Aalto University
School of Engineering

Land cover mapping with multi-temporal SAR and optical satellite data

Master's Thesis
Department of Built Environment
School of Engineering
Aalto University

Espoo, 4 February 2016

Bachelor of Science in Technology
Monica Sandberg

Supervisor: Professor Miina Rautiainen
Advisor: D.Sc. (Tech.) Oleg Antropov

Aalto University
School of Engineering

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Monica Sandberg		
Title of thesis Land cover mapping with multi-temporal SAR and optical satellite data		
Degree programme Degree Programme in Geomatics		
Major Photogrammetry and Remote Sensing		**Code** M3006
Thesis supervisor Professor Miina Rautiainen		
Thesis advisor D.Sc. (Tech.) Oleg Antropov		
Date 4.2.2016	**Number of pages** 74	**Language** English

Abstract

Satellite data are widely used within remote sensing to respond to the growing need for a deeper understanding of the Earth's bio- and geophysical parameters. Applications, such as land cover classification has for long been an important task within the field. Optical satellite data have proven to be efficient tools, however, they are unavailable in some conditions, such as cloudy weather. This deficit can be addressed with synthetic aperture radars (SAR), and recently, improvements have been made in their spatial and temporal coverage. Furthermore, a fusion of these data takes advantage of their different characteristics and can lead to even improved outcomes. The aim of this study was to develop and implement an effective land cover classification approach for the boreal forest zone by using multi-temporal SAR and optical data.

Optical and SAR satellite data were collected from the area around Hyytiälä, Finland. One Landsat 8 scene and a time series of Sentinel-1 data spanning over a year were used. Co- and cross-polarized data were available. A very high resolution (VHR) reference image was manually interpreted to form training and test data. Features were extracted from both data sets and those from the SAR data were reduced using feature selection. A land cover classification was then performed separately on each data set and with a fused data set. Different features were tested to find an optimal combination. The classifications were performed with the nearest neighbor rule and the maximum likelihood classifier. This resulted in several classification maps which were validated with the test plots.

The results showed that the single-sensor classifications were noisy. Classifications with only optical imagery performed better. Additionally, removing some of the original data from the calculations, which can speed up the process, led to worse results. The multi-sensor classifications with the fused data improved the results significantly. Much of the noise was no longer present. The best classification was reached with a fused data set of four SAR features from VH polarized data and four optical features, which gained a final accuracy of 89.8 %. This classification was done with the maximum likelihood classifier. Accuracies up to 97.3 % were also reached but this result had clear flaws in the visual interpretation. It was concluded that fusing optical and SAR data for land cover classification in the boreal zone is a very promising strategy and should be investigated further to reach even better results.

Keywords data fusion, satellite data, land cover classification, boreal zone, multi-temporal, Landsat 8, Sentinel-1, remote sensing

Aalto-universitetet
Högskolan för
ingenjörsvetenskaper

Aalto-universitetet, PB 11000, 00076 AALTO
www.aalto.fi

Sammandrag av diplomarbetet

Författare Monica Sandberg

Titel Klassificering av marktäcke med multi-temporal SAR och optisk satellitdata

Utbildningsprogram Utbildningsprogrammet för geomatik

Huvudämne Fotogrammetri och fjärranalys **Kod** M3006

Övervakare Professor Miina Rautiainen

Handledare TkD Oleg Antropov

Datum 4.2.2016 **Sidantal** 74 **Språk** Engelska

Sammandrag

Satellitdata används i stor utsträckning inom fjärranalys för att fylla det ständiga behovet av mer ingående kännedom av jordens bio- och geofysiska parametrar. Applikationer, såsom klassificering av marktäcke, har redan länge varit en viktig uppgift inom studieområdet. Optisk satellitdata har visat sig vara ett effektivt redskap, men den är inte tillgänglig i vissa situationer, såsom molnigt väder. Denna brist kan övervinnas med syntetisk aperturradar (SAR) och nyligen har förbättringar skett inom den spatiala och temporala täckningen. Därutöver utnyttjar en fusion av dessa data de olika karaktärerna vilket kan leda till även förbättrade resultat. Denna studies syfte var att utveckla och tillämpa en effektiv metod för klassificering av marktäcke inom boreala skogar med hjälp av multi-temporal SAR och optisk data.

Optisk och SAR data samlades från området omkring Hyytiälä, Finland. En Landsat 8 scen och en tidsserie av Sentinel-1 data över ett år användes. Data med olika polarisationer var tillgängliga. En referensbild med hög resolution tolkades manuellt för att bilda tränings- och testdata. Variabler togs fram från båda datauppsättningarna varefter variablerna från SAR datan reducerades genom att välja de bästa. En klassificering av marktäcket utfördes sedan skiljt för de olika datauppsättningarna samt med sammanslagen data. Olika variabler testades för att hitta den bästa kombinationen. Klassificeringen gjordes med regeln för den närmaste grannen samt med maximum likelihood-metoden. Detta resulterade i flera klassificeringskartor vilka sedan validerades med testdatan.

Resultaten av klassificeringarna med data från en sensor hade mycket störningar. Den optiska bilden gav bättre resultat. Då en del av den ursprungliga datan togs bort från uträkningarna, vilket kan effektivera processen, blev resultaten sämre. En fusion av datan förbättrade resultaten betydligt då en stor del av störningarna försvann. Den bästa klassificeringen nåddes med en sammanslagen datauppsättning av fyra SAR-variabler från VH polariserad data och fyra optiska variabler med en slutlig noggrannhet på 89.8 %. Denna klassificering gjordes med maximum likelihood-metoden. Noggrannheter upp till 97.3 % nåddes även, men detta resultat hade stora brister i den visuella tolkningen. En slutsats drogs att en sammanslagning av optisk och SAR data för klassificering av marktäcke i boreala områden är en väldigt lovande strategi och bör undersökas vidare för att nå även bättre resultat.

Nyckelord data fusion, satellitdata, klassificering av marktäcke, borealt område, multi-temporal, Landsat 8, Sentinel-1, fjärranalys

Aalto-yliopisto
Insinööritieteiden
korkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Diplomityön tiivistelmä

Tekijä Monica Sandberg

Työn nimi Maanpeiteluokittelu multitemporaalisten SAR- ja optisten satelliittiaineistojen avulla

Koulutusohjelma Geomatiikan koulutusohjelma

Pääaine Fotogrammetria ja kaukokartoitus **Koodi** M3006

Työn valvoja Professori Miina Rautiainen

Työn ohjaaja TkT Oleg Antropov

Päivämäärä 4.2.2016 **Sivumäärä** 74 **Kieli** Englanti

Tiivistelmä

Satelliittiaineistot ovat laajasti käytettyjä kaukokartoituksessa maan bio- ja geofysikaalisten parametrien tutkinnassa. Sovellukset, kuten maanpeiteluokittelu, ovat jo kauan olleet tärkeitä alan tehtäviä. Optinen satelliittiaineisto on osoittautunut tehokkaaksi työkaluksi, mutta ei valitettavasti ole saatavilla kaikissa olosuhteissa kuten pilvisessä säässä. Tämä puute voidaan kiertää synteettisen apertuurin tutkilla (SAR) joiden spatiaalinen ja temporaalinen kattavuus on parantunut viime aikoina. Näiden kahden aineistojen fuusio hyödyntää niiden eri ominaisuuksia ja voi tuottaa vieläkin parempia tuloksia. Tämän tutkimuksen tavoite oli kehittää ja toteuttaa tehokas maanpeiteluokittelumenetelmä boreaaliselle metsäalueelle käyttäen multitemporaalisia SAR- ja optisia aineistoja.

