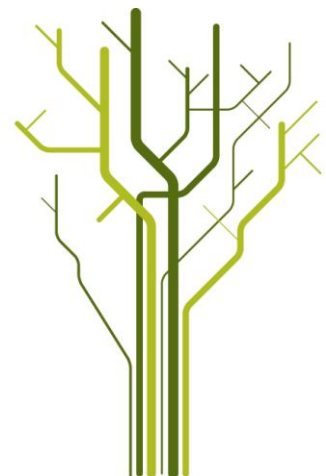# Multi-sensor Data Fusion and Feature Extraction for Forest Applications

Temesgen Gebrie Yitayew

FYS-3900 Master's Thesis in Physics

May 2012

Abstract

This thesis focusses on extracting, evaluating and selecting features to study multi-sensor data fusion for forest applications. Due to the difference in the underlying sensor technology, data from synthetic aperture radar (SAR) and optical sensors refer to different properties of the observed scene and it is believed that when they are fused together, they complement each other to improve the performance of a particular application. Improved fusion results and useful physical interpretation of the observed scene can also be obtained by extracting and selecting sensor specific features. The two primary aims of this study are investigating the benefit of Polarimetric SAR and multi-spectral optical data fusion for forest applications by using different features extracted from each of these datasets, and identifying those few best Polarimetric SAR and multi-spectral optical features that jointly perform best for forest classification. Two secondary aims are comparing the potential of four different datasets for forest applications based on their individual classification performances, and comparing two feature selection algorithms in selecting the first few best features.

Multi-frequency fully Polarimetric SAR data at P-, L- and C-band and multi-spectral Landsat TM data acquired over the Nezer forest in France were used for demonstration. The scene is composed of homogeneous fields of either bare soil or maritime pine trees of different ages, and the application was discriminating the bare soil, and the trees in terms of their ages. On the one hand, different combinations of the polarimetric channels were used to extract simple polarimetric SAR features. On the other hand, different combinations of the available multi-spectral bands were used to extract different vegetation indices. A Supervised maximum likelihood Bayesian classification scheme was applied to evaluate and compare the classification performances of each of the four datasets and their different combinations. The classification accuracy (%) was used for a quantitative comparison. The two standard, sequential forward and sequential backward, feature selection algorithms were applied and compared in selecting the best features.

A literature review of data fusion methods found that feature level fusion is the best approach for our application. A total of twenty-six features; six from each of the three Polarimetric SAR datasets and eight from the optical dataset were extracted. A number of features from the extracted set were found useful to interpret the scene in terms of its physical parameters. In comparing the classification performances of the four datasets, it was found that P-band is the best whereas C-band is the poorest. The L-band and the Landsat TM datasets were found to have moderate performances. Therefore, P-band is potentially the best band for forest applications, and whenever it is available, priority should be given to the use of it. Significant classification accuracy improvement (up to 12%) was achieved by fusing the polarimetric SAR and the multi-spectral optical datasets. Therefore, attention should be given to the combined use of them whenever they are available.

The sequential forward feature selection approach gave slightly better results in selecting the few best features than the backward one. Therefore, whenever the objective is to choose the few best features, the forward approach should be used. Five features were found to jointly preserve 98.5% of the classification information of the available set. This shows the incredible advantage of feature selection in preserving most of the classification information and at the same time reducing the size of the data. Two of the best features, namely the *mean radar backscatter* and the *cross-pol ratio*, were identified from the polarimetric SAR features. The *wetness* and the *soil brightness index* were found to be the two best optical features in complementing the

polarimetric SAR features irrespective of the SAR frequency used. The *normalized difference vegetation index* was found specifically useful in complementing the P-band whereas the *greenness* was best in complementing the L-and C-band features. In addition to retaining most of the valuable information, these few identified features were found useful to interpret the scene in terms of the different forest scattering mechanisms. Therefore, they can be reasonably expected to be used for other forest applications too.

Acknowledgements

I am pleased to acknowledge the people who assisted me for completing this thesis. First and foremost, my deepest gratitude goes to my supervisors Camilla Brekke and Anthony Paul Doulgeris for tirelessly working with me from the beginning up to the end and sharing me their broad knowledge and experience. They have been following up the progress of my thesis with great care and commitment. Their discussions taught me a lot beyond the thesis and their ingenious suggestions are evident in every section of this thesis. They have spent a great deal of their valuable time reading and commenting each and every page. They earn my gratitude more than anybody else.

Thanks to Gökhan Kasapoğlu for his help and valuable suggestions on certain issues in the thesis.

I would like to thank Robert Jenssen for being an excellent teacher and giving me the foundation of data analysis techniques

Thanks to Prof. Laurent Ferro-Famil at Univeristy of Rennes-1 for the Nezer forest dataset.

Special thanks to the Norwegian state educational loan fund-Lånekassen for the scholarship grant during my study at the university of Tromsø

Last but not least, I would like to thank my friends who helped and encouraged me in doing this project, specially Vidar and Thomas for helping me with Matlab related problems, Desalegn and Firehun for their wonderful help on LaTeX .

Contents

List of Figures

List of Tables

Abbreviations

SAR	Synthetic aperture radar
PolSAR	Polarimetric synthetic aperture radar
RAR	Real aperture radar
AIRSAR	Air-borne synthetic aperture radar
TM	Thematic Mapper
MSS	Multi-spectral Scanner
EMS	Electromagnetic spectrum
IR	Infrared
NIR	Near infrared
SWIR	Short wave infrared
IHS	Intensity hue saturation
PCA	Principal component analysis
PCS	Principal component substitution
PAN	Panchromatic
HPF	High pass filter
CPs	Control points
ROI	Region of interest
RST	Rotation-scaling-translation
QAC	Quick atmospheric correction
SLC	Single-look complex
MLC	Multi-look complex
VI	Vegetation index
NDVI	Normalized difference vegetation index
B	Soil brightness index
G	Greenness
W	Wetness
PVI	Perpendicular vegetation index
TVI	Triangular vegetation index
SAVI	Soil adjusted vegetation index
ARVI	Atmospherically resistant vegetation index
SFFS	Sequential forward feature selection
SBFS	Sequential backward feature selection

Chapter 1

Introduction

These days, many remote sensing satellite sensors are acquiring information at different spatial, spectral and temporal resolutions, and hence a wide spectrum of data is available for the same observed site. However, the information provided by the individual sensors might be incomplete or imprecise for a given application [2, 3, 4]. Hence, a combined use of data from two or more sources may provide complementary information which could help to understand the observed scene better or improve the result of a particular application [5, 6]. This is the main motivation for this study.

Data from synthetic aperture radar (SAR) and optical sensors refer to different characteristics of the observed scene, and it is believed that when combined, they offer complementary information that helps to distinguish the different classes of a particular observation. Optical data contains information on the reflective and emissive characteristics of the targets in terms of spectral intensity. This spectral information is related the chemical composition and moisture content of the observed target. On the other hand, SAR data contains information on the geometric structure, surface roughness and dielectric properties of natural and man-made objects. As an example, spectral signature is the information inferred from optical data, which is used to characterize ground targets. However, some vegetation species may not be discriminated as they can have similar spectral responses. Therefore, radar images can help in discriminating these vegetation species as they contain additional information about the geometric structure and dielectric properties of the vegetation cover.

Data fusion refers to combining information from two or more sources together to improve the quality and interpretability of the source data. This can be achieved at any one of the three different processing levels of the image information: pixel, feature or decision levels.

- Pixel level fusion is a low level fusion where different source images are combined to produce a single fused image

- Feature level fusion is an intermediate level of fusion, which requires the merging

of extracted features

- Decision level fusion is a high level fusion which is used to integrate separately processed image information using decision rules

Regardless of the fusion techniques used, the fused image is believed to contain greater information content for a given scene than any one of the individual image sources alone [3].

The process of transforming an input dataset into a set of representative features which accurately and concisely represent the original information is referred to as feature extraction. When dealing with data from multiple sensors, feature extraction is the first process undertaken to preserve sensor specific information. When wisely implemented, the extracted features can be helpful to interpret the observed target in terms of its physical parameters. A related process, i.e., feature selection, refers to selecting a subset of features from the available set. It is an important process to identify those best features that contain most of the valuable information for a particular application. An additional benefit of feature extraction, when it is accompanied by a systematic feature selection process, is that a smaller amount of memory and processing time will be required in the feature space because of the removal of redundant information in the process.

Many studies have been conducted to combine SAR and optical data for a number of applications [7, 8, 9, 10, 11]. To mention some of the applications, two SAR datasets from ERS-1/2 (C-band) and JERS-1 (L-band) are fused with a multi-spectral dataset from the SPOT satellite for the purpose of urban land cover classification [7]. In [8] a study was conducted to integrate images from ERS-1 satellite and Landsat thematic mapper sensors for geological study purposes. Snow cover mapping using these two different datasets was demonstrated in [9]. Another application example where SAR and optical data sets are integrated is land cover mapping, [10, 12].

Even though they are not many, a number of studies have been conducted to integrate SAR and optical data for forest applications. However, most of these studies concentrated on estimating different forest parameters, mainly biomass and some related variables such as canopy height [13, 14, 15, 16]. Very recently, classification of forest in terms of tree species has been studied in [17]. In most of these studies, the raw data from the different sources is directly used for either forest variable estimation or classification.

However, only few studies have been conducted on extracting and combining features from SAR and optical datasets for forest applications [13, 16]. In [16], statistical metrics derived from different features extracted from four data sources (LiDAR, SAR/InSAR, ETM+ and Quickbird) are used to compare the performance of the combined datasets with that of the individual ones in mapping forest biomass and canopy height. In [13], some selected polarimetric channels from multi-frequency polarimetric SAR, coherence amplitude from a single frequency interferometric SAR, and some selected bands from Landsat TM dataset are combined for the purpose of mapping foliage biomass. The selection among the channels and the bands was performed by looking at the corre-

lation information in a scatter plots. Apart from these two related efforts, no work is done to integrate different polarimetric SAR and multi-spectral optical features for forest classification.

This study focuses on extracting, evaluating and selecting different multi-frequency polarimetric SAR (PolSAR) and multi-spectral optical features to demonstrate the benefit of SAR and optical data fusion for forest applications. The primary aim of the feature evaluation and selection process is to identify those features, from the different datasets considered, which are best for forest applications, with a secondary aim of comparing the potential of two feature selection algorithms in selecting a few best features. In addition, the study will look in to the individual classification performances of four different datasets with respect to forest classification. The main motivation of extracting and using different multi-frequency PolSAR and multi-spectral optical features is that sensor specific information can be retained in the fusion process and could be helpful to improve the fusion results by using different features that are unique to the sensors.

A data fusion approach, which takes in to account the sensor specific information in the fusion process and is suitable for the evaluation and selection of features, is chosen by an extensive literature review. The application is illustrated by using datasets acquired over the Nezer forest, France, under the objective of discriminating maritime pine trees of different ages and bare soil. Four datasets are considered; three air-borne fully polarimetric SAR datasets acquired with three different frequencies, C-, L- and P-bands, and one multi-spectral Landsat Thematic Mapper (TM) dataset. As a preprocessing step of the datasets, atmospheric correction is applied to the multi-spectral Landsat TM dataset, then all the bands of the Landsat TM dataset are co-registered with the PolSAR images by manual image tie-point selection and resampling, and the SAR images are multi-looked to reduce speckle and to adapt to the Landsat TM resolution. After processing the original single-look complex polarimetric SAR data, different combinations of the polarimetric channels are used to extract simple polarimetric features [18], and different combinations of the available multi-spectral bands are used to extract different vegetation indices [1]. A supervised maximum likelihood Bayesian classification scheme is applied to evaluate and compare the classification performance of each of the four datasets and their different combinations. Percentage classification accuracy is used for a quantitative comparison. Two standard, sequential forward and backward, feature selection algorithms [1] are applied and compared in selecting the few best features.

In total, twenty-six features; six from each of the three PolSAR datasets and eight from the Landsat TM dataset are extracted and used for this study. This number does not present all possible polarimetric and optical features, however, it includes the most commonly used features and is enough for our intended purposes, i.e., (a) to investigate the benefit of extracting and selecting sensor specific features to identify those most representative ones and reduce the volume of the data, and (b) to investigate the complementariness of SAR and optical features. There are a lot more Polarimetric features, especially those can be derived from the polarimetric target decomposition theorems and other more multi-spectral features. They are not considered here as it is beyond the

3

scope of this study. The different fusion approaches are not quantitatively compared here. This is also beyond the scope of the project and the selection among the fusion approaches is simply made by reviewing the literature. For the classification and future selection part, a commonly used classifier is applied, and the different classifiers that can be used to analyse multivariate datasets will not be investigated here for the same reason.

As it is pointed out above, this study encounters discriminating among the different age groups of single-species pine trees. This task is actually harder than other forest applications that involve mixed-species trees. Therefore, the results of this study could potentially be used for other forest applications.

1.1 Structure of the thesis

The thesis consists of both theoretical concepts and experimental parts. The theoretical concepts are covered in chapters 2 and 3, while the experimental parts are covered in chapters 4, 5, 6 and 7.

The second chapter introduces two key concepts. The first one is the sensor technology of SAR and its optical multi-spectral counterpart, and the second fundamental concept is the interaction of electromagnetic waves with the forest in the microwave and visible/infrared regions of the electromagnetic spectrum.

The third chapter looks into the concept of data fusion. Here, the basic concepts of data fusion, as it is applied to remote sensing purposes, are discussed and the most widely used fusion techniques are reviewed. In addition, by comparing the different data fusion approaches, an appropriate one for our application is selected.

The fourth chapter first introduces the datasets used in this study, and then addresses the preprocessing of the datasets. Atmospheric correction, image to image registration and speckle filtering are the main preprocessing tasks considered.

Feature extraction from the datasets is presented in chapter five. Here, different polarimetric SAR and multi-spectral features are extracted and discussed in terms of the physical parameters of the scene.

In chapter six, a maximum likelihood Bayesian classification scheme is applied on the extracted feature sets. Here, the benefit of multi-source data fusion is demonstrated by combining the features from the different sources. The classification results from the individual and the combined datasets are quantitatively compared.

Feature evaluation and selection is covered in the seventh chapter. Here, the whole feature set from the different datasets are analysed and features, which jointly perform best with respect to classification accuracy, are selected. Two feature selection approaches are used and compared.

Lastly, chapter eight concludes the whole work.

Chapter 2

SAR and optical sensors and electromagnetic interactions with forest

A lot of literatures and textbooks can be referred for a detailed description of polarimetric synthetic aperture radar [19, 20, 21, 22] and optical multi-spectral sensors [1, 23]. In this chapter, some of the relevant concepts about these sensors are addressed to give a background information about the PolSAR and optical multi-spectral datasets used in this project. In addition, the interaction of electromagnetic radiation with the forest is among the topics discussed. A few basic characteristics of the images acquired by these sensors are also described, to give some insight about the issues that should be accounted for in the preprocessing step of the datasets considered. Finally some important comparisons between the datasets acquired by these sensors will be given.

2.1 The potential of remote sensing for forest applications

Due to their extended area coverage and complex nature, information about forests is not easily accessible. Remote sensing is an important tool for forest applications as it is capable of acquiring information over a wide area in a repetitive manner. The ability of satellite remote sensing systems to image extensive areas makes them preferable for applications at the global scale, and air-born remote sensing systems can be used for forest monitoring at a local scale [24]. From their orbiting nature, space-born sensors are capable of repeat imaging every part of the earth's surface at a fixed interval of time. Hence, they can provide a large amount of multi-temporal data for the same observed forest site, that can be used for different purposes. Even though they can offer greater flexibility with respect to the time of data acquisition, air-born systems will cost much more to provide similar frequency of observation over extended areas. An important aspect

of radar systems is their ability to see through clouds and atmospheric moisture. This is particularly useful for forest applications, especially in the humid tropics. Moreover, the high resolution and polarimetric capabilities of recent SAR systems, makes remote sensing an important source of information for different forest applications.

Remote sensing information acquired over forest can be used for different applications such as [25]:

- Forest/non-forest mapping

- Forest type and species mapping

- Forest age discrimination

- Deforestation and forest regeneration mapping

- Forest fire detection

- Carbon and biomass mapping

As it is pointed out in the introduction chapter, this study encounters forest age discrimination. Such an information can be used for different purposes such as timber production and forest inventory [26, 27].

2.2   Synthetic aperture radar (SAR)

2.2.1   Basic concepts

Radar is an active sensor, which works by transmitting and receiving pulses of microwave energy. The sensor transmits a signal, which is directed toward the target area to be investigated. The radiation reflected from that target is detected and measured by the radar receiver. For earth observation purposes, these sensors can be installed on either an air-borne or space-borne platforms and can operate at various frequencies. Real aperture radar (RAR) is a radar system where the resolution in the flight direction is controlled by the physical length of the antenna. For these systems, only the amplitude (and not the phase) of each return echo is measured and processed. To determine the spatial resolution at any point in a radar image, it is necessary to compute the resolution in two dimensions, in the range (across track) and azimuth (along track) directions. The range resolution of a RAR is dependent on the effective length of the pulse in the slant range direction and mathematically it is given by [28]:

$$\triangle R = \frac{C\tau}{2} = \frac{C}{2\beta}$$

where $\triangle$R is the slant range resolution, C is the speed of light, τ is the pulse length and β is the bandwidth. This resolution could be improved by decreasing the pulse length.

But if we decrease the pulse length, the system requires much input power (in order to get a detectable amount of signal) as the signal energy is given by

$$E \equiv P\tau$$

where P is the instantaneous peak power. However, the maximum power is severely limited by the sensor hardware, particularly in the case of space borne sensors. Thus, in order to have high detection ability (large E) and a high resolution (large β), a pulse with the seemingly incompatible characteristics of large τ and large β is needed. This can be achieved by signal processing techniques, modulating and de-modulating the transmitted and received signals respectively [28].

The azimuth resolution $\triangle A$ is given by the beam width and is a function of the operating wavelength, position in range and the dimension of the antenna in that direction. Mathematically, it is given by

$$\triangle A = L = \frac{\lambda R}{D_a}$$

Where L is the beam width, λ is the wave length, D_a is the dimension of the antenna in the azimuth direction and R is the slant range distance to the target. One way of achieving better azimuth resolution is to boost frequency; another is to increase along track antenna length; a third is to decrease the target range. None of these options is very effective from space. This is where the concept of Synthetic aperture radar (SAR) comes into play.

SAR is a coherent mostly airborne or space borne side-looking radar system which utilizes the motion of the platform to simulate an extremely large antenna using advanced signal processing, and that generates high-resolution remote sensing imagery. The geometry of SAR is shown in figure 2.1 below. The platform travels forward in the flight direction with the nadir directly beneath the platform. The radar beam is transmitted obliquely at right angles to the direction of flight illuminating a swath, which is offset from nadir. That is the reason why it is also known as a side looking radar. The radial line of sight distance between the radar and each target on the surface is called the slant range distance.

As the radar moves, a pulse is transmitted at each position (shown by string of dots in the figure below). The amplitude and phase of the signals returned from objects are recorded and stored throughout the time period in which the objects are within the beam of the moving antenna. Advanced signal processing techniques are used to coherently combine the recorded information from each of the returned signals to achieve very high azimuth resolution. The final output could be a single image with the highest azimuth resolution possible (single-look data) or a multi-look processed data composed of several images with reduced azimuthal resolution. Generally, depending on the processing, resolutions achieved by SAR sensors are of the order of 1-3 meters for air-borne SAR and 5-50 meters for space-borne SAR [20, 28]. SAR achieves fine range resolution the same way as RAR does, by pulse modulation. As an example, the SAR data used in this project

has 3 meters spatial resolution in the azimuth and 6.7 meters of spatial resolution in the range.
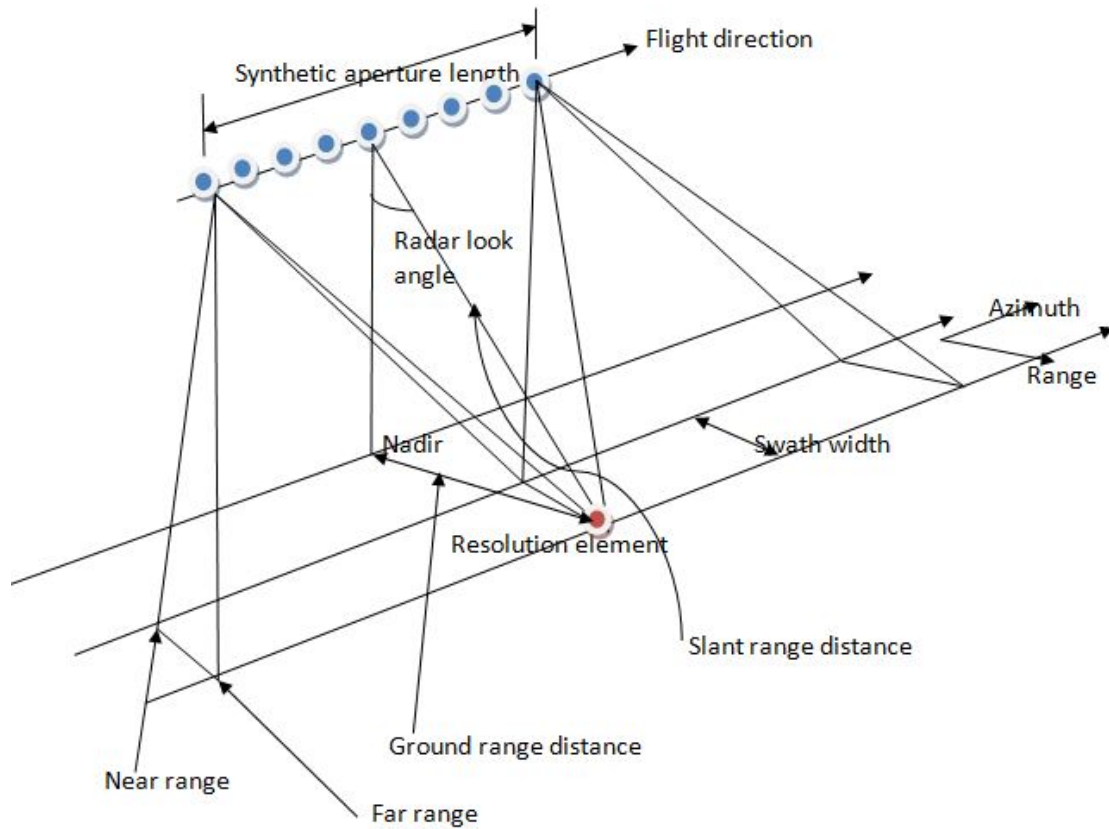
Figure 2.1: Geometry of SAR.

2.2.2   SAR Polarimetry

Polarization describes the orientation of the electric field component of electromagnetic waves in a plane perpendicular to the direction of propagation. In general, radar systems can have one, two, or all four of the following transmit-receive linear polarization combinations; HH, VV, HV, VH. Conventional SAR systems operate within a single, fixed-polarization antenna for both transmission and reception of microwave signals. In this way, a single radar reflectivity is measured, for a specific transmit and receive polarization combination, for every resolution element of the image. As a result, additional information about the scattering process contained in the polarization properties of the scattered signal is lost. To ensure that all the information of the scattered wave is retained, the polarization of the scattered wave must be measured. Polarimetric SAR is a radar system, which transmits microwave signals in two orthogonal directions (H and V) and records the backscattered signal in two or more separate channels. Depending of the

number of receiving channels, polarimetric SAR systems can be semi- or fully polarimetric. Fully polarimetric channels record the received signal in four separate channels. The target response is given by a 2×2 scattering matrix, where the diagonal elements are the co-polarization (HH, VV) terms, while the off diagonal elements are known as cross-polarization (HV, VH) terms. Mathematically the scattering matrix $\mathbf{S}$ is represented by:

$$\mathbf{S} = \left[\begin{array}{cc} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{array} \right]$$

SAR polarimetry has great advantage over conventional single-channel radar systems as it is capable of measuring the complete scattering matrix for each resolution element of a scene [21, 29]. This helps to obtain more information about the scattering mechanisms on surfaces or within volumes.

2.2.3   SAR image characteristics

A SAR image has several characteristics that make it unique. Images obtained from coherent sensors such as SAR systems are characterized by speckle. Speckle is a salt and pepper appearance of radar images which is caused by random constructive and destructive interferences from the multiple scattering returns that will occur within each resolution cell. In other words, speckle is a statistical fluctuation associated with the radar reflectivity (brightness) of each pixel in the image of a scene. It is a form of noise which degrades the quality of radar images and therefore reducing this effect could help for better discrimination of targets. One of the most common speckle suppression techniques is multi-looking. Single-look imaging uses all signal returns from a ground target to create a single image. Multi-looking is the dividing of the radar beam into several narrower sub-beams. Each sub-beam provides an independent look of the illuminated scene. Summing and averaging the images from the different looks will result an image with reduced speckle. In cases where only single-look processed images are available, the averaging can be done on the local neighbourhood of pixels. The later approach is used for this study, as the original datasets were supplied as single-look complex images. However, multi-looking in general is performed at the expense of the spatial resolution of radar images.

However, it is also important to note that speckle is not really noise in the classic sense because it is the radar signature of the target. It has some useful information that can be used to characterize the texture information in radar images. Texture in radar images refers to the natural variation of the average radar backscatter on a scale of several resolution cells. Therefore, its statistical characterization requires measurements from a finite sampling window rather than estimates at the individual pixel level [30].

Another property of radar images is that the position and proportions of objects in SAR images can appear distorted compared to a photograph. This distortion is a unique geometric characteristic of radar images resulted from the difference in sample spacing

between the slant range plane and ground plane. This slant range distortion causes variation in the scale of radar images from near to far range, and targets in the near range appears compressed relative to the far range. There are also other geometric distortions of SAR images and a more complete summary of them can be found in [30].

2.2.4   Interaction of microwaves with forest

When electromagnetic radiation strikes a surface, it may be transmitted, absorbed and re emitted or reflected. Transmission occurs when radiation passes through a target while reflection occurs when radiation "bounces" off the target and is redirected. In the case of absorption, the incident electromagnetic radiation will be converted to other forms of energy after being absorbed by the target. Remote sensing relies on measuring the reflected energy from targets. Depending on the surface roughness in comparison to the wavelength of the incoming radiation, earth surface features reflect either specularly or diffusely, or somewhere in between these two extremes. Usually the definition of smoothness or roughness for surface scattering is given by some criteria and is totally dependent on the wavelength of the electromagnetic radiation used, the incident angle and the surface standard deviation height. Such two criteria are the Rayleigh and the Fraunhofer criteria [22]. Specular reflection occurs when almost all of the incident energy is directed away from the surface in a single dominant direction and it is a property of smooth surfaces. On the other hand diffuse reflection occurs when the incident energy is reflected almost uniformly in all directions. If the wavelengths are much smaller than the surface variations or the particle sizes that make up the surface (rough surface), diffuse reflection will dominate [31].

SAR sensors record the backscatter signal (both the amplitude and the phase) from targets. The proportion of this backscattered signal as compared to the transmitted one is dependent on a number of factors including, surface roughness, slope of the surface, dielectric properties of the target, types of land cover (soil, vegetation, man-made objects), microwave frequency, polarization and incident angle. Let us closely examine each of these factors in the framework of land cover features in general and forest in particular.

In general, the backscatter intensity for rough surfaces is higher than smooth surfaces; hence the SAR image will look brighter. Trees and other vegetation are usually moderately rough on the wavelength scale. Hence, they appear as moderately bright features in SAR images. When a transmitted wave strikes a target and returns directly back to the sensor, it is known as single scattering or single bounce. In a forest media, this could happen if the wave directly bounces back after hitting the different tree structures or the ground. Figure 2.2 shows the different types of scatterings from a forest [32]. Single bounce is denoted by "A" and "C" in the figure. If the transmitted wave bounces off twice, it is termed as double bounce. This is a common property of a group of targets known as corner reflectors that include built-up areas, ships on the sea, high-rise buildings, metallic objects such as cargo containers etc. As it is shown in the figure (letter "B"), double

bounce could be a result of interactions between ground and trunk or between trunk and twigs in a forest. Such targets have brighter appearances in radar images. This is due to the fact that while the transmitted signal bounces off twice, the specular reflection component with its high energy will direct back to the sensor. The third type of scattering occurs when the transmitted wave bounces off more than twice, and it is known as multiple scattering or volume scattering. This is the most common type of scattering in environments such as dense forest canopies which occur due to interactions among leaves, branches, twigs and trunks. Therefore, in a forest, we usually have all these three scattering mechanisms, however the volume scattering is dominant. The proportion of each depends on the type, height and density of the trees.
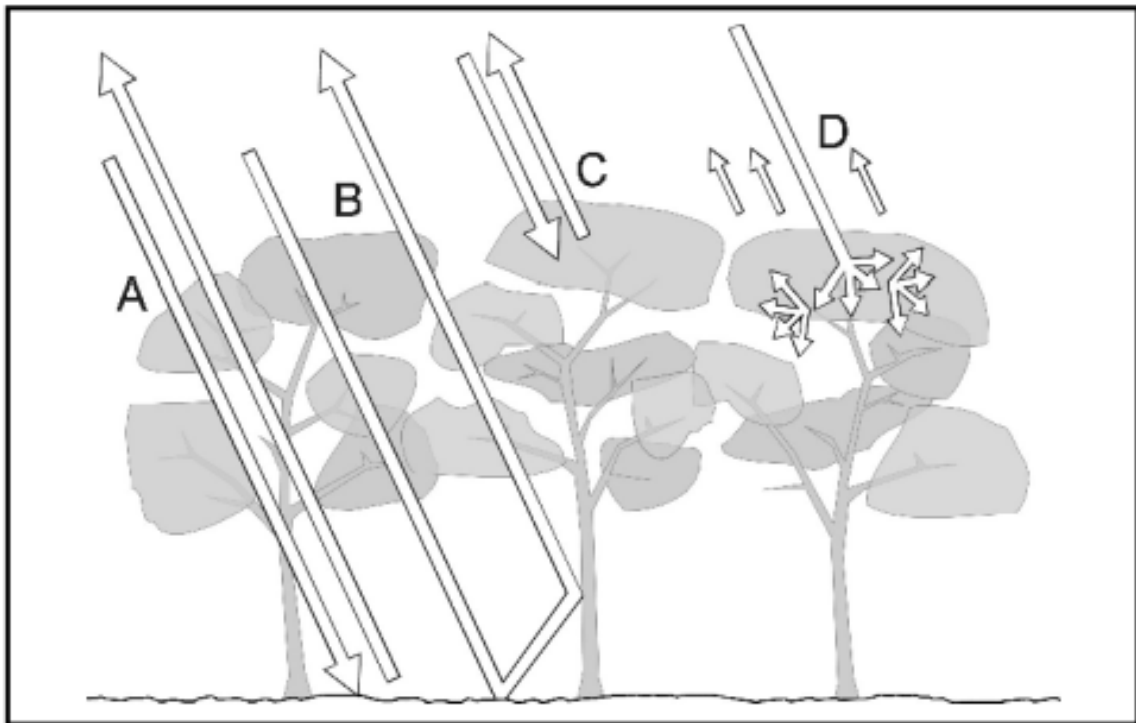
Figure 2.2: Scattering mechanisms in a forest.

Dielectric constant is another parameter which influences the interaction of radar waves with targets. It describes the response of a medium to the presence of electric field [33], and therefore it is related to its conductivity. For trees and natural targets, this information is related to their moisture content. A medium with a higher dielectric constant has a higher reflectivity. As an example, water has a dielectric constant of 80 whereas dry soil and rocks have 3-8 at radar wavelengths [33]. This direct relationship between water and dielectric constant and higher sensitivity of radar backscatter to dielectric constant can provide useful information when it comes to forest applications [34]. This is because information about moisture and volumetric water content of forest canopies can easily be inferred from radar signatures.

