used in our experiments. In the first experiment, the FWFS method was used over the three datasets using different number of features. The results can be seen in figure 6.3.

As can be seen in figure 6.3, the leaf-off image provided better results than the leaf-on image dataset except for the cases of 7, 8, 9 and 14 features for which the two images provided almost the same classification accuracy. The highest improvement was for the case of using 13 features. In this case a 4 percent improvement was reported.

Next, comparisons are made between the combined leaf-on-off dataset and the

#	3	4	5	6	7	8	9	10
p	0.019	0.001	0.001	0.179	0.046	0.214	0.329	0.019
#	11	12	13	14	15	16	17	18
p	0.001	0.005	0	0.311	0	0	0	0

Table 6.6: the calculated p-values for the leaf-on vs. leaf-off

#	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 6.7: the calculated p-values for the leaf-off vs. leaf-on-off

individual datasets. As is shown in figure 6.3, using the combined leaf-on-off dataset, the classification accuracy was significantly improved. The highest improvement reported was 15% which was for the case of using 14 features. This confirms the result obtained previously indicating that the PolSAR parameters obtained from leaf-on image and leaf-off image are not redundant at all but rather complementary for our goal of forest mapping. These results are in agreement with the results obtained from nonparametric methods.

A look at the p-values obtained from McNemar test at table 6.6 shows that in the cases of 6, 7, 8, 9 and 14 features the differences between leaf-off and leaf-on datasets are not significant. For these cases, the p-values are larger than 0.025. For all other cases the leaf-off dataset significantly provided better results. The same experiment was carried out for comparing the combined leaf-on-off dataset with leaf-off dataset. The obtained p-values in table 6.7 shows that for all cases the differences are significant.

Next, the performance of the FWFS and FWCBFS were compared to the baseline methods i.e. Wishart and SVM classifiers. Again, the comparison were made over the three datasets. The results are illustrated in figure 6.4. As can be seen, using the FWFS method, the classification accuracy has improved compared to the Wishart and SVM classifiers. Compared to Wishart classifier the amount of improvement was 10%, 14% and 12% for leaf-off, leaf-on and leaf-on-off datasets and compared to SVM

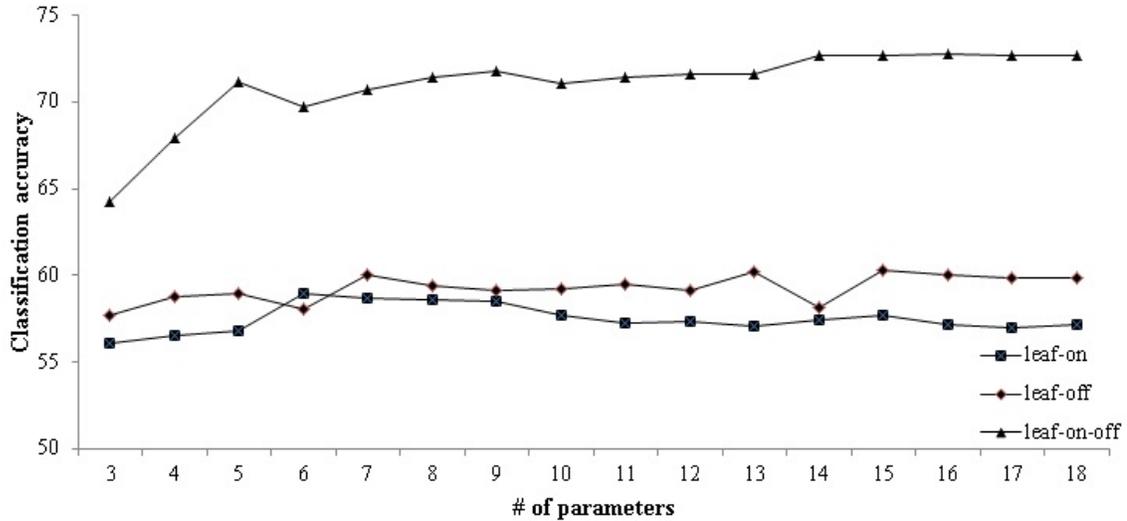


Figure 6.3: classification accuracy vs. number of features in the leaf-on, leaf-off and combined leaf-on-off datasets using the FWFS method

classifier the amount of improvement was 7%, 7% and 10% for the leaf-off, leaf-on and leaf-on-off datasets respectively.

As expected and shown in figure 6.5 and 6.4, the FWCBFS method provided the best results. In figure 6.5, the performance of the FWCBFS method is compared with the FWFS method. The combined leaf-on-off dataset was used in this case. According to figure 6.5, the FWCBFS method provided better results compared to FWFS method. The McNemar test was performed for comparing these two methods and the results showed that except for the case of 5 features for which a p-value of 0.044 was obtained, for all other feature subsets the p-values were in the order of 10^{-5} and hence the differences were significant. The best classification accuracy for the FWCBFS was obtained when using 9 features for each class. In this case, it improved the Wishart, SVM and FWFS results by 14%, 12% and 2.5% respectively.

In the next experiment, we listed the selected features. As is shown in figure 6.3, the best accuracy for the leaf-off, leaf-on and leaf-on-ff images were obtained by using 7 features, 7 features and 14 features respectively. These features are shown in table 6.8.

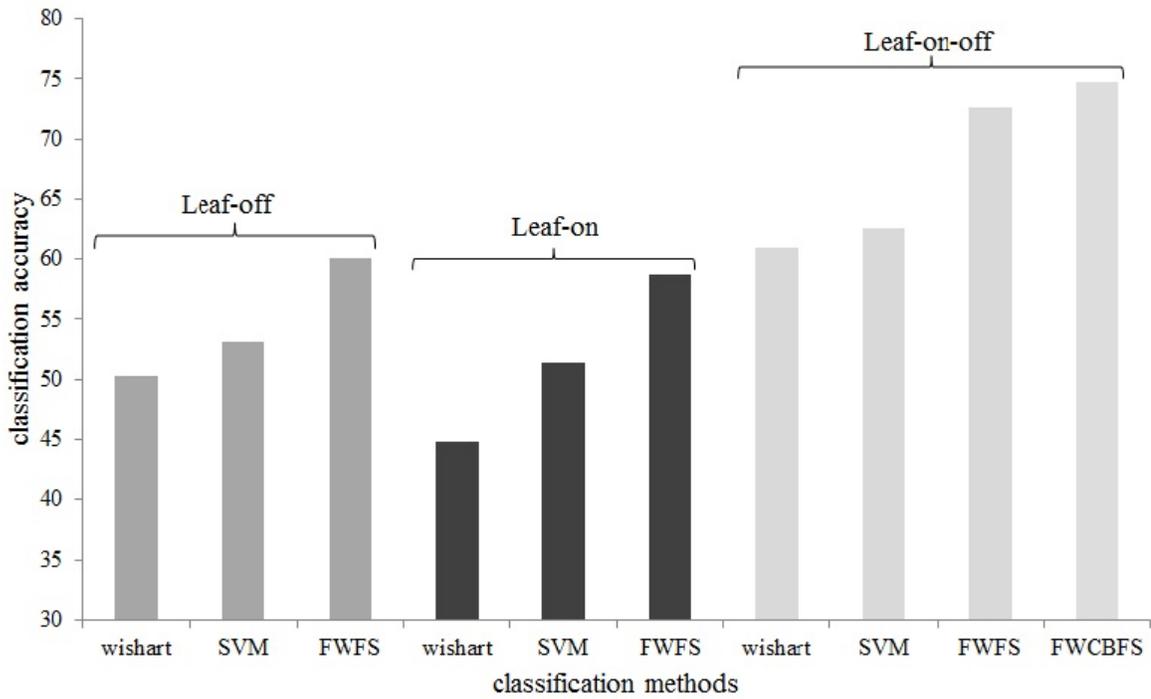


Figure 6.4: Quantitative comparison of wrapper methods with baseline classifiers using the three datasets