Optiset ja SAR-satelliittiaineistot kerättiin Hyytiälän ympäristöstä, Suomessa. Yksi Landsat 8 kuva sekä Sentinel-1 aikasarja vuoden ajalta oli käytettävissä. Sekä yhdensuuntais- että ristipolarisaatioaineistoja hyödynnettiin. Opetus- ja testiaineistot muodostettiin tulkitsemalla korkean resoluution referenssikuvaa manuaalisesti. Eri piirteet muodostettiin kummastakin aineistosta ja SAR-aineiston piirteiden määrää pienennettiin valitsemalla parhaat. Maanpeiteluokittelu suoritettiin erillisesti eri aineistoilla sekä yhdistetyllä aineistolla. Eri piirteitä kokeiltiin myös jotta paras yhdistelmä löydettiin. Luokittelu suoritettiin lähimmän naapurin menetelmällä sekä suurimman uskottavuuden menetelmällä. Tämä tuotti useita luokittelukarttoja jotka validoitiin testiaineiston avulla.

Luokittelutulokset yhden sensorin aineistoilla olivat hyvin kohinaisia. Optinen aineisto toimi paremmin yksinään. Kun osa alkuperäisestä aineistosta jätettiin pois laskuista, joka voi nopeuttaa prosessia, tulokset huononivat. Eri sensoreista yhdistetty aineisto paransi tuloksia merkittävästi. Suurin osa kohinasta hävisi. Paras luokittelu saatiin yhdistämällä neljä SAR-piirrettä VH polarisoidusta datasta ja neljä optista piirrettä, jolloin lopullinen tarkkuus oli 89.8 %. Jopa 97.3 % tarkkuuksiin päästiin, mutta tuolloin visuaalinen tulos ilmensi selviä virheitä. Lopulta todettiin, että optisten ja SAR-aineistojen yhdistäminen maanpeiteluokitteluun boreaalisella alueella on hyvin lupaava toimintatapa ja tätä pitäisi tutkia lisää vieläkin parempien tuloksien saamiseksi.

Avainsanat data fuusio, satelliittiaineisto, maanpeiteluokittelu, boreaalinen alue, multitemporaalinen, Landsat 8, Sentinel-1, kaukokartoitus

Contents

Abbreviations

ALOS	Advanced Land Observing Satellite
AOD	Aerosol optical density
ASAR	Advanced Synthetic Aperture Radar
ASTER	Advanced Spaceborne Therman Emission and Reflection Radiometer
CORINE	Co-ordination of Information on the Environment
CLC	CORINE Land Cover
DEM	Digital elevation model
DN	Digital number
ENVISAT	Environmental Satellite
EO	Earth Observation
ERS	European Remote Sensing Satellite
ESA	European Space Agency
ETM+	Enhanced Thematic Mapper Plus
FDR	Fisher's discriminant ratio
ISODATA	Iterative Self-Organizing Data Analysis Technique
JERS	Japanese Earth Resources Satellite
MDM	Minimum-Distance-to-Means
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NBR	Normalized Burn Ratio
NDBI	Normalized Difference Built-up Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-infrared
NN	Nearest neighbour rule
ML	Maximum likelihood classifier
OLI	Operational Land Imager
PALSAR	Phased Array type L-band Synthetic Aperture Radar
PCA	Principal component analysis
PNN	Probabilistic neural network
RMS	Root mean square
SAR	Synthetic aperture radar
SITS	Satellite Image Time Series
SMAC	Simplified Method for Atmospheric Correction
SPOT	Satellite Pour l'Observation de la Terre
STD	Standard deviation
SVM	Support vector machine
SWIR	Short-wave infrared
TIRS	Thermal Infrared Sensor
TM	Thematic Mapper
USGS	Department of the Interior U.S. Geological Survey
VHR	Very high resolution

1 Introduction

The growing need for a deeper understanding and modeling of the Earth's bio- and geophysical parameters can be addressed with the help of satellite data. Formerly, optical satellite data have been primarily employed and have proven to be successful in various tasks, such as land cover classification and change detection (Powell et al. 2007). However, these data are not usable in certain conditions, such as cloudy weather and during the night. The used wavelengths cannot penetrate clouds and as the collected data are reflected sunlight, no proper image is obtained in these conditions. Other satellites, such as synthetic aperture radars (SAR), use microwaves, which allow images to be acquired in most conditions (Lillesand et al. 2008, 626-726). Previously, these instruments have suffered from bad temporal and spatial coverage. Additionally, the processing of the data requires expert knowledge and advanced signal processing. With the launch of better Earth observation (EO) sensors, the spatial resolution has been improved and the revisit time is much higher.

Recently, the use of the available data has seen another uplift as data fusion of data from different satellite sensors has grown in popularity. Within remote sensing, data fusion is used to survey the Earth in a more accurate way than before. Many researchers have exploited the synergetic strengths, the interoperability and the complementarity of different sensors to achieve improved results in applications such as environmental monitoring, identification of threats, land cover classification and crop monitoring. For instance, it has been shown that, in a classification task, adding even one multi-spectral image to a SAR time series helps the classification accuracy (Waske and van der Linden 2008). The different physical nature of the SAR and optical sensors provide auxiliary information about the object under study and can improve the accuracy of the classification results significantly. Nevertheless, this particular field of study still remains open for more research. Especially the newly available enhanced data, such as longer SAR time series, provide many new research possibilities.

Land cover classification has already for a long time been an important task within remote sensing. Satellite imagery provides an effective and fast solution for monitoring the ground. On an ever-changing Earth, it is of great importance to be able to keep track of current land cover conditions. Human interaction has made changes which influence the climate, such as cutting down forests, and the effects need to be surveyed to avoid damage. This is especially important in the boreal forest areas, which store large amounts of carbon in forests and peatlands. In these areas, vegetation characteristics and seasonal changes differ from other areas. For example, these areas are covered by snow for long periods of the year and the growing season is short. At the same time, the seasonal changes allow to define new characteristic measures or descriptors which can be used in land cover classification. Therefore, it is important to develop methods for land cover mapping that are well suited for these areas.

As a response to these research needs, this master's thesis will aim to elaborate on these current topics concerning land cover mapping with fused data from different sensors and longer time series, and find a solution for a study area in boreal Finland. An optimal study methodology is sought after for a classification task where multi-sensor and multi-temporal

data will be employed. As a result, new and improved methods should be reached, which later can be developed even further.