Another very important parameter to consider in microwave-target interactions is frequency (or wavelength). Because the surface roughness is dependent on the operating frequency of SAR, it will have a direct effect on the appearance of a particular target in radar images. Penetration is the key factor for the selection of a specific band for a particular application: the longer the wavelength (the smaller the frequency), the stronger the penetration into vegetation and soil. This higher penetration capability of relatively longer wavelengths has a two fold advantage for forest applications. One is that longer wavelengths are capable of penetrating through clouds and atmospheric moisture. This is especially useful for forest monitoring in the tropical regions which are frequently under cloud covers throughout the year. One of the main advantages of SAR over optical sensors, that are to be discussed in the section to follow, is this penetration capability which enables them to acquire information in almost all weather conditions. The second key advantage of those longer wavelengths for forest applications is that they can penetrate through the vegetation canopy and interact with the many structures of the forest [33]. This may give useful distinguishing ability that may not be present in surface scattering alone. Most widely used bands by SAR sensors for different forest applications are listed in table 2.1.

Table 2.1: Commonly used bands by SAR sensors for forest applications.

Band designation	Wavelength	Frequency
P	30 - 130 cm	0.3 - 1 GHz
L	15 - 30 cm	1 - 2 GHz
C	3.75 - 7.5 cm	4 - 8 GHz
X	2.4 - 3.75 cm	8 - 12.5 GHz

The type of polarization of the radiation in polarimetric SAR has a major role in determining the type of signal interaction with the forest components. The probability that the like-polarised radiation interact with structures having similar orientation is very high, so vertical structures in a forest will interact strongly with VV polarization. Branches with horizontal orientation interact strongly with HH polarization. The cross-polarized backscatter (HV/VH) is highly sensitive to biomass as it is commonly originated from canopy volume scattering in a forested/vegetated media. This is a direct consequence of the higher depolarizing effect of the multiple scatterers in a forest. For flat bare soil surfaces with no significant moisture content, HH and VV have approximately similar responses [35].

Incident angle is another parameter which affects the radar backscatter. It refers to the angle between the incident radar beam and the direction perpendicular to the surface of the target at the point of contact. Its effect on the microwave-forest interaction is primarily due to its influence on the microwave vertical penetration depth. In general, the radar backscatter decreases with increasing incident angle.

2.3   Optical Multi-spectral sensors

Unlike active sensors such as radar, optical sensors are passive sensors, which measure naturally available energy. It is well known that the sun is a very convenient source of naturally available energy for remote sensing. The sun's energy is either reflected, as it is for visible and a portion of infrared (IR) wavelengths, or absorbed and then re-emitted, as it is for thermal infrared wavelengths. Therefore, reflected energy can only be detected during the time when the sun is illuminating the Earth. Energy that is naturally emitted (such as thermal infrared) can be detected day or night, as long as the amount of energy is large enough to be recorded.

Optical wavelength regions in the electromagnetic spectrum (EMS) which are mainly applicable for passive remote sensing are:

- The visible region (0.4-0.7μm)

- The infrared region (0.7 μm-1mm)

The visible region is a narrow band which is visible to the human eyes. It consists of the various color components of the EMS among which red ($610 - 700nm$), green ($500 - 570nm$) and blue ($450 - 500nm$) are the principal color components. The portion of the infrared band which is useful for passive remote sensing can be further divided into reflected IR and thermal IR. Radiation in the reflected IR region is used for remote sensing purposes in ways very similar to radiation in the visible portion. The thermal IR region is quite different from the visible and reflected IR portions, as this energy is essentially the radiation that is emitted from the Earth's surface in the form of heat. There is no clear distinction between these regions as there is radiation reflected and emitted from some portion of the region simultaneously.

Multi-spectral remote sensing systems record reflected or emitted energy from an object or region of interest in multiple bands of the electromagnetic spectrum. These sensors are implemented as either air-borne or space-borne systems. The Landsat Multi-spectral Scanner and Thematic Mappers are the two well known sensors, which acquire information about the Earth's surface from space [1]. The Multi-spectral Scanner is able to record the reflected energy in the visible and IR portion of the spectrum in four discrete bands, whereas the TM records reflected visible, reflected IR and emitted (thermal IR) energy in seven separate bands. The wavelength ranges of both Landsat MSS and Landsat TM bands are listed in table 2.2.

Table 2.2: Landsat instrument bands.

Band	MSS	Band	TM
1	0.5-0.6 μm GREEN	1	0.45-0.52 μm BLUE
2	0.6-0.7 μm RED	2	0.52-0.6 μm GREEN
3	0.7-0.8 μm IR	3	0.63-0.69 μm RED
4	0.8-1.1 μm IR	4	0.76-0.9 μm NIR
		5	1.55-1.75 μm SWIR1
		6	10.4-11.5 μm TIR
		7	2.08-2.35 μm SWIR2

Where, IR = infrared; NIR = near infrared; SWIR = short wavelength infrared; TIR = thermal infrared; and μm = micron or micrometer.

The output of multi-spectral imaging systems is a stacks of images, each associated with the different bands. Images acquired by optical sensors like any other images are susceptible to geometric distortions caused by variations in platform stability including changes in their speed, altitude, and angular orientation with respect to the ground during data acquisition. It is assumed that all of the system corrections have been made for the datasets used in this study.

2.3.1   Interaction of visible and IR electromagnetic waves with the forest

The amount of reflected visible and IR energy from land cover features is mainly influenced by the chemical composition and moisture content of the observed scene. This energy, which can be recorded by optical sensors, is usually expressed as a percentage of the amount of energy incident upon those features, and it is termed as *reflectance*. Across any range of wavelengths, the percent reflectance values for landscape features such as water, bare land, sand, vegetation, etc. can be plotted and compared. Such plots are called spectral signatures. Vegetation in general has a particular spectral signature form which enables it to be distinguished readily from other types of land cover features in an optical/near-infrared image. The spectral signature of a typical healthy green vegetation is shown in figure 2.3. It can be clearly seen from the figure that the reflectance for vegetation is low in both the blue and red regions of the spectrum. This is due to the fact that chlorophyll and other pigments in plants absorb the incoming radiation at these specific wavelength ranges for the purpose of photosynthesis, a food making process in plants. On the other hand, these pigments reflect the incident radiation at the green region, which gives rise to the green color of vegetation. In the near infrared (NIR) region, the reflectance is much higher than that in the visible band due to the cellular structure in the leaves. This is evidenced by the peak reflectance values in the figure. Therefore, the internal structure of healthy leaves act as excellent diffuse

reflectors of near-infrared wavelengths. Hence, vegetation can be identified by the high NIR but generally low visible reflectance values [1, 36, 37].
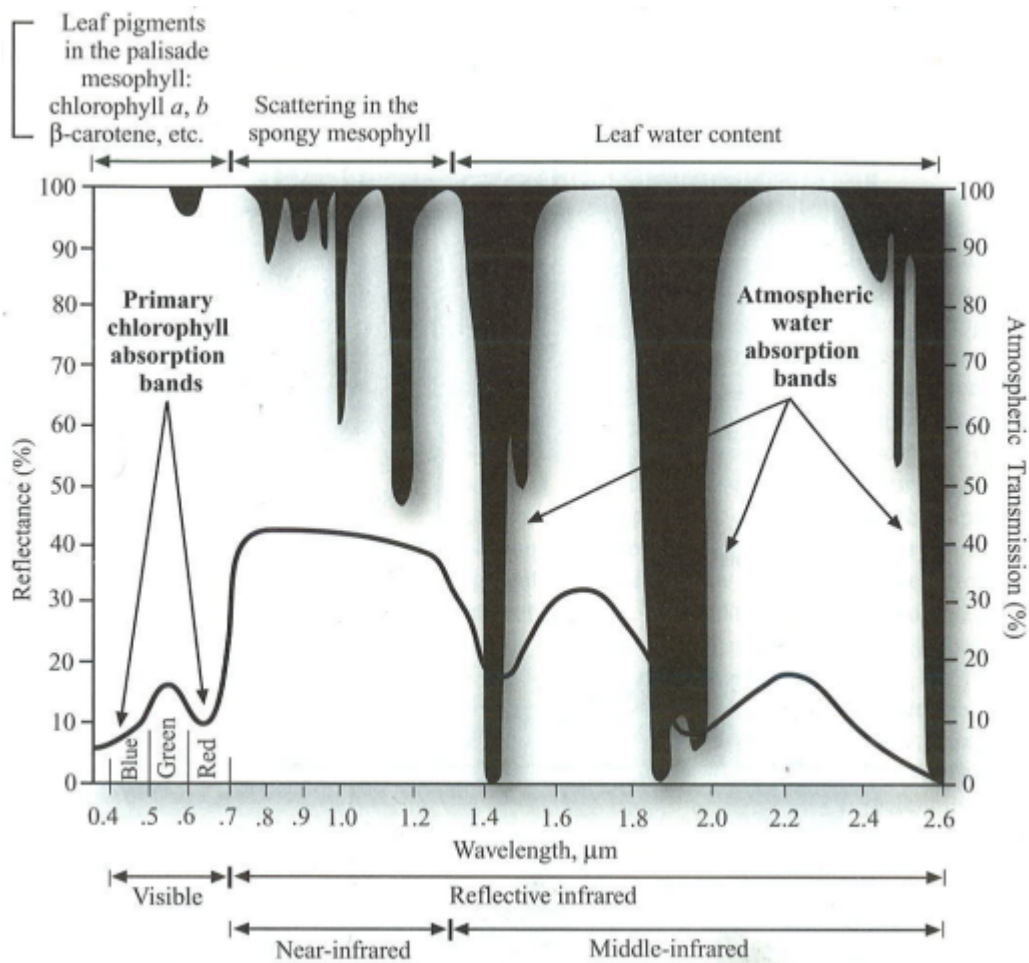
Figure 2.3: Spectral characteristics of healthy green vegetation [1].

In most cases, different landscape features have different spectral signatures. As an example, the spectral signature of some common landscape features is given in figure 2.4, [38]. This unique spectral absorption and reflection characteristics of the different landscape features in visible and near infrared region is the basis for multi-spectral and hyperspectral remote sensing. In principle, a particular landscape feature can be identified from its spectral signature if the spectral resolution of the sensing system is high enough to distinguish its spectrum from other landscape features [38]. As it is pointed out in the introduction chapter, this study encounters discriminating among the different tree age categories of a forest. Therefore, a relevant point for our study that can be inferred from the figure is that the shape of the reflectance spectrum can be used for identification of vegetation types. From the figure, vegetation types 1 and 2 can be easily distinguished

from one another as they have different responses to the visible and infrared regions. For example, this could be used to discriminate between oak and pine trees in a forest of heterogeneous species or younger and older trees in a mono-species forest. Therefore, even with in the same vegetation type, it is possible to discriminate among the different classes of vegetation based on age, moisture content, health etc. as these conditions affect the way the radiation interacts with vegetation.
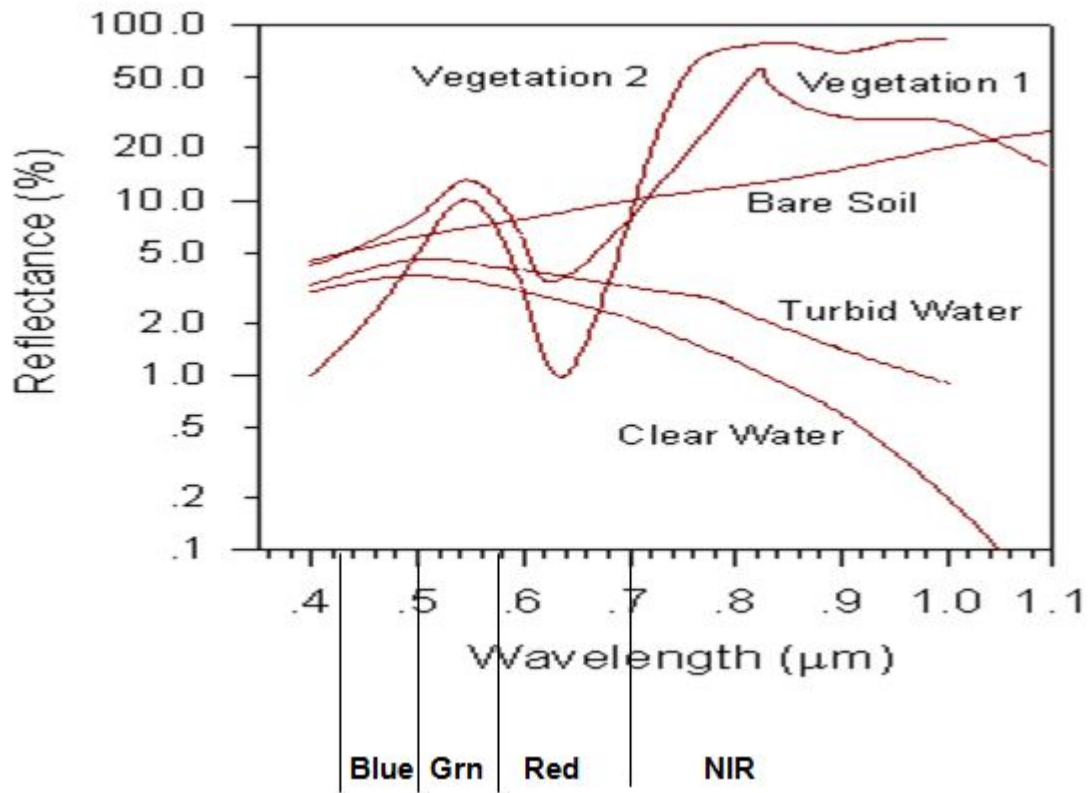
Figure 2.4: Spectral signature of some common land cover features.

Another point that has to be mentioned with respect to optical sensors is that they do not directly measure the spectral reflectance of targets, rather they measure the spectral radiance (up-welling radiance) at the sensors. Therefore, in addition to surface reflectance, the spectral radiance measured by these remote sensors depends on the interactions of input solar energy with the atmosphere (during its upward and downward passages). This is one of the limitations of optical remote sensing systems compared to microwave remote sensing systems. Since these additional factors affect our ability to retrieve accurate spectral reflectance values for ground features, an atmospheric correction has to be made before any further processing. This issue is addressed in section 4.2.1.

2.4   Comparison and Complementariness of SAR and optical datasets

Advantages of SAR sensors include the ability to obtain measurements under almost all weather and environmental conditions so that data can be collected at any time. In addition, their high resolution imaging and polarimetric capabilities are other important aspects of SAR sensors. Moreover, since they use their own source of energy, SAR sensors can be used to better control the way a target is illuminated. However, active remote sensing systems in general require the generation of a fairly large amount of energy to adequately illuminate targets. From the discussions of this chapter, some important points and comparisons of SAR and multi-spectral optical sensors are summarized in table 2.3.

Table 2.3: Comparison between SAR and optical Multi-spectral sensors.

	Optical multi-spectral	SAR
Platform	airborne/space-borne	airborne/space-borne
Radiation	Use reflected sunlight (passive)	Use Its own radiation (active)
Spectrum	visible/infrared	microwave
Frequency	multi-frequency	multi-frequency
Polarimetry	not available	is possible
Acquisition time	day time	Day/night
Weather	blocked by clouds	Can see through clouds and light rain

As it is discussed above, information acquired by SAR sensors contain information on the geometric structures, surface roughness and dielectric properties of natural and man-made objects. On the other hand, optical data contains information on the reflective and emissive characteristics of the Earth's surface. This information is highly dependent on the chemical composition and moisture content of the observed target. We have seen in the above discussions that this information can be effectively used to determine the type of feature that the imaged surface contains (water, vegetation, etc.). Even though it is affected by atmospheric attenuation, optical multi-spectral imagery take advantage of the lack of the speckle effect leading to images with a far better quality [39]. These different types of information retrieved from SAR and optical sensors are referring to different object qualities and when used together, they will complement each other. This will improve the result of a particular application such as classification. This is of course one of the motives of this study, taking advantage of this complementary information. Data fusion is treated next.

Chapter 3

Multi-Source Data Fusion

This chapter is aiming at introducing the concept of data fusion applied in the context of remote sensing purposes. The commonly used fusion approaches in the data fusion community are reviewed and some practical examples are provided. Finally, some comparisons among the fusion approaches are made to choose a convenient approach for our study.

3.1 Introduction

Data fusion refers to combining data from two or more sources or data from a single source acquired at two different times to increase the quality and improve the interpretation performance of the source data. Fused images may provide increased interpretation capabilities and more reliable results since data with different characteristics and complementary information are combined [3]. Preprocessing such as image registration and geocoding are applied to the component datasets prior to fusion to bring them into alignment [40].

In remote sensing, fusion of multi-temporal and multi-sensor datasets is of considerable importance to earth and space observation applications, such as environmental, agricultural and maritime monitoring. It is applied to integrate the information acquired with different spatial, spectral and temporal resolution sensors mounted on satellites, aircraft and ground platforms to produce fused data that contains more detailed information than each of the component inputs alone. In particular it has been successfully applied for land cover classification [10, 41]; urban area surface feature enhancement and mapping [42]. Specific application examples include: sharpen images [43]; enhance certain features not visible in either of the component datasets [44]; complement data sources for improved classification [10]; change detection using multi-temporal data [45]. Some examples of the many studies conducted to integrate a pair of images where the first is

21

acquired by a multi-spectral sensor of low spatial resolution while the second is acquired by a panchromatic sensor of high spatial resolution include [46, 47, 48, 49].

The datasets to be fused could be from:

- Two or more different sensors with different spatial or spectral resolutions, A good example in this regard is the fusion of images acquired by SAR sensors with optical data which is acquired by sensors sensitive to the visible/infrared portion of the electromagnetic spectrum.

- A single sensor, where the data is acquired at different times (multi-temporal). This helps to reveal the changes between datasets acquired at different times.

- It is also possible to combine remotely sensed data with ancillary data, for example fusion of optical images with geographic information system (GIS) data.

The fusion of optical and SAR data has received a tremendous attention by the remote sensing community[7, 8, 9, 10, 11, 50]. This is because, as it is discussed in chapter 2, the two datasets contain complementary information about the observed scene. Some application examples where these two different datasets applied include; land cover mapping [10, 12], geological study [8], snow cover mapping [9] and forest biomass estimation [14, 15]. As it is pointed out in the introduction chapter, this study investigates the fusion of multi-frequency Polarimetric SAR and multi-spectral optical datasets for the purpose of forest classification.

Different techniques have been proposed by many authors to integrate multi-source remote sensing data for the purpose of enhancing various features. The most commonly used data fusion techniques are discussed in the section to follow.

3.2   Data fusion techniques

In general remote sensing data fusion techniques can be classified into three different levels: pixel level, feature level and decision level [3]. This classification is according to the stage of processing at which the fusion takes place.

3.2.1   Pixel level fusion

This level of fusion refers to fusion at the lowest level, where multiple source images are combined to produce a single fused image. In many cases, this technique is applied to enhance the spatial resolution of one of the images while maintaining the spectral properties of the other images. A celebrated example in this regard is the fusion of a pair of images where the first acquired by a multi-spectral sensor has a pixel size greater than the pixel size of the second image acquired by a panchromatic sensor. The results will be a new multi-spectral image with a spatial resolution equal to the panchromatic

one. In general, precise co-registration of the datasets is required (at a sub-pixel level of accuracy) for this type of fusion.
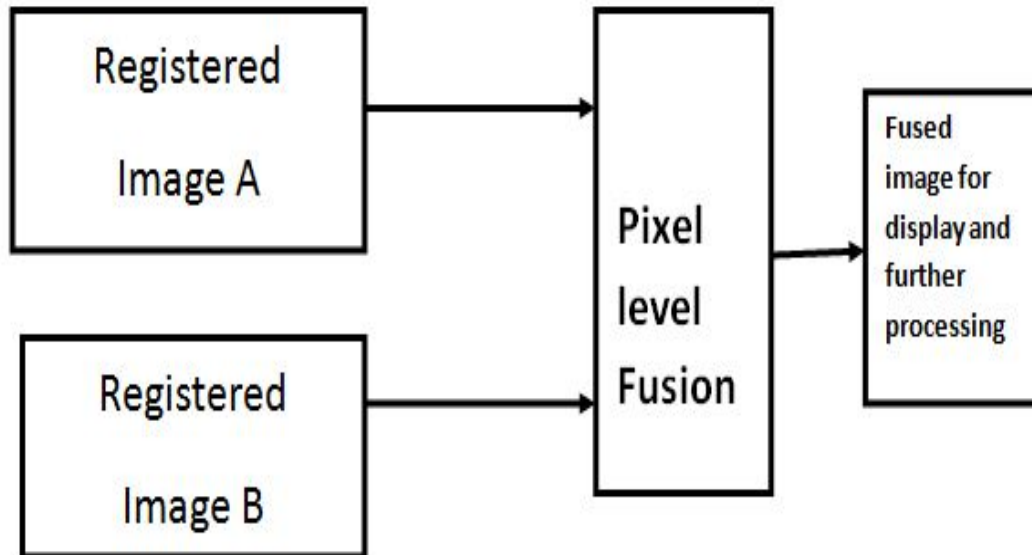
Figure 3.1: Flow chart of pixel level fusion.

Figure 3.1 shows the general structure of pixel level fusion of two images, where the component images undergo a registration process first before being merged. There are different techniques of pixel level fusion. A detailed review of these techniques can be found in [3, 51, 52]. The basic principles and particular application examples of the some commonly used pixel level fusion techniques as reviewed from the literature are presented as follows:

1. Arithmetic fusion algorithms

   These algorithms produce the fused image pixel by pixel, as an arithmetic combination of the corresponding pixels in the input images [51]. They are also the simplest and sometimes effective fusion methods. The arithmetic operations involved includes addition, multiplication, averaging, subtraction and division. Addition and multiplication of images is useful for contrast enhancement whereas difference and ratio images are particularly suitable for change detection [3]. Examples of studies in which this fusion approach is applied to combine high-resolution panchromatic images with lower resolution multi-spectral data to obtain high-resolution multi-spectral imagery include [53, 54, 55]. In their respective works, they used a combination of weighted addition and multiplication to combine the input images.

2. Color composite (RGB) methods

   These methods assign three different monochrome inputs as the red, green and blue

23

color channels of the fused image. The intensity information (grey scale value) of the input images is converted into color information (pseudo color). Prior to the fusion process, different image enhancement techniques such as contrast stretching and histogram equalization are usually applied to the monochrome input images to improve the color contrast of the fused image [51]. The difference in the grey scale values of each of the single channels results in variations in color which facilitates the interpretation of multi-channel image data. The RGB color composite technique is also commonly applied in combination with other fusion techniques such as intensity hue saturation (IHS) and principal component analysis (PCA), methods that will be discussed in the next subsections. This method has been applied to combine different datasets. As an example, it is applied to integrate images from SPOT and Seasat satelites for the purpose of improving the interpretability of geological features over temperate agricultural regions [56]. Some more examples where this method is applied to integrate optical and microwave data include [57, 58].

3. Intensity hue saturation (IHS) transformation fusion techniques

This method is one of the most widely used pixel level fusion technique. It is a class of component substitution techniques where the intensity component image of the multi-spectral data (lower spatial resolution) is replaced by an image with higher spatial resolution to improve the spatial resolution of the multi-spectral image. The method involves three steps. First, three bands of the lower spatial resolution dataset is transformed to IHS space. Second, the higher spatial resolution image replaces the intensity component. Third, backward transform from IHS space to the original space to construct the fused result. The higher spatial resolution image has a contrast stretching applied to it so that it approximately can have the same variance and mean value as the intensity component image, just to make sure that the two images are approximately equal spectrally [3, 43]. This method has been successfully applied to merge different datasets. The following are some of the examples. It has been used to merge Landsat TM and PAN data [59], hyperspectral and radar data to enhance urban surface features [42]. Another study was also carried out in [60] to merging IRS-1C multi-spectral data and panchromatic data using this method. In [61], the IHS fusion approach and other fusion techniques were applied to combine multi-spectral optical data with a panchromatic image for forest application.

4. Principal component analysis (PCA) fusion method

This is a method used to transform a multivariate data set of inter-correlated variables into a dataset of uncorrelated linear combination of the original variables [3]. The method generates a new set of orthogonal axes. It involves the computation of the eigenvalues/eigenvectors in decomposing the data into its principal components [62]. The principal components correspond to the dominant eigenvalues whose corresponding eigenvectors describe the direction that optimally retain most of the variance of the data. The uncorrelated data is the result of projecting the original

data on to the eigenvectors. There are two approaches of image fusion using PCA [3, 51]. The first involves performing PCA on a multichannel image and then replacing the first principal component by a different image (usually an image with higher spatial resolution like PAN). The procedure of this merging method is similar to the IHS method and it is also known as principal component substitution (PCS). The second approach integrates the disparate natures of multi-sensor input data in one image. The image channels of the different sensors are combined into one image file and a PCA is calculated from all the channels [36]. In both cases, inverse PCA is used to transform the data back to the original image space. A number of studies use this technique to fuse different datasets. In [50] the method is used to combine a SPOT XS image and an ERS-2 SAR image. The integration of ERS-2 and four bands of IRS-1C datasets using this fusion technique is also demonstrated in [63] for geological information enhancement. In [64], the performance of the PCS fusion method is compared with two other pixel level fusion approaches, i.e., IHS and Brovey's transformation ([65]) for forest applications. It is reported in their results that PCS provides better information for the discrimination of forest stand types than the other two.

5. High pass filtering (HPF)

This method is commonly used to enhance the spatial resolution of a lower spatial resolution image by using high pass filters. In this method, the higher spatial resolution data have a small high pass filter applied. The results of the filtering operation contain the high frequency information that is mostly related to the spatial information. The HPF results are added, pixel by pixel, to the lower spatial resolution, but higher spectral resolution, dataset. The result will be a fused image with both higher spatial and spectral resolutions. As an example, this method has been successfully applied to merge the IRS-1C multi-spectral and panchromatic bands [60].

3.2.2   Feature level fusion

Feature level fusion requires the extraction of various features from multiple data sources and then combining them into a single feature vector that can be used instead of the original data for further processing. The features are an abstraction of the raw data intended to either highlight a particular characteristics of the observed target or provide a reduced form of the raw data, which accurately and concisely represents the original information. Representative features for imagery data includes [66]:

- Geometrical characteristics of image segments, such as edges, lines, line length, line relationships (parallel, perpendicular), arcs, circles, conic shapes, size, area.

- Structural features, such as surface area, relative orientation, orientation to vertical and horizontal ground plane.

- Statistical features, such as number of surfaces, moments/mean, variance, kurtosis, skewness, Fourier descriptors, wavelet coefficients, entropy.

- Spectral features, such as color coefficients, spectral peaks, spectral signature and vegetation indices.

- Polarimetric features, such as co-pol and cross-pol ratios.

- Contextual features, e.g., texture.

Transformation of the raw data into feature vectors is termed as *feature extraction*. Methods applied to extract features usually depend on the characteristics of the individual source data and the application [52]. Segmentation procedures, region characterization, band combinations, polarimetric decomposition and principal component analysis are some of the methods that can be used for the feature extraction process. The output of this process is a list of feature vectors describing the main characteristics of the original data. These extracted features can use the separate sensor specific characteristics, which pixel level does not. Feature level fusion is then implemented on these feature sets. Figure 3.2 shows a general scheme of fusion at feature level for two registered images. As can be seen from the figure, the extracted features from each of the images are fed to a feature level fusion algorithm to form a single feature vector. When the feature sets are homogeneous, e.g., features obtained from data of identical sensors, feature level fusion can be achieved by computing a single resultant feature vector from the weighted average of the individual feature vectors. However, in the case of non-homogeneous feature sets, e.g., features obtained from data of different sensors, this level of fusion can be attained by combining them into a single concatenated feature vector.

In cases where there are differences in the range of values and distribution of the individual feature vectors, a feature normalization procedure that consists of modifying the scale of the values of the features is applied in order to map them into a common domain.
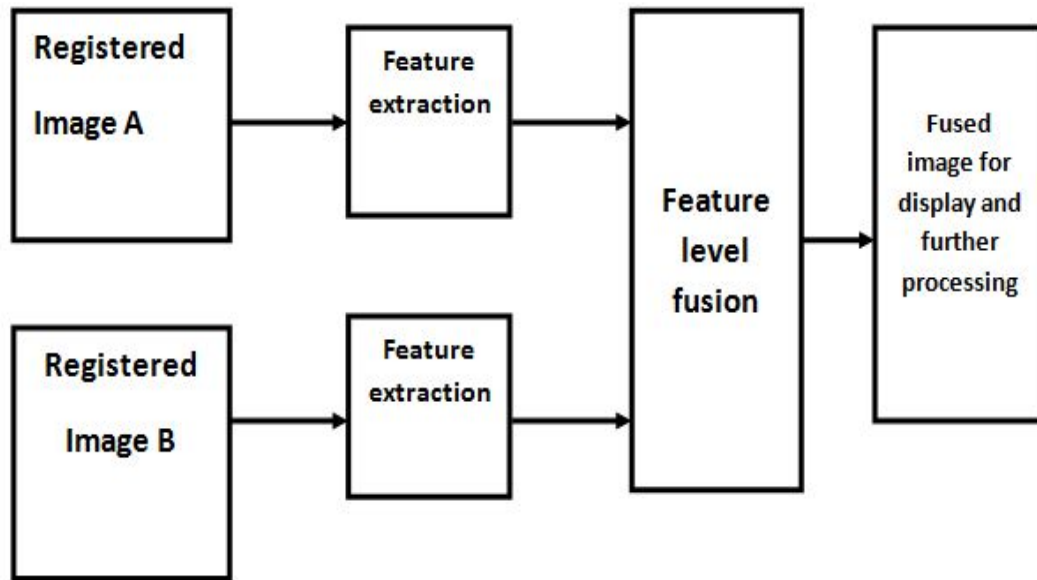
Figure 3.2: Flow chart of feature level fusion.

As an example, in [41], this level of data fusion is used to integrate polarimetric SAR (PolSAR)and Polarimetric Interferometric SAR (PolInSAR) for land cover classification. In their study, they used polarimetric target decomposition techniques [67] to extract a number of different features from the available PolSAR (L-and P-band) and PolInSAR (L-and P-band) datasets. Then they have applied logistic regression technique to fuse the features. A similar study was conducted in [68] to combine different polarimetric SAR features for land cover classification. They have considered two sets of polarimetric SAR features: one is based on polarimetric target decomposition theorems and the second set of features is computed from the different combinations of the available polarimetric channels.

3.2.3   Decision level fusion

In this type of fusion, the component datasets are transformed into feature vectors first, and then they undergo a preliminary classification process independently. Finally, the preliminary classification results from multiple sensors are combined applying decision rules to form a final classified image. The framework of this level of fusion is depicted in figure 3.3. Some examples of decision level fusion are classical inference, Bayesian inference, Dempster–Shafer's method, voting Strategies, expert Systems, logical templates, neural Networks, Support Vector Machines (SVM) and fuzzy Logic. A more complete list of decision level fusion techniques and their detailed description can be found in [66, 69]. In the data fusion community, these methods have been applied either independently or

in combination for integrating different datasets. Some examples of these methods as applied for remote sensing purposes are given below.