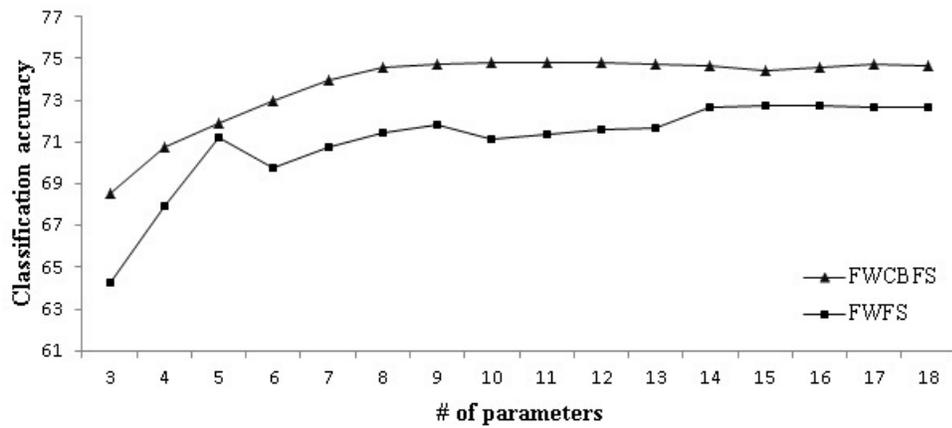


Figure 6.5: FWCBFS vs. FWFS using the combined leaf-on-off dataset

	selected parameters		
leaf-off	$[T_huy]_{11}$ $[T_holm]_{12}$ d	$[Krog]_{kd}$ $[pol_deg]_{max}$	$[T_bar]_{23}$ $[rec_pow]_{max}$
leaf-on	$[T_huy]_{11}$ $[Free]_{Pv}$ F	H $[T_bar]_{33}$	$[Krog]_{ks}$ $[T_holm]_{11}$
leaf-on-off	$[T_huy]_{12}^1$ $[T_holm]_{12}^1$ $[C]_{12}^2$ $[Touzi]_{\psi_s}^2$ F^2	$[H(1-A)]^1$ $[pol_deg]_{max}^1$ H $[pol_deg]_{max}^2$ d^2	$[Krog]_{kd}^1$ d^1 $[T_bar]_{23}^2$ $[pol_deg]_{min}^2$

Table 6.8: selected features (the superscript 1 and 2 in the third row refer to leaf-on and leaf-off features respectively)

It was shown earlier that the best classification results for FWCBFS are obtained using 9 features for each class. These features are listed in table 6.9. Again, the superscript 1 and 2 refer to leaf-on and leaf-off features respectively. The first column shows whether the species is a hardwood (h) or softwood (s) class. In order to better understand the importance of each selected feature for softwood and hardwood classes, they are divided into softwood and hardwood features. Figure 6.6 and 6.7 show the selected softwood and the hardwood features for the leaf-on and leaf-off seasons respectively. They show how frequent the features are selected for the hardwood and the softwood classes. The more the frequency of a feature, the more the importance of that feature. Some remarks on the selected features are given below:

- On the whole, 32 features were selected from leaf-on dataset and 40 features were selected from leaf-off image. This confirms the higher usefulness of the leaf-off image for forest mapping.
- The most significant features for the softwood during the leaf-on season were $[T_huy]_{12}$, $[T_huy]_{22}$ and $[Free]_d$. Each of these features were selected twice for the softwood species during leaf-on season (figure 6.6).

- $[T_bar]_{11}$ was the most significant feature for the hardwood during leaf-on season. It was selected twice for the hardwood species during leaf-on season (figure 6.6).
- $[S]_{hh}$, H , $[Free]_d$, $[T_bar]_{13}$, $[T_bar]_{22}$ and $[T_bar]_{11}$ were the most significant PolSAR parameters for the softwood during the leaf-off season. The first five were selected twice and the last one was selected three times for different hardwood species (figure 6.7). There is not a significant feature during this season for the hardwood species, as all the selected parameters are considered only once.
- $[T_huu]_{22}$ is the only feature which was found to be useful for both softwood and hardwood in both leaf-on and leaf-off seasons.

Thus far, the overall accuracy was used to compare different methods. In order to see the classification accuracies in different classes, producer accuracy was used. Table 6.10 shows the producer accuracies for the three different methods using the leaf-on-off dataset. Notice that the Wishart classifier, here, uses 18 elements, i.e. the 2×9 elements of the covariance matrix of the leaf-on and leaf-off datasets, SVM uses the full set of features, FWFS uses 14 selected features and FWCBFS uses 9 features for each class. A look at table 6.10 reveals that except class Sw, the classification accuracies were improved for the rest of the classes. Although, the FWCBFS method improved the classification accuracy for class Sw compared to FWFS, it is still lower than the Wishart classifier for this class.

6.3 Context-based Method: CCBFS

The method was described in section 4.3. In order to incorporate the neighboring information, the method extends the simple combination rule i.e. equation 4.19 to equation 4.23 which is repeated below

type	class	selected parameters		
hardwood	Or	$[T_huy]_{13}^1$ $[T_bar]_{11}^1$ $[ro]_{ccc}^1$	$[T_huy]_{22}^1$ $[T_holm]_{22}^1$ $[C]_{23}^2$	$[Free]_d^1$ $[ro]_{12}^1$ $[T_huy]_{22}^2$
	Po	$[S]_{hv}^1$ $[T_bar]_{11}^1$ $[pol]_{max}^2$	$[T_huy]_{12}^1$ $[S]_{hh}^2$ $[unpol]_{max}^2$	$[krog]_{Ks}^1$ $[Touzi]_{lan}^2$ F^2
	Hx	$[S]_{hh}^1$ d^1 $[krog]_{Kh}^2$	α^1 $[T_huy]_{23}^2$ $[T_bar]_{12}^2$	$[krog]_{Kh}^1$ $H(1 - A)^2$ $[T_holm]_{12}^2$
softwood	Pr	$[Free]_d^1$ $(1 - H)A^2$ $[T_bar]_{12}^2$	$[S]_{hh}^2$ $[Free]_d^2$ $[T_bar]_{22}^2$	$[C]_{23}^2$ $[T_bar]_{11}^2$ $[T_holm]_{11}^2$
	Pj	$[C]_{13}^1$ H^2 $[Touzi]_{\psi_s}^2$	$[T_huy]_{13}^1$ α^2 $[ro]_{ccc}^2$	$[Touzi]_{lan}^1$ $[T_bar]_{13}^2$ d^2
	Pw	$[S]_{vv}^1$ $[T_bar]_{23}^1$ $[T_huy]_{22}^2$	$[T_huy]_{12}^1$ $[T_holm]_{12}^1$ $[T_bar]_{22}^2$	$[krog]_{Kd}^1$ ped^1 $[pol]_{max}^2$
	Sb	$[T]_{11}^1$ $[Free]_d^1$ $[T_bar]_{13}^2$	$[T_huy]_{11}^1$ $[T_huy]_{12}^2$ $[T_holm]_{22}^2$	$[T_huy]_{22}^1$ $[T_bar]_{11}^2$ $[ro]_{12}^2$
	Sw	$[T_huy]_{12}^1$ $[S]_{hh}^2$ $[Free]_d^2$	$[T_huy]_{22}^1$ $[T]_{22}^2$ $[T_bar]_{11}^2$	$[Free]_s^1$ H^2 $[T_holm]_{13}^2$
vegetation	GV	$[T]_{23}^1$ $[T]_{23}^2$ $[T_bar]_{23}^2$	$[T_bar]_{22}^1$ $[C]_{23}^2$ $[T_holm]_{23}^2$	$[T]_{22}^2$ $[T_bar]_{12}^2$ $[ro]_{13}^2$