1.1 Research questions and objectives

The purpose of this thesis was to develop and implement a new effective land cover classification approach by using multi-temporal synthetic aperture radar data and optical data, acquired with Sentinel-1 and Landsat 8. A particular focus was to assess the potential of a new European Space Agency (ESA) instrument, Sentinel-1, which has not been studied so far. The main focus of the study was on land cover mapping in the boreal forest zone with fused data from the available satellite imagery. The goal of the thesis can be expressed with the following research questions:

1. *How can multi-temporal SAR and optical satellite data be used in land cover classification in boreal forest areas and how accurate are the classification results?*

2. *Which features and classification methods are the most suitable for boreal forest areas?*

3. *How is multi-temporal SAR and optical satellite data fused optimally?*

4. *Does data fusion of multi-temporal SAR and optical satellite data improve the classification compared to single-sensor classification and how?*

The work flow consists of many steps (Figure 1). The right area was first extracted from the given data. Training and test plots were then defined in the reference data. Next, features were chosen from the data sets used for classification. With the help of the training data, these features were used for land cover classification. A second classification was carried out by fusing the two data sets. Finally, all classifications were validated with the help of the test plots.

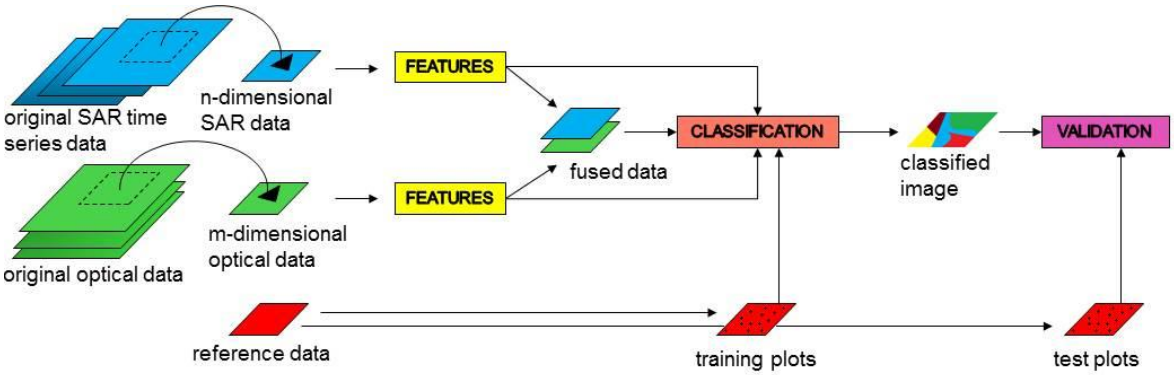

Figure 1 *The work flow of the experiments conducted for this study*

The work flow presented above is a general overview of the process, but actually consists of many more steps and sub-tasks. For example, the features were chosen in various ways and different classification algorithms were examined. In the end, the aim was to find the optimal methods for this specific research problem.

1.2 Scope of thesis and outline

It is important to define the scope of the research problem and the objectives. As this is was methodological study, the study site was relatively small. The idea was to later use the same methodology on a larger area, for example the whole Finland. This might require some further development of the methodology as the study area will then be more heterogeneous. However, this thesis only focused on the smaller area and tried to optimize the methods for that area.

Some limitations concerned the used Earth observation images and data. First, the used imagery was pre-processed, using in-house VTT software, and the pre-processing stage was not investigated very deeply. Second, the primary reference data was created using visual interpretation of a very high resolution (VHR) image and did not cover all pixels of the study area. The available full-coverage reference data (CORINE Land Cover) lacked in precision and could only be used for an approximate validation.

Finally, in the current situation where so many classification algorithms and techniques are available, a few had to be chosen. Two different algorithms were used in the process of developing a new methodology for classifications of fused SAR and optical data.

This thesis is divided into six chapters. This first chapter serves as an introduction to the subject and presents the research problem. Chapter 2 gives an overview of remote sensing as a field of study and of the theory behind the used methods for extracting features, for the classification and for the validation. It also introduces previous studies. Chapter 3 presents the data that was used in this thesis, including the reference data and the test data from two different sensors: Sentinel-1 and Landsat 8. Chapter 4 explains the different methods and strategies that were used to conduct the different experiments, which mainly included how features were formed, how classifications were done and how the results were validated. Chapter 5 introduces the results of these experiments, primarily leading to different classification maps. Finally, chapter 6 evaluates and discusses the results, and gives recommendations about future work.

2 Theory

Remote sensing, including land cover mapping, is a current method which interests many researchers. New methodologies and instruments are constantly developed to satisfy the need for a deeper understanding of the Earth. To be able to understand these processes, some fundamental knowledge about the field is needed.

The aim of this chapter is to shortly present the basic concepts of remote sensing and then more thoroughly examine tasks within remote sensing that are more relevant for this study. An overview of previous research related to this study is also given.

2.1 Remote sensing

Remote sensing is the science of obtaining information about an object, area or phenomenon using data collected with a device that is not in contact with the investigated object, area or phenomenon. Commonly this means that the study object is investigated from airborne or spaceborne platforms by collecting information of the object's emitted or reflected electromagnetic radiation.

Electromagnetic energy can be categorized according to its wavelength location in the electromagnetic spectrum (Figure 2). The visible light only covers a small part of the whole spectrum. Other forms of electromagnetic energy are, for example, X-rays, near-infrared or microwaves. They all radiate according to basic wave theory, i.e. they travel at the velocity of light in a sinusoidal manner.

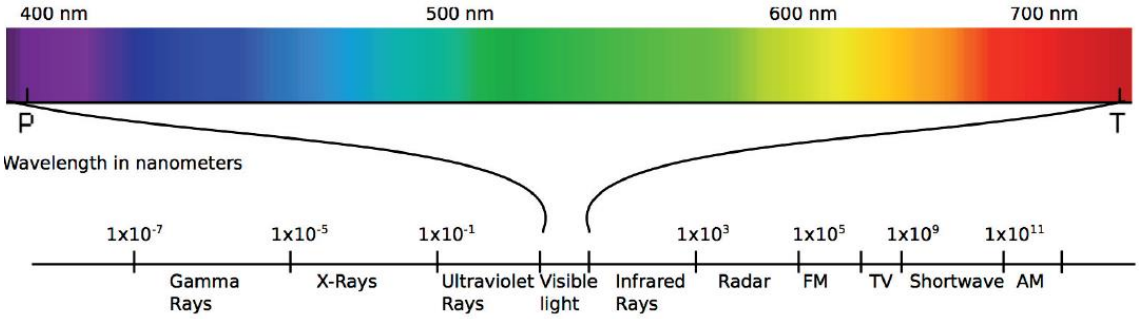

Figure 2 *The electromagnetic spectrum and the spectrum of visible light (Pölönen 2013)*

There are a few important things to consider when studying the electromagnetic energy. First, the atmosphere affects the signal while moving through it. Some frequency bands might get completely absorbed by the atmosphere and should be avoided. Second, the interaction of electromagnetic energy and the Earth's surface should be considered. When the energy reaches an object, it can be reflected, absorbed and/or transmitted. Depending on the object, the proportion of reflected, absorbed and transmitted energy differs. This proportion will also vary with the change of the wavelength of the incoming radiation. Knowledge of this phenomenon can be useful when trying to identify objects.