In [70] decision level fusion is utilized to combine hyperspectral and polarimetric SAR data to enhance urban features. Supervised classification methods are applied on the features extracted from the hyperspectral and SAR data separately and the results of these classifications are fused together. Feature images were derived from the SAR data using speckle reduction and polarimetric decomposition techniques. PCA is used to extract features from the hyperspectral data (the different PCA band images are used as feature images). Three decision level fusion methods were applied and the results were compared: a weighted majority vote, a support vector machine and a decision tree method. According to their comparison, the decision tree fusion approach gives the best classification result. Another decision level fusion method for statistical classification of multi-source data was proposed by [71]. The method was applied to combine Landsat TM image and ERS-1 SAR image, acquired at different times, for land use classification. It incorporates a priori information about the likelihood of changes between the acquisitions of the different images to be fused.
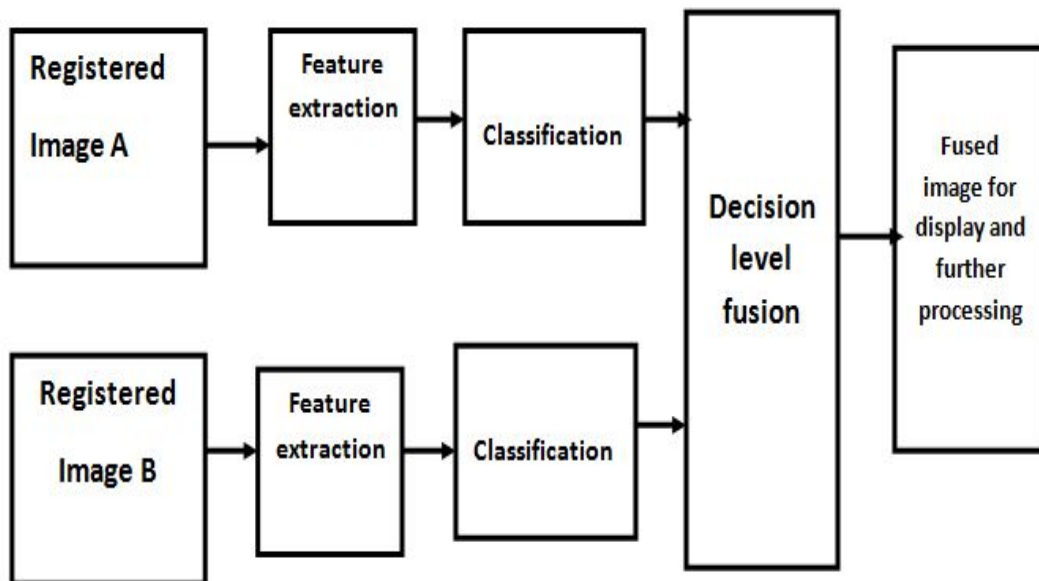
Figure 3.3: Decision level fusion.

In [72] Dempster–Shafer's fusion method is applied to combine the different bands of a Landsat TM dataset to produce a single classified image of a forested area. This fusion method is based on evidential reasoning, which can represent both imprecision and uncertainly through the definition of belief and plausibility functions [66]. This method is also applied to integrate data from optical (Spot) and the SAR image (Radarsat) for the purpose of classifying a vegetated area [73].

Before moving th the next section, which is about the comparison of the above three data fusion techniques, it is important to mention that the above classification does not present all the available data fusion techniques. This is because, there are also multi-level fusion approaches which are based up on the different combinations of the above three fusion levels [74]. However, they are not discussed here as it is beyond the scope of this study.

3.3   Comparison among the different levels of data fusion

The selection among the three different data fusion levels discussed above is dependent on a number of factors including the available information type and the application. As can be seen from figures 3.1, 3.2 and 3.3, registration of the component images is an important step in all the three levels of fusion. This preprocessing step will be discussed in detail in section 4.2 of chapter 4. In general, fusing information that is closer to the source provides better accuracy [75]. Thus, pixel level fusion is potentially more accurate than feature level fusion, which in turn is potentially more accurate than decision level fusion. However, pixel level fusion has some drawbacks, such as a requirement of an accurate data alignment among the sensors [3], it is only feasible for sensors producing the same type of observation and requires more time and memory as the whole raw data need to be processed. In feature level fusion, the feature extraction and selection scheme ensures that redundant or correlated features can be detected and removed before any further processing. Therefore, feature level fusion does not require much more space compared to the pixel level fusion. Another very important advantage of feature level fusion over pixel level is its potential of retaining sensor specific information, that is unique to the datasets considered. In addition, it has a potential of handling information from different sensors with a reasonable accuracy. However, in both pixel and feature level approaches, an assumption of a common distribution of the component datasets is required. Decision level fusion has a potential of handling information from arbitrary sources as it doesn't assume anything about the distribution of each of the input datasets. However, it has a potential of reducing the available statistical information to discrete levels and therefore it causes a loss of useful information [75]. In addition, decision level fusion is more complicated compared to the other two and usually it requires a special software. The selection of the best fusion level for a particular application is also dependent on other factors such as characteristics of the sensors and computational resources available. The three levels of fusion are compared and contrasted in table 3.1.

Table 3.1: Comparison of the three levels of fusion.

Fusion levels	Pros	Cons
Pixel level	• Accurate for data from identical sources • Simple; and no special classifier software needed	• Precise image registration is required (at sub-pixel level of accuracy) • Needs much memory, and longer processing time as the whole available data is used in the fusion and classification • Only feasible for sensors producing the same type of observation and assume that the combined data can be modelled using a common distribution
Feature level	• Sensor-specific features gives advantage over pixel level fusion • Simple; and no special classifier software needed • Capable of handling data from different sources • Does not need much memory as compared to pixel level	• Often assumes that the resulting feature data can be modelled using a common probability density function
Decision Level	• Suited for data with different probability densities • Does not need much memory as compared to pixel level fusion	• Complex, and therefore special software is often needed • Might cause loss of some information

As it is pointed out in the introduction section, this study is dealing with four datasets where three of them are polarimetric SAR datasets at three different frequencies and the fourth one is an optical dataset. It was discussed in chapter 2 that, because of the underlying sensor technology, datasets from SAR and optical sensors have different characteristics. From the previous discussions and table 3.1 above, feature level fusion has many interesting features. Its ability to handle data from different sources with reasonable accuracy, its capability of taking advantage of sensor-specific information and its simplicity to be implemented are some of them. Even though it has a potential of handling data from different distributions, decision level fusion is too complicated to implement and has a potential of "killing" some of the information contained in the data. In addition, as it is pointed out in the introduction chapter, another objective of this study is to evaluate and select different features that have the best joint performance,

and it is straight forward to see that feature level fusion is the most convenient approach to investigate the combined performance of features. **Therefore, feature level fusion is chosen for this study.**

Chapter 4

Data characteristics and preprocessing

This chapter briefly introduces the datasets used in this study. In addition, different preprocessing tasks are considered and applied to the datasets.

4.1 Study area and data characteristics

In this study, data fusion and feature extraction and selection are demonstrated using multi-sensor datasets consisting of polarimetric C-, L- and P-band airborne SAR data and multi-spectral Landsat TM data over the Nezer forest, France. The Nezer forest is a well monitored forest site with mono-species maritime pine trees of large rectangular production plots. The area is flat with no significant slope. The scene is composed of bare soil and forested areas with trees of different ages. Six tree-age groups are included from 5-8 years to more than 41 years of age. The ground truth map is shown in figure 4.1. Table 4.1 shows the proportion of each of the classes. As it can be seen from the table, the bare soil is the dominant class of all the classes.

The polarimetric SAR data is acquired in August 1989 by the NASA/JPL airborne synthetic aperture radar (AIRSAR) which operates in fully polarimetric mode at P-, L- and C-band simultaneously. The radar look angle is 45 degrees. It consists of three polarimetric SAR datasets acquired with the three different frequencies, C- (5.3 GHz), L- (1.2 GHz) and P-band (440 MHz). Each of these three datasets has four co-registered polarimetric channels of single-look complex data. However, the two cross-pol terms are equal and therefore we have only three distinct scattering components; HH, VV, HV=VH. The scene covers a small area and has a size of 1968 pixels in the azimuth and 445 pixels in the range. The pixel spacing is 3m in the azimuth and 6.7m in the range direction. P-, L- and C-band color composite images with red for |HH|, Green for |HV| and blue for |VV| are shown in figure 4.2.
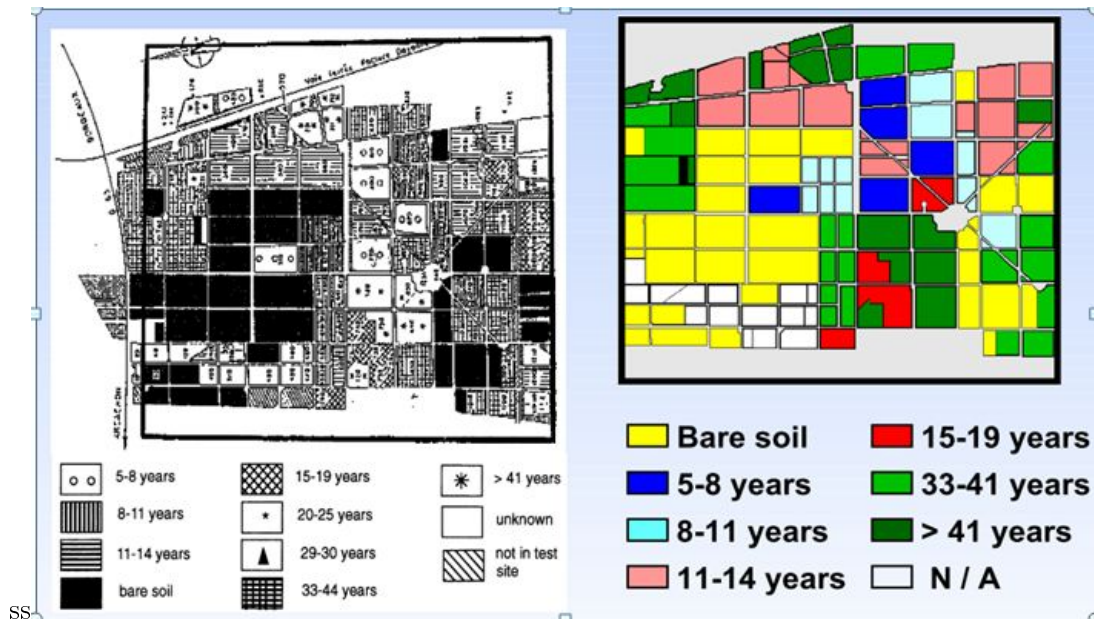
Figure 4.1: Ground truth maps of the Nezer forest.

Table 4.1: The proportion of each of the classes in the dataset.

Classes	Proportion (%)
Bare soil	36.76
5-8 years	7.65
8-11 years	8.33
11-14 years	14.18
15-19 years	4.06
33-41 years	18.86
>41 years	10.16
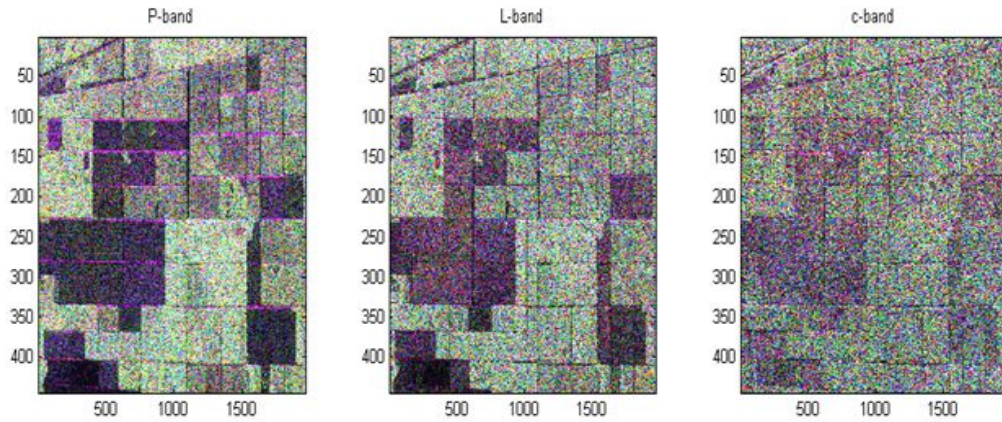
Figure 4.2: Color composite P-, L-, and C-band images of the SAR data, HH(Red), HV(Green) and VV(blue).

The multi-spectral Landsat TM dataset is acquired by the Landsat TM sensor of Landsat-4 satellite on 22 July 1991 and has seven co-registered spectral channels. The range of wavelengths corresponding to these seven spectral bands of the Landsat TM sensor are already listed in table 2.2 of chapter 2. Originally, all except the 6th band were at $30m \times 30m$ resolution; the thermal band was at a resolution of $120m \times 120m$. However, it is provided after all the seven bands are co-registered to a common resolution of $30m \times 30m$. A slightly bigger portion of the original Landsat image, which contains the area of interest, is cut and shown as color composite image in figure 4.3, with band-4(Red), band-2(Green) and band-3(Blue).

Figure 4.3: Color composite image of a portion of the Landsat data, band-4(Red), band-2(Green) and band-3(Blue).

4.2 Preprocessing of the datasets

Preprocessing refers the different processes applied on datasets before any further actions, such as classification and fusion are implemented on them. Some of the objectives of applying preprocessing include:

- To correct distorted or degraded remote sensing data, so that a more faithful representation of the original scene can be obtained. Typical processes in this category include atmospheric correction and speckle reduction.

- To align two or more images so that they are overlaid in such a way that corresponding pixels on each image represent the same location on the Earth, where geocoding and image to image registration are examples in this category.

- To cut out an area of interest from a bigger image, which is also known as cropping.

It is mentioned in chapter 3 that proper alignment of the component datasets is a key factor for the success of any data fusion approach. This can be achieved by registering the component images. Generally, image to image registration can be defined as the process of aligning two or more images, or one or more images with another data source, for example, a map [76]. It is an important step in data fusion, as the component images to be fused should contain overlapping views of the same ground features. Image registration makes sure that the component datasets will cover the same geographic area, have the same pixel size, have the same image size, and have the same orientation. In

image to image registration, one image (called input image) need to be scaled, translated or rotated, to align it with the other (called base or reference image). It involves locating and matching similar regions in the two images to be registered. This can be done either manually, where a human carries out these tasks visually using interactive software or automatically where algorithms perform these tasks. In remote sensing, manual registration is frequently used, as automated registration techniques do not always offer the required reliability and accuracy. In the case of manual registration, the user selects control points (CPs) also called tie-points from both of the images. First, the CPs in both images are interactively matched pair wise to achieve correspondence. Then, corresponding CPs are used to compute the parameters of a geometric transformation to perform the desired alignment. The main difficulty with the manual registration technique is that it is very laborious, time taking and probably impractical to manually select CPs from large amounts of data. A related but more general term is geocoding, which refers to geometric transformations of one or more images into a common cartographic map projection.

In this study, a number of preprocessing steps are applied to the datasets. They are depicted in the schematic diagram of figure 4.4, and each of them are discussed below. ENVI and Matlab are entirely used for the preprocessing.
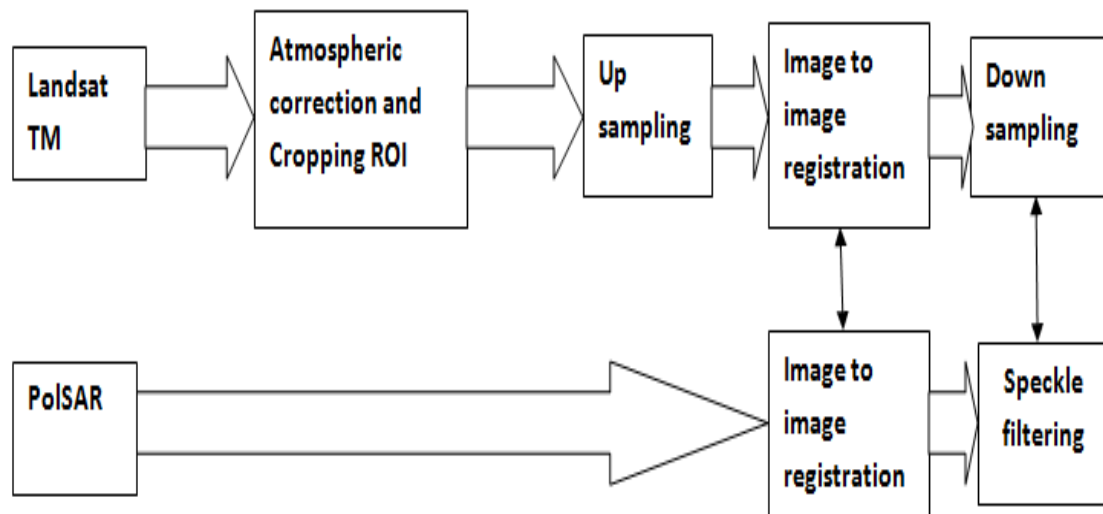
Figure 4.4: A schematic diagram showing the preprocessing applied to the datasets.

4.2.1 Atmospheric correction

As it is pointed out in chapter 2, one of the problems with data from optical sensors is that it is affected by atmospheric effects. These atmospheric effects are caused by the interaction of electromagnetic waves with gaseous molecules and particles in the atmosphere through absorption and scattering. As these atmospheric effects bear upon

the information extraction process of multi-spectral and hyperspectral images, they are usually corrected before any further processing. There are a number of atmospheric correction techniques for multi-spectral and hyperspectral remote sensing images. A review of the commonly used atmospheric correction methods for remote sensing images can be found in [77]. In this study, the quick atmospheric correction (QAC) module in ENVI is applied to the Landsat TM dataset as it is simple and give reasonably good results [78]. After applying the QAC, the cropped and corrected image is shown in figure 4.5.
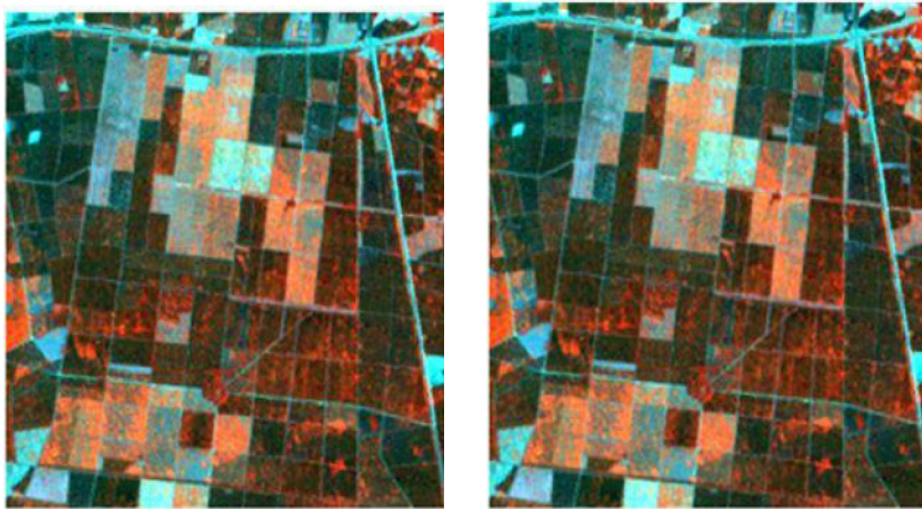
Figure 4.5: Atmospheric correction of the Landsat TM dataset; left: uncorrected, right: corrected; band-4(Red), band-2(Green) and band-3(Blue).

From the figure, it can be clearly seen that no significant change is observed for these channels after the correction. A potential reason for this could be, as the image is acquired in summer, it may not have aerosol and significant cloud cover.

4.2.2   Image-to-image registration

The image to image registration is performed manually by locating and matching tie points. However, before staring the tie point selection, the two component images should be at the same resolution. This can be achieved by resampling either the multi-spectral TM data or the SAR data. At this point, it is necessary to decide which one of the datasets should be the base image and which one is the input one, as this decision determines the image to be resampled. It is usually advisable to make radar images as base images upon registration when combining them with other data sources [40]. This is due to the fact that resampling an image affects the spatial statistics of the neighbouring pixels, which is of importance for many radar image feature extraction methods that

might use speckle statistics or texture. After choosing the SAR image as a base image, each of the multi-spectral band images are resampled (up-sampled) in ENVI so that their resolution will be approximately equal to the original resolution of the single-look complex SAR data $3m \times 6.7m$. Using the SAR image as a base image has an extra advantage in our case as it enables us to perform the registration at a higher resolution. The reason is; if we had used the multi-spectral TM data as our base image and down sampled the SAR image to the resolution of multi-spectral TM data, we would have faced a reduction in the information content of the data which would have in turn complicated the control point selection in the registration process.

There are three options of resampling in ENVI; nearest neighbour, bilinear and cubic convolution. The nearest neighbour resampling method is used throughout the whole work as it has the least effect with respect to radiometric distortion. Now the two datasets are ready for control point selection. A total of 172 control points are selected. Then each of the multi-spectral TM bands are warped about the SAR image so that the two; the SAR and TM images, are able to exactly align with each other. ENVI has three warping methods; rotation-scaling-translation (RST), polynomial and triangulation. The best result is obtained using the triangulation warping method and the results are shown in figure 4.6. After the registration, it can be checked in ENVI that the two images are exactly aligned with each other, with the same number of pixels 1968×445, with the same resolution and orientation.
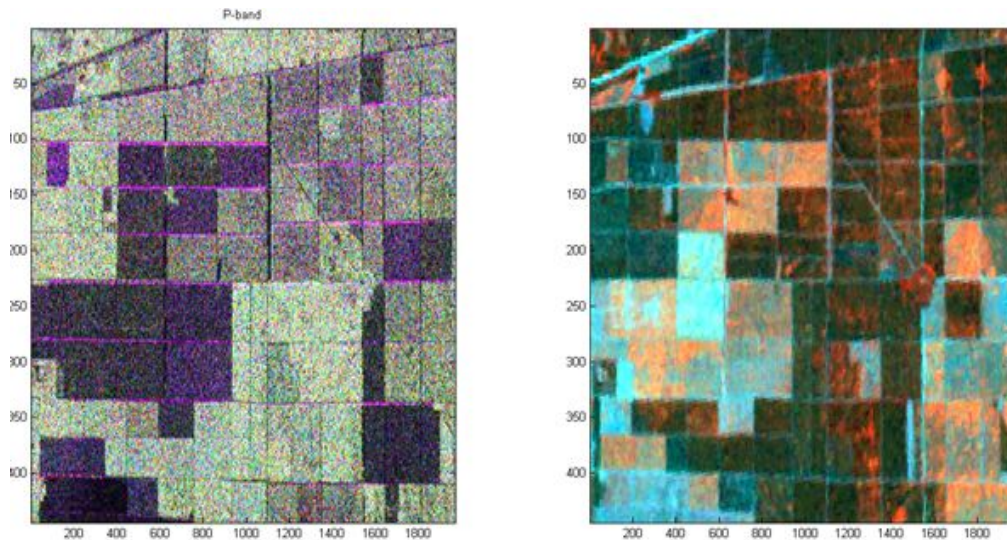
Figure 4.6: Registered images at the SAR resolution; left: P band, HH (Red), HV (Green) and VV (blue). Multi-spectral TM data, band-4(Red), band-2(Green) and band-3(Blue) (right).

4.2.3   Speckle suppression

It is mentioned in section 2.2.3 that multi-look averaging of the local neighbourhood pixels of single-look complex images is one of the methods applied to radar images to reduce speckle noise. Here, the polarimetric SAR data is processed to a 32-look averaged data (8 in the azimuth and 4 in the range). After the multi-looking, the SAR data end up with a resolution of 24mx26.8m and has a size of 246 pixels in the azimuth and 111 pixels in the range. The speckle filtered P-band color composed image is shown in figure 4.7 together with the unfiltered SLC image for comparison.
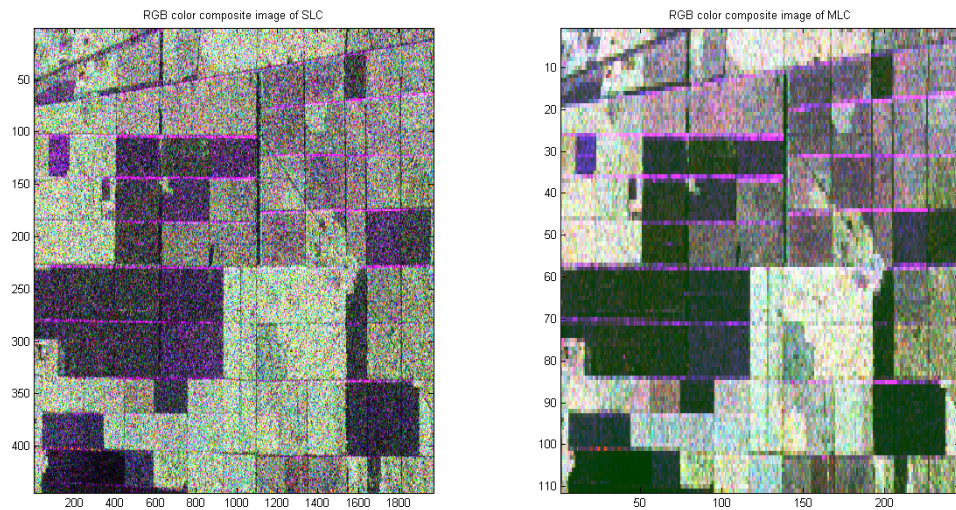
Figure 4.7: Unfiltered P-band SLC image (left) and Speckle filtered image (right); color coding: $|HH|$(Red), $|HV|$ (Green) and $|VV|$ (blue).

It can be clearly seen from the figure that the multi-look averaged image has less speckle than the original SLC image.

Then, the co-registered bands of the multi-spectral TM data (figure 4.6) are down sampled to the multi-looked resolution. Since the SAR image is chosen as a reference image, the registration is done on the slant range plane, and now all the registered datasets are on this plane. It has been mentioned that the original ground resolution of the TM data is $30x30m$. It is possible to compute the original resolution of the TM data on the slant range plane, using trigonometry for comparison. From the radar look angle of 45 degrees of our SAR datasets, the slant range resolution of the unregistered TM data is $21.2mx21.2m$. This shows that the multi-looked resolution ($24mx26.8m$) and the original Landsat TM resolution have small differences only.

Now both of the datasets:

- Cover the same geographic area

- Have approximately equal resolution; $24m \times 26.8m$

- Have the same number of pixels; 246×111

- Have the same orientation

Therefore, the datasets are ready for further processing, such as fusion and classification.

Chapter 5

Feature extraction

Feature extraction refers to the process of transforming an input dataset into a new representation set of features which accurately and concisely represents the original information. An obvious advantage of feature extraction when assisted with feature selection is that only small amount of memory and processing time will be required in the feature space as redundant information is removed in the process. As it is pointed out in chapter 3, feature extraction is also the first and fundamental step in taking advantage of sensor specific information when dealing with data from multiple sensors. Another basic advantage when it is wisely implemented is that the extracted features can have useful physical interpretation with respect to the observed scene.

This chapter is devoted to the extraction and discussions of the different features considered in this study. In total, twenty-six features; six from each of the three PolSAR datasets and eight from the Landsat TM dataset are extracted and used for this study. This number does not present all possible polarimetric and optical features, however, it includes the most commonly used features and is enough for our intended purposes. More features will be considered in future work. Feature extraction from the PolSAR datasets is based on processing the intensity, polarimetric and texture information, while feature extraction from the TM dataset is based on different band combinations to take advantage of the unique spectral responses of vegetation in the visible and near-infrared bands.

5.1 Feature extraction from the PolSAR datasets

Polarimetric features can be generally grouped into two broad categories, the first contains features processed from a simple transform of the original dataset (from the covariance matrix for example), [18], and the second category contains features extracted based on the different target decomposition theorems, [67]. Features from both categories can provide useful physical interpretation of the observed scene. A number of features from

both categories are tested for different applications in [18, 68, 79]. In [18], six features from the first category were tested for land cover (lake, some forest, crop land and some urban areas) classification. In their results, it is found that the six features considered have a rich distinguishing power for the different classes. In [68], seven features from the first category and twenty-six features from the second category were used for the identification of four different land cover classes; water, wetland, woodland and farmland. Four features from the first category are considered in [79] for sea ice classification. Some more examples of studies that features from the second category are applied for land cover classification include [41] and [80]. In all of these studies, part or all of the PolSAR features considered are found useful for the respective applications.

Only features from the first category are considered here. The fact that the information from the different polarimetric channels represent different characteristics of the observed forest (section 2.2.4), the correlation among the channels and texture, a common property in radar images are the basis for the polarimetric SAR features extracted here. Based on [18], six PolSAR features are chosen for this study. These polarimetric features are mean radar backscatter, non-Gaussianity measure, cross-pol ratio, co-pol ratio, co-pol correlation magnitude, and correlation phase. A mathematical definition of each is given below. They are calculated from a local covariance matrix which is computed from the multi-look processed data.

The mean radar backscatter represents the total intensity of the backscattered signal. The non-Gaussianity measure represents the texture information of the observed scene. As can be observed from their mathematical expressions below, the rest of the four features are computed by ratioing elements of the covariance matrix, and are intended to enhance different physical structures of the scene by using correlation information from the separate polarimetric channels. A local covariance matrix $\mathbf{C}$, from which the six features are derived, is calculated by passing over the full size image, using a sliding window of size 4 x 8. This window size is the number of looks to be averaged to reduce the speckle effect at the same time, section 4.2.3. The multivariate covariance matrix from a collection of L local $\mathbf{x}_i$ inside the window, is given by:

$$\mathbf{C} = \sum_{i=0}^{i=L} \mathbf{x}_i \mathbf{x}_i^H$$

where the $\mathbf{x}_i$ are the SLC vectors and the superscript H denotes the complex conjugate transpose. Mathematically, each of the six features are defined as follows:

1. Mean radar backscatter is given by

$$\boldsymbol{\mu} = det(\mathbf{C})^{\frac{1}{d}}$$

2. Non-Gaussianity measure is given by

$$\boldsymbol{\nu} = \frac{mean(\mathbf{x}_i^H \mathbf{C}^{-1} \mathbf{x}_i)^2}{d(d+1)}$$

3. Cross-pol ratio, the ratio of cross-polarized returns to co-polarized returns, is given

by

$$\mathbf{f_x} = \frac{2 * \mathbf{C}_{HV}}{\mathbf{C}_{HH} + \mathbf{C}_{VV}}$$

4. Co-pol ratio, the ratio of co-polarized returns, is given by

$$\mathbf{f_c} = \frac{\mathbf{C}_{HH}}{\mathbf{C}_{VV}}$$

5. co-polarized correlation magnitude, a measure of the correlation of the magnitudes between co-polarized returns, is given by

$$\mathbf{f_m} = magnitude(\frac{\mathbf{C}_{HHVV}}{\sqrt{\mathbf{C}_{HH} * \mathbf{C}_{VV}}})$$

6. co-polarized correlation phase, a measure of the correlation of the phases between co-polarized returns, is given by

$$\mathbf{f_p} = angle(\frac{\mathbf{C}_{HHVV}}{\sqrt{\mathbf{C}_{HH} * \mathbf{C}_{VV}}})$$

where L is the number of looks to be averaged, thirty-two in this case, d is the number of distinct elements in the scattering matrix, three in this case (HH, HV=VH, VV) and det() is determinant of the matrix.