Table 6.9: selected features for each class (the superscript 1 and 2 in the third row refer to leaf-on and leaf-off features respectively)

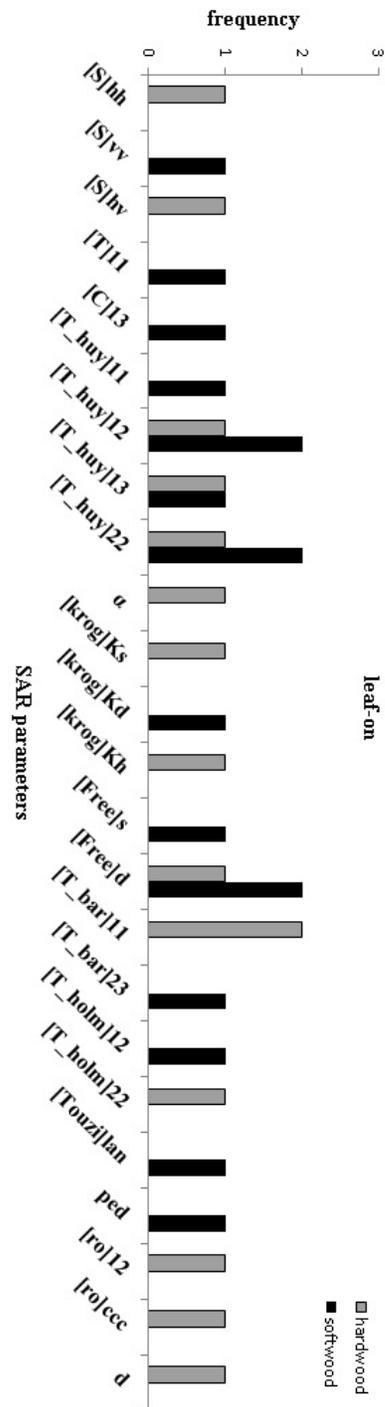


Figure 6.6: the leaf-on features selected by FWCBFS and their frequency

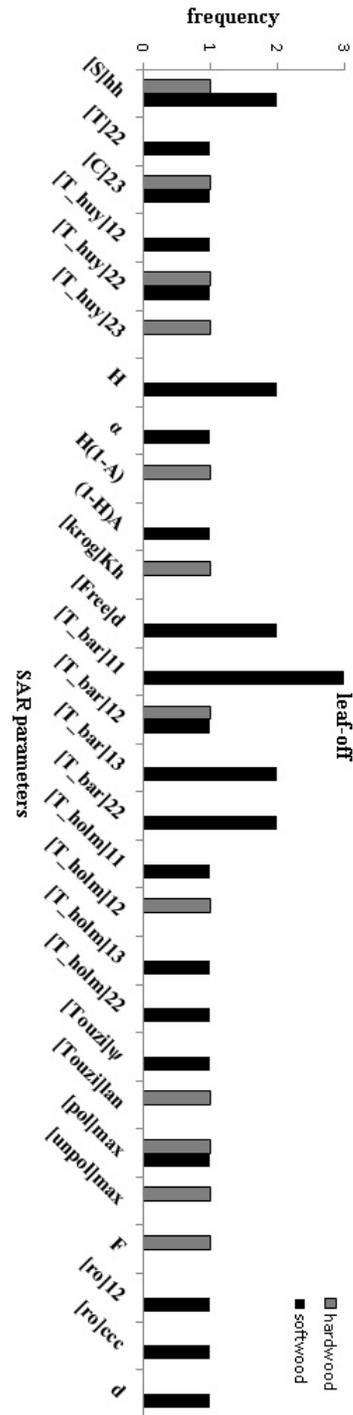


Figure 6.7: the leaf-off features selected by FWCBFS and their frequency

name	Wishart	SVM	FWFS	FWCBFS
Or	43.2	60.5	64.4	65.7
Po	51.5	56.2	65.2	70.9
Hx	45.5	30.4	59.7	48.5
Pr	52.2	37.7	58.2	57.2
Pj	34.5	37.7	56.9	57.8
Pw	62.8	78.1	84.9	86.6
Sb	53.3	57.6	68.1	73.1
Sw	57.1	40.0	32.3	41.9
GV	90.7	81.7	94.5	96.1
overall	60.9	62.62	72.6	74.8

Table 6.10: classification accuracies in different classes using the combined leaf-on-off dataset

$$\sum_{j=1}^N \sum_{l=1}^n w(l) p(x_{lj} | \omega_i) = \max_{k=1}^M \sum_{j=1}^N \sum_{l=1}^n w(l) p(x_{lj} | \omega_k)$$

All the experiments in this section were carried out over the best set of features obtained previously. As was shown in section 6.2, using the FWCBFS method, the best classification results were obtained when using 9 features for each class. The same subsets of features were used in the CCBFS method.

The size of the weighting function $w(l)$ is a critical value in the CCBFS method. It should be large enough to incorporate a fair contribution of the adjacent pixels but small enough to minimize the transitional effects of the boundaries between adjacent classes. Therefore, window size should be selected based on the size and adjacency of existing classes in the study area. So, the question that raises here is how the window size affect the forest classification accuracy and what is the preferred window size? To answer this question, a number of window sizes ranging from 3×3 to 41×41 are used. The results are illustrated in figure 6.9. The figure shows the classification accuracy as a function of the window size. As can be seen, increasing the window size, the classification accuracy increases to a certain point and then it deteriorates. This can be interpreted as follows: in the beginning, the incorporated pixels are more

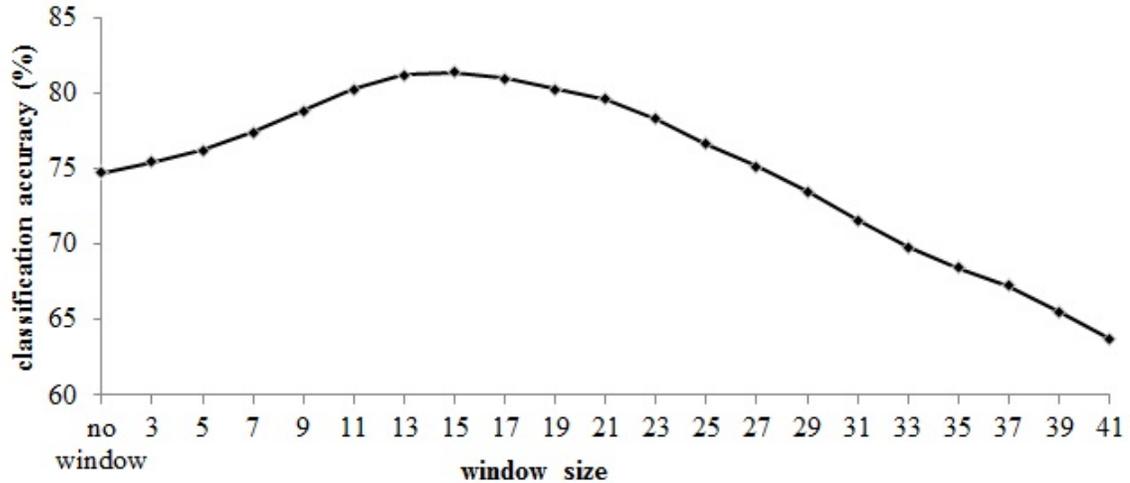


Figure 6.8: increasing the window size in CCBFS method

likely from the same class type. However, when the size of the weighting window gets larger, it is highly likely that the included pixels are from the classes of different type. The deterioration is mainly due to the incorporation of non-related areas in the windowing region. Analysis indicated 13 as the best window size.