Remote sensing typically uses visible, infrared or microwave radiation in its applications. Depending whether the sensor supplies its own source of energy or not, it is called an active or passive system. Active sensors, such as radars, illuminate the investigated object

or area with their own energy and then measure the response. Passive sensors only sense naturally available energy, such as sunlight. This can be either emitted or reflected energy.

Optical remote sensing uses the optical spectrum (approximately 0.3 to 14 µm) to study the Earth. Within this range are UV, visible, near- mid-, and thermal infrared wavelengths. An example of a satellite using optical sensors is the Landsat series. The optical sensors detect the solar radiation reflected from targets on the ground and form an image of the view. The system can also be a multi-spectral imaging system, saving measurements from different wavelength bands to separate channels, which creates a multi-layer image.

Microwave remote sensing uses wavelengths of approximately 1 mm to 1 m. An important advantage of using microwave energy is its capability of penetrating the atmosphere in all conditions (Henderson et al. 2002). This removes, for example, the issue of cloudy weather (Figure 3). Microwave reflections or emissions also appear very different from the response an optical sensor would receive and a totally different view can therefore be achieved (Lillesand et al. 2008, 626-726).

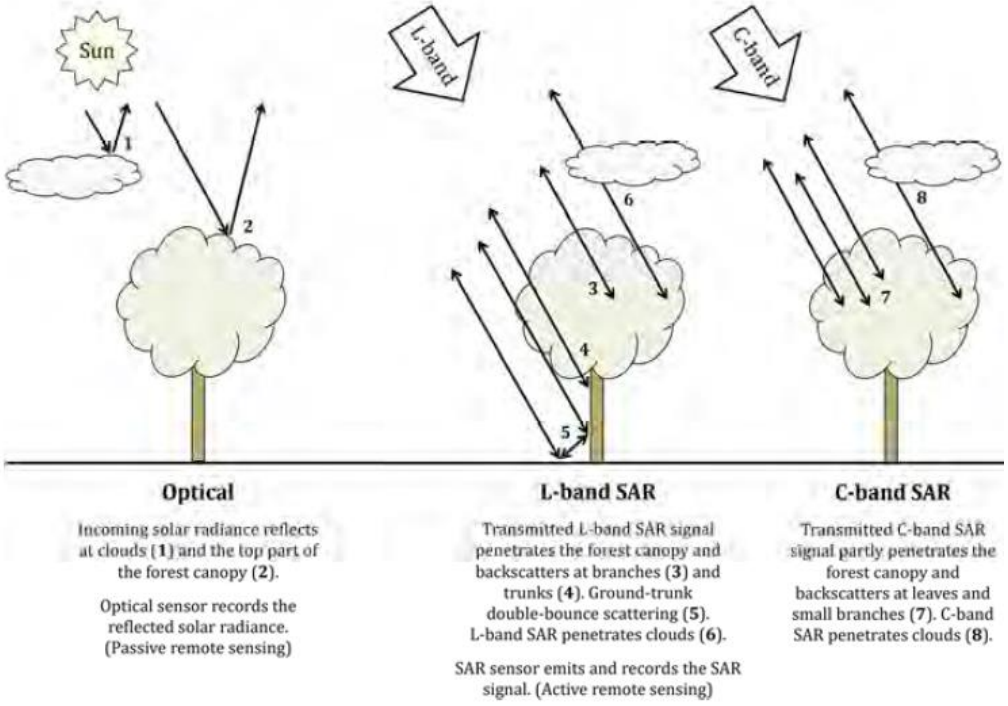

Figure 3 *The difference in penetration abilities of optical remote sensing and microwave remote sensing. The L-band has a longer wavelength than the C-band. (Reiche 2015)*

A radar is an active microwave sensor and stands for radio detection and ranging. The original purpose was to detect how far an object is. There are different kinds of radar, but in remote sensing, imaging radars are most common. These systems have an antenna fixed to the spacecraft, pointing to the side. The antenna transmits microwave energy as pulses and then measures the time and strength of the "echo". Nowadays, synthetic aperture radars (SAR) have been developed which have a short antenna, but work like a very long antenna due to modified recording and processing techniques. This results in a better resolution and finer details on the ground can be investigated.

Two important factors of the transmitted radar signal are the wavelength and the polarization. A longer wavelength results in a better ability to penetrate clouds, different land covers, etc. For example, the longer L-band (15 to 30 cm wavelength) penetrates the forest canopy and the backscatter could be from the trunk or ground, whereas the shorter C-band (3.8 to 7.5 cm wavelength) only partly penetrates the canopy and the backscatter is rather from the leafs and small branches (Figure 3). Therefore, L-band is usually a good option for monitoring forested areas, whereas C-band is more suitable for areas with short vegetation (Antropov et al. 2014). Radar signals can also be transmitted or/and received in different polarization modes. The polarization of an electromagnetic wave refers to the geometric plane in which the electrical field is oscillating. Radar signals are typically transmitted either parallel to the antenna axis (horizontal polarization, H) or perpendicular to the axis (vertical polarization, V). The antenna can also be set to only receive a specific polarization. As objects modify the polarization of the reflected energy in different ways, the chosen polarization mode can influence a lot how the resulting imagery looks like. (Lillesand et al. 2008, 626-726) A system using the VV polarization, where the signal is both transmitted and received in a vertical polarization, separates small grain crops and broadleaf plants well, whereas HH is a better source for soil conditions. The cross-polarizations HV and VH contain information about total biomass. (Idol et al. 2015)

The surface interaction of optical sensors and radars is very different. The optical sensors provide information about the surface reflection for the different wavelengths, whereas radar data tell more about the geometry of the object. The backscatter has information about surface roughness, geometry and internal structure and it can vary, for example, when changes in incidence angle, look direction, moisture conditions appear.

2.2 Time series

During the past years, remote sensing has seen a progress in satellites where the frequency of images of the same area has increased. Instead of only using two images to compare the change or study an area, these Satellite Image Time Series (SITS) provide a multitude of imagery on a long time range and are a new way to study the Earth (Petitjean et al. 2010).

A time series can reveal much more than only two images could. Changes, such as urbanization, last for many years and have many steps during the construction which can be out of interest. Further, some changes might be seasonal, such as crop cycles. Two images cannot catch this whole dynamics. (Petitjean et al. 2010) An example of a time series can be seen in Figure 4. This figure shows the values of a spectral index (Normalized Difference Vegetation Index, NDVI) of five pixels, representing different vegetation classes, during one growing season. During most of the year, classes can be distinguished with these profiles. Especially from July until October, the profiles are separated. However, the different classes have very similar values for example in December and May and some classes have similar values in between those months.