A total of eighteen polarimetric features; six from each of the three SAR datasets (C-, L-, and P-band) are extracted. A logarithm transformation is applied to $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, $\mathbf{f_x}$ and $\mathbf{f_c}$ for improved visualization. The values of all the features are scaled to a common range, [0,1], for ease of comparison. This feature normalization is also a basic preprocessing step for feature combination. Each of the six features from the P-, L- and C-bands are shown as images in figures 5.1, 5.3 and 5.5 respectively. The corresponding histograms are shown in figures 5.2, 5.4 and 5.6.
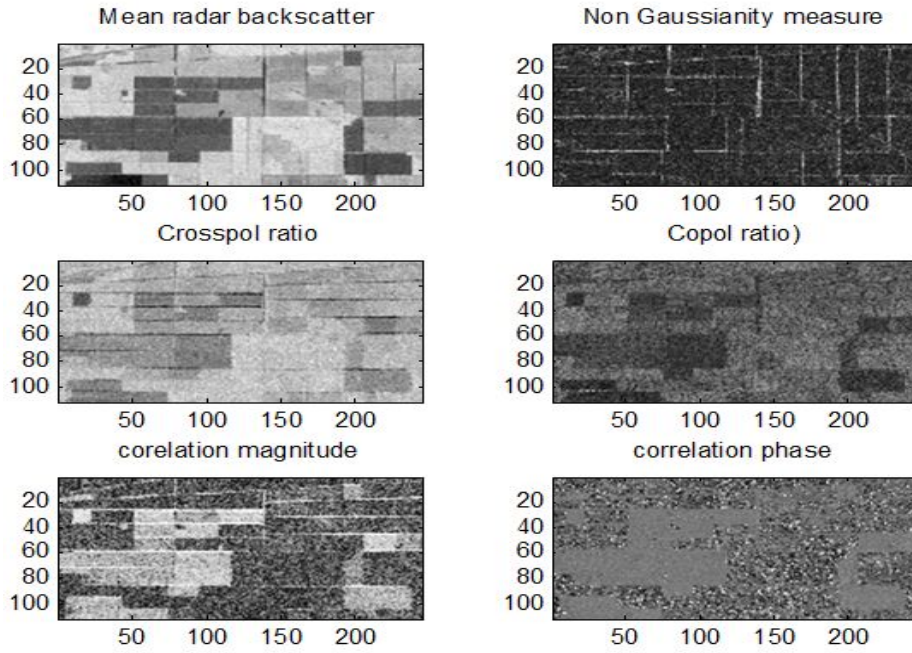
Figure 5.1: The 6 feature images of P band PolSAR data.
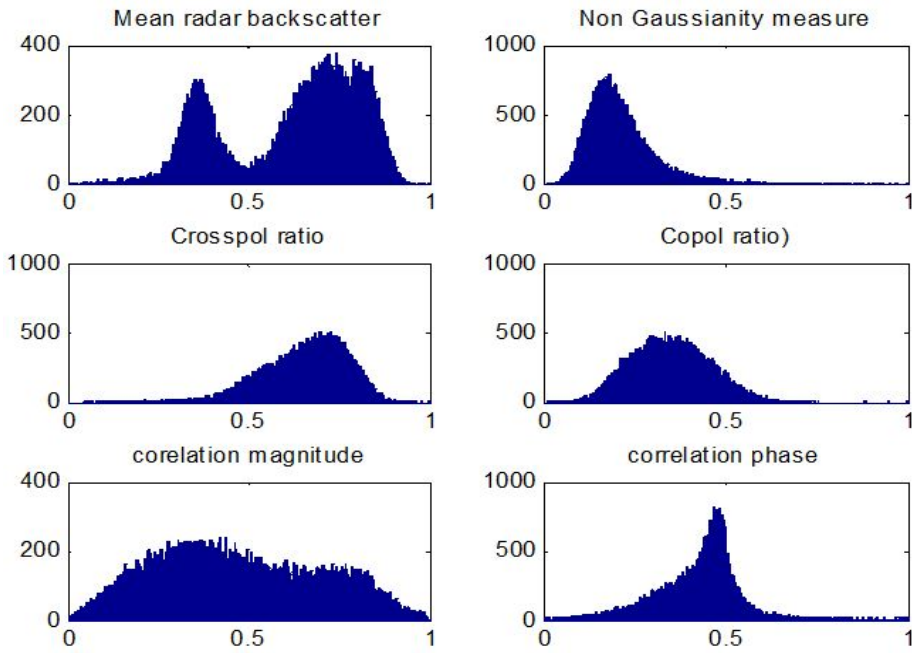
Figure 5.2: Histogram plots the 6 features of P band PolSAR data.
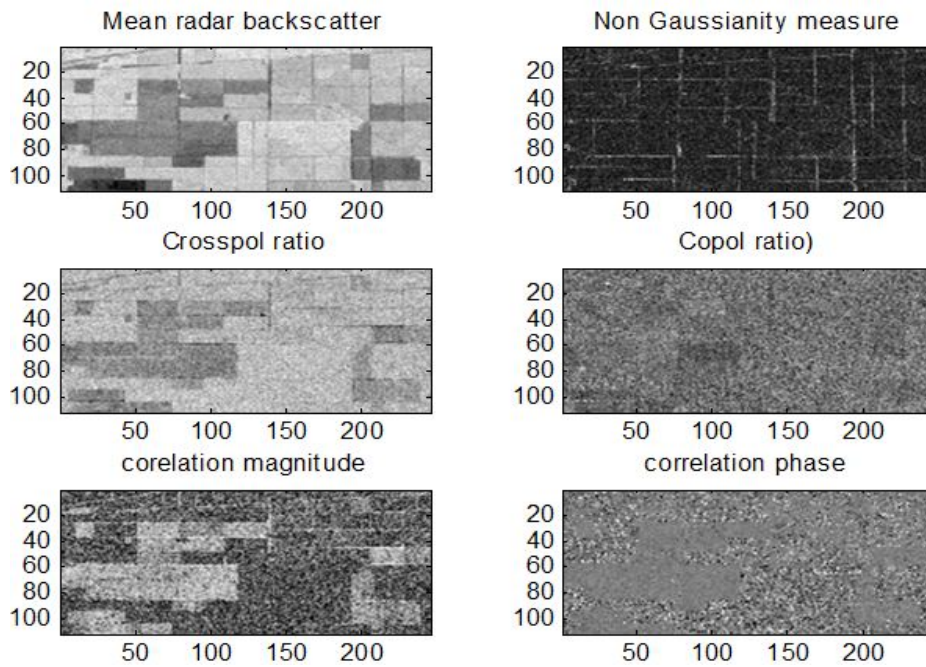
Figure 5.3: The 6 feature images of L band PolSAR data.
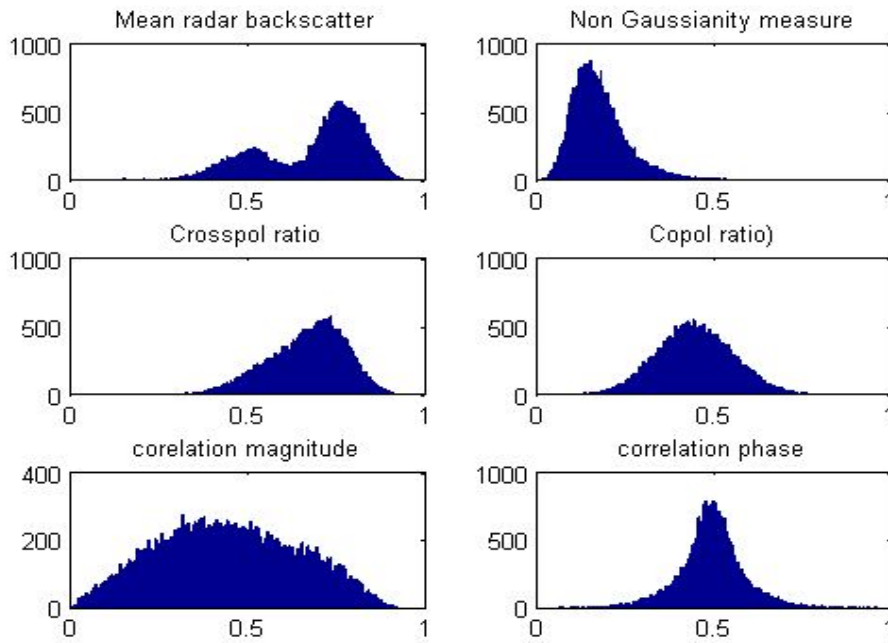
Figure 5.4: Histogram plots the 6 features of L band PolSAR data.
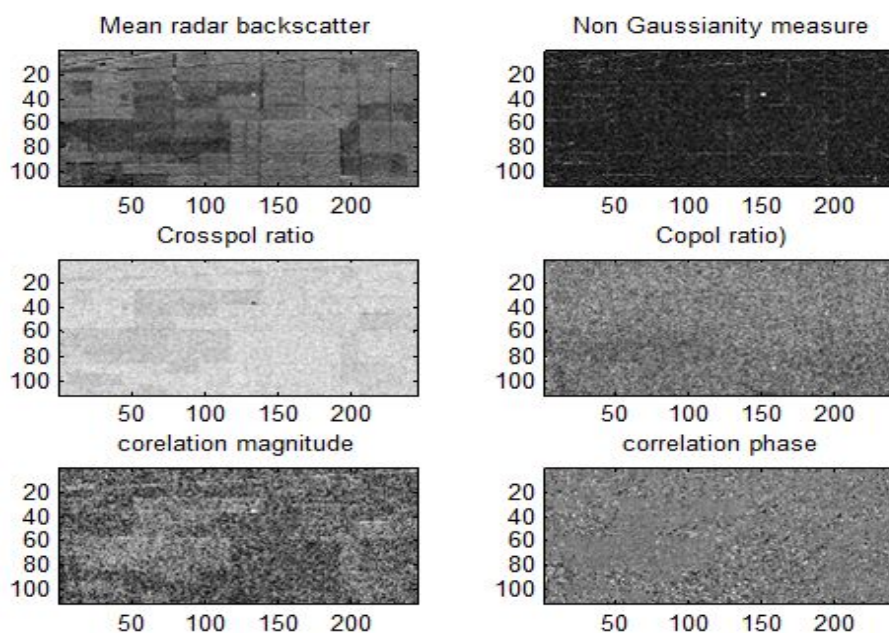
Figure 5.5: The 6 feature images of C band PolSAR data.
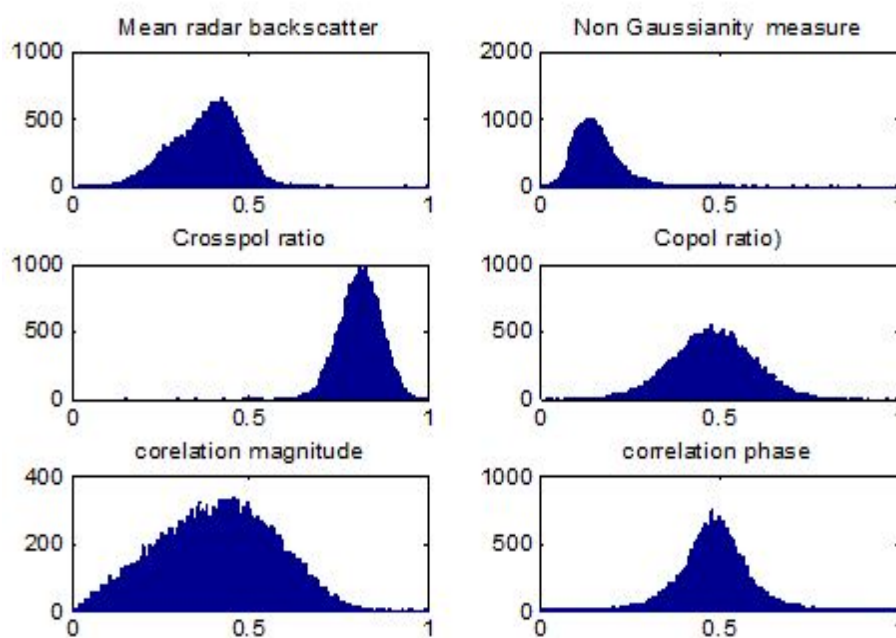
Figure 5.6: Histogram plots the 6 features of C band PolSAR data.

Referring to each of the six different feature images of the three datasets and their respective histograms, the level of contrast and detail decreases from P- through L- to C-band features. Another point that can be observed from the feature images is that there is a relative difference in contrast even among features within the same dataset. This indicates that some of the features might perform better than others in discriminating among the different classes. The physical interpretations with respect to the P-band features is given below. Even though the level of detail decreases from the P- to C-bands, the discussions are also valid for the L- and C-band features.

The mean radar backscatter, figure 5.1, shows relatively higher contrast than the rest of the features. The bare soil appears darker, while the older trees have brighter appearances. The young and middle age trees (5-8, 8-11, 11-14 and 15-19 years of age) have different shades of grey appearances. Generally, it can be seen from the figure that the intensity increases with tree age. This is not far from the expected, because as the age of the tree increases it will have much more branches and twigs which in turn result in higher amounts of backscatter.

The non-Gaussianity measure has low values (dark) almost everywhere except at the boundaries of the different classes, where it has higher values with brighter appearances, figure 5.1. It is acting like an edge detector, which enhances edge mixtures. This edges are highlighted primarily due to high contrast mixing. Such mixtures "measure" a high variance relative to the average, which appears as a high texture measure, but it is due to two classes and not "true" radar texture.

Even though it is not as high as the mean radar backscatter, the cross-pol ratio also shows reasonably good contrast among the different classes. Here also, the bare soil appears dark, and the older trees appear brighter. This is from the fact that the cross polarized return (HV) in general is weaker than HH or VV but is strengthened by depolarization caused by the different tree structures. In the co-pol image, it is possible to identify the bare soil from the forested area as the bare soil appears relatively darker. However, it only shows little contrast among the different forest age groups compared to the cross-pol image.

Again referring to figure 5.1, the bare soil and the edges appeared brighter than the rest of the scene in the correlation magnitude image. This is opposite to the other features where in most of them, the bare soil appears darker. The darker appearances in the forested part in this case, indicates that there exists low correlation between HH and VV returns in the forest than the bare soil part. This is due to the fact that, as it is discussed in section 2.2.4, the HH and VV polarizations interact with the tree structures differently.

In the correlation phase image, the bare soil appeared much smoother than the rest of the scene, which is characterised by granular structures of different shades of grey dots. The smooth appearance of the correlation phase in the bare soil part than the forested region is from the fact that the variation in the phase change between the transmitted and received polarizations of HH and VV is smaller for the bare soil than the forest. This

is of course reasonable because, a rapid phase change is expected in the forest part (due to the many interactions with the different tree structures) of the scene than the bare soil part.

The relatively higher contrast among the different forest age groups in the cross-pol image than the other three, co-pol ratio, correlation magnitude and correlation phase, is due to the fact that the cross-polarized return from forested/vegetated area is highly sensitive to vegetation biomass, and this biomass is expected to vary with tree age.

Figure 5.7 shows the scatter plots of all possible pairwise combinations of P-band features.
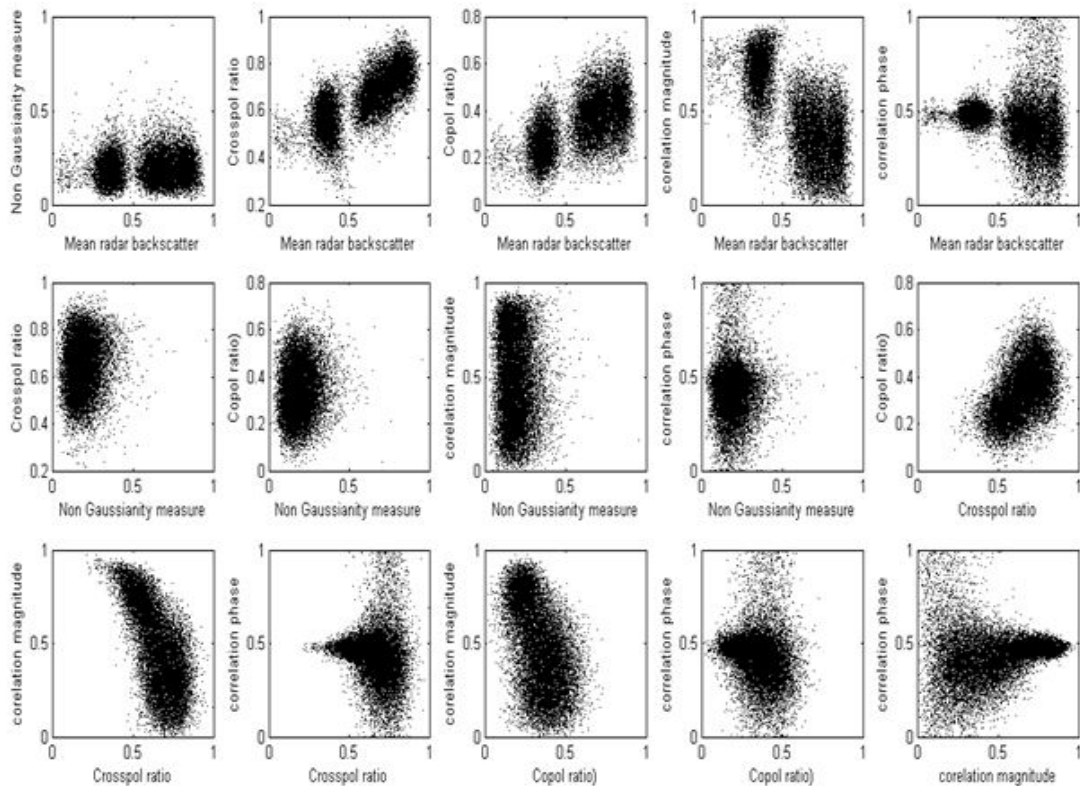
Figure 5.7: Scatter plots of all possible pairwise combinations P-band features.

It can be clearly seen in the first upper row of figure 5.7 that the scatter plots show two separable main clusters and potentially four globular clusters. In contrast, the rest of the plots show lots of overlap/mixing. This shows that the mean radar backscatter has a relatively better quality compared to the other five features. In reference to the ground truth map, figure 4.1, one can guess that the above two separable main clusters (blobs) could correspond to the bare soil and the forest classes.

5.2   Feature extraction from the multi-spectral TM dataset

Vegetation indices (VIs) are the most widely used multi-spectral/hyperspectral features for vegetation applications. They are combinations of surface reflectance at two or more wavelengths in a multi-spectral or hyperspectral optical data designed to highlight a particular property of vegetation, such as relative abundance and activity of green vegetation, percentage green cover, chlorophyll content and absorbed photosynthetically active radiation, [1]. Most of the vegetation indices take advantage of the inverse relation of the reflective properties of healthy green vegetation in the infrared and visible regions. This is from the fact that, in case of healthy green vegetation, most of the energy in the near infrared wavelengths is transmitted and reflected, with little absorbed, in contrast to the visible wavelengths where absorption is predominant, with some reflected and little transmitted.

As it is pointed out in section 2.3.1 of chapter 2, the higher absorption in the visible (blue and red bands) region is due to the presence of chlorophyll pigments in green plants. They use this absorbed energy for photosynthesis, a food making process in green plants. A relative lack of absorption between the blue and red bands produces a trough in the absorption curve at approximately 0.54μm, which corresponds to the green portion of the electromagnetic spectrum [1]. At the near infrared (NIR) portion of the EMS on the other hand, the reflectance and transmittance of plant leaves is very high; because of the structure of the leaf. When there is a bulk of incident sunlight in this region, and if plants absorb this energy with the same efficiency as they do in the visible region, they could become much too warm and the proteins would be irreversibly denatured. Therefore, they reflect or transmit it.

VIs are computed either by ratioing any two or more spectral bands, or by taking the ratio of sums, differences or products of any number of bands. It is also important to note that, even though there are some VIs which provide unique information, many of them are functionally equivalent (redundant), [1]. The following eight vegetation indices are extracted from the Landsat TM dataset for this project, and their mathematical definitions are taken from [1]. The historical development and detailed discussion of a large number of VIs, including those considered here, can be found in [1].

1. The Normalized Difference Vegetation Index (NDVI)

   The NDVI is one of the oldest, most well known, and most frequently used VIs. It is defined by the ratio of the difference of the near-infrared and red reflectance, over the sum of those. Its values ranges from -1 (no vegetation) to +1 (abundant vegetation) [31]. It is found to be related to many properties of plants. It can be used to identify the health status of plants, to give a description of phenological changes and to estimate green biomass and crop yield. For Landsat TM data, it is computed from the following band combinations:

   $$NDVI = \frac{(NIR-RED)}{(NIR+RED)}$$

where the band designation are based on table 2.2 of chapter 2.

2. Soil brightness index (B), Greenness (G) and wetness (W) indices

These three features are derived on the basis of an orthogonal transformation of the original Landsat Multi-spectral data space, called tasseled cap or Kauth-Thomas transformation, into a new feature space. The basic idea in tasseled cap transformation is that, if information about a particular scene characteristics is represented by more than one band in a multi-spectral information acquiring system, the total information about this scene characteristics can be best captured by a systematic combination of the bands, [81]. The soil brightness index is intended to measure the soil characteristics, the greenness vegetation index is intended to highlight the green biomass and the wetness is intended to highlight the moisture content of a vegetated area [1]. In contrast to the NDVI, which is formed from the combination of red and infrared bands, these three features are formed by including the SWIR bands to account soil and moisture information. For a Landsat TM data, the linear combinations of the visible, near-infrared and middle-infrared bands in deriving these three features are given by:

$$B = 0.2909BLUE + 0.2493GREEN + 0.4806RED + 0.5568NIR + \\ 0.4438SWIR1 + 0.1706SWIR2$$

$$G = -0.2728BLUE - 0.2174GREEN - 0.5508RED + 0.7221NIR + \\ 0.0733SWIR1 + 0.1648SWIR2$$

$$W = 0.1446BLUE + 0.1761GREEN + 0.3322RED + 0.3396NIR + \\ 0.6210SWIR1 + 0.4186SWIR2$$

3. Perpendicular vegetation index (PVI)

This vegetation index is intended to highlight plant development by using the perpendicular distance from the soil line. The longer the perpendicular distance from the soil line in the direction of high canopy closure inside the triangular shaded area, see figure 5.8, the higher the values of PVI and the more matured the plant is. It can be computed from the following band combinations:

$$PVI = \sqrt{(0.355NIR - 0.149GREEN)^2 + (0.355GREEN - 0.852NIR)^2}$$

4. Triangular vegetation index (TVI)

The triangular vegetation index uses the relative difference between red and near-infrared reflectance together with the magnitude of green reflectance to describe the radiative energy absorbed by plant pigments. It is computed as an area of a triangle in spectral space, where the vertices of the triangle are determined by the maximum of the green reflectance, the minimum of the red absorption and maximum of the near-infrared reflectance. Mathematically it is given by:

$$TVI = 0.5 * (120(NIR - GREEN) - 200(RED - GREEN))$$
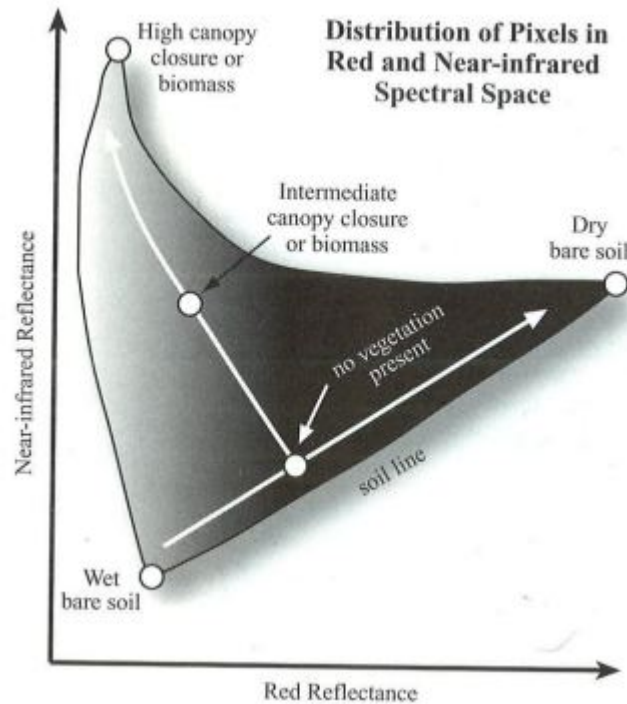
Figure 5.8: The soil line and distribution of pixels in a red and near/infrared space, [1].

5. Soil adjusted vegetation index (SAVI) and atmospherically resistant vegetation index (ARVI)

Even though the NDVI is a useful and universally used VI, its efficiency is highly influenced by a number of factors including soil background variations, moisture, atmospheric conditions and the presence of dead material in the canopy itself. Therefore a number of other vegetation indices have been developed by correcting NDVI for different factors, where SAVI and ARVI are two of such indices. In the SAVI, an adjustment factor, L is incorporated into the NDVI to account for soil background variations. L is usually determined iteratively, for example for a dark and light soil background, the search for L continues until an appropriate value that give equal vegetation index results for the dark and light soils is found. Soil brightness variations are found to be minimized by using an L value of 0.5, [82], and this value of L is used for this study. The ARVI is also a simple modification of the NDVI to reduce its sensitivity to atmospheric effects. Landsat TM band combinations for these two indices are given by:

$$SAVI = \frac{(1+L)(NIR-RED)}{L+NIR+RED}$$

$$ARVI = \frac{NIR-(RED-C*(BLUE-RED))}{NIR+(RED-C*(BLUE-RED))}$$

Where C is a constant. Different values of C can be used depending on the appli-

cation, whether it is maritime, desert or highly vegetated area [83]. A value of 0.3
is used here.

It is found that a square root transformation of the perpendicular vegetation index im-
proved its visualization. All the values are normalized to the same common scales, [0,1],
as it is done for the PolSAR features. All of the above eight vegetation indices extracted
from the multi-spectral TM data, together with their respective histograms are shown in
figures 5.9 and 5.10 respectively.
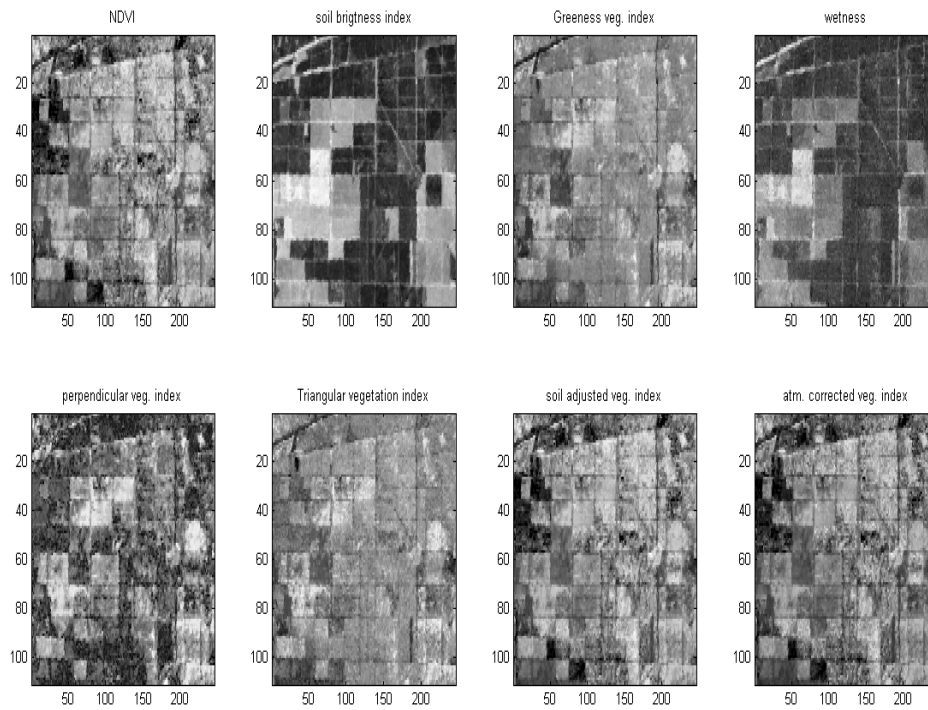
Figure 5.9: The 8 features of multi-spectral TM dataset.
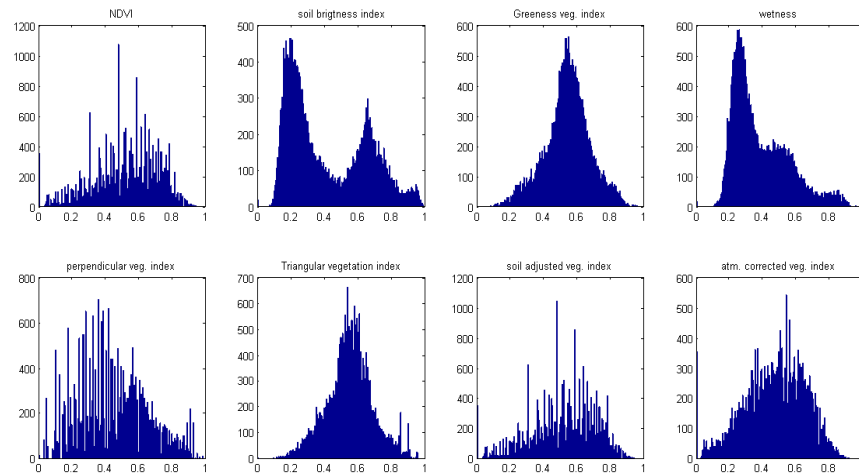
Figure 5.10: Histogram plots the 8 features of multi-spectral TM dataset.

It can be clearly seen from figure 5.9 that the bare soil and the forested parts can be easily identified from the soil brightness index and wetness images. The soil brightness index has higher values for the soil part than the forest as expected. Beyond highlighting the soil part, it has also identified an extra class within the bare soil, indicated as a brighter rectangle in the middle portion of the bare soil region. However, this information is not included in the ground truth map and might cause some confusion in the classification task. The same is true for the wetness index. In the wetness image higher values correspond to the bare soil part, and the forest appears darker. This means that the moisture content of the bare soil is higher than that of the forest. For an image acquired during the summer time, this could happen due to rain.

It is somewhat difficult to relate the remaining six vegetation indices to the physical property of the scene, see figure 5.9. As an example, in the NDVI image, it is expected that higher values (bright) correspond to the forested part and the brightness is expected to increase with forest age. However, there are high values of NDVI in both the bare soil parts as well as in younger tree regions, and the same is true for the other five indices. Even though it is difficult to associate to the different tree age categories, the high contrast in the images of these six features in the forested region might help to discriminate some of the different tree age categories.

The scatter plots of all the vegetation indices (all possible pair-wise combination) is shown in figure 5.11.
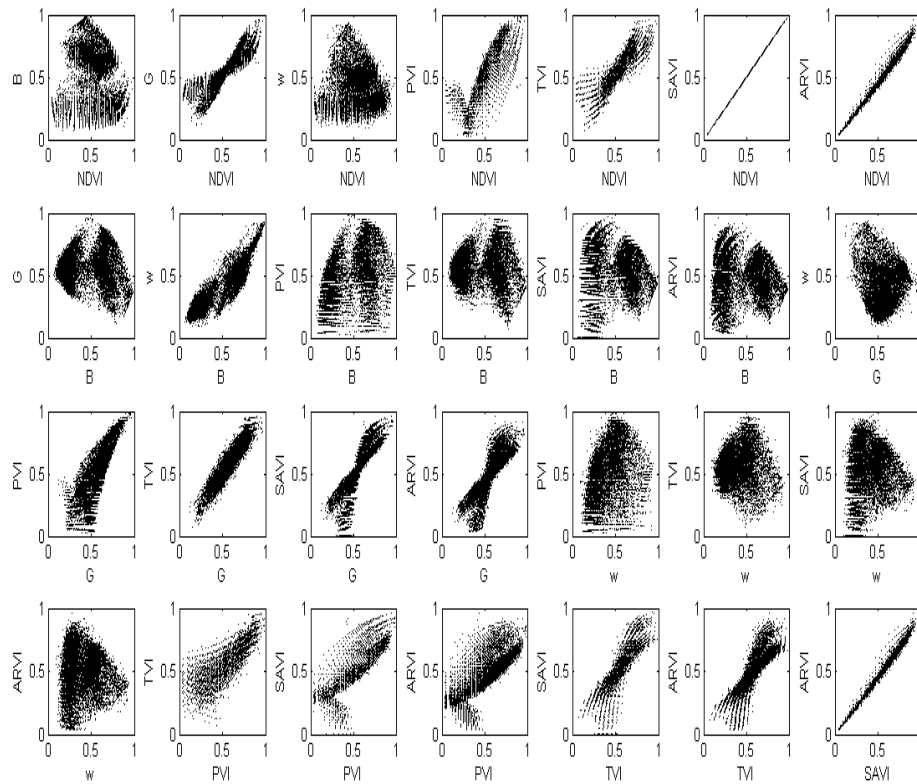
Figure 5.11: Scatter plots of all pair-wise combinations of the Landsat TM features.

As can be clearly seen from the scatter plots, there are some kinds of patterns which look like "dotted lines". These patterns are resulted from the 256 discrete level of Landsat data. As expected, a lot of correlation is observed among the different vegetation indices. Specially three of them, namely NDVI, soil adjusted vegetation index and atmospherically resistant vegetation indices show high correlation. This is due to the fact that two of them is a simple modifications of the NDVI, as explained above. It can be observed from the scatter plots that the two features namely soil brightness index and wetness are able to form some clusters, and hence it is possible to expect some better classification performances from them than from the others.

In summing up, a total of 26 features are extracted from the four datasets, six from each of the three PolSAR datasets and eight from the Landsat TM dataset. In general, it can be seen from their respective feature images that the P-band shows better contrast with respect to the different classes than the L- and C-band datasets. The C-band feature images relatively show the least detail of the four datasets. Even though there is a lot of contrast in many of the Landsat TM feature images, it is a bit difficult to associate this

56

information to the different tree age categories. Moreover, there exists a lot of redundant information in the TM dataset as expected. The different features considered in this study are listed in table 5.1 for ease of reference.

Table 5.1: List of extracted features.