In order to see how the weighting window affects different classes, the above experiment was repeated on different classes. The results are shown in figure 6.9. As can be seen, the same behaviour as above was observed in most of the classes. However, the decline in classification accuracies has taken place in different points. The classes with very small number of samples such as Pj, Hx and Sw tend to show lower accuracy when window size increases whereas the larger classes such as GV, Pw, Sb, Or and Po have kept the high classification accuracies even in larger window sizes. This difference can be explained by the size and characteristic of the different classes. If the windows are too large, however, misclassification occurs at the borders of the adjacent classes.

In the next experiment, in order to see how the spatial information can improve the classification accuracy, the CCBFS method is compared with the FWCBFS algorithm. These two methods are the same except that the latter does not incorporate the

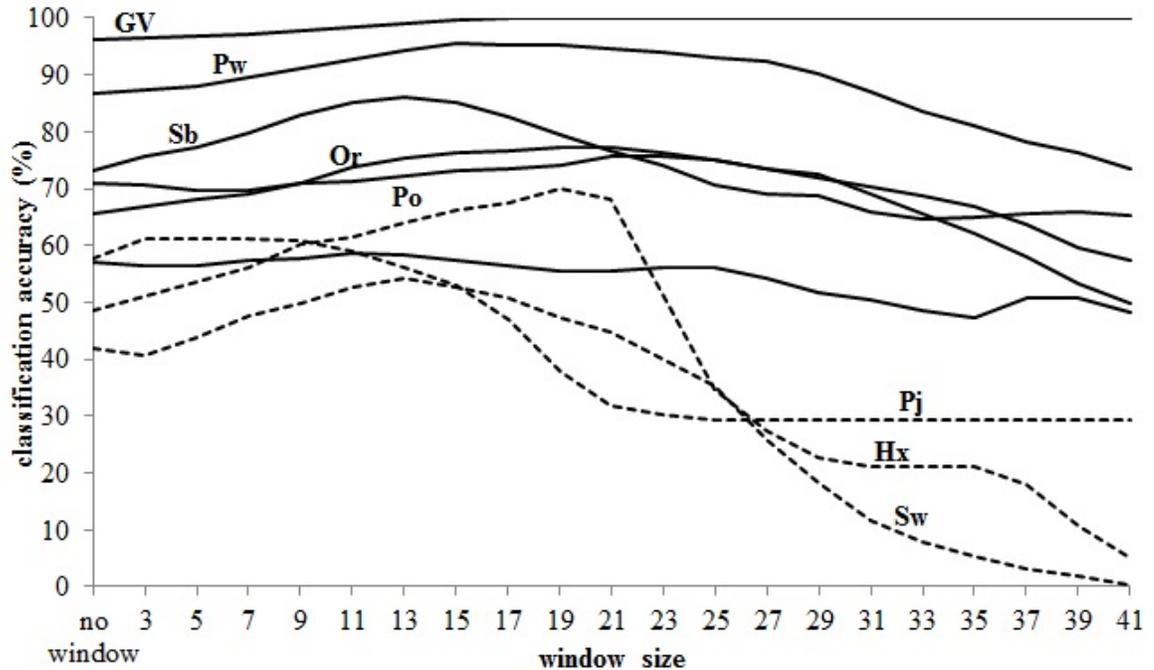


Figure 6.9: the window size effect over different classes: solid lines represent the classes with fairly large number of samples and dotted lines indicate the classes with small number of samples

spatial information. Table 6.11 shows the classification accuracies of the methods. An improvement is observed in almost all the classes. The CCBFS method outperformed the FWCBFS method by 7%. Compared to Wishart classifier and SVM method, the improvement is 21% and 19%. This is the highest improvement obtained in this research. Finally, a McNemar test was performed to compare the FWCBFS with CCBFS and the obtained p-value was close to zero. This indicates that the the CCBFS method is significantly better than FWCBFS method.

6.4 Summary

In this chapter two categories of methods were proposed for the forest classification of polarimetric SAR data: nonparametric methods (i.e. NFS and CBFS), wrapper methods (FWFS and FWCBFS). Also, an extension of the FWCBFS method which incorporates the spatial context (i.e. CCBFS) was proposed. The common framework

name	Wishart	SVM	FWCBFS	CCBFS
Or	43.2	60.5	65.7	75.5
Po	51.5	56.2	70.9	72.3
Hx	45.5	30.4	48.5	63.9
Pr	52.2	37.7	57.2	58.4
Pj	34.5	37.7	57.8	56.0
Pw	62.8	78.1	86.6	94.4
Sb	53.3	57.6	73.1	86.1
Sw	57.1	40.0	41.9	54.1
GV	90.7	81.7	96.1	99.1
overall	60.9	62.62	74.8	81.2

Table 6.11: classification accuracies in different classes using the combined leaf-on-off dataset

of all the proposed algorithms was feature selection and classifier ensemble.

In this section, a summary of the classification results in different classes obtained from nonparametric methods, wrapper methods and context-based method is given. Figure 6.10 illustrated this summary. In this figure, Wishart classification results in different classes served as the baseline and the differences in classification accuracies with respect to the baseline are plotted.

As can be seen, an improvement is achieved in most of the classes. The nonparametric methods i.e. the NFS and the CBFS performed well on the Or, Po, Pw, Sb, and GV classes, whereas a decline is observed for the Hx, Pr, Pj and Sw classes. The Wishart classifier outperforms the NFS by 13%, 12%, 8% and 27% for the classes of Hx, Pr, Pj and Sw respectively. Although, the CBFS method improved the results, the Wishart classifier outperformed the CBFS 3%, 5%, 3%, and 23% for the classes of Hx, Pr, Pj and Sw respectively. This shows the better performance of the parametric method (Wishart) for the classes of small training sample size namely Hx, Pr and Pj compared to the nonparametric methods (NFS and CBFS).

Figure 6.10 illustrates that the wrapper methods i.e. the FWFS and the FWCBFS have significantly improved the classification accuracies compared to Wishart classifier for most of the classes. They, particularly, performed quite well on the classes with

small sample size. The FWFS improved the classification accuracies by 14%, 6%, and 22% for the classes of Hx, Pr, and Pj respectively. These improvements were 3%, 5%, and 24% for the FWCBFS method. Sw was the only class for which the Wishart classifier outperformed all the proposed methods. Further investigations are required to elucidate the cause of this behaviour.

Results obtained from CCBFS method demonstrated markedly improvement compared to all other methods. This is more evident for the classes of large sample size such as Or, Pw and Sb. For these classes an improvement of 32%, 32% and 33% were observed compared to Wishart classifier. Even for the class Sw for which the Wishart classifier significantly outperformed the other methods, the CCBFS method presented almost the same performance as the Wishart classifier.