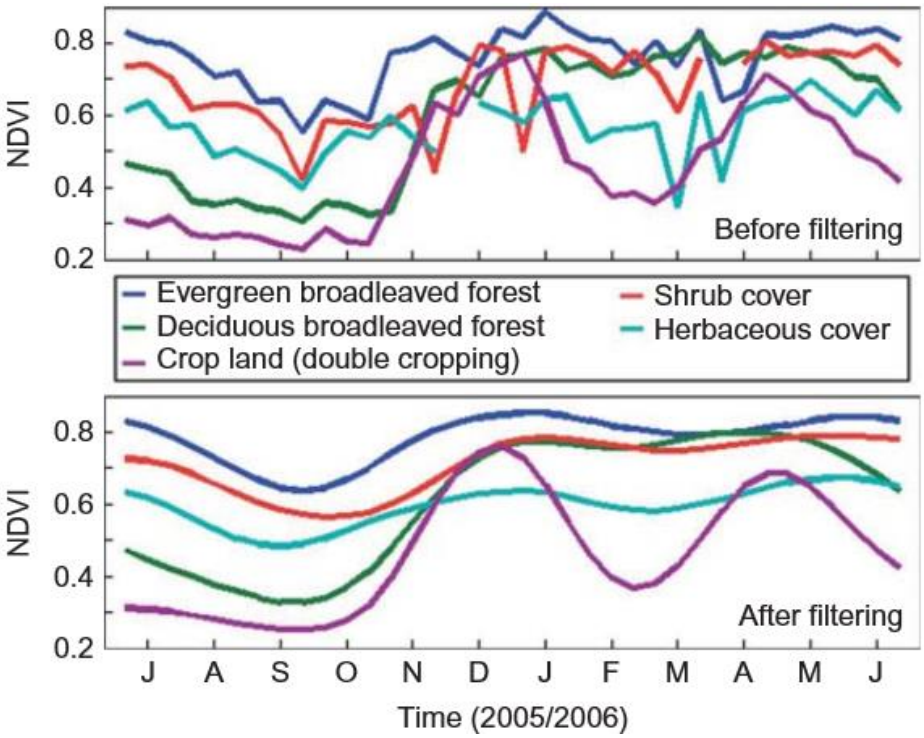

Figure 4 *Time series with NDVI profiles of five randomly chosen pixels in Minas Gerais in Brazil in the 2005-2006 growing season. Different pixels represent different vegetation classes. (Atzberger and Eilers 2011)*

As a result of the better temporal coverage of time series data, more data will be available. This will lead to some new challenges. First, reference data sets, such as ground truth and training samples, are not able to stay updated in the higher temporal resolution. This new data might have to be compared to older reference data. Alternatively, unsupervised classification methods can be used. Second, irregular sampling in terms of the temporal coverage could cause problems. When the sampling rate becomes higher, i.e. the temporal resolution increases and samples are taken more frequently, a change in the sampling frequency can affect the results. This is not a problem with a lower sampling rate where a change does not affect the result significantly. Third, some of the phenomena under investigation can have a yearly shift. For example, the vegetation cycle is quite similar from year to year, but weather conditions can cause changes in the start and ending of different changes. (Petitjean et al. 2012) All of these possible problems require the system that deal with the time series to adapt to these conditions.

2.3 Land cover classification

Remote sensing has been proven to be a good way of examining and monitoring land cover, both on a local and global scale (Foody 2002). There are numerous ways of extracting different land cover classes. These methods consist of algorithms that have already for a long time been regarded as useful, but the field continuously produces newer and better methods. Some examples of classification methods used in land cover classification and how they can be grouped are presented by Lu and Weng (2007).

The main goal of image classification is to automatically assign a land cover class or theme to all pixels in an image. Normally, multi-spectral data are used for this purpose. The classification is typically based on the spectral pattern of the pixel, that is the digital numbers (DN), given to a pixel according to its spectral reflectance and emittance

properties, or features derived from it. Spectral pattern recognition is the set of classification methods that use this pixel-by-pixel spectral information to perform an automatic land cover classification.

Another classification method is temporal pattern recognition. Here, the time is used for identifying the class. Some classes, such as agricultural fields, cannot be distinguished by using just one image. However, by using images from different dates, the class can be identified using knowledge of the seasonal changes. On one single date, an area might be very similar to another one but when comparing two dates, the differences can be found.

In land cover classification, it is important to know the difference between land cover and land use. Land cover is the physical condition of the ground surface. This can include forest, grassland, concrete pavement, and so on. Land use is more of a reflection of the human actions on the land. These could be residential areas, industrial areas, agricultural areas, and so on. A land use class can be composed of many land covers. As remote sensing measures the physical conditions, the provided information is the land cover. (Japan Association on Remote Sensing 1993, 226) When performing land cover classification, it might actually be that the used classes are a mix of land cover and land use classes.

2.3.1 Training and test data

The training and test data are an important part of the classification process. The training data is used to train the classifier, i.e. to use areas with a known classification to statistically characterize that class which then can be used when classifying unknown pixels. The test data is used to validate the results, i.e. compare these areas with a known classification to the classification done with the developed methodology.

To reach a successful classification, it is important to have complete and representative training data. This means that all classes should be described with some statistics that separate the classes. If a class is very uniform, only one statistic is needed but if it is not, all the variations should be described. It is also important to consider the size and spatial distribution of the training sites to increase the chance of finding more representative distributions (Jin et al. 2014).

2.3.2 Features

A classification task requires the definition of features. Features are used to describe and compare the images. They are a pattern that should be able to describe the images and pixels in a way that separates the desired classes. Features can be the original values in the image, for example the reflectance values, or derived values, for example the mean value. Features are typically organized as a feature vector. When using m features, they form the feature vector

$$x = [x_1, x_2, \dots, x_m]^T.$$

Each of these feature vectors is a unique descriptor of a single pattern, for example a pixel. When placing the different feature vectors in the feature space, it should be possible to draw decision lines to separate points belonging to the different classes. (Theodoridis and Koutroumbas 2009, 4-7) Feature vectors with a known class that can be used to place these lines are called training feature vectors. When a point with an unknown class, but with a

known feature vector, is placed in the feature space, its location around these lines will tell which class it should belong to.

The feature generation, the choosing and forming of features, should be considered separately for each classification task. It is important to find the features that best work in the task in hand. (Theodoridis and Koutroumbas 2009, 4-7) Sometimes it is enough to use the original pixel values in the image, whereas some more refined extracted features might be useful in other situations. Feature extraction means that you use the available values to calculate some new features which could reveal better discriminatory abilities.

The amount of features should be limited. Theoretically, when using an infinite number of training samples, the classification accuracy increases when adding more features. However, in practice, only a limited number of samples are available. When adding more features, the accuracy increases until a certain point. After this, it starts to decrease. However, increasing the amount of training samples would increase the accuracy. (Landgrebe 2003) Figure 5 illustrates this concept, called the Hughes phenomenon. Increasing the measurements complexity (amount of features), increases the accuracy to a certain point. For more training samples, this point can be found further to the right. The accuracy does not decrease only when using an infinite amount of training samples.