Feature number	SAR features from each of P-, L- and C-bands	Optical features from Landsat TM
1	Mean radar backscatter	Normalized difference VI
2	Non-Gaussianity measure	Soil brightness index
3	Cross-pol ratio	Greenness
4	Co-pol ratio	Wetness
5	Correlation magnitude	Perpendicular VI
6	Correlation phase	Triangular VI
7		Soil adjusted VI
8		Atmospherically resistant VI

Chapter 6

Experiment 1: Fusion of the datasets and classification

This chapter aims to quantitatively evaluate and compare the classification performances of each of the four datasets and their different combinations, using their respective features, in discriminating the different classes of our scene.

As can be seen from the histogram plots of many of the features in figures 5.2 and 5.10, their distribution approximates a Gaussian distribution. In addition, from the scatter plots of figures 5.7 and 5.11, some feature combinations are able to form clearly separable compact clusters. The nature of their distribution and their capability of forming natural compact clusters indicate that multivariate Gaussian distribution assumption can possibly be made in the multi-dimensional feature space. This mixture of Gaussian assumption can then be used in a supervised or unsupervised fashion for the classification task of the different classes [62, 84].

In this study, a supervised maximum likelihood Bayesian classification scheme is applied on the feature sets using training samples taken based on the ground truth. The method is a statistical approach, which assumes class dependant multivariate normal distributions of pixels from feature images. This classification technique has been commonly applied for remote sensing purposes [85, 86, 87]. Linear discriminant functions are computed from the mean and covariance estimates of the training data points. Each pixel in the feature image is then assigned to the class for which it has the highest probability of membership determined by the discriminant functions. Mathematically, it can be stated as follows [62]:

For M classes $w_1, w_2, ..., w_M$, the Bayesian classifier assigns an unknown feature vector $\mathbf{x}$ to class w_i if

$$P(w_i|\mathbf{x}) > P(w_j|\mathbf{x}) \quad \forall \text{ i} \neq \text{j}$$

where the conditional probability $P(w_i|\mathbf{x})$ is obtained from Bayes rule given by

$$P(w_i|\mathbf{x}) = \frac{p(\mathbf{X}|w_i)P(w_i)}{p(\mathbf{X})}$$

where $P(w_i)$, $i = 1, 2, ..., M$ are the prior probabilities of each of the classes that can be estimated from the available training data points. $p(\mathbf{x})$ is the pdf of $\mathbf{x}$ which is given by $p(\mathbf{x}) = \sum_{i=1}^{i=M} p(\mathbf{x}|w_i)P(w_i)$. The remaining unknown parameters are the class-conditional probability density functions $p(\mathbf{x}|w_i)$, $i = 1, 2, ..., M$, describing the distribution of the feature vectors in each of the available classes. These class conditional probability density functions in the l-dimensional feature space can be computed by making an assumption of multivariate normal distributions for pixels of feature images, and they are given by

$$p(\mathbf{x}|w_i) = \frac{1}{(2\pi)^{l/2}\Sigma_i^{1/2}} \exp(-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i))$$

where $\boldsymbol{\mu}_i$ and Σ_i are the mean and covariance structures of each of the classes that can be estimated from the available training data points.

6.1   Ground truth

The ground truth map is repeated here for ease of reference. It can be clearly seen from the left and right figures of figure 6.1 that there are some discrepancies between the two provided ground truth maps. Therefore, another ground truth map is created based on fusing the information from both maps. This new map contains only regions that agree with both of the ground truth maps and is shown in figure 6.2.
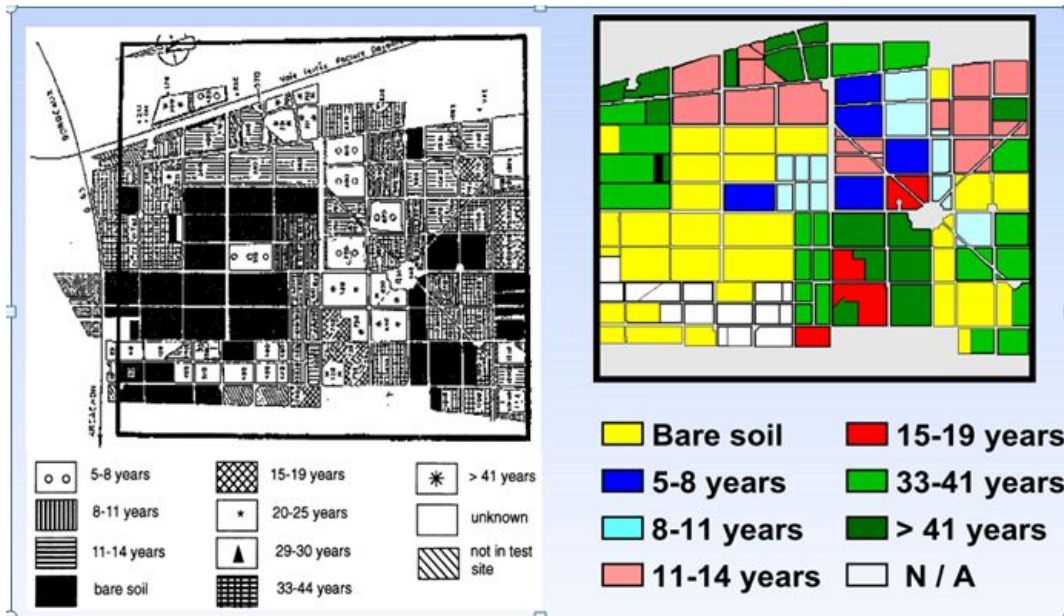
Figure 6.1: Ground truth maps of the Nezer forest.
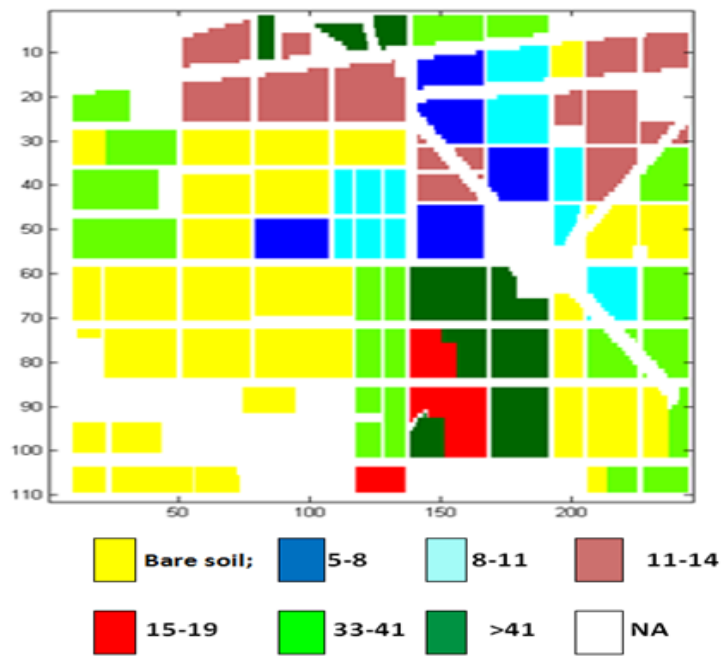
Figure 6.2: The prepared ground truth map with its color-coding.

6.2 Training and testing data points

In a supervised classification task, whenever separate training and testing datasets are not available, it is a usual practice to split the available data into two parts: one for training and the other for testing based on the ground truth map. These training and testing pairs are made not to overlap each other for cross-validation. Cross-validation is applied to get a more accurate estimate of the performance of the trained classifier by using independent training and testing pairs. However, the results from a single training and testing experiment might be misleading as the results are highly dependent on the choice of the training/testing split. Therefore, it is reasonable to select training datasets randomly a number of times (say N times) and run the experiment N times using the selected training and testing pairs independently. Then the classification accuracies will be computed as the average of the separate estimates. In this study, 150 training data points are selected from each of the classes considered for a single random selection experiment and the remaining data points are used for testing. All classification results are computed by averaging ten estimates of the classification accuracy from ten independent training and testing experiments. Figure 6.3 shows a pair of training and- testing data points for a single random selection experiment.
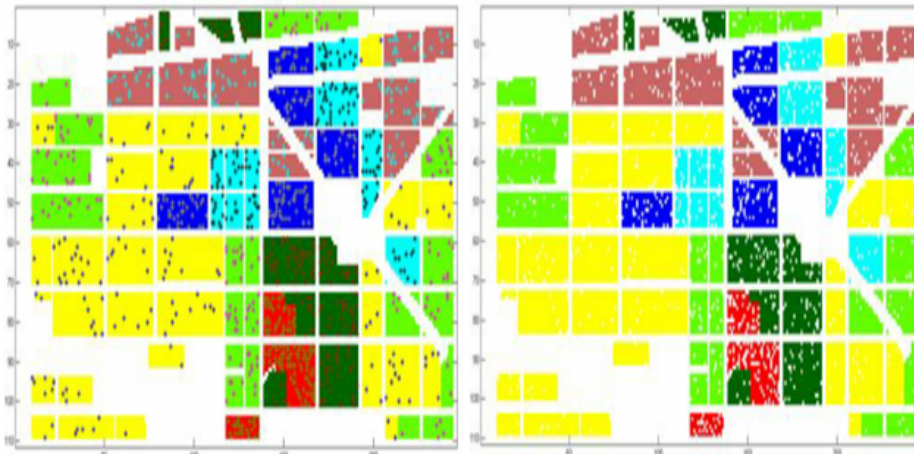
Figure 6.3: Training data points, shown as dotted (left) and testing data points (right).

6.3 Fusion of the datasets at feature level

It is pointed out in section 3.2.2 that when the available features are from data of different sensors, a suitable way of fusing them is to concatenate them into a single feature vector. Therefore, in this study, feature level fusion is achieved by concatenating the features from the four different datasets. To evaluate the discriminating capability of the different combinations of the four datasets, the features from each of them are first inserted into

separate feature vectors. The six PolSAR features from each of the P-, L- and C-bands are inserted into three separate feature vectors, F_P, F_L and F_C respectively. The same is done for the eight Landsat TM features where they are inserted into feature vector F_{TM}. Then, feature level multi-frequency PolSAR data fusion is demonstrated by concatenating subsets (or all) of $\{F_P, F_L, F_C\}$. For example, fusion of P- and L-band datasets is attained by concatenating feature vectors F_P and F_L. Fusion of the Landsat TM dataset with the PolSAR datasets is achieved the same way by concatenating F_{TM} with subsets (or all) of the three PolSAR feature vectors.

6.4   Classification results from the individual and fused datasets

In evaluating the classification performance of the individual datasets, the four feature vectors, F_P, F_L, F_C and F_{TM}, are separately applied as input to the classifier, whereas different subsets of $\{F_P, F_L, F_C, F_{TM}\}$ with two or more feature vectors are used in the case of the fused datasets. Classification accuracy (in percentage) is used to evaluate and compare the classification performances of the datasets and their combinations. For a single training-and-testing experiment, it is computed using the expression:

$$\%\textbf{Classif.accuracy} = \frac{\textbf{Number of correctly classif. data points}}{\textbf{Total number of testing data points}} \times \textbf{100}\%$$

The values reported as percentage classification accuracy throughout this study are computed by averaging ten estimates of the classification accuracies obtained from equation 6.4 for cross validation.

In this chapter, classification is applied for three different cases. First, the whole scene with all the seven classes is considered to evaluate the performance of each of the four datasets and their combinations in discriminating among the bare soil and the different tree age categories. Second, classification is applied to discriminate the bare soil from the forest. Finally, only the forested part is treated to discriminate among the six different tree age categories. This final approach will help to remove the effect of the dominant bare soil part in evaluating the capability of the datasets for tree age discrimination.

6.4.1   Discriminating among bare soil and the different forest ages; seven classes

In this case, classification is applied to discriminate among the available seven classes. The ground truth map used for the evaluation of the classification accuracy is the one shown in figure 6.2. The classification accuracies are reported in table 6.1.

Table 6.1: Classification results from all individual and combined datasets.

	The different feature vectors and their different combinations	Average % classification accuracies with their corresponding standard deviation values
1	SAR, P-band (six features)	67.89 ± 0.30
2	SAR, L-band (six features)	54.76± 0.44
3	SAR, C-band (six features)	43.97 ± 0.94
4	Vegetation indices (all eight features)	53.34 ± 0.47
5	Six TM bands (taken as six features)	52.07 ± 0.72
6	SAR, P-and L-band (12 features)	68.35 ± 0.48
7	SAR, P-and C-band (12 features)	68.67 ± 0.43
8	SAR, L-and C-band (12 features)	56.63 ± 0.48
9	SAR, P-, L-and C-band (18 features)	68.90 ± 0.46
10	SAR, P-band and eight VI (14 features)	75.34 ± 0.38
11	SAR, L-band and eight VI (14 features)	64.82 ± 0.47
12	SAR, C-band and eight VI (14 features)	58.19 ± 0.41
13	SAR, P-,L-bands and eight VI (20 features)	75.60 ± 0.39
14	SAR, P-,C-bands and eight VI (20 features)	75.91 ± 0.41
15	SAR, L-,C-bands and eight VI (20 features)	66.13 ± 0.53
16	SAR, P-,L-, C-bands and eight VI (26 features)	75.92 ± 0.27

6.4.1.1   Discussions

The individual classification performances of the four datasets are reported in the first four rows of table 6.1. As expected, P-band has the best classification performance of all the datasets, whereas C-bands has the poorest classification performance of all. L-band is the second best from the four datasets considered. The relatively higher performance of the P- and L bands as compared to that of the C-band is because of the ability of longer wavelength microwave bands to penetrate through the vegetation canopy and interact with the different structures of the trees in the forest and the ground. The performance of the Landsat TM dataset is slightly lower than the performance of the L-band, but far better than the performance of the C-band. Even though it is not a big improvement, it can be seen from the table that using the eight extracted Landsat TM features gives a slightly higher classification accuracy than directly using the six bands (excluding the thermal band). This indicates that in addition to enhancing the interpretability of the observed scene in terms of the physical parameters, feature extraction could also improve classification results.

The results of multi-frequency PolSAR data fusion are listed from the sixth to ninth rows of the table. The results show that no big improvement is achieved from any of the combinations. Both the fusion of P- and L-bands, and P- and C-bands give approximately equal results compared to the performance of the P-band alone. The performance of the L-band dataset is slightly improved (about 2%) by fusing it with C-band. Only about one

percent improvement is achieved by combining all the three P-, L- and C-band datasets (68.90%), compared to the performance of the P-band alone (67.89%).

In the case of combining data from the PolSAR and Landsat TM datasets, we can see many outstanding performances. ***It can be seen from the table that combining all the eight vegetation indices with any of the three PolSAR datasets (P-, L- or C-band) gives a much better performance than any of them used alone***. The classification accuracy of P-band is improved from 67.89% to 75.34%, the L-band performance is improved from 54.74% to 64.82% and that of the C-band jumped from 43.97% to 58.19% when fused with the Landsat TM dataset. The four best results around 75% are within two standard deviations of each other and are therefore effectively equivalent. As can be clearly seen from the table, this best classification performance involves the combination of Landsat TM and P-band datasets. Adding the other two (L- and C-band) datasets brings bout no significant improvement to the joint performance of Landsat TM and P-band datasets.

This big classification performance improvement upon the combined use of SAR and optical datasets is due to the fact that each of them captures different properties of the observed forest as well as bare soil. Optical sensors record the reflected energy where this energy is highly dependent on the chemical composition of leaves, the internal structure of leaves and the moisture content of the leaves and the bare soil. In the case of PolSAR sensors, the microwave scattering of single bounce, double bounce and multiple scattering is highly sensitive to the volume, structure and moisture content of the canopy, and surface roughness and moisture content of the bare soil. These different characteristics of the two sensors are the basis for their complementariness.

In [88], a slightly different study was done on the Nezer forest dataset and it was found that fusing all the three PolaSAR datasets (P-, L- and C-band) gives a relatively bigger performance improvement than our results. However, the fusion is applied directly on the multi-looked data without extracting any features. Moreover, a slightly different classification approach was used and fusion of SAR and optical datasets was not included in their study. Therefore, some further improvement may be gained by exploring different features and other classifiers for PolSAR data in the future.

Figure 6.4 compares the classification result achieved by all the 26 features with the ground truth map. It can be clearly seen that the bare soil is effectively discriminated from the forest group. However, there are some mixing in the different forest age categories. For example, there is much mixing between forest ages of 5-8 years and 8-11 years (dark blue and cyan) and between 33-41 years and >41 years (light green and dark green). This gives a clue that there is some challenge with respect to discriminating some of the different forest age categories.
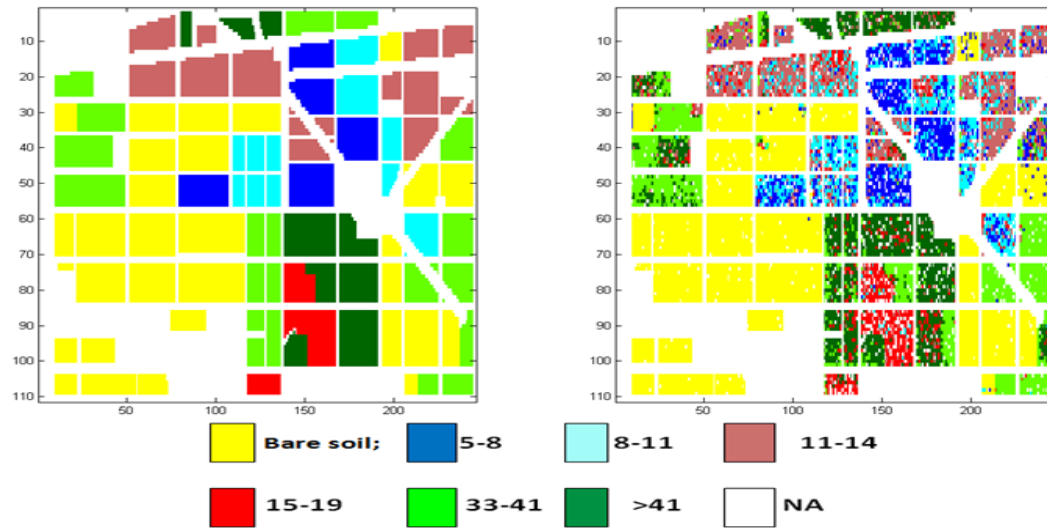
Figure 6.4: Ground truth (left) and classification result (right) using all 26 features.

6.4.2   Bare soil versus forest; two classes

Here, the performance of each of the four datasets and their different combinations is evaluated in discriminating between bare soil and forest as a whole. This is simply a two-class classification problem where the bare soil forms one class and the different forest age groups are merged together to form the second class. After merging all the different forest age groups into one, the ground truth map is shown in figure 6.5. Training data points are randomly chosen from each of the two classes and all the single and fused datasets are fed to the classifier the same way as explained previously. The classification results are reported in table 6.2.

Figure 6.5: Ground truth for the two-class case.

Table 6.2: Classification results from all individual and combined datasets in discriminating bare soil from forest.

	The different feature vectors and their different combinations	Average % classification accuracies with their corresponding standard deviation values
1	SAR, P-band (six features)	98.94 ± 0.11
2	SAR, L-band (six features)	98.85 ± 0.21
3	SAR, C-band (six features)	89.92 ± 0.36
4	Vegetation indices (all eight features)	91.59 ± 0.37
5	SAR, P-and L-band (12 features)	99.30 ± 0.15
6	SAR, P-and C-band (12 features)	99.30 ± 0.09
7	SAR, L-and C-band (12 features)	99.01 ± 0.21
8	SAR, P-, L-and C-band (18 features)	99.42 ± 0.11
9	SAR, P-band and eight VI (14 features)	99.33 ± 0.17
10	SAR, L-band and eight VI (14 features)	97.87 ± 0.73
11	SAR, C-band and eight VI (14 features)	93.61 ± 0.50
12	SAR, P-,L-bands and eight VI (20 features)	99.40 ± 0.16
13	SAR, P-,C-bands and eight VI (20 features)	99.42 ± 0.14
14	SAR, L-,C-bands and eight VI (20 features)	98.12 ± 0.67
15	SAR, P-,L-, C-bands and eight VI (26 features)	99.46 ± 0.14

67

6.4.2.1   Discussions

As can be seen from table 6.2, the P-band and the L-band PolSAR datasets perform better than the C-band and Landsat TM datasets. Generally, it can be seen from the table that most of the performances from the combined datasets (except the combination of C-band with the Landsat TM features) are approximately equal compared to the P-band or L-band performance alone. Therefore, data fusion may not be necessary in this case, as a single longer frequency PolSAR dataset (either P-or L-band), can discriminate the bare soil from the forest with a very high accuracy. As can be seen from the table, even the Landsat TM or the C-band dataset alone, or their combination can be used to discriminate the bare soil from the forest with an acceptable accuracy. This might be interesting for forest/non-forest mapping purposes whenever data at longer wavelengths is not available. Figure 6.6 compares the classification result achieved by all the 26 features with the ground truth map.

Figure 6.6: Ground truth (left) and classification result (right) using all 26 features in discriminating between bare soil and forest.

6.4.3   Discriminating among the different forest ages; six classes

In this case, the bare soil is masked out and the classification is applied only on the forested part to discriminate among the different forest age classes. The ground truth map is shown in figures 6.7 and the classification results are reported in table 6.3.
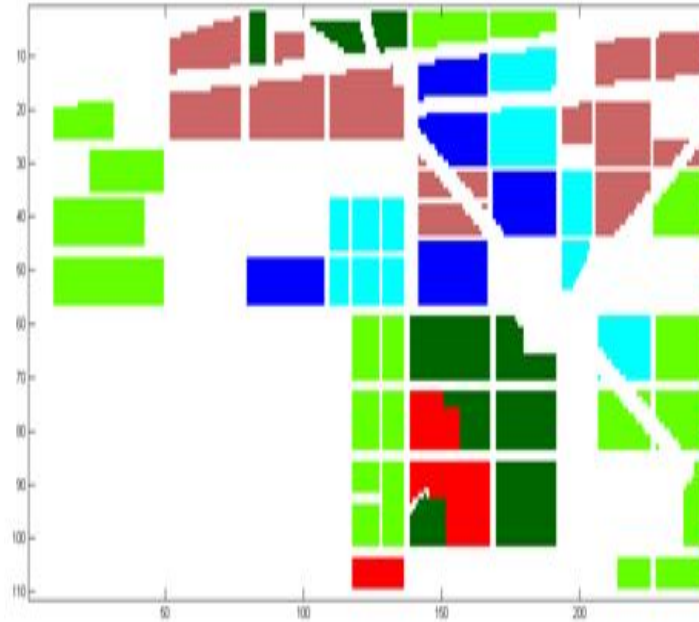
Figure 6.7: Ground truth for the different tree age categories.

Table 6.3: Classification results from all individual and combined datasets in discriminating the six tree age categories.

	The different feature vectors and their different combinations	Average % classification accuracies with their corresponding standard deviation values
1	SAR, P-band (six features)	49.12 ± 0.45
2	SAR, L-band (six features)	29.76 ± 0.64
3	SAR, C-band (six features)	20.70 ± 0.55
4	Vegetation indices (all eight features)	35.66 ± 1.21
5	SAR, P-and L-band (12 features)	49.79 ± 0.52
6	SAR, P-and C-band (12 features)	50.03 ± 0.47
7	SAR, L-and C-band (12 features)	32.33 ± 0.53
8	SAR, P-, L-and C-band (18 features)	50.54 ± 0.54
9	SAR, P-band and eight VI (14 features)	61.67 ± 0.89
10	SAR, L-band and eight VI (14 features)	45.58 ± 0.80
11	SAR, C-band and eight VI (14 features)	38.21 ± 1.21
12	SAR, P-,L-bands and eight VI (20 features)	62.44 ± 0.74
13	SAR, P-,C-bands and eight VI (20 features)	62.48 ± 0.82
14	SAR, L-,C-bands and eight VI (20 features)	47.28 ± 0.85
15	SAR, P-,L-, C-bands and eight VI (26 features)	63.08 ± 0.66

6.4.3.1 Discussions

Interesting results are obtained from this classification stage as well, and important points can be inferred from table 6.3. In general it can be seen from the table that lower classification accuracy values are obtained compared to the first experiment where all the seven classes are considered. This is due to the fact that discriminating the bare soil from the forest is done at a higher efficiency than that of discriminating among the different tree age categories, and this contributed to the relatively higher accuracy values in the first experiment. When it comes to the performance of the individual datasets, the P-band and the optical features perform better than the L- and C-bands. This is not far from the expected as the P-band has higher penetration capability into the forest, and the vegetation indices are intended to highlight different biophysical properties of plants. More importantly, it can be noted that all the performances of using the single datasets are below 50%, and this forces to look for other possibilities of improving the classification accuracy (as it is difficult to draw any information from such poor classification performance, $< 50\%$). It is obvious that one of the possibilities is a combined use of data from different sources. It can be seen from the table that the classification accuracies of all the three single frequency PolSAR datasets are increased by a significant amount when combined with the Landsat TM dataset. The classification accuracy of P-band is improved from 49.12% to 61.67%, the L-band performance is improved from 29.76% to 45.58% and that of the C-band improved from 20.70% to 38.21%. For tree age discrimination also, the best classification performance involves the P-band and the Landsat TM datasets, and adding the other two datasets does not bring about significant improvement.

Figure 6.8 compares the classification result achieved by all the 26 features with the ground truth map. Here also, it is clear to see from the figure that there is some challenge of the datasets to discriminate between the younger trees (5-8 and 8-11 years) and also between the older trees (33-41 and >41 years). Let us further investigate this issue using the information from the confusion matrix. For ease of comparison of the elements of the confusion matrix, classification is applied for an equal number of testing points after modifying the ground truth map by taking equal number of samples (636 in this case) from each of the classes. This issue of equal number of sampling will be further discussed in section 7.2 of chapter 7. 100 data points are used for training and the remaining 536 data points are used for testing. The resulting confusion matrix is shown in table 6.4. The numbers reported in the confusion matrix are computed by ratioing the number of correctly classified and misclassified data points in each class to the total number of testing points in each of the classes (536).

Table 6.4: Confusion matrix for tree age classification.

		Actual classes					
		5-8	8-11	11-14	15-19	33-41	>41
Predicted classes	5-8	**0.6828**	0.3022	0.0541	0	0.0056	0
	8-11	0.2817	**0.5093**	0.1791	0.0522	0.0055	0
	11-14	0.0317	0.1119	**0.5933**	0.0727	0.0410	0.0111
	15-19	0.0018	0.0727	0.1119	**0.8041**	0.0448	0.0970
	33-41	0.0018	0.0018	0.0466	0.0280	**0.5578**	0.1175
	>41	0	0.0018	0.0149	0.0429	0.3451	**0.7742**

From the confusion matrix, it can be clearly seen that about 30% of the trees from the group $8-11$ years are misclassified into the group $5-8$ years, and about 28% of the trees from group $5-8$ years are misclassified into the age group $8-11$ years. Also, about 18% and 11% of the trees from the group $11-14$ years are misclassified into the classes of tree age $8-11$ and $15-19$, respectively. 34.5% of the trees from the group $33-41$ years are misclassified into the group >41 years, however, only 11.75% of the trees are misclassified on the other way round. On the other hand, more trees from the groups $15-19$ and >41 years are correctly classified to their respective classes compared to the other four tree age categories.

From the above observations, we can see that there is more challenge in discriminating between tree age groups $5-8$ and $8-11$, and between groups $33-41$ and >41 years compared to the others. One potential reason for this could be the expected similarity in size and shape of trees around the same age, and this could be related to the phenological changes of pine trees in these specified age limits. Another potential reason could be the two year difference in the acquisition dates of the SAR and optical datasets. This is due to the fact that some of the trees from the group $5-8$, $8-11$ and $33-41$ will grow to $8-11$, $11-14$ and >41 years, respectively in this two years time. It can be seen from the confusion matrix that most of the trees from the group $15-19$ years are correctly classified. This supports both of the reasons explained as there is at least a one year gap between $11-14$ and $15-19$ and at least 14 years of gap between $15-19$ and $33-41$ years of tree age categories.

Another very interesting point to note from the first and this experiments is that when the Landsat TM dataset is combined with the L-band dataset, it achieves a comparable value to the performance of the P-band alone (slightly lower), and when the Landsat TM dataset is combined with the c-band dataset, a slightly bigger accuracy value is obtained compared to the performance of the L-band alone. It looks that the combined use of Landsat TM and L-band is equivalent to the use of P-band alone, and the combined use of Landsat TM and C-band is equivalent to the use of L band alone. Therefore, whenever data at a longer wavelength is not available, which is often the case, equivalent accuracy may be obtained by combining an optical dataset with the available shorter SAR band.
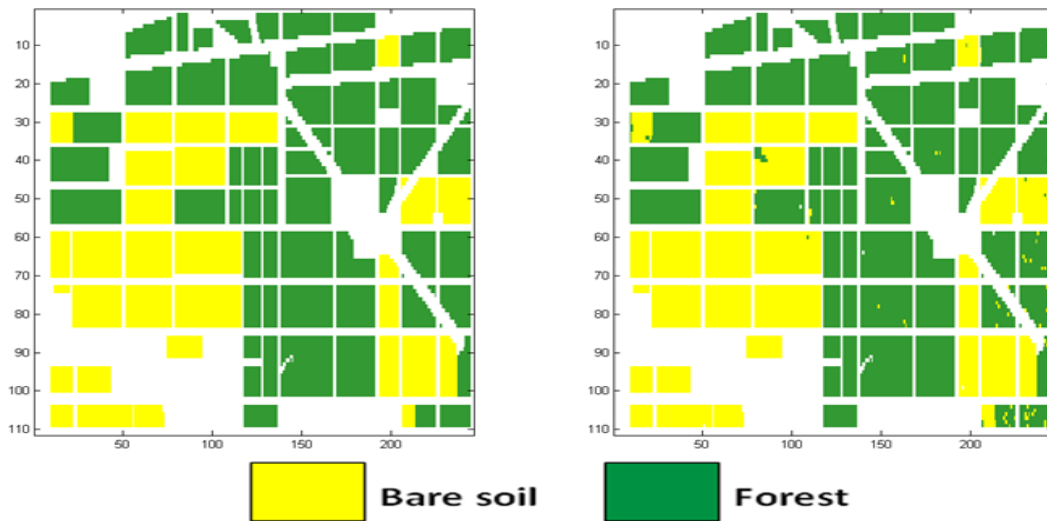
71
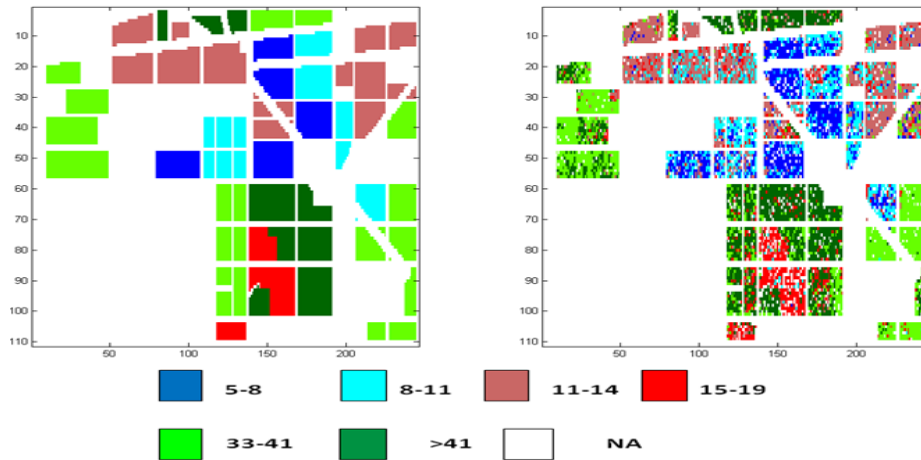
Figure 6.8: Ground truth (left) and classification result (right) using all 26 features in discriminating among the tree age categories.