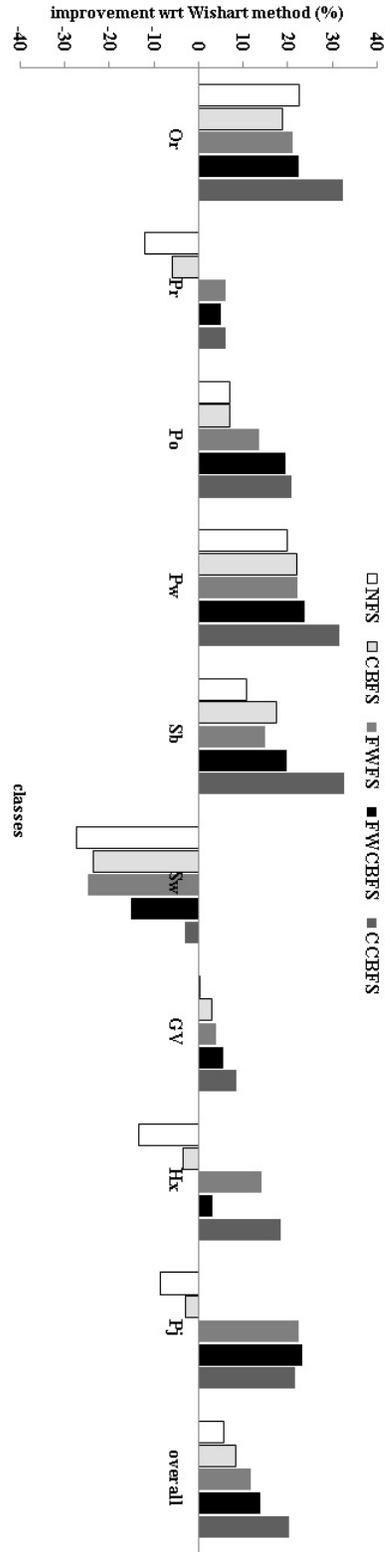


Figure 6.10: summary of the results: the classification accuracies in nonparametric, wrapper and context-based methods in different classes. Wishart classification results in different classes served as the baseline and the differences in classification accuracies with respect to the baseline are plotted. 119

Chapter 7

Conclusions and Further Works

7.1 Conclusions

Two series of algorithms based on the feature selection and classifier ensemble have been devised, developed and implemented to classify polarimetric SAR images for forest mapping. nonparametric feature selection (NFS) method used a nonparametric definition of evaluation function for selecting the best parameters from the available feature set. class-based feature selection (CBFS) method provided a class-based definition of the evaluation function to generate multiple features sets. These two methods, though simple and fast, the evaluation function in feature selection is not directly related to the ultimate goal of classification. To overcome this, fast wrapper feature selection (FWFS) and fast wrapper class-based feature selection (FWCBFS) were proposed. They significantly improved the classification accuracy compared to the nonparametric methods. Also, in comparison to the Wishart classifier, which is a commonly used method for the classification of PolSAR data, a significant improvement was achieved. Finally, an extension of the FWCBFS method which incorporated the information from neighboring pixels was proposed. Results obtained from this contextual class-based feature selection (CCBFS) method demonstrated the highest improvement compared to all other methods. This proved that the neighboring information provides valuable information for the classification of PolSAR data.

The three datasets were then compared in terms of the classification accuracy and the selected parameters. In general, the leaf-off dataset provided better results than leaf-on image. This may be related to higher penetration of the SAR signal in the leaf-off season especially for the hardwoods and greater contrast between hardwood and softwood. However, the best results were obtained when using the combined dataset.

This is mainly because the combined dataset may have complementary information for forest mapping. The backscattering mechanisms are different in each season especially for the hardwood classes and this add to the effectiveness in distinguishing different forest species using the combined dataset.

7.2 Further Works

The proposed algorithms generally provided good performance. However, there is still room for improvement. Several different directions for future research are listed below

- In this thesis, the combination of leaf-on with leaf-off image significantly improved the classification accuracy compared to the individual datasets. However, the employed images had the same incidence angle. It would be useful to examine the combination of images of multiple incidence angles. Each incidence angle presents its own backscattering scattering mechanisms. Combining the SAR parameters obtained from these images can provide complementary information for forest mapping.
- All the proposed algorithms in this research, adopts the classification in a species by species manner. Perhaps an extension of this is to employ a decision tree to initially categorize the species into softwood and hardwood and then subdivide each category into its constituting species.
- While this dissertation has focused primarily on the use of PolSAR data for the forest mapping, the framework is general and may be applied to other applications which involves discriminating different classes.
- In this research, for the extraction of the coherent parameters, the averaging was avoided at the first processing stage and coherent observables were extracted

from unfiltered coherency matrix. It would be desirable to investigate the effect of applying filter before and after the extraction of coherent parameters for forest mapping.

- A more in depth look into the behaviour of different species in the leaf-on and leaf-off seasons should be investigated.

Bibliography

- [1] A. Agrawal and W. Boerner. Redevelopment of kennaught's target characteristics polarization state theory using the polarization ratio formalism for the coherent case. *IEEE Transaction on Geoscience and Remote Sensing*, 27:2–14, 1989.
- [2] A. Al-Ani and M. Deriche. A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence. *Journal of Artificial Intelligence Research*, 17:333–361, 2002.
- [3] V. Alberga. *Comparison of polarimetric methods in image classification and SAR interferometry applications*. PhD thesis, Fakultat fir Elektrotechnik und Informationstechnik der Technischen Universitat Chemnitz genehmigte, 2003.
- [4] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft. Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery. *IEEE Transaction on Geoscience and Remote Sensing*, 47(11):3795–3809, November 2009.
- [5] R. M. Barnes. Roll invariant decompositions for the polarization covariance matrix. Internal report, Lincoln Laboratory, MIT, Lex., MA 02173-0073, 1988.
- [6] J. A. Benediktsson, J. Chanussot, and M. Fauvel. Multiple classifier systems in remote sensing: From basics to recent developments. In Kittler J. Roli Haindl, M., editor, *MCS 2007*, volume 4472, pages 501–512. Springer, Heidelberg, 2007.
- [7] U. Benz. Supervised fuzzy analysis of single- and multi-channel SAR data. *IEEE Transaction on Geoscience and Remote Sensing*, 37:1023–1037, 1999.
- [8] H. Bischof, W. Schneider, and A. J. Pinz. Multispectral classification of landsat-images using neural networks. *IEEE Transaction on Geoscience and Remote*

Sensing, 30(3):482–490, 1992.

- [9] W. Boerner. *Polarimetry in radar remote sensing. Basic and applied concepts, in Manual of Remote Sensing: Principles and Applications of Imaging Radar, 3 ed.*, volume 8. Wiley, 1998.
- [10] M. Borgeaud. Theoretical models for polarimetric microwave remote sensing of earth terrain. *Journal of Electromagnetic Waves and Applications*, 3:6181, 1989.
- [11] D. Borghys and C. Perneel. A supervised classification of multi-channel high-resolution SAR data. In *Earsel*, volume 6, pages 26–37, 2007.
- [12] D. Borghys, C. Perneel, M. Keller, A. Pizurica, and W. Philips. Supervised feature-based classification of multi-channel SAR images using logistic regression. In *EUSAR 04, Ulm*, May 2004.
- [13] D. Borghys, Y. Yvinec, C. Perneel, A. Pizurica, and W. Philips. Hierarchical supervised classification of multi-channel SAR images. In *3rd Internat. Workshop on Pattern Recognition in Remote Sensing, Kingston-Upon-Thames*, August 2004.
- [14] D. Borghys, Y. Yvinec, C. Perneel, A. Pizurica, and W. Philips. Supervised feature-based classification of multi-channel SAR images. *Pattern Recognition Letters*, 27(4):252–258, March 2006.
- [15] M. Born and E. Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 7 edition, October 1999.
- [16] B. A. M. Bouman and D. Hoekman. Multi-temporal, multi-frequency radar measurements of agricultural crops during the agriscatt-88 campaign in the