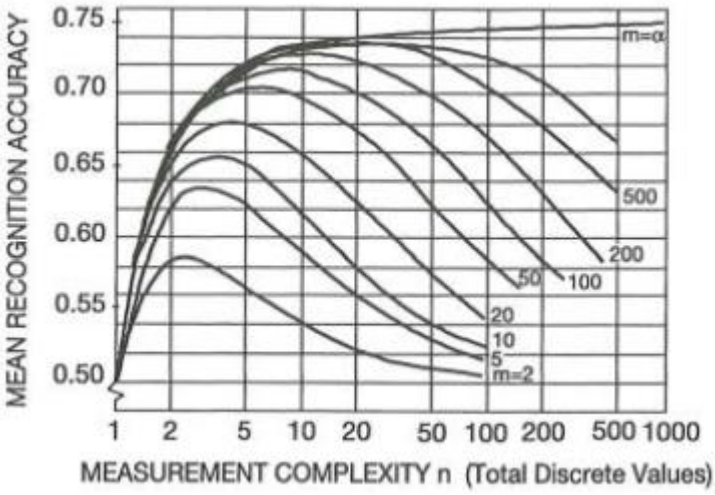

Figure 5 *The recognition accuracy for two classes when using m training samples. (Landgrebe 2003)*

In addition to the Hughes phenomenon, the amount of features should be limited for processing reasons. A larger number of features will lead to longer processing times due to bigger amounts of data. The optimal solution is to find an amount that is sufficient for the classification task, but does not lead to lower accuracy or longer processing times.

The amount of features can be reduced with feature selection. It is a way to reduce the amount of features by choosing a subset of the available features. Different methods are available. Some methods look at the single features and their predictive power, whereas other methods focus on how different feature combinations work together. Prior knowledge of the available features is also a way of choosing the final features. Two different feature selection methods are presented below: Fisher's discriminant ratio and stepwise regression.

Fisher's discriminant ratio (FDR) is commonly used for calculating the discriminatory power of different features between two or more equiprobable classes. It is not affected by the type of class distribution. (Theodoridis et al. 2010, 114-17) The simplest way for calculating the FDR is to use two classes. When using more than two classes, the calculations can be done by combining many two-class problems or simply by moving to a multi-class case. An advantage of the multi-class calculation is that it is less likely that a "random" set of features result is a good class separation. (Guyon 2003)

The FDR is a measure of separability. It can be used, for example, in linear discriminant analysis where the goal is to reduce the dimensionality of a data set by projecting it on a single line. (Duda et al. 2012, 44-51) Figure 6 shows how the location of this line can result in different success in the separation of the data classes. The aim of discriminant analysis is to find the line for an optimal separation and the FDR can be used to measure this separation.

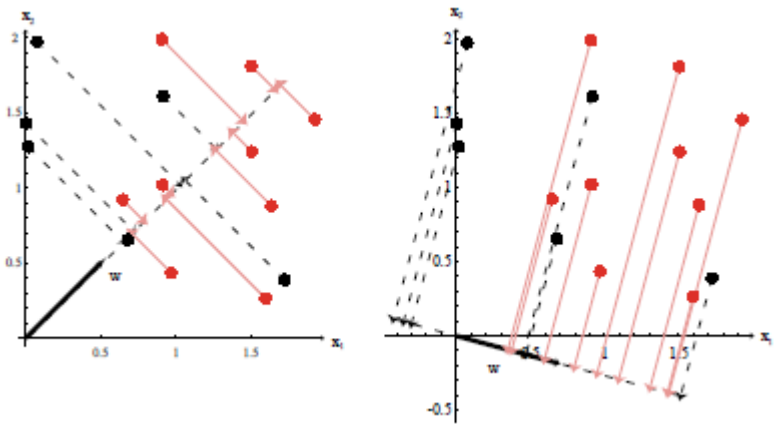

Figure 6 *Two-dimensional data is projected onto two different lines. By moving the line, a better separation between the red and black dots can be reached. (Duda et al. 2012, 45)*

Larger FDR values are a result of a good separability. (Duda et al. 2012, 44-51) When using samples of different features in the calculations, the FDR for these features is calculated. With this result, the more meaningful and more separable features can be found. However, the FDR only tells about the different features' discriminatory power but does not take into consideration how different combinations of features work.

Stepwise regression is another way to select features. This method adds and removes terms from a multilinear model by comparing their statistical significance in a regression. In feature selection, the features are the terms. After choosing an initial model, the explanatory power of models including more or less features are investigated. This is done by calculating the p-value of an F-statistic for the different features using varying feature combinations. (Draper and Smith 1998, 335-338)

Terms are added or rejected with the help of a null hypothesis, depending on if the term is already in the model or not. If a term is not currently in the model, the null hypothesis is that adding the term to the model would lead to a zero coefficient for the term. The term is added to the model if the null hypothesis can be rejected. In practice, the terms with p-values less than the defined entrance tolerance are compared and the one with the smallest value is added to the model. This step is repeated as long as the p-values are low enough. If the term already is in the model, the null hypothesis is that it has a zero coefficient. The

term is removed from the model if the null hypothesis can be rejected. In practice, the terms with a p-value larger than the defined exit tolerance are compared and the term with the highest value is removed from the model. Next, new p-values are calculated and the ones below the entrance value are again compared, then moving to the ones above the exit value. This continues until no p-values fit the criteria above and the model cannot be improved anymore. (Draper and Smith 1998, 335-338)

The advantage of stepwise regression, compared to Fisher's discriminant ratio, is that it takes into consideration the combinations of features. Even if single features have good discriminatory power, they might not work well together. Stepwise regression tries the different feature combinations to find an optimal model which is advantageous in the situation where features are combined for the classification.

2.3.3 Dimensionality reduction

Multi-sensor and multi-temporal data bring out the problem of very many dimensions in the used data set. Too many dimensions can cause problems during the calculations. In addition to the previously presented feature extraction and selection methods, this problem can be tackled with dimensionality reduction. The dimensions of the original data set can be reduced, while keeping only the most important information, and the following processing is then done with the new set. This section will mainly cover one dimensionality reduction method, principal component analysis.

Dimensionality reduction is especially important when using time series. Long time series contain a huge amount data, which can be a problem in the calculations. Shape-based approaches for assessing the similarity of time series, where the time series is compared to another one directly with all original values, might be impossible. Feature-based or model-based approaches might be more feasible. The feature-based approach converts the time series to a lower dimensional feature vector, whereas the model-based approach transforms the time series into model parameters, i.e. a parametric model is created for each time series. (Aghabozorgi et al. 2015)

Principal component analysis (PCA) is an effective way to reduce the dimensionality of the used data. The aim is to reduce the variance and the dimensionality of the data set by projecting the data along new non-correlated axes (Gupta et al. 2013). The total system variability is explained by p components, but most of this variability can often be explained with a smaller number of components, k principal components. By using the principal components, the dimensionality of the data set can be reduced. (Johnson and Wichern 2007, 430-480) The technique transforms the data in a way that brings out similarities and differences, but no data loss is present until k components are used to reduce the dimensionality (Gupta et al. 2013).