6.4.4   Summary of the results

In summing-up, we have seen from the above three cases that when used alone, the P-band PolSAR dataset gives the highest classification performances of all the performance of the individual datasets. In addition, we found that all the best performances from the fused datasets involve the P-band dataset. Therefore, this leads to a conclusion that P-band is the best band for our applications. The L-band dataset is the next best from the PolSAR datasets. The C-band dataset is the poorest of all the datasets with respect to our application. With such poor performance, it can not be trusted for many forest applications. However, it can be used for limited forest applications such as deforestation mapping as it can identify forest from non-forest with reasonable accuracy. The classification accuracy significantly improved by using a combination of the polarimetric SAR data and the multi-spectral optical data. Therefore, whenever available, a combined use of SAR and optical data provides more accurate and reliable results than using any of them alone. Finally, it is important to mention that the above comparisons are done based on all the features of the respective datasets. No comparison is done among the features of the datasets. Maybe some of the individual features from the datasets of lower performances might perform better than some of the individual features from the best datasets. Moreover, the same level of accuracy might be achieved by using fewer best features. This is the concern of the next chapter.

Chapter 7

Experiment 2: Feature evaluation and selection

7.1   Introduction

Feature selection is a mechanism that we use to reduce the number of features with little or no loss in performance compared to the performance of the available features treated all together. In other words, it is a process of choosing and eliminating features with little or no predictive information. Therefore, in a classification task, the objective of feature selection is to select the most important features from the available set so as to reduce their number and at the same time retain their class discriminatory information as much as possible. One of the main reasons to reduce the number of features is to minimize the computational complexity. Another major reason of performing feature selection in a supervised classification task is to improve the generalization capability of the classifier [62]. This is the ability of a trained classifier to classify correctly data outside the training set. Feature selection could also be performed under the objective of improving the classification performance of the available features.

For two main reasons, feature selection is important for this study. The first one is that there are a number of features in our extracted feature set with little detail and contrast compared to others. As it is shown from the feature images and their respective histograms in chapter 5, some of the features within the same dataset have poorer contrast and less detail than others. A good example is that the P-band mean radar backscatter has a higher contrast than the remaining P-band features, see figure 5.1 of chapter 5. As it is discussed in chapter 5, there is also a difference in contrast and level of detail among features of the different datasets. As an example, the C-band features show poor contrast compared to the P- and L-band features, see figures 5.1 and 5.5. This indicates that some of the features can have better discriminating capability than others. This has also reflected in section 6.4 that when used together, the C-band features have poor classifi-

cation performances than the features in the remaining three datasets. Therefore, there may not be a significant benefit of using features with such poor predictive capability except adding complexity. The second main reason is that there exists high correlation among some of the extracted features. For example, some of the vegetation indices are highly correlated as it is shown in the scatter plots of figure 5.11. From the plots, it can be clearly seen that there exists a high correlation among NDVI, soil adjusted vegetation index and atmospherically resistant vegetation index. However, it is a difficult task to choose features by simply looking at their correlation in a scatter plot. This is because, with all of the twenty-six extracted features, we will have 325 possible combinations of two features in a scatter plot, a difficult task to manage. Therefore it is important to see other ways of systematically selecting best performing subset of features.

There are a number of approaches for feature selection, where many of them have arisen from the field of Statistics and Pattern recognition. One simple and straight forward approach for a feature selection problem is to evaluate the class discriminatory capability of each of the available features independently, rank them accordingly and select the first few best features. The major advantage of such methods is that they are computationally simple. However, they neglect to take into account the correlation that unavoidably exists among features and influences the classification capability of the selected subset. This is from the fact that even though it is possible to get good classification information by treating two correlated features separately, there might be little gain if they are combined into a feature vector. This will be elaborated more with examples in section 7.3.

Another approach is that the selection of the best subset of features is achieved by transforming the available feature space into a new space. When applied to the original data space, such as multi-spectral or hyperspectral data space, such an approach can also be used to extract features at the same time. In addition to reducing the size of the available data, such transformations have additional objectives such as removing redundant information by decorrelation or discriminating better in the new space. A celebrated example in this category is the Principal component analysis (PCA). In PCA, a reduced, new uncorrelated feature set is obtained by projecting the original space into the first few principal axes, where the directions of these principal axes are determined from an eigenvalue problem [62]. This approach could be one solution for the correlation problem pointed out above. However, beyond leaving the selected features uncorrelated, PCA does not guarantee features with higher class distinguishing ability, a big limitation when used for classification purposes. Another limitation of feature extraction and selection using PCA for remote sensing purposes is that it is no longer possible to interpret feature images with respect to the physical parameters of the observed scene in the transformed space.

In other feature selection approaches, the selection of a subset of best features is carried out directly in the feature space without doing any kind of transformation by evaluating different feature combinations to account correlation. The methods in this category are the most widely used feature selection approaches. In such approaches an optimal subset of features can be obtained by performing an exhaustive search in the space, i.e.,

evaluating all possible combinations of m out of K features to select m features. However, the number of possible subsets grows exponentially with the number of candidate features as it is required to evaluate $\frac{K!}{(K-m)!m!}$ different subsets. For example, in our case, if we decided to select five out of the available twenty-six features, we need to evaluate 65,780 different subsets and this number grows into 230,230 for six features. Thus, the method is impractical even for a moderate number of features. A more realistic approach relies on a greedy search method to traverse the space.

Two commonly used and simple approaches in this category are the *sequential forward* and *sequential backward* feature selection techniques, [62, 89]. Both are from the class of Hill-climbing (also known as stepwise selection and elimination) techniques, [89, 90], where the algorithm greedily searches for feature subsets by adding or removing features iteratively while moving towards a state which is better than the current one with respect a particular optimality rule. The sequential forward selection approach starts from nothing and in each step adds one new feature to the current selected subset whereas the backward method starts with all the available features and discards one at the time. Despite their simplicity, both the sequential forward and backward selection methods have some limitations. In the case of the backward method, once a feature is discarded, there is no possibility for it to be reconsidered again. The opposite is true for the forward procedure; there is no way of removing any of the selected features once it is included in the subset. This is called the nesting effect. Therefore, they are suboptimal by their nature. There are also other suboptimal and optimal (under different assumptions) feature selection methods, where some of them are variants of the methods discussed above while others are more advanced techniques. A more complete summary of generally used feature selection methods can be found in [62, 89].

In a classification task, different optimality criteria can be used to evaluate the performance of the different subsets. One possibility is using some kind of class separability measure which is independent of the classifier used. Another possibility is using the classification performance of the subsets in discriminating the available classes. Feature selection techniques which use the former approach are called *filter techniques*, while those used the later approach are called *wrapper techniques* [62, 89].

7.2   Method

In this study, both the standard sequential forward and backward feature selection approaches are applied and compared to select the best subset of features. Besides their simplicity, these two approaches examine the joint performance of the features by forming different subsets and this has two important advantages. One is that joint performance is what we are really interested in demonstrating the benefit of data fusion by combing features from the four different datasets. The other advantage is that since the performance of the combination of features is evaluated, correlation will automatically be accounted in the selection process. This is due to the fact that if a feature is highly correlated with

any of the features in a preselected subset, it is more likely that adding it to the subset will bring about no significant improvement to the overall performance of the subset except adding complexity. The criteria used to evaluate the performance of the different subsets in the selection process is the average classification accuracy achieved by each of them in discriminating the available classes. Therefore, both of them are implemented as a wrapper techniques.

As a preprocessing step for our feature selection process, all the 26 features from the four datasets are concatenated to form a single feature vector $\mathbf{X}$; $\mathbf{X} = \{F_P, F_L, F_C, F_{TM}\}$, where F_P, F_L, F_C and F_{TM} are the feature vectors from the P-band, L-band, C-band and Landsat TM datasets, respectively. Next a new ground truth map is created by taking an equal number of samples from each of the classes. Hence, each of the classes will be represented by an equal number of testing samples in the computation of the classification accuracy, and thus the variation in size of the available classes will no longer influence the feature selection process. The sample size is determined by the class with the smallest number of pixels. For this data, the class with tree age 15-19 years was the smallest with 636 samples. Therefore, the same number of pixels are randomly chosen from the remaining six classes. Figure 7.1 shows the newly created ground truth map with an equal number of samples.
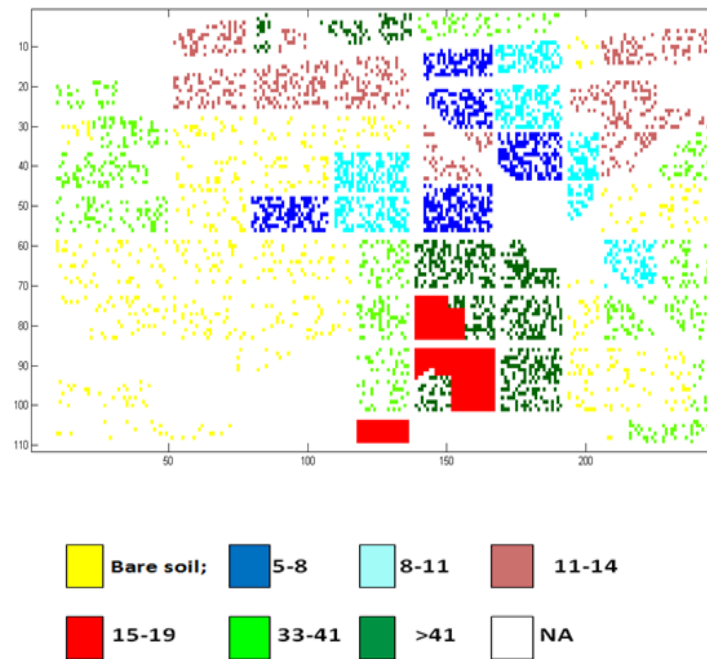
Figure 7.1: Ground truth map with equal number of samples.

The sequential forward feature selection (SFFS) algorithm implemented here starts from an empty set and adds one feature at a time until all the available features have been added. At each step, the algorithm iteratively evaluates new subsets and adds the feature

that, when added to the current subset, yields the highest classification accuracy. Then, the final subset that first achieves no less than 95% of the classification accuracy obtained by all the 26 features is selected.

Let F be the feature vector formed from the collection of best features added at each step, then our SFFS algorithm is:

1. Start with empty set $F_0 = \emptyset$; classification accuracy is zero

2. Calculate the classification accuracy for all combinations of F_k and each remaining feature R, i.e., $F_k + R$

3. Choose the feature, R_{max} that corresponds to the highest joint classification accuracy and update the feature selection set to $F_{k+1} = F_k + R_{max}$

4. Repeat for all remaining features

5. Select the final number of features that preserves 95% of the classification accuracy obtained by all the 26 features

The sequential backward feature selection (SBFS) algorithm works in the opposite direction of SFFS. It starts with all the 26 features and sequentially removes one feature at a time that least reduces the classification accuracy of the subset. This time F will be the feature vector formed from the remaining best features after removing one "bad" feature at each step. The SBFS algorithm is:

1. Start with the set with all the 26 features $F_0 = \mathbf{X}$

2. Calculate the classification accuracy of all possible subsets, $F_k - R$ by removing one feature (R) at a time

3. Find the subset with the highest classification accuracy. This subset is $F_{k+1} = F_k - R_{min}$, where R_{min} is the feature that reduces least the joint classification accuracy.

4. Continue until the last feature is removed

5. Select the final number of features that preserves 95% of the classification accuracy obtained by all the 26 features

In both approaches, it is possible to halt the algorithm as soon as the predefined requirement (95% of the classification accuracy obtained by all the 26 features) is achieved (at step 4) without evaluating any more candidate subsets by traversing to the other end of the search space. Here, the evaluation is done until the last feature is added/removed for two reasons. The first one is that it might be interesting to see the performance of not only the best few features but all the available 26 features. The second reason is that it might also be interesting to check whether the two feature selection approaches considered here give exactly identical results all the way up to the last feature or not.

7.3   Feature selection results for the discrimination of all the seven classes

The results of both the SFFS and SBFS for all the twenty-six features in discriminating all the available classes is presented in tables 7.2 and 7.3 respectively. The labelling of all the 26 features is given in table 7.1 for ease of reference.

Table 7.1: Feature labels.

Features names	Labels	Features names	Labels
P-band Mean radar backscatter	μ_P	C-band Non-Gaussianity measure	ν_C
P-band Non-Gaussianity measure	ν_P	C-band Cross-pol ratio	f_{xC}
P-band Cross-pol ratio	f_{xP}	C-band Co-pol ratio	f_{cC}
P-band Co-pol ratio	f_{cP}	C-band Correlation magnitude	f_{mC}
P-band Correlation magnitude	f_{mP}	C-band Correlation phase	f_{pC}
P-band Correlation phase	f_{pP}	Normalized difference vegetation index	NDVI
L-band Mean radar backscatter	μ_L	Soil brightness index	B
L-band Non-Gaussianity measure	ν_L	Greenness vegetation index	G
L-band Cross-pol ratio	f_{xL}	Wetness	W
L-band Co-pol ratio	f_{cL}	Perpendicular vegetation index	PVI
L-band Correlation magnitude	f_{mL}	Triangular vegetation index	TVI
L-band Correlation phase	f_{pL}	Soil adjusted vegetation index	SAVI
C-band Mean radar backscatter	μ_C	Atmospherically resistant vegetation index	ARVI

Table 7.2: Forward feature selection results.

Subset by adding one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
0	0	0	$+ f_{mC}$	69.79±0.64	1.017
$+ \boldsymbol{\mu_P}$	**53.94 ± 0.41**	**0.786**	$+ \nu_L$	69.80±0.70	1.017
$+ \mathbf{B}$	**59.91±0.44**	**0.873**	$+ \nu_C$	69.78±0.61	1.017
$+ \mathbf{NDVI}$	**64.09±0.46**	**0.934**	$+ \nu_P$	69.65±0.81	1.015
$+ \mathbf{f_{xP}}$	**66.68±0.56**	**0.972**	$+$ PVI	69.59±0.75	1.014
$+$ W	67.55±0.67	0.985	$+ f_{cL}$	69.33±0.67	1.011
$+ f_{cC}$	68.22±0.65	0.994	$+ \mu_L$	69.24±0.80	1.009
$+ f_{cP}$	68.70±0.79	1.001	$+ f_{pL}$	69.16±0.92	1.008
$+$ TVI	68.79±0.70	1.003	$+ f_{xC}$	69.14±0.69	1.008
$+$ SAVI	69.35±0.60	1.011	$+ f_{xL}$	69.06±0.69	1.007
$+ \mu_C$	69.57±0.63	1.014	$+$ ARVI	68.97±0.75	1.005
$+ f_{mL}$	69.74±0.61	1.016	$+ f_{pC}$	68.81±0.78	1.003
$+$ G	69.79±0.64	1.017	$+ f_{pP}$	68.59±0.80	100
$+ f_{mP}$	69.87±0.64	1.019			

Table 7.3: Backward feature selection results.

Subset by removing one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
All (26)	68.59±0.80	1.000	- f_{mP}	69.79±0.64	1.017
- PVI	68.94±0.82	1.005	- G	69.74±0.62	1.017
- ν_C	69.09±0.91	1.007	- f_{mL}	69.57±0.63	1.014
- f_{pP}	69.23±0.99	1.009	- μ_C	69.35±0.60	1.011
- f_{xC}	69.29±0.99	1.010	- f_{cP}	68.81±0.61	1.003
- f_{pC}	69.37±0.99	1.011	- SAVI	68.33±0.56	0.996
- ARVI	69.46±0.93	1.013	- TVI	68.21±0.64	0.994
- f_{xL}	69.56±1.08	1.014	- f_{cC}	67.54±0.67	0.985
- f_{pL}	69.62±0.99	1.015	- **W**	**66.68±0.56**	**0.972**
- ν_L	69.73±0.87	1.016	- **f_{xP}**	**64.09±0.46**	**0.934**
- μ_L	69.73±0.74	1.016	- **NDVI**	**59.91±0.44**	**0.873**
- ν_P	69.76±0.75	1.017	- **B**	**53.94 ± 0.41**	**0.786**
- f_{cL}	69.79±0.64	1.017	- **μ_P**	0	0
- f_{mC}	69.87±0.64	1.019			

As can be seen from tables 7.2 and 7.3, all the greedily searched subsets, by adding/removing a single feature at a time, are listed in the first and fourth columns. The corresponding average classification accuracies are reported in the second and fifth columns of the tables. These average classification accuracies are also reported as a fraction of the total classification accuracy achieved by all the twenty-six features, which is 68.59%.

In the case of the SFFS approach, when the first best feature, which is the P-band mean radar backscatter, is added to the empty set, the classification accuracy has jumped from 0 to 53.94% (table 7.2). The step-wise increase and decrease of the classification accuracies for both approaches are plotted in figure 7.2 for ease of visualization. Then, at each step, the next best feature with respect to improving the joint classification accuracy is added. As can be clearly seen from table 7.2, the joint performance has continued increasing until the P-band correlation magnitude (f_{mP}) is added. Then it starts to decrease until the last feature (f_{pP}) is added. For the decreasing trend, the algorithm still finds the feature that gives the maximum classification accuracy, but this corresponds to the least drop in accuracy. ***Here an important point can be noted that classification performance can actually be improved by systematically selecting features***. We can see from the table that the subset from μ_P to f_{mP} has achieved bigger accuracy (69.87%) compared to the accuracy of the total twenty-six features (68.59%). Referring to the third column of the same table, no less than 95% of the classification accuracy can be retained by choosing a subset containing the first four best features. Therefore, these are the selected features using the SFFS algorithm which jointly perform best in discriminating among the seven classes. The features are listed in table 7.4.

In the SBFS approach the opposite is done by removing the worst (least important) features one at a time starting from all until the last feature. As can be seen from table 7.3 and the right panel of figure 7.2, removing some of the "bad" features results in a slight improvement to the classification accuracy. In this approach also, we can see that 95% of the classification accuracy can be retained by keeping the last four features. These features are the same features that are selected using the SFFS approach (table 7.4).
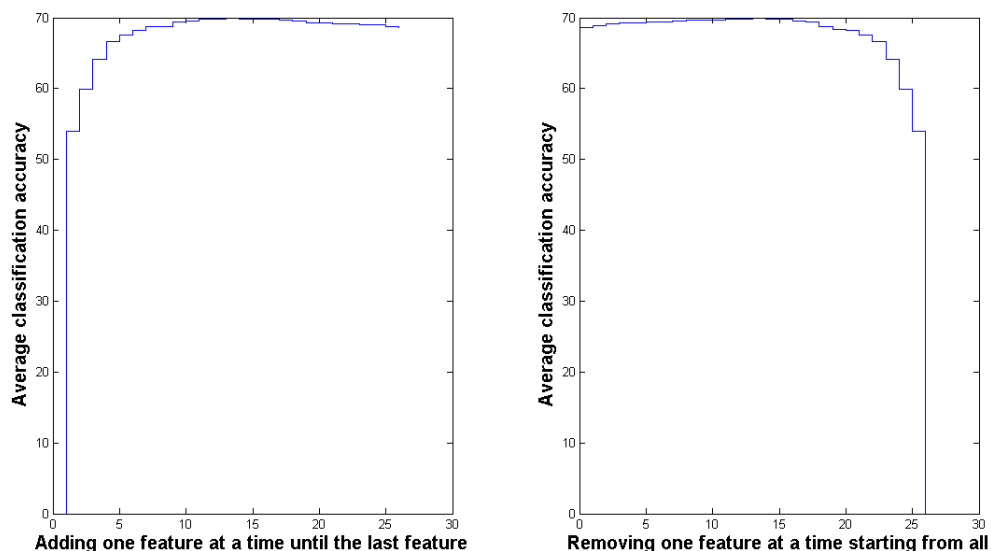
Figure 7.2: Step-wise increase and decrease of the classification accuracies, SFFS (Left) and SBFS (Right).

Table 7.4: Features selected using SFFS and SBFS for seven class discrimination.

	Selected features
1	P-band Mean radar backscatter (μ_P)
2	Soil brightness index (B)
3	Normalized difference vegetation index (NDVI)
4	P-band Cross-pol ratio (f_{xP})

Comparing tables 7.2 and 7.3, it can be clearly seen that the two feature selection algorithms give similar results in selecting only the first six best features. However, for more number of features (more than six), the results from these two algorithms will no longer be the same. This is due to the fact that SBFS could be fully influenced by the interactions among features in the selection process while these interactions are ignored in SFFS. Even though the ordering and the features are not the same, the average clas-

sification accuracy is approximately equal in selecting identical number of features using these two approaches.

The classification result of the four selected features in discriminating the seven classes, together with the ground truth map for comparison, is shown in figure 7.3. These best four features jointly achieve 74.15% of classification accuracy when applied for the *whole scene*. Comparing to the classification accuracy of all the 26 features of table 6.1 of chapter 6, which was at 75.92%, the accuracy of the four features accounts about 97.67%.
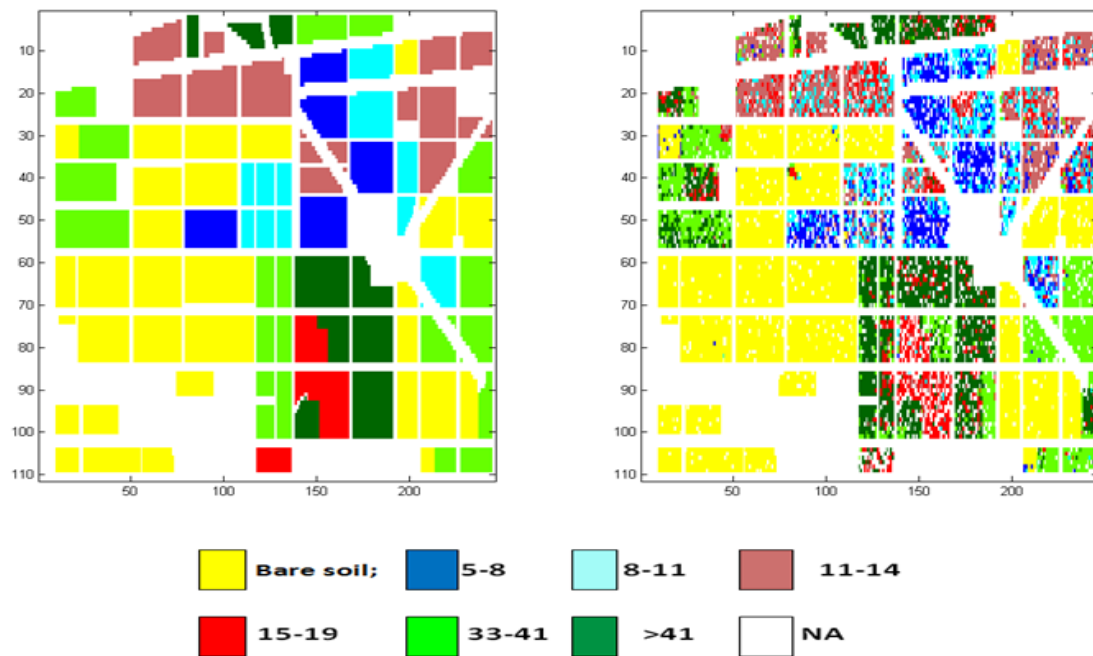
Figure 7.3: Ground truth (left) and seven class classification result (right) from the selected four features (right) and ground truth map (left).

To easily visualize the joint performance of two features, the classification result is also illustrated in color-coded scatter plots with possible pair-wise combinations of the selected four features. This is shown in figure 7.4, and will be discussed in section 7.3.1.
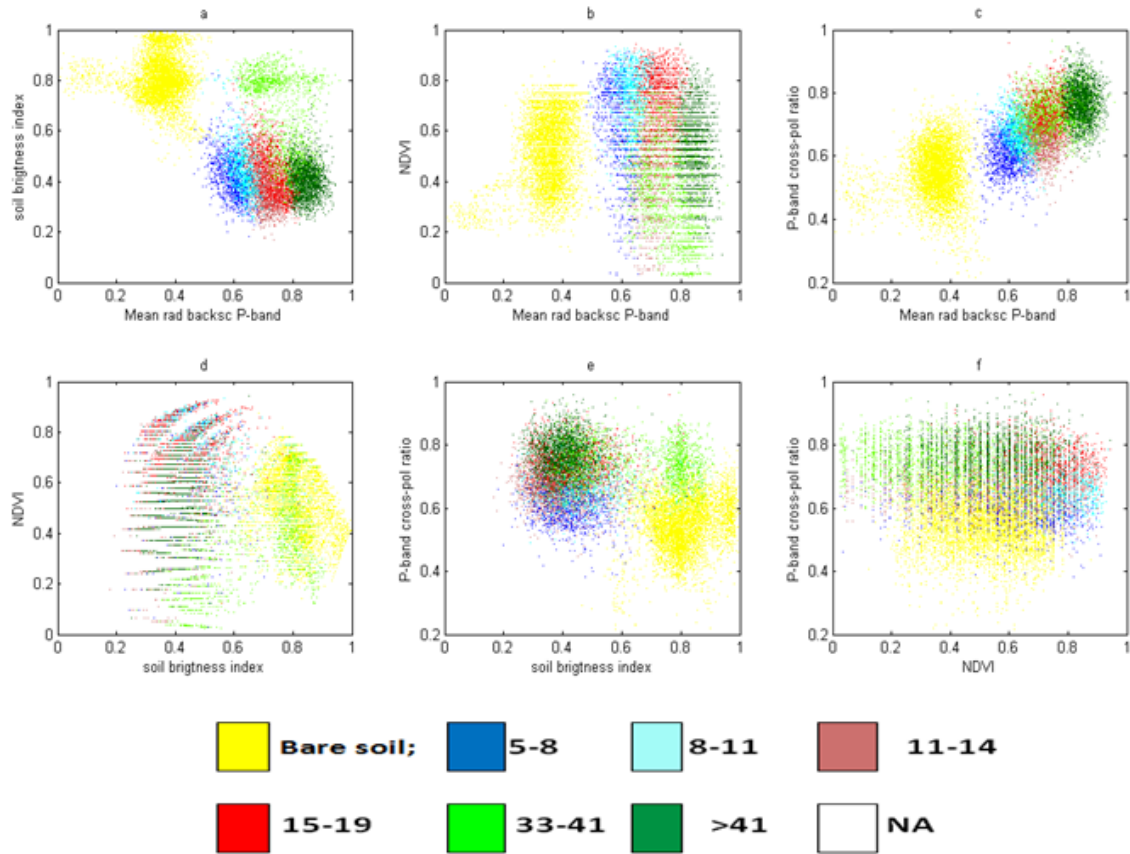
Figure 7.4: Color-coded scatter plots of possible pair-wise combinations of the four selected features.

7.3.1 Discussions

In the discussions to follow, it should be kept in mind that the feature selection is done based on evaluating the joint performance, and not on the basis of individual performances. Therefore, other features might be better than some of the selected features in terms of individual performances, but we do not really care about the individual performances as our objective is to choose a subset with features that complement each other best. As it is pointed out earlier, this also helps to automatically account for feature dependence in the selection process. To strengthen this idea, all the twenty-six features are ranked based on their individual performances and the result is shown in table 7.5. The same training and testing pairs used for the selection process are also used here for the ranking purpose, just for a fair comparison.

Table 7.5: Individual rankings of all the twenty-six features.

Ranking	Features	Ranking	Features
1	P-band mean radar backscatter (μ_P)	14	Soil adjusted vegetation index (SAVI)
2	L-band mean radar backscatter (μ_L)	15	Normalized difference vegetation index (NDVI)
3	P-band Cross-pol ratio (f_{xP})	16	Triangular vegetation index (TVI)
4	Wetness (W)	17	C-band Cross-pol ratio (f_{xC})
5	P-band Correlation magnitude (f_{mP})	18	C-band Correlation magnitude (f_{mC})
6	Soil brightness index (B)	19	P-band Correlation phase (f_{pP})
7	C-band Mean radar backscatter (μ_C)	20	C-band Co-pol ratio (f_{cC})
8	L-band Cross-pol ratio (f_{xL})	21	L-band correlation phase (f_{pL})
9	L-band Correlation magnitude (f_{mL})	22	L-band Non-Gaussianity measure (ν_L)
10	P-band Co-pol ratio (f_{cP})	23	L-band Co-pol ratio (f_{cL})
11	Perpendicular vegetation index (PVI)	24	P-band Non-Gaussianity measure (ν_P)
12	Atmospherically resistant vegetation index (ARVI)	25	C-band Non-Gaussianity measure (ν_C)
13	Greenness vegetation index (G)	26	C-band Correlation phase (f_{pC})

It can be clearly seen from the table that the soil brightness index, the second best feature in the selected subset, is ranked at the sixth position. The NDVI, the third best in the selected subset is ranked at the fifteenth position and the P-band cross-pol ratio is ranked at the third position. Therefore, when we say that the selected features are "best", it is based on their joint performances. Now, if we have to choose four features based on the individual rankings, the first four features of table 7.5 are our candidates, and their joint classification accuracy is about 61.32%, which is smaller than the one achieved by our selected subset (66.68%). Therefore, features with relatively higher individual performances does not always mean that they have higher collective performance, and this is primarily due to mutual correlation. This explains the reason why feature selection results based on the individual performances are far from optimal.

Four out of the twenty-six features are found particularly useful to discriminate among the seven classes (table 7.4). *Two of the selected features, namely P-band mean radar backscatter and P-band cross-pol ratio are from the P-band PolSAR dataset.* This is an expected result because the P-band dataset is potentially the best for forest applications as it is discussed in chapter 6. *The remaining two features, namely the soil brightness index and the normalized difference vegetation index are from the Landsat TM dataset.*

As it can be clearly seen from the top row scatter plots of figure 7.4, the P-band mean radar backscatter shows a remarkably good distinguishing capability among the different classes. It can effectively discriminate the bare soil from the forest. Except for some tree age groups, it also shows relatively higher distinguishing ability among the different tree age categories. It is pointed out in section 5.1 that the mean radar backscatter represents the intensity information from the observed target. This information is dependent on the magnitude of the backscatter signal in radar systems, which in turn is highly dependent on the surface roughness, the dielectric properties and the moisture content of the observed target. Even though the scattering mechanisms are highly dependent on the surface roughness, single bounce is the dominant type of scattering from the bare soil part, while volume scattering is dominant from the forest. Because of these different mechanisms, it turns out that the bare soil is generally darker than forest in radar images. Therefore, these different scattering mechanisms in the two different regions could potentially be the basis for the higher performance of the mean radar backscatter in discriminating the bare soil from the forest. The magnitude of the backscatter is also expected to vary across the different tree age categories. This is because, as the age of the tree increases, there will be much more leaves, branches and twigs which results in an increase of the backscatter with tree age. These variations across the different tree age categories could possibly be the potential reasons for the relatively higher performance of the mean radar backscatter in tree age discrimination.

The next two features from our selected subset that are found useful in complementing the P-band mean radar backscatter are the soil brightness index and the normalized difference vegetation index respectively. This is an expected result for the reason that SAR and optical data are highly complementary to each other, as it is briefly discussed in chapters 2 and 6. Referring to the scatter plots of figure 7.4, even though it is not as effective as the P-band mean radar backscatter, the soil brightness index can discriminate the bare soil from the forest with relatively higher accuracy. This is due to the fact that it is more sensitive to the soil part than the forest. From the scatter plot of P-band mean radar backscatter and soil brightness index (a), it can be seen that the soil brightness index can discriminate the light green cluster (part of the class with 33-41 years of old trees) whereas the P-band mean radar backscatter cannot. Therefore it is selected as it provides this important information that the P-band mean radar backscatter lacks. The usefulness of the NDVI arises from the unique inverse relation relationship between the RED and NIR reflectance of vegetation (section 5.2). This unique information, which is obtained from optical data only, when it is combined with SAR data, it has a potential of adding value to the combined performance.

The last feature in the selected subset is the cross-pol ratio. Its relatively good performance is related to its higher sensitivity to vegetation biomass. As the age of the tree increases, more depolarization will be resulted from the many structures of trees; leaves, branches and twigs. This in turn means an increase in the cross-pol return with tree age, and this variation is the basis for its relatively good performance compared to the other four PolSAR features. Generally it has low values in the bare soil part than that of the

forest (figure 5.1), and this information could be useful to discriminate the bare soil from the forest too.