- netherlands. *International Journal of Remote Sensing*, 14:1595–1694, 1993.
- [17] J. V. Bradley. *Distribution-Free Statistical Tests*. Prentice-Hall, Englewood Cliffs, New Jersey, 1968.
- [18] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
- [19] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [20] G. J. Briem, J. A. Benediktsson, and J. R. Sveinsson. Multiple classifiers applied to multisource remote sensing data. *IEEE Transaction on Geoscience and Remote Sensing*, 10:2291–2299, 2002.
- [21] L. Bruzzone, D. F. Prieto, and S. B. Serpico. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. *IEEE Transaction on Geoscience and Remote Sensing*, 37:1350–1359, 1999.
- [22] C. J. Burges. A tutorial on support vector machines for pattern recognition. In *Data Mining Knowledge Discovery*, volume 2, pages 121–167, 1998.
- [23] Chih C. Chang and Chih J. Lin. *LIBSVM: a library for support vector machines*, 2001.
- [24] C. T. Chen, K. S. Chen, and J. S. Lee. The use of fully polarimetric information for the fuzzy neural classification of SAR images. *IEEE Transaction on Geoscience and Remote Sensing*, 41(9):2089–2100, 2003.
- [25] J. Chen, Y. Chen, , and J. Yang. A novel supervised classification scheme based on adaboost for polarimetric SAR signal processing. In *9th International Conference on Signal Processing (ICSP 2008) Beijing, China*, pages 2400–2403, October 2008.

- [26] K. S. Chen, W. P. Huang, and D. H. Tsay. Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network. *IEEE Transaction on Geoscience and Remote Sensing*, 34(3):814–820, 1996.
- [27] D. A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28(1):45–62, 2002.
- [28] S. Cloude. *Polarimetry: The Characterization of Polarimetric Effects in EM Scattering*. PhD thesis, University of Birmingham, UK, 1999.
- [29] S. Cloude and E. Pottier. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transaction on Geoscience and Remote Sensing*, 35:68–78, 1997.
- [30] S. R. Cloude. Group theory and polarisation algebra. *Optik*, 75:26–36, 1986.
- [31] S. R. Cloude. Uniqueness of target decomposition theorems in radar polarimetry. In W.M. Boerner et al, editor, *Direct and Inverse Methods in Radar Polarimetry, Part I, NATO-ARW*. Kluwer Academic Publishers, 1992.
- [32] S. R. Cloude, K. P. Papathanassiou, A. Riegber, and W. M. Boerner. Multifrequency polarimetric SAR interferometry for vegetation structure extraction. In *International Geoscience and Remote Sensing Symposium, IGARSS*, volume 1, pages 129–131, July 2000.
- [33] S. R. Cloude and E. Pottier. A review of target decomposition theorems in radar polarimetry. *IEEE Transaction on Geoscience and Remote Sensing*, 34:498–518, 1996.
- [34] R. Collobert and S. Bengio. Svmtorch: Support vector machines for large-scale regression problems. *Journal of Machine Learning Research*, 1:143–160, 2001.

- [35] M. M. Crawford, S. Kumar, M.R. Ricard, J. C. Gibeaut, and A. Neuenschwander. Fusion of airborne polarimetric and interferometric SAR for classification of coastal environments. *IEEE Transaction on Geoscience and Remote Sensing*, 3(1):1306–1315, 1999.
- [36] T. G. Dietterich and R. Bakiri. Solving multiclass learning problems using error correcting output codes. *journal of Artificial Intelligence Research*, 2(1):263–286, 1995.
- [37] L. Du and J S. Lee. classification of earth terrain covers using complex polarimetric SAR data. *International Journal of Remote Sensing*, 17(4):809–826, 1996.
- [38] L. J. Du and J. S. Lee. Polarimetric SAR image classification based on target decomposition theorem and complex Wishart distribution. In *International Geoscience and Remote Sensing Symposium, IGARSS, Lincoln*, volume 1, pages 439–441, May 1996.
- [39] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2001.
- [40] D.L. Evans, T.G. Farr, J.J. van Zyl, and H.A. Zebker. Radar polarimetry: analysis tools and applications. *IEEE Transaction on Geoscience and Remote Sensing*, 26:774 –789, 1988.
- [41] L. Ferro-Famil, E. Pottier, and J. S. Lee. Unsupervised classification of multi-frequency and fully polarimetric SAR images based on the H/A/Alpha- Wishart classifier. *IEEE Transaction on Geoscience and Remote Sensing*, 39:2332–2342, November 2001.
- [42] Y. Freund and R.E. Schapire. Experiments with a new boosting algorithm. In

13th International Conference on Machine Learning, Morgan Kaufman, pages 148–156, 1996.

- [43] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin. Unsupervised classification of radar images using hidden markov chains and hidden random fields. *IEEE Transaction on Geoscience and Remote Sensing*, 41:675–686, 2003.
- [44] G. M. Foody and A. Mathur. A relative evaluation of multi class image classification by support vector machines. *IEEE Transaction on Geoscience and Remote Sensing*, 42(6):1335–1343, 2004.
- [45] S. Foucher and C. Lopez-Martinez. An evaluation of PolSAR speckle filters. In *International Geoscience and Remote Sensing Symposium, IGARSS*, pages 845–848, 2009.
- [46] A. Freeman. Fitting a two-component scattering model to polarimetric SAR data from forests. *IEEE Transaction on Geoscience and Remote Sensing*, 45(8):2583–2592, 2007.
- [47] A. Freeman and S. Durden. A three component scattering model to describe polarimetric sar data. In *SPIE, Radar Polarimetry*, volume 1748, pages 213–224, 1992.
- [48] S. Fukuda and H. Hirosawa. Support vector machine classification of land cover: application to polarimetric SAR data. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 1, pages 187–189, Sydney, Australia, July 2001.
- [49] S. Fukuda, R. Kataqiri, and H. S. Hirosawa. Unsupervised approach for polarimetric SAR image classification using support vector machine. In *International*

Geoscience and Remote Sensing Symposium, IGARSS, Toronto, Canada, number 5, pages 2599–2601, 2002.

- [50] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, New York, 2 edition, 1990.
- [51] P. Gong and P.J. Howarth. Performance analyses of probabilistic relaxation methods for land-cover classification. *Remote Sensing of Environment*, 30:33–42, 1989.
- [52] A. Guissard. Radar polarimetry with applications to satellite remote sensing. *TELE, U.C.L.*, 1:51–64, 2003.
- [53] R. Harris. Contextual classification post-processing of landsat data using a probabilistic relaxation model. *International Journal of Remote Sensing*, 6:847–866, 1985.
- [54] M. Hellmann. A new approach for interpretation of SAR-data using polarimetric techniques. In *International Geoscience and Remote Sensing Symposium, IGARSS*, volume 4, pages 2195–2197, July 1998.
- [55] M. Hellmann. *Classification of fully polarimetric SAR-data for cartographic applications*. PhD thesis, Fakultät Elektrotechnik der Technischen Universität Dresden, Dresden, June 2000.
- [56] T. K. Ho. The random subspace method for constructing decision forests. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.
- [57] D. Hoekman. A new polarimetric classification approach evaluated on agricultural crops. In *ESA POLINSAR Workshop*, Italy, 2003.