The main idea of the principal component analysis is that it extracts and displays the greatest variance in the given data on the first axis, the first principal component, the second greatest variance on the second axis, and so on. This is an orthogonal linear transformation. The steps to perform this transformation and the dimensionality reduction are the following.

1) Normalize the data by subtracting the mean. This mean is calculated and subtracted from each dimension separately. This results in a data set with a zero mean.

2) Calculate the covariance matrix. This matrix has the same amount of rows and columns as the dimension of the original data set.
3) Calculate the eigenvalues and eigenvectors from the covariance matrix. The eigenvectors characterizes the data, whereas the corresponding eigenvalue show the importance of the principal component.
4) Order the eigenvectors from biggest to smallest and choose the corresponding k highest eigenvectors. The eigenvector with the largest eigenvalue is the first principal component, and so on.
5) Multiply these eigenvectors with the original data set to get a lower dimensional new data set containing the most important information from the original data set. For example, for an m-dimensional data set, the pixel value at (i, j) for the first principal component will be

$$\boldsymbol{pc1_{ij} = a_1 p_{ij1} + a_2 p_{ij2} + \cdots + a_m p_{ijm'}} \quad (1)$$

where $pc1_{ij}$ is the pixel value at (i, j) in the first principal image, $a_1 \ldots a_m$ are the elements of the first eigenvector for different dimensions in the original data and $p_{ij1} \ldots p_{ijm}$ are the original pixel values at (i, j) in the different dimensions. (Gupta et al. 2013)

2.3.4 Supervised and unsupervised classification

Classification performed on individual pixels can be divided into supervised and unsupervised classification. In short, supervised classification uses prior knowledge of the classes to classify pixels to the wanted classes, whereas unsupervised classification groups the data according to the natural spectral groups which then has to be combined to the desired classes.

The supervised classification uses a priori knowledge of the desired classes to describe the numerical descriptors of them. This information is then used in the pixel categorization and therefore, it can be said that the classification is "supervised". The prior knowledge is some sample plots, training areas, where the land cover is known. In the training stage, these areas are identified and a numerical description of each class is formed. In the classification stage, the pixels that need to be classified can then be compared to these training areas and classified to the most similar class. Many different methods for this comparison exist. In the output stage, all pixels have been categorized and the results can be shown in the preferred manner, such as thematic maps or tables.

The unsupervised classification does not use any prior information, training data, in the classification. It works with the principle stating that pixels belonging to the same class should have similar values. This means that the algorithm looks for natural groupings or clusters with the help of a clustering algorithm. Normally, the amount of classes cannot be defined by the user. Once the classification is done, the analyst compares the results to some reference data to identify which class these groups belong to. The analyst can also choose to merge some of these groups to a class he defines. The advantage of this method is that classes that might not be apparent in the training phase could be discovered.

2.3.5 Classification algorithms

Throughout the years, a wide variety of classification algorithms have been developed. Many of these have not been developed for remote sensing, but are in extensive use within the field. An example of this would be support vector machines (SVM), which originally were developed in machine learning (Mountrakis et al. 2011). However, some algorithms have been made for a specific remote sensing application. Two commonly used algorithms are presented next: the nearest neighbour rule and the maximum likelihood classification.

The nearest neighbour rule (NN or k-NN) is a nonlinear classifier which classifies a given point according to the nearest neighbours with the help of a defined distance measure. Before applying the nearest neighbour rule, the data should be normalized. This can be done by restricting the range of the data between a minimum and maximum, for example [0,1]. For a point p_i in an image m, the normalized value is calculated

$$\boldsymbol{p_m} = \frac{\boldsymbol{p_m} - \boldsymbol{min_m}}{\boldsymbol{max_m} - \boldsymbol{min_m}} \quad (2)$$

For an unknown feature vector x, N training samples and a given distance measure, the nearest neighbour method consists of the following steps:

1) In the feature space, find the k nearest neighbours to x.
2) Identify the class of the training samples and count the amount per class.
3) The unknown vector x is assigned to the class with the highest number of training samples. (Theodoridis and Koutroumbas 2009, 61-64)

An illustration of the method can be seen in Figure 7. The eleven closest neighbours to the star are found and the star is classified to the blue class as most of the neighbours belong to that class.

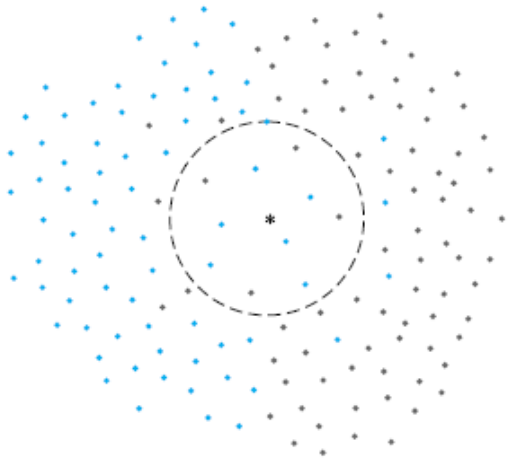

Figure 7 *Example of 11-NN. The star needs to be classified to the blue or black class. The circle represents the eleven closest neighbours. Seven are blue which results in a classification to the blue class. (Theodoridis and Koutroumbas 2009, 61)*

Before using the nearest neighbour rule, the value of k needs to be chosen. The simplest way is to choose $k = 1$. In this case, you only look for one nearest neighbour and assign the class of that feature vector to the unknown vector. Hence, there is no need to compare

the amounts of classes of the different feature vectors. However, k can also be a higher number. In those cases, more than one neighbour is investigated and the class amounts are compared to decide the classification. In general, k is not chosen as a multiple of the amount of available classes (Theodoridis and Koutroumbas 2009, 61-64). When using a large number of training samples, a larger k brings the k-NN classifier closer to the expected optimal Bayesian classifier. However, with a small number of training samples, a larger k may not result in a better classification. (Theodoridis et al. 2010, 25-26)

The unknown feature vector is compared to its neighbours with a distance measure. This measure has to be defined. The simplest distance measure is the Euclidean distance:

$$\boldsymbol{d(x,z) = \sqrt{\sum_{i=1}^{m}(x_i - z_i)^2},} \quad \textbf{(3)}$$

where x is the unknown feature vector, z is one of the training samples and m is the dimensionality of the feature vectors. When using time series, the Euclidean distance has been widely used and proven to be efficient (Aghabozorgi et al. 2015).

Even if the nearest neighbour rule is one of the simplest classification methods, its computational complexity is a major problem (Theodoridis et al. 2010, 25-26). The distance measure has to be calculated between every unknown feature vector and every training sample, which increases the computation time a lot. Especially with high-dimensional spaces, it can decrease the effectivity of the algorithm.