7.4   Feature selection results for forest age discrimination only

As it has been discussed in chapter 6 and can be clearly seen from the scatter plots of 7.4, discriminating the bare soil from the forest is more effective than that of discriminating among the forest age categories. Therefore, the feature selection experiment using both SFFS and SBFS algorithms is repeated here by masking out the bare soil part to check whether we could end up with the same set of features or not. The ground truth map after masking out the bare soil part is shown in figure 7.5 for an equal number of samples from each of the six classes.
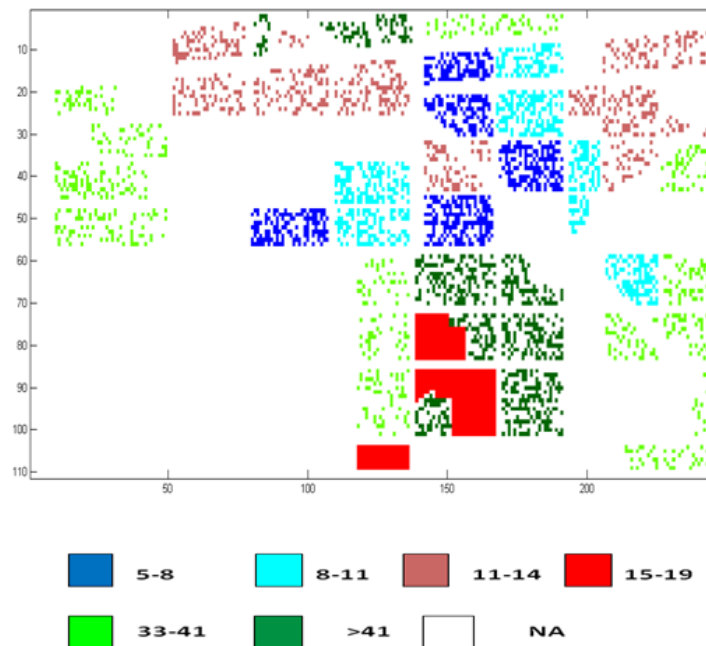
Figure 7.5: Ground truth map with equal number of samples for tree age categories.

All the subsets generated using SFFS and SBFS approaches together with their corresponding classification accuracies are presented in tables 7.6 and 7.7, respectively.

Table 7.6: Forward feature selection results.

Subset by adding one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
0	0	0	$+f_{mL}$	66.30±0.74	1.012
$+\boldsymbol{\mu}_P$	**46.50 ± 0.54**	**0.710**	$+f_{mP}$	66.23±0.87	1.011
$+\mathbf{B}$	**53.62±0.42**	**0.819**	$+$PVI	66.28±0.83	1.012
$+\mathbf{NDVI}$	**58.52±0.69**	**0.893**	$+f_{pC}$	66.28±0.83	1.012
$+\mathbf{f_{xP}}$	**61.79±0.60**	**0.943**	$+\nu_C$	66.30±0.72	1.012
$+\mathbf{W}$	**63.69±0.72**	**0.972**	$+f_{pP}$	66.25±0.63	1.011
$+\mu_C$	64.45±0.70	0.984	$+f_{xL}$	66.14±0.70	1.010
$+f_{cC}$	64.74±0.73	0.988	$+f_{mC}$	66.11±0.76	1.009
$+f_{cP}$	65.09±0.62	0.994	$+\nu_L$	65.03±0.82	1.008
$+$TVI	65.33±0.52	0.997	$+f_{xC}$	65.93±0.73	1.007
$+$SAVI	65.72±0.41	1.003	$+$ARVI	65.74±0.82	1.004
$+ f_{cL}$	65.91±0.43	1.006	$+f_{pL}$	65.57±0.83	1.001
$+ \mu_L$	66.01±0.52	1.008	$+\nu_P$	65.49±0.78	1.000
$+$G	66.12±0.56	1.009			

Table 7.7: Backward feature selection results.

Subset by removing one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
All	65.49±0.78	1.000	$- f_{xL}$	66.03±0.55	1.008
$- f_{pC}$	65.57 ± 0.83	1.001	$- \mu_L$	65.89±64	1.006
$- \nu_L$	65.66±0.88	1.002	$-$ G	65.72±41	1.003
$-$ ARVI	65.79±0.0.85	1.004	$- f_{cP}$	65.40±0.64	0.998
$- f_{xC}$	65.94±0.83	1.007	$- f_{cC}$	64.91±0.68	0.991
$- f_{pL}$	65.99±0.80	1.008	$-$ SAVI	64.46±0.71	0.984
$- f_{mC}$	66.13±0.83	1.010	$-$ TVI	64.45±0.70	0.984
$- \nu_P$	66.20±0.87	1.011	$- \mu_C$	**63.69±0.72**	**0.972**
$- \nu_C$	66.30±0.83	1.012	$-$ **W**	**61.79±0.60**	**0.943**
$- f_{mL}$	66.30±0.66	1.012	$- \mathbf{f_{xP}}$	**58.52±0.68**	**0.893**
$- f_{pP}$	66.23±0.77	1.011	$-$ **NDVI**	**53.62±0.42**	**0.819**
$- f_{cL}$	66.07±0.75	1.009	$-$ **B**	**46.50±0.54**	**0.710**
$- f_{mP}$	66.24±0.76	1.011	$- \boldsymbol{\mu}_P$	0	0
$-$ PVI	66.16±0.74	1.010			

From tables 7.6 and 7.7, we can see that 95% of the classification accuracy can be retained by choosing a subset of five features. These five best features are identical in both the SFFS and SBFS approaches and are listed in table 7.8. The selected five best features are the same features ranked in the same order compared to the first five features of the first experiment where all the seven classes treated. However, for tree age discrimination, the first four best features are incapable of retaining 95% of the classification accuracy of

the total as they did in the first experiment. An additional fifth best feature, which is the wetness, is required to do so. Therefore one more feature is found useful in retaining most of the classification information and discriminate among the tree age categories.

Table 7.8: Features selected using SFFS and SBFS for tree age discrimination.

	Selected features
1	P-band Mean radar backscatter (μ_P)
2	Soil brightness index (B)
3	Normalized difference vegetation index (NDVI)
4	P-band Cross-pol ratio (f_{xP})
5	Wetness (W)

The classification result of the five selected features in discriminating the tree age categories, together with the ground truth map for comparison, is shown in figure 7.6. These best five features jointly achieve 60.35% of classification accuracy when applied for the *whole scene*. Comparing to the classification accuracy of all the 26 features of table 6.3 of chapter 6, which was at 63.08%, the accuracy of the five features accounts about 95.68%.
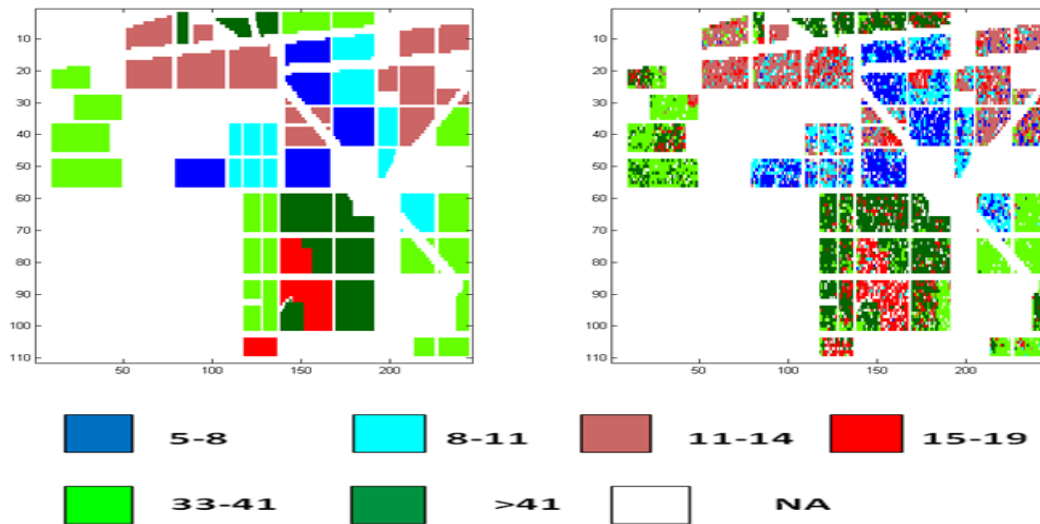
Figure 7.6: Ground truth (left) and tree age classification result (right) from the selected five features (right) and ground truth map (left).

As four of the selected features for tree age discrimination are the same as those selected for the seven class case of the first experiment, the discussions given in section 7.3.1 with respect to discriminating among the tree age categories are also valid here. Moisture content difference among the different tree age categories is the potential reason for the

relatively higher performance of the wetness. As the age of the tree increases, it is natural that more green biomass will be present in a given area or volume, and hence the moisture content will increase with tree age.

7.5   Feature selection for soil versus forest only

In this case also, both the SFFS and SBFS algorithms are applied to all the twenty-six features to select features which are best in discriminating the bare soil from the forest. The ground truth map after merging all the tree age categories is shown in figure 7.7 for an equal number of samples from each of the two classes.
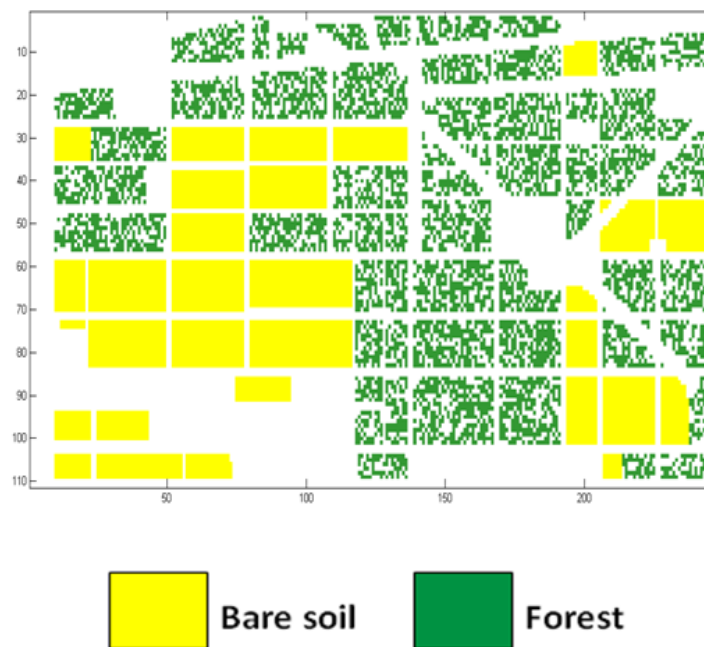
Figure 7.7: Ground truth map with equal number of samples for bare soil versus forest.

All the greedily searched subsets using the SFFS and SBFS algorithms are listed in tables 7.9 and 7.10 respectively.

Table 7.9: Forward feature selection results.

Subset by adding one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
0	0	0	$+ \nu_L$	99.573±0.070	1.0005
$+\boldsymbol{\mu_P}$	**99.047 ± 0.132**	0.9952	$+ \nu_C$	99.561±0.069	1.0004
+W	99.535±0.044	1.0001	+ TVI	99.557±0.061	1.0003
$+\mu_L$	99.582±0.028	1.0006	$+ \mu_C$	99.571±0.052	1.0005
$+f_{pP}$	99.584±0.026	1.0006	$+ f_{mL}$	99.582±0.053	1.0006
+ARVI	99.589±0.045	1.0007	$+ f_{mP}$	99.588±0.061	1.0006
$+f_{xC}$	99.609±0.057	1.0009	+ NDVI	99.589±0.064	1.0007
$+f_{pL}$	99.601±0.059	1.0008	+ PVI	99.596±0.050	1.0007
$+f_{pC}$	99.599±0.061	1.0008	+ G	99.589±0.047	1.0007
$+f_{mC}$	99.601±0.056	1.0008	$+ f_{xL}$	99.588±0.054	1.0006
$+ f_{cL}$	99.595±0.045	1.0007	$+ f_{xP}$	99.577±0.064	1.0005
+ SAVI	99.595±0.049	1.0007	$+ f_{cP}$	99.577±0.064	1.0005
$+ \nu_P$	99.590±0.054	1.0007	+ B	99.524±0.103	1.0000
$+ f_{cC}$	99.581±0.058	1.0006			

Table 7.10: Backward feature selection results.

Subset by removing one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
All	99.524 ± 0.103	1.0000	$- f_{pC}$	99.621±0.034	1.0010
- B	99.557 ± 0.064	1.0003	- PVI	99.621±0.030	1.0010
$- f_{cP}$	99.557±0.064	1.0003	$- f_{xP}$	99.622±0.029	1.0010
$- \nu_L$	99.589±0.056	1.0007	$- f_{mC}$	99.616±0.024	1.0009
- G	99.600±0.048	1.0008	$- \mu_L$	99.614±0.024	1.0009
$- f_{xC}$	99.605±0.043	1.0008	- SAVI	99.605±0.024	1.0008
$- \nu_C$	99.606±0.035	1.0008	$- f_{pP}$	99.587±0.034	1.0006
- ARVI	99.613±0.039	1.0009	$- f_{mP}$	99.567±0.039	1.0004
$- \nu_P$	99.614±0.040	1.0009	- TVI	99.549±0.033	1.0003
$- f_{pL}$	99.613±0.039	1.0009	$- \mu_C$	99.557±0.024	1.0003
$- f_{cL}$	99.618±0.041	1.0009	- NDVI	**99.535±0.044**	1.0001
$- f_{xL}$	99.619±0.023	1.0010	- W	**99.047±0.132**	0.9952
$- f_{pC}$	99.627±0.019	1.0010	$- \boldsymbol{\mu_P}$	0	
$- f_{mL}$	99.623±0.034	1.0010			

From the two tables above we can see that only the first best feature is more than enough to achieve the required 95% accuracy. The P-band mean radar backscatter is the best feature from both SFFS and SBFS approaches which achieves an accuracy of 99.047%, and is hence selected. This is about 99.5% of the total accuracy achieved by all the 26 features, which is 99.524%. Interestingly, the wetness is the next best feature identified which best complements the P-band mean radar backscatter. Together, they achieved

the same classification accuracy (99.535%) as compared to the accuracy of all (99.524%). Therefore, the P-band mean radar backscatter and the wetness can effectively replace all the twenty-six features in discriminating the bare soil from the forest. The classification result using the selected feature (μ_P) in discriminating the bare soil from the forest, together with the ground truth map for comparison, is shown in figure 7.8. Even though it is not included in the selected subset, the wetness is also the fifth best feature for the seven class case of the first experiment. Its relatively good performance in distinguishing the bare soil from the forest is from the expected moisture content difference between them. This was also reflected in the wetness image of figure 5.9 of chapter 5 that the bare soil appeared brighter than the forest.
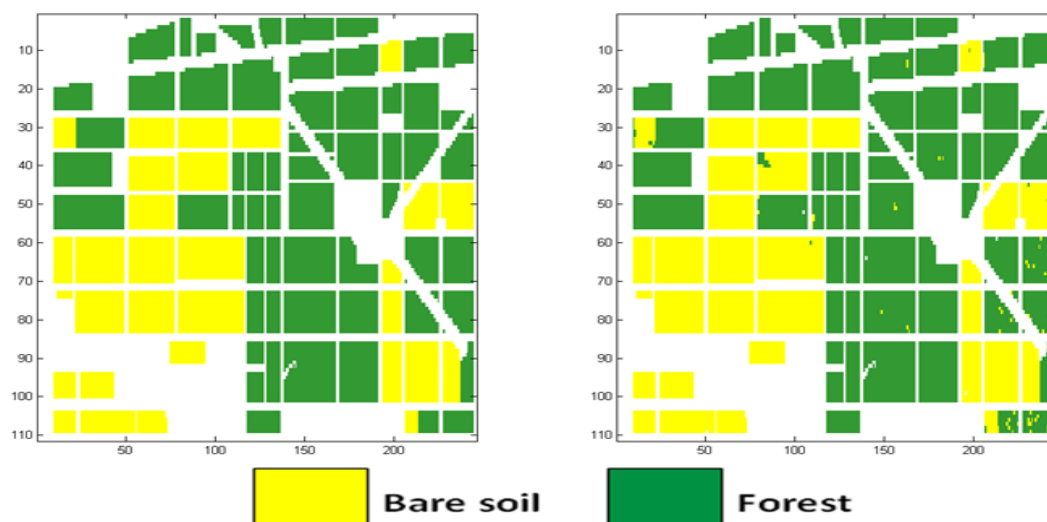
Figure 7.8: Ground truth (left) and classification result (right) using the P-band.

7.6   Combining the three feature selection results

From the above three feature selection experiments, five best features are identified in retaining most of the classification information for our application. These five features are those listed in table 7.8. Four of them are identified in the first experiment, and one more feature, the wetness, is found useful for tree age discrimination. Even though the wetness is not among the selected features in the first experiment, it was the fifth best feature for the seven class case, see tables 7.2 and 7.3. Together with the P-band mean radar backscatter, the wetness is also found to replace all the twenty-six features to discriminate the bare soil from the forest. Therefore, these five features are our final selection for the discrimination of all the available classes. The classification result of these five selected features in discriminating the seven classes, together with the ground truth map for comparison, is shown in figure 7.9. These best five features jointly achieve 74.78% of classification accuracy when applied for the *whole scene*. Comparing to the

classification accuracy of all the 26 features of table 6.1 of chapter 6, which was at 75.92%, the accuracy of the five features accounts about 98.50%. ***This is a big achievement with respect to reducing the number of features and at the same time retaining most of the classification information***.
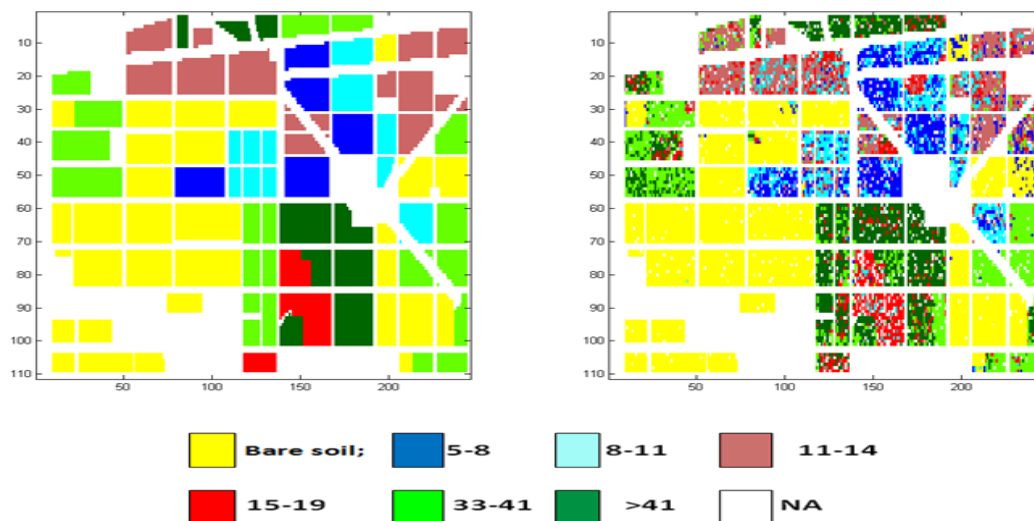
Figure 7.9: Ground truth map (left) and seven class classification result (right) from the selected five features.

7.7 Features selection for the combination of single frequency PolSAR and optical dataset

From our five selected features, three of them are from the Landsat TM dataset, and two of them are from the P-band PolSAR dataset. Therefore, no features are selected either from the L- or C-band dataset. As it is pointed out in chapter 2, remote sensing from space is the most cost effective and efficient way of monitoring the earth. However, we do not have P-band from space and therefore its usage has been limited to experimental purposes. Currently, L-band is also unavailable from space. C-band is better in terms of availability compared to the other two. Therefore, it is reasonable and might be interesting to see whether the selected features will remain the same across the three frequencies considered here or not. This might help to identify features from both the PolSAR and Landsat TM datasets, which jointly perform best for forest applications irrespective of the SAR frequency used. Therefore, feature selection is applied by combining each of the P-, L- and C- band features at a time with the optical ones.

7.7.1 Feature selection results for P-band and Landsat TM features

Both the SFFS and SBFS algorithms are applied to all the fourteen features from the P-band and Landsat TM datasets. All the subsets generated from the SFFS and SBFS algorithms together with their corresponding classification accuracies are presented in tables 7.11 and 7.12 respectively. The results from the two approaches are compared in table 7.13.

Table 7.11: Forward feature selection results.

Subset by adding one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
0	0	0	+ SAVI	68.70±0.61	1.0111
+$\boldsymbol{\mu}_P$	**53.94 ± 0.41**	**0.7938**	+ f_{mP}	68.67±0.63	1.0106
+**B**	**59.91±0.44**	**0.8817**	+ f_{pP}	68.67±0.52	1.0106
+**NDVI**	**64.09±0.46**	**0.9432**	+ G	68.61±0.66	1.0097
+**f$_{xP}$**	**66.68±0.55**	**0.9813**	+ν_P	68.48±0.69	1.0078
+W	67.54±0.67	0.9939	+ ARVI	68.22±0.74	1.0040
+ f_{cP}	68.15±0.77	1.0029	+PVI	67.95±0.71	1.0000
+ TVI	68.28±0.58	1.0048			

Table 7.12: Backward feature selection results.

Subset by removing one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
All (14)	67.95±0.71	1.0000	- TVI	68.16±0.76	1.0031
- PVI	68.22±0.74	1.0040	- f_{cP}	67.55±0.67	0.9941
- ARVI	68.48±0.69	1.0078	- W	**66.68±0.56**	**0.9814**
- ν_P	68.61±0.66	1.0097	- **f$_{xP}$**	**64.09±0.46**	**00.9432**
- G	68.67±0.52	1.0106	-SAVI	**59.91±0.44**	**0.8817**
- f_{pP}	68.67±0.63	1.0106	- B	**53.94 ± 0.41**	**0.7938**
- f_{mP}	68.70±0.61	1.0111	- $\boldsymbol{\mu}_P$	0	0
- NDVI	68.28±0.58	1.0049			

Table 7.13: Comparing SFFS and SBFS results for P-band and Landsat TM features.

Ordering	Forward (SFFS)	Backward (SBFS)
1	P-band mean radar backscatter	P-band mean radar backscatter
2	Soil brightness index	Soil brightness index
3	NDVI	Soil adjusted vegetation index
4	P-band cross-pol ratio	P-band cross-pol ratio
Total accuracy	66.68±0.55(0.9813)	66.68±0.56(0.9814)

7.7.2 Feature selection results for L-band and Landsat TM features

Both the SFFS and SBFS algorithms are applied to all the fourteen features from the L-band and Landsat TM datasets. All the subsets generated using SFFS and SBFS approaches together with their corresponding classification accuracies are presented in tables 7.14 and 7.15 respectively. The results from the two approaches are compared in table 7.16.

Table 7.14: Forward feature selection results.

Subset by adding one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
0	0	0	+ f_{mL}	52.37±0.62	1.0148
+$\boldsymbol{\mu}_L$	**41.11 ± 0.59**	**0.7533**	+ f_{pL}	52.29±0.76	1.0133
+**G**	**47.09±0.68**	**0.9126**	+ ν_L	52.18±0.86	1.0112
+**W**	**49.99±0.63**	**0.9686**	+ NDVI	51.80±0.74	1.0037
+B	50.98±1.11	0.9879	+ SAVI	52.28±0.75	1.0130
+f_{xL}	51.27±1.16	0.9934	+ f_{cL}	51.95±0.76	1.0068
+TVI	51.45±1.09	0.9969	+PVI	51.61±0.85	1.0000
+ ARVI	52.06±0.72	1.0088			

Table 7.15: Backward feature selection results.

Subset by removing one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
All (14)	51.61±0.85	1.0000	- TVI	51.40±0.86	0.9959
- PVI	51.95±0.76	1.0068	- NDVI	51.28±1.08	0.9936
- f_{cL}	52.28±0.75	1.0130	- f_{xL}	50.48±0.95	0.9781
- f_{pL}	52.46±0.79	1.0165	- W	**49.80±0.75**	**0.9650**
- ARVI	52.20±0.91	1.0115	- SAVI	**45.98±0.66**	**0.8910**
- ν_L	52.21±0.64	1.0117	- B	**41.11 ± 0.59**	**0.7533**
- G	54.90±0.78	1.0079	- $\boldsymbol{\mu}_L$	0	0
- f_{mL}	51.76±0.77	1.0030			

Table 7.16: Comparing SFFS and SBFS results for L-band and Landsat TM features.

Ordering	**Forward (SFFS)**	**Backward (SBFS)**
1	L-band mean radar backscatter	L-band mean radar backscatter
2	Greenness	Soil brightness index
3	Wetness	Soil adjusted vegetation index
Total accuracy	49.99±0.63(0.9686)	49.80±0.75(0.96504)

7.7.3   Feature selection results for C-band and Landsat TM features

Both the SFFS and SBFS algorithms are applied to all the fourteen features from the C-band and Landsat TM datasets. All the greedily searched subsets using the SFFS and SBFS algorithms together with their corresponding classification accuracies are presented in tables 7.17 and 7.18 respectively. The results from the two approaches are compared in table 7.19.

Table 7.17: Forward feature selection results.

Subset by adding one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
0	0	0	$+f_{mC}$	44.41±0.41	0.9839
+W	**30.17 ± 0.19**	**0.6684**	+ARVI	44.73±0.42	0.9909
+G	**36.93±0.68**	**0.8182**	$+\nu_C$	45.06±0.41	0.9982
+$\boldsymbol{\mu}_C$	**41.54±0.48**	**0.9203**	$+f_{pC}$	45.09±0.41	0.9989
+$\mathbf{f_{xC}}$	**42.08±0.49**	**0.9323**	+ NDVI	44.95±0.48	0.9958
+$\mathbf{f_{cC}}$	**42.68±0.52**	**0.9455**	+ SAVI	45.55±0.45	1.0090
+B	**43.41±0.59**	**0.96161**	+PVI	45.14±0.61	1.0000
+TVI	43.93±0.64	0.9732			

Table 7.18: Backward feature selection results.

Subset by removing one at a time	Average classification accuracy	Fraction of the total	Subset by adding one at a time	Average classification accuracy	Fraction of the total
All (14)	45.14±0.61	1.000	- f_{cC}	43.99±0.49	0.9746
- PVI	45.55±0.45	1.0091	- f_{xC}	**43.30±0.61**	**0.9592**
- f_{pC}	55.54±0.49	1.0089	**- TVI**	**41.84±0.33**	**0.9268**
- ARVI	45.46±0.42	1.0072	**- SAVI**	**40.79±0.74**	**0.9037**
- ν_C	45.32±0.48	1.0040	**- NDVI**	**36.35±0.58**	**0.8054**
- B	44.95±0.64	0.9958	**- $\boldsymbol{\mu}_C$**	**30.17±0.19**	**0.6684**
- G	44.91±0.51	0.9950	**- W**	0	0
- f_{mC}	44.78±0.40	0.9921			

Table 7.19: Comparing SFFS and SBFS results for C-band and Landsat TM features.

Ordering	Forward (SFFS)	Backward (SBFS)
1	Wetness	Wetness
2	Greenness	C-band mean radar backscatter
3	C-band mean radar backscatter	NDVI
4	C-band cross-pol ratio	Soil adjusted vegetation index
5	C-band co-pol ratio	Triangular vegetation index
6	Soil brightness index	
Total accuracy	43.41±0.59(0.9616)	43.30±0.61(0.95.92)

7.7.4   Discussions

From the above three cases, it can be clearly seen from their respective results that each of the SFFS and SBFS algorithms give different subsets of features. Therefore it is a good time to critically investigate these two approaches to be able to make the decision easier.

For the first case, where the selection is applied on the P-band and Landsat TM features, we can see from table 7.13 that four features are suggested by both the SFFS and SBFS approaches in preserving 95% of the classification information. However, the SAVI is chosen by the SBFS algorithm instead of the NDVI, which is included in the subset proposed by the SFFS approach. Except for this difference, the remaining features are identical with the same order and the same classification accuracy value. It looks that the NDVI and SAVI can be replaced each other without affecting the classification accuracy. This is from the fact that these two features are highly correlated. This high correlation is depicted in the color-coded scatter plots (i) of figure 7.10. Therefore it can be said that the two approaches are still giving the same result in choosing the few best features and we can take either of the subsets proposed by the two approaches as our final choice.
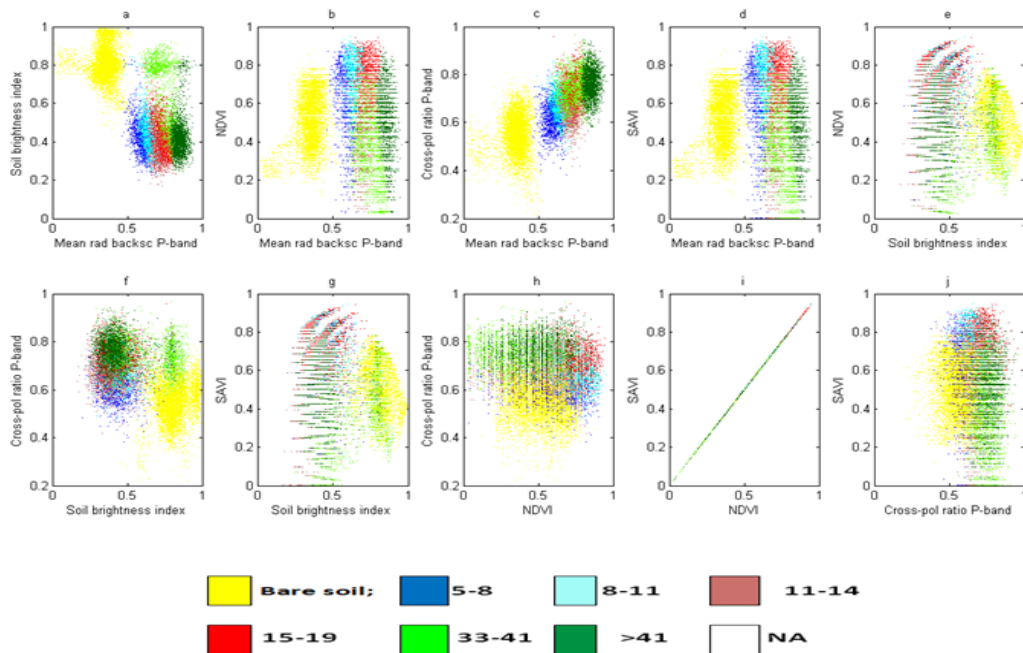
Figure 7.10: Color-coded scatter plots of possible pair-wise combinations of five features.

For the second case where the selection is applied on the L-band and Landsat TM features, we can see from table 7.16 that the SFFS and SBFS methods give different results (except μ_L). Even though the features identified by the two approaches are different, the average classification accuracy is approximately equal (very small difference, in favour of the forward) in preserving 95% of the classification information. The scatter plots of possible pair-wise combinations of the five features of table 7.16 are plotted in figure 7.11 for ease of visualization.
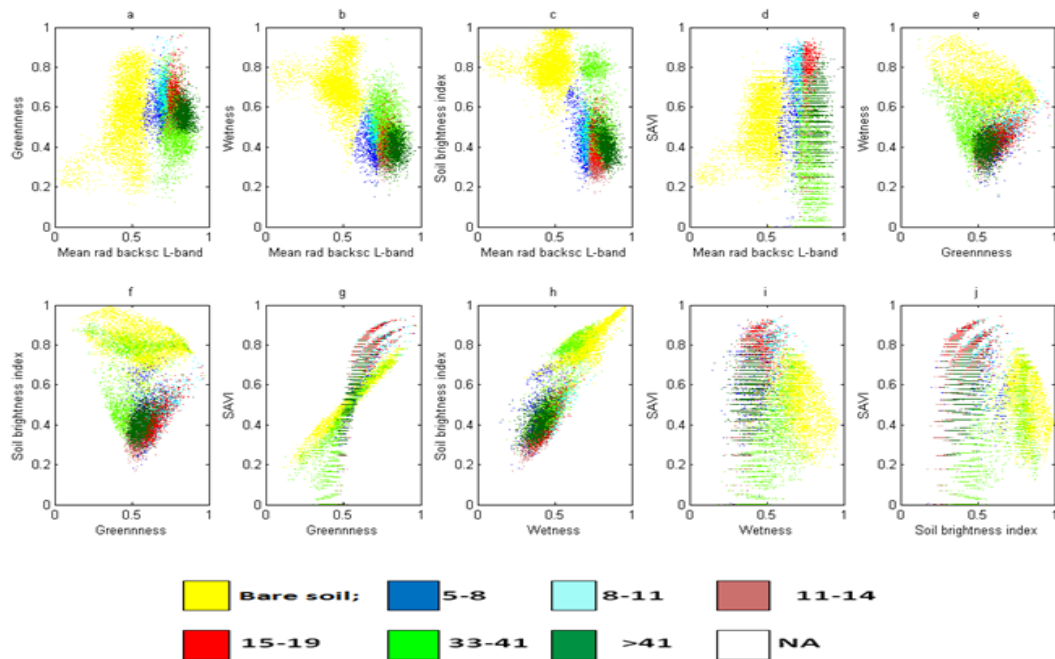
Figure 7.11: Color-coded scatter plots of possible pair-wise combinations of the five features.