- [58] D. Hoekman and M. Quinones. Biophysical forest type characterization in the colombian amazon by airborne polarimetric SAR. *IEEE Transaction on Geoscience and Remote Sensing*, 40(6):1288–1300, Jun 2002.
- [59] W.A. Holm and R.M. Barnes. On radar polarization mixed target state decomposition techniques. In *Proceedings of the 1988 National Radar Conference*, pages 248–254, April 1988.
- [60] C.W. Hsu and C.-J. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 15(2):415–425, March 2002.
- [61] C. Huang, L.S. Davis, and J. R. G. Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23:725–749, 2002.
- [62] J. Huynen. *Phenomenological Theory of Radar Targets*. PhD thesis, University, Delft, The Netherlands, 1970.
- [63] Y. Ito and S. Omatu. Polarimetric SAR data classification using competitive neural networks. *International Journal of Remote Sensing*, 19(14):2665–2684, 1998.
- [64] A. Jain and D. Zongker. Feature selection: Evaluation, application and small sample performance. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19:153–158, 1997.
- [65] R. Jones. A new calculus for the treatment of optical systems, description and discussions. *Journal of Optical Society of America*, 31:488–493, 1941.
- [66] T. Kavzoglu and P. M. Mather. The role of feature selection in artificial neural network applications. *International Journal of Remote Sensing*, 23(15):2919–2937, 2002.

- [67] J. Kittler. *Feature Selection and Extraction. Handbook of Pattern Recognition and Image Processing*. Academic Press, New York, 1986.
- [68] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 20:226–239, 1998.
- [69] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence.*, 97(1-2):273–324, 1997.
- [70] Y. Kouskoulas, F. Ulaby, and M. C. Dobson. Classification of short vegetation using multifrequency SAR. In *Proc. IEEE*, volume 1, pages 103–105, 1998.
- [71] E. Krogager. *Aspects of Polarimetric Radar Imaging*. PhD thesis, Technical University of Denmark, Copenhagen, Denmark, 1993.
- [72] E. Krogager, J. Dall, and S. Madsen. Properties of sphere, diplane and helix decomposition. In *Proceedings of the 3rd International Workshop on Radar Polarimetry*, 1995.
- [73] S. Kumar, J. Ghosh, and M. Crawford. Best-bases feature extraction algorithms for classification of hyperspectral data. *IEEE Transaction on Geoscience and Remote Sensing*, 39:1368–1379, 2001.
- [74] B.-C. Kuo and D. A. Landgrebe. A robust classification procedure based on mixture classifiers and nonparametric weighted feature extraction. *IEEE Transaction on Geoscience and Remote Sensing*, 40(11):2486–2494, November 2002.
- [75] D. A. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley and Sons, New Jersey, 2003.

- [76] C. Lardeux, P-L. Frison, C. Tison, D. Deleffie, J-C. Souyris, J-P. Rudant, and B. Stoll. Comparison of compact polarimetric with full polarimetric radar data for land use discrimination based on svm classification. In *POLinSAR*, Frascati, Italy, January 2007.
- [77] D. G. Leckie, D. Hill, A.S. Bhogal, W. Burt, and E. Malta. Radarsat imagery of forest terrain and the role of environmental factors, detection of partial disturbances and complementarity with landsat data. Final report to canadian space agency, Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada Victoria, B.C., April 2003.
- [78] J. Lee, M. Grunes, T. Ainsworth, L. Du, D. Schuler, and S. Cloude. Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Transaction on Geoscience and Remote Sensing*, 37:2249–2258, 1999.
- [79] J. S. Lee. SAR polarimetry: basics, processing techniques and applications. In *International Geoscience and Remote Sensing Symposium, IGARSS, Tutorial*, 2006.
- [80] J. S. Lee, M. Grunes, E. Pottier, and L. Ferro-Famil. Unsupervised terrain classification preserving polarimetric scattering characteristics. *IEEE Transaction on Geoscience and Remote Sensing*, 42:722–731, 2004.
- [81] J. S. Lee, M. R. Grunes, and E. Pottier. Quantitative comparison of classification capability: fully polarimetric versus dual and single-polarisation SAR. *IEEE Transaction on Geoscience and Remote Sensing*, 39(11):2343–2351, 2001.
- [82] J S Lee, Grunes M R, and Kwok R. Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution. *International Journal of Remote Sensing*, 15(11):2299–2311, 1994.

- [83] Jong-Sen Lee and Eric Pottier. *Polarimetric Radar Imaging From Basic To Applications*. CRC Press, 2009.
- [84] Y.J. Lee and O.L. Mangasarian. Rsvm: reduced support vector machine. In *Proceedings of the 1st SIAM International Conference on Data Mining*, Chicago, 2001.
- [85] C. Lopez-Martinez, E. Pottier, and S. Cloude. Statistical assessment of eigenvector based target decomposition theorems in radar polarimetry. *IEEE Transaction on Geoscience and Remote Sensing*, 43(9):2058–2074, September 2005.
- [86] Y. Maghsoudi, M. J. Valadan Zoej, and M. J. Collins. Using a class-based feature selection for the classification of hyperspectral data. *International Journal of Remote Sensing*, 32:4311–4326, 2011. In Press.
- [87] Yasser Maghsoudi, Michael Collins, and Don Leckie. On the use of feature selection for classifying multi-temporal radarsat-1 images for the forest mapping. *IEEE Geoscience and Remote sensing letters*, 8:904–908, 2011.
- [88] Yasser Maghsoudi, Michael Collins, and Don Leckie. Speckle reduction for the forest mapping analysis of multi-temporal radarsat-1 images. *International Journal of Remote Sensing*, 33(5):1349 – 1359, November 2011.
- [89] Quinn McNemar. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157, 1947.
- [90] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transaction on Geoscience and Remote Sensing*, 42:1778–1790, 2004.
- [91] G. Mercier and F. Girard-Ardhuin. Unsupervised oil slick detection by SAR imagery using kernel expansion. In *International Geoscience and Remote Sensing*

Symposium, IGARSS, volume 1, pages 494–497, July 2005.

- [92] M. Hellmann, G. Jaeger, E. Kraetzschmar, and M. Habermeyer. Classification of full polarimetric SAR-data using artificial neural networks and fuzzy algorithms. In *International Geoscience and Remote Sensing Symposium, IGARSS*, volume 4, pages 1995–1997,, Hamburg, Germany, July 1999.
- [93] R. Min, X. Yang, and Z. Zhao. Application of adaboost in polarimetric SAR image classification. In *Radar Conference, IEEE*, 2009.
- [94] C.J. Oliver and S. Quegan. *Understanding Synthetic Aperture Radar Images*. Artech House, 1998.
- [95] M. Pal and P. M. Mather. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4):554–565, 2003.
- [96] K. Papathanassiou. *Polarimetric SAR interferometry*. PhD thesis, DLR, Oberpfafenhofen, Germany, 1999.
- [97] J. Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In A. J. Smola, P. L. Bartlett, B. Scholkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 61–74, Cambridge, 2000. MIT Press.
- [98] E. Pottier. Dr. j. r. huynen’s main contributions in the development of polarimetric radar techniques and how the ‘radar targets phenomenological concept’ becomes a theory. In *SPIE*, volume 1748, pages 72–85, 1992.
- [99] E. Pottier. The h/a/alpha polarimetric decomposition approach applied to Pol-SAR data processing. In *in Proceedings of the PIERS-Workshop on Advances in Radar Methods*, 1998.