As can be seen from the above scatter plot (a), the L-band mean radar backscatter has a deficiency of effectively discriminating some of the tree age categories. Especially, it has a difficulty in isolating the light green cluster $(33 - 41)$ from the dark green (>41), and the red $(15 - 19)$ from the purple (11-14) ones. Therefore, in the feature selection process, it seeks for best complementary features that help in solving this problem. The first feature suggested by the SFFS algorithm in solving this problem is the greenness while the soil brightness index is the one suggested by the SBFS algorithm. From tables 7.14 and 7.15, the greenness adds more value to the accuracy of μ_L than the soil brightness index did. Therefore, in this case, the forward is better in selecting two features than the backward approach. The relatively higher performance of the greenness has also reflected in the scatter plots. It can be clearly seen that the size of the light green cluster that can be distinguished from the remaining tree age groups (a) using the greenness is bigger than that of the top light green cluster in (c) that can be distinguished using the soil brightness index. It is also reasonable for the greenness to have relatively good performance with respect to distinguishing the tree age categories than the soil brightness index as it is intended to highlight the green biomass. As the age of the tree increases so does the amount of green biomass, and therefore the value of the greenness is expected to increase with tree age.

From its nature, the forward approach has the potential of evaluating a large number of states near the empty set, however towards the full set only small number of subsets

will be evaluated since most features have already been selected. Therefore, the SFFS approach is better in selecting relatively smaller number of features (near the empty set). On the other hand, the backward approach has more potential of evaluating a large number of states near the full set than near the empty set. Therefore, the SBFS approach is better in selecting relatively large number of features (near the full set). This means that it may have relatively lower performance in selecting very few features compared to the SFFS algorithm. The opposite is true for the SFFS approach in selecting large number of features. The limitation of the SBFS approach in selecting a few best features is due to the fact that it is evaluating different subsets starting from the back by removing one at a time. There is a possibility of some of the individually "good" features to be thrown away at early stages in favour of a stronger "mutual benefit" combination, and which are themselves removed later on. As it can be clearly seen from table 7.15, this is what exactly happened to the greenness in our SBFS experiment. The greenness is removed at an early stage while it can perform better than any of the remaining features when it is combined with the first best feature (μ_L). As it is discussed in the introduction section of this chapter, both the forward and backward approaches are suffering from this "nesting" limitation of being unable to reconsider previous decisions. However, it will not be a big problem for the forward approach in selecting the first few best features for the reason that the algorithm is able to make better decisions near the empty set by evaluating many different combinations of the available features. Comparing tables 7.14 and 7.15, we can see that slightly better or approximately equal accuracy values are achieved by the SFFS algorithm in selecting the first few features compared to the SBFS algorithm. Therefore, from the above discussions and from our objective, selecting only few features which jointly perform best, the SFFS approach is better for our application.

Also in the last case where the selection is applied on the C-band and Landsat TM features, we can see from table 7.19 that the SFFS and SBFS methods give different results. In the case of SFFS approach, six features are identified in preserving 95% of the classification information whereas five features are identified in the SBFS approach. However, it can be seen from tables 7.17 and 7.18 that slightly higher accuracy values are obtained in the SFFS than the SBFS in choosing smaller number of features (up to four features). This agrees with the above explanation that the SFFS approach is better in identifying those few best features than the SBFS approach. Here, unlike the first two cases, two of the optical features, wetness and greenness ranked on top of the mean radar backscatter. This is due to the fact that, as it is found in chapter 6, the C-band dataset generally has poorer performance than the optical one. However, the mean radar backscatter is still the best SAR feature in complementing with the optical features.

Empirical comparisons between these two sequential feature selection algorithms can be found in [91, 92]. In [91], it is reported that the SFFS approach is better when the optimal number of selected features is smaller while the backward is better if otherwise. This agrees with our results. In [92], an initial investigation was done to compare these two features selection algorithms. Their preliminary results show that the performance

of the two approaches is dependent on the number of the available features. It is reported that the forward performs better than the backward when the number of the available features is small (e.g., 10) where as the backward is better for 30 or more features. As it is mentioned in these two references, results from a related work in [93] showed that the backward approach often outperforms the forward one. However, our results show that, irrespective of the number of the available features used, the two approaches sometimes give identical results in selecting the few best features and sometimes not. When there is a difference in the outputs from these two algorithms, the SFFS approach gives slightly better classification accuracy values than the SBFS approach.

Since we have now found that the SFFS approach fits to our objective, the features that are identified using this approach are taken as our final selection results with respect to the above three cases. The features selected are listed in tables 7.20, 7.21 and 7.22 for each of the three cases.

Table 7.20: The selected features for P-band and Landsat TM.

	Selected features
1	P-band mean radar backscatter
2	Soil brightness index
3	NDVI
4	P-band cross-pol ratio

Table 7.21: The selected features for L-band and Landsat TM.

	Selected features
1	L-band Mean radar backscatter
2	Greenness vegetation index
3	Wetness

Table 7.22: The selected features for C-band and Landsat TM.

	Selected features
1	Wetness
2	Greenness
3	C-band mean radar backscatter
4	C-band cross-pol ratio
5	C-band co-pol ratio
6	Soil brightness index

Comparing the above three results, we can see that the mean radar backscatter is the first feature selected in the P-and L-band cases, and also it has got a third place in the C-band case. Therefore, it is the best SAR feature irrespective of the SAR frequency

used. In the P- and C-band cases, the cross-pol ratio is the forth best feature among the selected features. Even though it is not included in selected subset, it can be clearly seen from table 7.14 that it is the next best PolSAR feature in the L-band case too. Therefore, it is the best PolSAR feature next to the mean radar backscatter. The reasons for the relatively higher performance of these two PolSAR features has already explained in section 7.3.1.

In general, four optical features, namely the wetness, the soil brightness index, the greenness and the normalized difference vegetation index are identified. The wetness is the first best feature identified for the C-band case and also the third best features for the L-band case. Even though it is not included in the selected subset, it can be clearly seen from table 7.11 that it is the fifth best feature in the ranking for the P-band case too. Therefore, it is one of the best optical features found useful for our application. The soil brightness index is one of the selected features for the P-and C-band cases. It is also the fourth best feature for the L-band case (table 7.14). Therefore, it is the other best optical feature found useful in complementing the PolSAR features. The NDVI is identified as one of the best complementary features for the P-band case only. As it can be seen from tables 7.14 and 7.17, it is ranked in the tails for the L- and C-band cases. On the other hand the greenness is found particularly useful for the L- and C- band cases than P-band. The reasons for the relatively higher performance of the soil brightness index and the NDVI is explained in section 7.3.1 where as the explanation for the wetness is given in sections 7.4 and 7.5. One more point to add about the soil brightness index is that, in addition to highlighting and discriminating the bare soil part from the forest, it is observed from the scatter plots of figures 7.4, 7.10 and 7.11 that this feature has a potential of discriminating among some of the tree age categories. One potential reason for this is that the pixels are fairly large and are able to measure a mixture of tree and ground. The proportion of each will change as the trees grow and spread branches with age, and this might cause a change in the value of the soil brightness index with tree age. The greenness has already been discussed at the beginning of this section. Finally, for the C-band and Landsat TM case, a sixth feature, namely the co-pol ratio is included in the selected subset in preserving 95% of the classification accuracy. However, this feature is not consistently appearing in the best subset for all the other cases, and it looks that it is included for this particular case to fulfil the 95% classification accuracy requirement.

7.8 Chapter summary

In summary, two feature selection approaches namely the sequential forward and backward feature selection approaches are applied and compared all the way up to the end of the chapter. In some cases, these two approaches give identical results in selecting the few best features. However, in other cases they give different results even in selecting the few best features. When they give different results, the forward approach provides small

improvements to the classification accuracies of the first few best features compared to the backward approach. As this is somewhat advantageous with respect to identifying the first few best features, only the features selection results from the forward approach are considered.

From the results above, we have seen that systematic feature selection is an important process in identifying and selecting the few best features. In all the experiments, six or fewer features are identified that jointly perform best to preserve most of the classification information. We have also seen that when systematically applied, feature selection can improve the classification performance.

The mean radar backscatter is the best feature from all the PolSAR features. The cross-pol ratio is the next best feature found useful from the PolSAR dataset. Two optical features, namely the wetness and the soil brightness index are found best in complementing the selected PolSAR features, irrespective of the SAR frequency used. However, the normalized difference vegetation index is found specifically useful in complementing the P-band features. On the other hand the greenness is found particularly useful in complementing the L-and C-band PolSAR features.

Chapter 8

Conclusion and future work

8.1 Conclusion

In this study, the benefit of SAR and optical data fusion for forest applications is investigated by extracting and selecting different polarimetric SAR and multi-spectral optical features.

The work started by choosing a convenient data fusion level that can retain the sensor specific information from each of the datasets considered. It is found that feature level fusion is a convenient fusion method to retain the sensor specific information and to evaluate and select the best combination of features from different data sources.

In the preprocessing stage, the datasets are co-aligned and different radiometric corrections are applied. In general, it is of interest to note that image registration is a fundamental step prior to fusing datasets from several sources. In particular, since SAR and optical sensors have different imaging geometries and resolutions, image registration is a vital step in combining datasets from these two sensors.

In this study, a total of twenty-six features are extracted; six from each of the three PolSAR datasets and eight from the optical one. Of all the twenty-six features, some of them show appreciable contrast among the different classes and are capable of forming separable clusters and some of them do not. Several of them are also found useful to easily interpret the scene visually and in terms of its biophysical parameters. Moreover, high correlation is found among some of the features, particularly among some of the vegetation indices. Therefore, at the feature extraction stage, a clue has been found that most of the class discriminatory information might be embedded in few of the features.

A supervised maximum likelihood Bayesian classification scheme is applied to evaluate and compare the classification performance of each of the four datasets and their different combinations. The classification accuracy (in percentage) is used to quantitatively eval-

uate the performances. From the PolSAR datasets, it is found that P-band has the best whereas C-band has the poorest classification performances with respect to our application. L-band has performed moderately. The Landsat TM dataset has slightly lower performance than the L-band dataset but a far better performance than the C-band dataset. In evaluating the classification performances of the fusion results, it is found that combining the Landsat TM features with any of the three PolSAR datasets (P-, L- or C-band) results in a much better performance than any of them used alone. The improvement to the classification accuracy is as big as 12%. All the best classification performances from the fusion datasets involve P-band. This shows that P-band is the one that can be trusted for forest applications. Therefore when it is available, priority should be given to the use of it. Moreover, our results show that the combined use of Landsat TM and L-band is as good as the P-band, and the combined use of Landsat TM and C-band is as good as L-band. Therefore, whenever data at a longer wavelength is not available, which is often the case, equivalent accuracy may be obtained by combining an optical dataset with the available shorter SAR band. This strong complementariness of SAR and optical datasets leads to a recommendation that attention should be given to the combined use of them when they are available.

It is found that discriminating the bare soil from the forest is performed at a higher accuracy than discriminating among the different tree age categories. Hence, data fusion might not be necessary for this forest/non-forest mapping case. Our results show that it can be done with a very high accuracy by using P- or L-bands. Even, the C-band or the Landsat TM dataset alone or their combination showed a great potential to discriminate the bare soil from the forest. Therefore, they can be used whenever the longer wavelength bands are not available.

In the feature selection stage, the two standard sequential forward and backward feature selection algorithms are applied and compared to select the best subset of features from the available set. In some cases, these two approaches give identical results in selecting the few best features. However, in other cases they give different results even in selecting the few best features. When they give different results, slightly better classification accuracy values are obtained from the forward approach than the backward one. As this is somewhat advantageous whenever the goal is to identify the first few best features, only the feature selection results from the forward approach are considered.

Two groups of feature selection experiments have been done for two different objectives. The first group of experiments is done to investigate those few best features from the available set that can discriminate the different classes. In this case, our main objective is to reduce the number of features as much as possible and retain most of the classification information at the same time. In our case "most of the classification information" is defined as 95% of the classification performance obtained by all the available features. Five features; two P-band features and three Landsat TM features are identified. The two P-band features are the P-band mean radar backscatter and the P-band cross-pol ratio, whereas the three optical features are the NDVI, the wetness and the soil brightness index. The joint classification performance of these five features accounts 98.50% of the

classification performance of all the twenty-six features. This is a big achievement with respect to reducing the number of features and at the same time retaining most of the classification information. Another point to mention is that the P-band mean radar backscatter alone is found to discriminate the bare soil from the forest with almost the same classification accuracy obtained by all the twenty-six features. Its classification accuracy accounts about 99.50% of the classification accuracy of the whole set. This also shows the incredible advantage of extracting and selecting features. Therefore, whenever a very large volume of data is available, feature extraction and selection not only helpful to better understand the scene but also can reduce processing time and memory. Even though it is not our primary interest, an additional advantage of feature selection found in the process is that systematically selecting features can also improve the classification performance.

In the second group of experiments, feature selection is performed for a related but slightly different objective; to identify those few best PolSAR and optical features that jointly perform best *independent of the SAR frequency used.* It is found that two PolSAR features and two optical features are best complementary to each other irrespective of the frequency of SAR used. The first best PolSAR feature found is the mean radar backscatter, and the second one is the cross-pol ratio. The two Landsat TM features are the wetness and the soil brightness index. Two other Landsat features, namely the NDVI and the greenness are found useful in complementing specific SAR frequencies. It is found that the NDVI best complements the P-band features whereas the greenness best complements the L- and C-band features. These best features identified from both datasets are found useful to interpret the forest in terms of its scattering mechanisms. Therefore, they can be reasonably expected to be used for other forest applications too.

8.2   Future work

In the future we believe, based on the insight gained from this work, several experimental works can be done as an easy extension of this study to improve the existing results and to investigate more new results. The Nezer forest is a well monitored mono-species forest with no significant slope. Therefore, first, it is of our primary interest to test the results obtained from this work for forest of more complex nature, such as tropical forests or any other forest of heterogeneous tree species. In terms of type, only fourteen; six PolSAR and eight optical features are studied in this project. However there are more Polarimetric SAR features, particularly those features that can be derived from the different polarimetric target decomposition theorems, and more vegetation indices that can be derived from both multi-spectral and hyperspectral datasets. Therefore, it is also of our interest to investigate the combination of these PolSAR and optical features for different forest applications. It is pointed out in the feature selection chapter that both the sequential forward and backward feature selection algorithms are suffering from

the so called the "nesting effect", and therefore they are suboptimal by their natures. Therefore, it is also of interest to investigate more reliable feature selection algorithms; both new and existing ones. Finally, both the fusion and the feature selection results might also be improved by using more robust classifiers. Therefore. it is also of interest to investigate other classifiers in the future.

Bibliography

[1] J.R. Jensen. *Remote sensing of the environment.* Pearson Education India, 2009.

[2] D.L. Hall and J. Llinas. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6–23, 1997.

[3] C. Pohl and JL Van Genderen. Review article multisensor image fusion in remote sensing: concepts, methods and applications. *International journal of remote sensing*, 19(5):823–854, 1998.

[4] PK Varshney. Multisensor data fusion. *Electronics & Communication Engineering Journal*, 9(6):245–253, 1997.

[5] G. Simone, A. Farina, FC Morabito, S.B. Serpico, and L. Bruzzone. Image fusion techniques for remote sensing applications. *Information Fusion*, 3(1):3–15, 2002.

[6] V. Clement, G. Giraudon, S. Houzelle, and F. Sandakly. Interpretation of remotely sensed images in a context of multisensor fusion using a multispecialist architecture. *Geoscience and Remote Sensing, IEEE Transactions on*, 31(4):779–791, 1993.

[7] D. Amarsaikhan, M. Ganzorig, P. Ache, and H. Blotevogel. The integrated use of optical and InSAR data for urban land-cover mapping. *International Journal of Remote Sensing*, 28(6):1161–1171, 2007.

[8] E. Ricchetti. Visible? infrared and radar imagery fusion for geological application: A new approach using DEM and sun-illumination model. *International Journal of Remote Sensing*, 22(11):2219–2230, 2001.

[9] G. Venkataraman, B.C. Mahato, S. Ravi, YS Rao, P. Mathur, et al. Fusion of optical and microwave remote sensing data for snow cover mapping. In *Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International*, volume 4, pages 2554–2557. IEEE, 2004.

[10] A.H.S. Solberg, A.K. Jain, and T. Taxt. Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images. *Geoscience and Remote Sensing, IEEE Transactions on*, 32(4):768–778, 1994.

[11] KS Kierein-Young. The integration of optical and radar data to characterize mineralogy and morphology of surfaces in death valley, california, usa. *International Journal of Remote Sensing*, 18(7):1517–1541, 1997.

[12] TM Kuplich, CC Freitas, and JV Soares. The study of ERS-1 SAR and Landsat TM synergism for land use classification. *International Journal of Remote Sensing*, 21(10):2101–2111, 2000.

[13] M. Moghaddam, J.L. Dungan, and S. Acker. Forest variable estimation from fusion of SAR and multispectral optical data. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(10):2176–2187, 2002.

[14] S. Huang, R.L. Crabtree, C. Potter, and P. Gross. Estimating the quantity and quality of coarse woody debris in yellowstone post-fire forest ecosystem from fusion of SAR and optical data. *Remote Sensing of Environment*, 113(9):1926–1938, 2009.

[15] N. Ackermann, C. Thiel, M. Borgeaud, and C. Schmullius. Potential of fusion of SAR and optical satellite imagery for biomass estimation in temperate forested areas. In *Proceedings of ESA Living Planet Symposium, Bergen*, 2010.

[16] P. Hyde, R. Dubayah, W. Walker, J.B. Blair, M. Hofton, and C. Hunsaker. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, 102(1):63–73, 2006.

[17] P.T. Wolter and P.A. Townsend. Multi-sensor data fusion for estimating forest species composition and abundance in northern Minnesota. *Remote Sensing of Environment*, 115(2):671–691, 2011.

[18] A.P. Doulgeris and T. Eltoft. Scale mixture of Gaussian modelling of polarimetric SAR data. *EURASIP Journal on Advances in Signal Processing*, 2010:1–1, 2010.

[19] D. Massonnet and J.C. Souyris. *Imaging with synthetic aperture radar*. EFPL Press, 2008.

[20] C. Oliver and S. Quegan. *Understanding synthetic aperture radar images*. SciTech Publishing, 2004.

[21] J.J. van Zyl. *Synthetic Aperture Radar Polarimetry*, volume 2. Wiley, 2011.

[22] G. Rees. *Physical principles of remote sensing*, volume 1. Cambridge Univ Pr, 2001.

[23] D.A. Landgrebe. *Signal theory methods in multispectral remote sensing*, volume 1. Wiley-Interscience, 2003.

[24] J.J. van der Sanden and D.H. Hoekman. Potential of airborne radar to support the assessment of land cover in a tropical rain forest environment. *Remote Sensing of Environment*, 68(1):26–40, 1999.

[25] van der Sanden. Radar remote sensing to support tropical rain forest management, ph.d. dissertation, wageningen agri. univ., wageningen, the netherlands. Tropenbos Guyana Series 5, 1997.

[26] FM Gemmell. Effects of forest cover, terrain, and scale on timber volume estimation with thematic mapper data in a rocky mountain site. *Remote Sensing of Environment*, 51(2):291–305, 1995.

[27] M.S. Chubey, S.E. Franklin, and M.A. Wulder. Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric engineering and remote sensing*, 72(4):383, 2006.

[28] C. Elachi and J. Van Zyl. *Introduction to the physics and techniques of remote sensing*, volume 28. John Wiley and Sons, 2006.

[29] D.L. Evans, T.G. Farr, J.J. Van Zyl, and H.A. Zebker. Radar polarimetry: analysis tools and applications. *Geoscience and Remote Sensing, IEEE Transactions on*, 26(6):774–789, 1988.

[30] Natural Resources Canada. Tutorial: Fundamentals of remote sensing. *Multimedia Applications Section of the Canada Centre for Remote Sensing, Natural Resource,http://ccrs.nrcan.gc.ca/resource/tutor/fundam/indexe.php*, 1986.

[31] J. Weier and D. Herring. Measuring vegetation (NDVI and EVI). *NASA's Earth Observatory http://earthobservatory. nasa. gov*, 2004.

[32] I. Woodhouse and A. Behan. Synthetic aperture radar polarimetry for forestry applications, survey Ireland - winter 1999.

[33] D.P. Lusch. Introduction to microwave remote sensing. *Center for Remote Sensing and Geographic Information Science Michigan State University*, 1999.

[34] M. Zink and R. Zimmermann. Microwave remote sensing for monitoring forest vitality (for). In *Third ERS Symposium on Space at the service of our Environment*, volume 414, page 1891, 1997.

[35] Natural Resources Canada. Soil moisture mapping, earth sciences. *Multimedia Applications Section of the Canada Centre for Remote Sensing, http://www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/radar/1707*, 2008.

[36] H. Yesou, Y. Besnus, and J. Rolet. Extraction of spectral information from Landsat TM data and merger with SPOT panchromatic imagery–a contribution to the study of geological structures. *ISPRS Journal of Photogrammetry and Remote Sensing*, 48(5):23–36, 1993.

[37] M. Govender, K. Chetty, and H. Bulcock. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA*, 33(2), 2009.

[38] S. C. Liew. Principles of remote sensing, tutorial, centre for remote imaging, sensing and processing national university of singapore. 1997.

[39] O. Hellwich, M. Günzl, and C. Wiedemann. Fusion of optical imagery and SAR/INSAR data for object extraction. *International Archives of Photogrammetry and Remote Sensing*, 33(B3/1; PART 3):389–396, 2000.

[40] A.H.S. Solberg. Data fusion for remote sensing applications. 2006.

[41] 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.

[42] C.M. Chen, GF Hepner, and RR Forster. Fusion of hyperspectral and radar data using the IHS transformation to enhance urban surface features. *ISPRS Journal of photogrammetry and Remote Sensing*, 58(1-2):19–30, 2003.

[43] P.S. Chavez, S.C. Sides, and J.A. Anderson. Comparison of three different methods to merge multiresolution and multispectral data- Landsat TM and SPOT panchromatic. *Photogrammetric Engineering and remote sensing*, 57(3):295–303, 1991.

[44] DG Leckie. Synergism of synthetic aperture radar and visible/infrared data for forest type discrimination. *Photogrammetric Engineering and Remote Sensing*, 56:1237–1246, 1990.

[45] G. Duguay, G. Holder, P. Howarth, and E. LeDrew. Integrating remotely sensed data from different sensors for change detection. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS87), Ann Arbor, USA, May 1987*, pages 333–000, 1987.

[46] M. González-Audícana, J.L. Saleta, R.G. Catalán, and R. García. Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(6):1291–1299, 2004.

[47] Z. Li and H. Leung. Fusion of multispectral and panchromatic images using a restoration-based method. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(5):1482–1491, 2009.

[48] S. Teggi, R. Cecchi, and F. Serafini. Tm and IRS-1C-PAN data fusion using multiresolution decomposition methods based on the'a trous' algorithm. *International Journal of Remote Sensing*, 24(6):1287–1301, 2003.

[49] M. Ehlers, S. Klonus, and PJ Astrand. Quality assessment for multi-sensor multi-date image fusion. *Proc XXIth Int Congr ISPRS Beijing China*, 37(B4):499–506, 2008.

[50] D. Amarsaikhan and T. Douglas. Data fusion and multisource image classification. *International journal of remote sensing*, 25(17):3529–3539, 2004.

[51] V.S. Petrovic. Multisensor pixel-level image fusion. *doktorska teza, Manchester University, UK*, 2001.

[52] J. Zhang. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1(1):5–24, 2010.

[53] M. Ehlers, S. Klonus, P.J. Åstrand, and P. Rosso. Multi-sensor image fusion for pansharpening in remote sensing. *International Journal of Image and Data Fusion*, 1(1):25–45, 2010.

[54] G. Cliche, F. Bonn, and P. Teillet. Integration of the SPOT panchromatic channel into its multispectral mode for image sharpness enhancement. *Photogrammetric Engineering and Remote Sensing*, 51:311–316, 1985.

[55] P. CHAVEZ. Digital merging of Landsat TM and digitized NHAP data for 1: 24, 000-scale image mapping((National High Altitude Program)). *Photogrammetric Engineering and Remote Sensing*, 52:1637–1646, 1986.

[56] H. Yesou, Y. Besnus, J. Rolet, JC Pion, and A. Aing. Merging Seasat and SPOT imagery for the study of geological structures in a temperate agricultural region. *Remote sensing of environment*, 43(3):265–279, 1993.

[57] JF Dallemand, J. Lichtenegger, V. Kaufmann, D. Paudyal, and P. Reichert. Combined analysis of ERS-1 SAR and visible/infrared remote sensing data for land cover/land use mapping in a tropical zone: A case study in guinea. In *ESA, Proceedings of 1 st ERS-1 Symposium on Space at the Service of Our Environment,*, volume 2, 1993.

[58] K.H. Marek and K. Schmidt. Preliminary results of the comparative analysis of ERS-1 and ALMAZ-1 SAR data. *ISPRS journal of photogrammetry and remote sensing*, 49(3):12–18, 1994.

[59] R. Welch and M. Ehlers. Merging multiresolution SPOT HRV and Landsat TM data. *Photogrammetric Engineering and Remote Sensing*, 53:301–303, 1987.

[60] R. Ahmad and R.P. Singh. Comparison of various data fusion for surface features extraction using IRS pan and LISS-III data. *Advances in Space Research*, 29(1):73–78, 2002.

[61] J. Hill, C. Diemer, O. Stöver, and T. Udelhoven. A local correlation approach for the fusion of remote sensing data with different spatial resolutions in forestry applications. *International Archives of Photogrammetry and Remote Sensing*, 32(Part 7):4–3, 1999.

[62] S. Theodoridis and K. Koutroumbas. *Pattern recognition.* Academic Press, 2008.

[63] SK Pal, TJ Majumdar, and A.K. Bhattacharya. ERS-2 SAR and IRS-1C LISS III data fusion: A PCA approach to improve remote sensing based geological interpretation. *ISPRS journal of photogrammetry and remote sensing*, 61(5):281–297, 2007.

[64] N. Prasad, S. Saran, SPS Kushwaha, and PS Roy. Evaluation of various image fusion techniques and imaging scales for forest features interpretation. *Current Science*, 81(9):1218–1223, 2001.

[65] J. Zhou, DL Civco, and JA Silander. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. *International Journal of Remote Sensing*, 19(4):743–757, 1998.

[66] D.L. Hall and S.A.H. McMullen. *Mathematical techniques in multisensor data fusion*. Artech House Publishers, 2004.

[67] S.R. Cloude and E. Pottier. A review of target decomposition theorems in radar polarimetry. *Geoscience and Remote Sensing, IEEE Transactions on*, 34(2):498–518, 1996.

[68] 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*, 2010:4, 2010.

[69] Y. Zeng, J. Zhang, and JL Van Genderen. Comparison and analysis of remote sensing data fusion techniques at feature and decision levels. In *ISPRS Commission VII Mid-term Symposium Remote Sensing: From Pixels to Processes*, 2006.

[70] D. Borghys, M. Shimoni, G. Degueldre, and C. Perneel. Improved object recognition by fusion of hyperspectral and SAR data. In *5th EARSeL SIG IS workshop on Imaging Spectroscopy:" Innovation in environmental research*, 2007.

[71] AH Schistad Solberg. Texture fusion and classification based on flexible discriminant analysis. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 2, pages 596–600. IEEE, 1996.

[72] M. Germain, J.M. Boucher, and G.B. Benie. Multiband image fusion using an unsupervised contextual method. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 4, pages IV–3341. IEEE, 2002.

[73] S. Foucher, M. Germain, J.M. Boucher, and G.B. Bénié. Multisource classification using ICM and Dempster-Shafer theory. *Instrumentation and Measurement, IEEE Transactions on*, 51(2):277–281, 2002.

[74] V. Petrovi. Multi-level image fusion. In *Proc. of the SPIE*, volume 5099, pages 87–96, 2003.

[75] R. Mayer, A. Verma, A. Ibragimov, and P.M. Koola. A framework for fusion application in non-destructive inspection.

[76] J. Le Moigne, N.S. Netanyahu, and R.D. Eastman. *Image Registration for Remote Sensing*. Cambridge Univ Pr, 2011.

[77] W. Zheng and Z. ZENG. A review on methods of atmospheric correction for remote sensing images [j]. *Remote Sensing Information*, 4:66–67, 2004.

[78] L.S. Bernstein, S.M. Adler-Golden, R.L. Sundberg, R.Y. Levine, T.C. Perkins, A. Berk, A.J. Ratkowski, G. Felde, and ML Hoke. A new method for atmospheric correction and aerosol optical property retrieval for VIS-SWIR multi-and hyper-spectral imaging sensors: QUAC (Quick atmospheric correction). Technical report, DTIC Document, 2005.

[79] W. Dierking, H. Skriver, and P. Gudmandsen. SAR polarimetry for sea ice classification. In *Applications of SAR Polarimetry and Polarimetric Interferometry*, volume 529, page 18, 2003.

[80] V. Turkar and YS Rao. Supervised and unsupervised classification of PolSAR images from SIR-C and ALOS/PALSAR using PolSARPro.

[81] EP Crist and RJ Kauth. The Tasseled Cap de-mystified((transformations of MSS and TM data)). *Photogrammetric Engineering and Remote Sensing*, 52:81–86, 1986.

[82] A.R. Huete and H.Q. Liu. An error and sensitivity analysis of the atmospheric-and soil-correcting variants of the NDVI for the MODIS-EOS. *Geoscience and Remote Sensing, IEEE Transactions on*, 32(4):897–905, 1994.

[83] Y.J. Kaufman and D. Tanre. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *Geoscience and Remote Sensing, IEEE Transactions on*, 30(2):261–270, 1992.

[84] A.W. Moore. Clustering with Gaussian mixtures. *School of Computer Science, Carnegie Mellon University, http://www. autonlab. org/tutorials/gmm14. pdf, Tutorial Slides*, 2001.

[85] JA Kong, AA Swartz, HA Yueh, LM Novak, and RT Shin. Identification of terrain cover using the optimum polarimetric classifier. 1988.

[86] HH Lim, AA Swartz, HA Yueh, J.A. Kong, RT Shin, and JJ Van Zyl. Classification of earth terrain using polarimetric synthetic aperture radar images. *Journal of Geophysical Research*, 94(B6):7049–7057, 1989.

[87] M.S. Majd, E. Simonetto, and L. Polidori. Maximum likelihood classification of high-resolution SAR images in urban area. *Trees (trees, small shrubs)*, 4281(3719):3719, 1990.

[88] E. Pottier. Advanced concepts in polarimetry. part 2: Polarimetric target classification. Technical report, DTIC Document, 2005.

[89] A.L. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1-2):245–271, 1997.

[90] R. Caruana and D. Freitag. Greedy attribute selection. In *Proceedings of the eleventh international conference on machine learning*, pages 28–36. Citeseer, 1994.

[91] D.W. Aha and R.L. Bankert. A comparative evaluation of sequential feature selection algorithms. *LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG-*, pages 199–206, 1996.

[92] Y.H. Chan, W.W.Y. Ng, D.S. Yeung, and P.P.K. Chan. Empirical comparison of forward and backward search strategies in L-GEM based feature selection with RBFNN. In *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, volume 3, pages 1524–1527. IEEE, 2010.

[93] J. Doak. An evaluation of feature-selection methods and their application to computer security (technical report cse-92-18). *Davis: University of California, Department of Computer Science*, 1992.