- [100] E. Pottier, L. Ferro-Famil, and J. S. Lee J.S. . PolSarpro v3.0 -lecture notes: Advanced concepts. Technical report, <http://earth.esa.int/polsarpro/>, 2008.
- [101] E. Pottier and J. Saillard. On radar polarization target decomposition theorems with application to target classification by using network method. In *International Conference on Antennas and Propagation (ICAP 91)*, pages 265–268, York, UK, April 1991.
- [102] C. Proisy, E. Mougin, E. Dufrene, and V. Le Dantec. Monitoring seasonal changes of a mixed temperature forest using ers SAR observations. *IEEE Transaction on Geoscience and Remote Sensing*, 38(1):540–552, 2000.
- [103] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.
- [104] S. Quegan, T. Le Toan, J. J. Yu, F. Ribbes, and N. Floury. Multitemporal ERS SAR analysis applied to forest mapping. *IEEE Transaction on Geoscience and Remote Sensing*, 38:741–753, 2000.
- [105] A. Reigber, M. Jager, M. Neumann, and L. Ferro-Famil. Polarimetric fuzzy k-means classification with consideration of spatial context. In *POLINSAR-Applications of SAR Polarimetry and Polarimetric Interferometry*, 2007.
- [106] J. A. Richards and X. Jia. *Remote Sensing Digital Image Analysis*. Berlin, Germany: Springer-Verlag, 4 edition, 2006.
- [107] E. Rignot and R. Chellappa. Segmentation of polarimetric synthetic aperture radar data. *IEEE Transaction on Image Processing*, 1(3):281–300, July 1992.
- [108] A. Rodriguez, D. G. Corr, E. Pottier, L. Ferro-Famil, and D. Hoekman. Land cover classification using polarimetric SAR data. In *Proc. PolInSAR Conference, ESA-ESRIN, Frascati*, January 2003.

- [109] A. Rosenfeld, R.A. Hummel, and S.W. Zucker. Scene labeling by relaxation operations. *IEEE Transaction on System, Man, and Cybernetics*, 6:420–433, 1976.
- [110] J.S. Rowe. *Forest Regions of Canada*. Canadian Forestry Service, Ottawa, Ontario, 1972.
- [111] S. B. Serpico and L. Bruzzone. A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Transaction on Geoscience and Remote Sensing*, 39(7):1360–1367, 2000. Special Issue on Analysis of Hyperspectral Image Data.
- [112] X. L. She, J. Yang, and W. J. Zhang. The boosting algorithm with application to polarimetric SAR image classification. In *The 1st Asian and Pacific Conference on Synthetic Aperture Radar (APSAR 2007) Huangshan, China*, pages 779–783, November 2007.
- [113] D. Sheffer and Y. Ultchin. Comparison of band selection results using different class separation measures in various day and night conditions. In *SPIE Conf. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, volume 5093, pages 452–461, 2003.
- [114] 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*, 11(3):169–180, 2009.
- [115] M. Simard, S. S. Saatchi, and G. De Grandi. The use of decision tree and multiscale texture for classification of jers-1 SAR data over tropical forest. *IEEE Transaction on Geoscience and Remote Sensing*, 38(5):2310–2321, 2000.
- [116] G. Sinclair. Modifications of the radar range equation for arbitrary target and

- arbitrary polarizations. Technical Report 302-19, Antenna Laboratory, The Ohio State University Research Foundation, September 1948.
- [117] B. Solaiman, L. Pierce, and F. T. Ulaby. Multisensor data fusion using fuzzy concepts: Application to land-cover classification using ERS-1/JERS-1 SAR composites. *IEEE Transaction on Geoscience and Remote Sensing*, 37:1316–1326, 1999.
- [118] A.H. Schistad Solberg, T. Taxt, and A.K. Jain. A markov random field model for classification of multisource satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 34(1):100–113, January 1996.
- [119] K. Soren. *Bismuth iron garnet films for magneto-optical photonic Crystals*. PhD thesis, Laboratory of Solid State Devices, IMIT, Stockholm, 2004.
- [120] B.M. Steele. Combining multiple classifiers: An application using spatial and remotely sensed information for land cover type mapping. *Remote Sensing of Environment*, 74:545–556, 2000.
- [121] C. Tison, J. Nicolas, F. Tupin, and H. Maire. A new statistical model of urban areas in high resolution SAR images for markovian segmentation. *IEEE Transaction on Geoscience and Remote Sensing*, 42:2046–2057, 2004.
- [122] S. Tong and D. KollerHuang. Support vector machine active learning with applications to text classification. In *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pages 999–1006, 2000.
- [123] R. Touzi. Target scattering decomposition of one-look and multi-look SAR data using a new coherent scattering model: The tsvm. In *International Geoscience and Remote Sensing Symposium, IGARSS*, pages 2491–2494, Anchorage, AK, 2004.

- [124] R. Touzi. Speckle effect on polarimetric target scattering decomposition of SAR imagery. *Canadian Journal of Remote Sensing*, 33(1):60–68, 2007.
- [125] R. Touzi. Target scattering decomposition in terms of roll-invariant target parameters. *IEEE Transaction on Geoscience and Remote Sensing*, 45:73–84, 2007.
- [126] R. Touzi, A. Deschamps, and A.M. Demers. Phase of target scattering for wetland characterization using polarimetric c-band SAR. *IEEE Transaction on Geoscience and Remote Sensing*, 47(9):3241–3261, 2009.
- [127] R. Touzi, S. Goze, T. Le Toan, A. Lopes, and E. Mougin. Polarimetric discriminators for SAR images. *IEEE Transaction on Geoscience and Remote Sensing*, 30(5):973–980, September 1992.
- [128] R. Touzi, R. Landry, and F. J. Charbonneau. Forest type discrimination using calibrated c-band polarimetric SAR data. *Canadian Journal of Remote Sensing*, 30(3):543–551, 2004.
- [129] K. Tragl. Polarimetric radar backscattering from reciprocal random targets. *IEEE Transaction on Geoscience and Remote Sensing*, 28:590–597, 1990.
- [130] Ince Turker, Serkan Kiranyaz, and Moncef Gabbouj. Classification of polarimetric SAR images using evolutionary rbf networks. In *ICPR'2010*, pages 4324–4327, 2010.
- [131] F. Ulaby and C. Elachi. *Radar Polarimetry for Geoscience Applications*. Artech House, 1990.
- [132] V. N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, New York, USA, 1998.

- [133] V. N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, 1999.
- [134] O. Watanabe, J.L. Balczar, and Y. Dai. A random sampling technique for training support vector machines. In *Proceedings of the IEEE International Conference on Data Mining (ICDM 2001)*, pages 43–50, San Jose, CA, 2001.
- [135] U. Wegmuller, A. Wiesmann, T. Strozzi, and C. L. Werner. Forest mapping with multitemporal SAR. In *ForestSAT2002 Conference, Edinburgh, U.K*, 2002.
- [136] T.-F. Wu, C.-J. Lin, and R.-C. Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5:975–1005, 2004.
- [137] S. Y. Yang, M. Wang, and L. C. Jiao. Radar target recognition using contourlet packet transform and neural network approach. *Signal Processing*, 89(4):394–409, 2009.
- [138] Lamei Zhang, Bin Zou, Junping Zhang, and Ye Zhang. Classification of polarimetric SAR image based on support vector machine using multiple-component scattering model and texture features. *EURASIP Journal on Advances in Signal Processing*, 2010:9 pages, 2010.
- [139] Y-D. Zhang, L-N. Wu, and G. Wei. A new classifier for polarimetric SAR images. *Progress In Electromagnetics Research, PIER 94*, 94:83–104, 2009.
- [140] Q. Zhixin, A. G. O. Yeh, L. Xia, and L. Zheng. Integrating object-oriented image analysis and decision tree algorithm for land use and land cover classification using radarsat-2 polarimetric SAR imagery. In *International Geoscience and Remote Sensing Symposium, IGARSS*, pages 3098 –3101, July 2010.

- [141] T. Zou, W. Yang, D. Dai, and H. Sun. Polarimetric SAR image classification using multifeatures combination and extremely randomized clustering forests. *EURASIP Journal on Advances in Signal Processing*, 2010:1–9, 2010.