The maximum likelihood classifier (ML) is a supervised classification method based on the Bayes theorem. This classifier has been one of the most popular methods and still remains as a benchmark against newer methods (Richards 2005). The main idea is to assign pixels to the class with the highest likelihood, with the help of a discriminant function. According to the Bayes theorem, the a posteriori distribution $P(c|x)$, which is the probability that a pixel with the feature vector x belongs to class c, is

$$\boldsymbol{L_c = P(c|x) = \frac{P(x|c)\,P(c)}{P(x)},} \quad \textbf{(4)}$$

where $P(x|c)$ is the likelihood function, $P(c)$ is the a priori information (the probability that a class exists in the test area) and $P(x)$ is the probability that x is observed. Furthermore, $P(x)$ can be expressed as

$$\boldsymbol{P(x) = \sum_{i=0}^{M} P(x|c)\,P(c),} \quad \textbf{(5)}$$

where M is the number of classes. The feature vector x is assigned to the class c with the following rule.

$$\boldsymbol{x \in c \; if \; P(c|x) > P(d|x) \; for \; all \; c \neq d.} \text{ (Asmala 2012)} \quad \textbf{(6)}$$

It is common to assume that $P(c)$ is the same for each class and as $P(x)$ is common for all classes, the likelihood only depends on $P(x|c)$, also called the probability density function. A multivariate normal distribution is used as the probability density function which leads to the following likelihood function:

$$L_c(x) = \frac{1}{(2\pi)^{\frac{m}{2}}|C_c|^{\frac{1}{2}}} e^{\left(-\frac{1}{2}(x-\mu_c)^t C_c^{-1}(x-\mu_c)\right)}, \quad (7)$$

where m is the number of features, and μ_c is the mean value and C_c the covariance matrix of the class c. (Japan Association on Remote Sensing 1993, 220-221)

The main idea of the maximum likelihood classifier is to assign a pixel to the class with the highest likelihood. The method goes through the following steps:

1) The mean vector and covariance matrix are calculated for each class, using the training set pixels.
2) For each unknown pixel with feature vector x, the likelihood to belong to each class is calculated.
3) The pixel is assigned to the class with the highest likelihood. (Asmala 2012)

The maximum likelihood classifier has some characteristics that one should be aware of when using the method. It is important to have enough training samples to be able to estimate the mean vectors and covariance matrices. Additionally, if two bands correlate too much or the training samples are too homogeneous, the inverse of the covariance matrix becomes very unstable. In this case, a dimensionality reduction would be good. The maximum likelihood classifier is also not applicable if the distribution of the data set does not follow the normal distribution. (Japan Association on Remote Sensing 1993, 220-221)

2.3.6 Validation of classification results

A map is a model or generalization of reality and, therefore, usually has some errors. The mapping process mostly leads to loss of information and completeness. Consequently, it is important to assess the quality of the derived map and express it in a meaningful way. (Foody 2002) This means that once classifications are made, the success of the results should be examined with some validation methods.

A common validation method for classification results is to use a confusion matrix. This matrix uses test data and the known classes of the pixels in this data, and compares this information to which class those pixels are classified. It basically tells to which class the pixels of known class have been classified by showing the known classes in the columns and the predicted classes on the rows of a matrix.

In addition to a confusion matrix, some accuracies can be calculated. The final accuracy is calculated as the ratio of pixels classified to the right class and the total amount of pixels. The producer's accuracy is the ratio of pixels classified to the right class and the total amount of pixels belonging to that class. The user's accuracy is the ratio of pixels classified to the right class and the total amount of pixels that are classified to that class.

2.4 Data fusion

Data fusion combines data from many sources or time periods in an attempt to improve the information, leading to a better result in proceeding actions, such as a classification (Hall and McMullen 2004, 1-35). The combined data can be obtained from different platforms, such as satellites, aircrafts or ground-based measurements. In the case of multi-sensor data fusion, the combined data is from different sensors, for instance synthetic aperture radars (SAR) and optical sensors. Data fusion can be compared to the human cognitive process of combining information we get from our different senses. For example, by combining what we see, what we hear and what we smell, we can form a better interpretation of a situation than if we only used one of these senses.

The aim of data fusion is to enhance the information from many sources by combining them and to improve the accuracy. Pohl and Van Genderen (1998) have presented some more specific goals. Data fusion can sharpen images, improve geometric corrections or provide stereo-viewing capabilities for stereophotogrammetry. Data fusion can also be a tool for enhancing some features which cannot be found in data from only one sensor. Other important aspects of data fusion are the possibility to complement data sets to get a better classification result and to detect changes with the use of multi-temporal data. Finally, data fusion can be used to substitute missing information in one image by using information from another image and to replace defective data.

Data fusion, or in the case of remote sensing, image fusion can be performed on three different processing levels, defined by the stage at which the fusion is done. These stages are the pixel level, the feature level and the decision level (Pohl and Van Genderen 1998). Image fusion at pixel level is done by combining the measured physical parameters, such as surface reflectance and radar backscatter. An important aspect of this is the co-registration or geocoding of the images. Resampling of the image data to a common pixel size is also needed. The fusion is performed by taking the values of corresponding pixels in the data sets that are to be fused and giving all those values to that pixel in the fused data set. This leads to a higher dimensionality. The evaluation of the data, for example a classification, is then done with the fused data. Image fusion at feature level requires that some features are extracted or selected from the original images. After this, those features are fused, for example, with some statistical approaches. The fusion is similar to the fusion at pixel level; only the used values have been modified and changed to features. The further evaluation is again done after the fusion. The last type of image fusion is fusion on decision level. This means that the separate data sources are first evaluated separately, possibly using features, and then those results are fused. This is done by using some decision rules. Common results are accepted and differences have to be solved.

Image fusion can be performed with different kinds of data sets. One possibility is to use data acquired at different times. This data can be either from the same sensor or from different sensors. Another possibility is to fuse data with different spatial resolution. Once again, the fusion can be performed with single or multi-sensor data. Data from the same date but from different sensors is another possibility for data fusion. Finally, combining remote sensing data with some ancillary data, such as topographic maps, is another way. (Pohl and Van Genderen 1998)

Many of the applications used on fused data are possible also without data fusion, but the combined data does bring some benefits. Some interpretations can be hard or almost

impossible without the fusion. Overall, the goal with data fusion is to improve the accuracy of the desired task. The fusion brings benefits such as larger spatial and temporal coverage. This means that a target that cannot be measured by one sensor at some time might be able to be measured by another sensor. Data fusion also leads to increased confidence when a second sensor can confirm the results from the first one. Another benefit is a possible higher spatial resolution. An integration of multiple sensors can geometrically form a synthetic aperture which can lead to a better resolution. (Hall and McMullen 2004, 1-35) Finally, sensors with different physical nature, such as optical and SAR sensors, provide auxiliary information about the study area which should improve further processing, such as classification results.

2.5 Land cover classification in the boreal zone using optical and SAR satellite imagery

The boreal forests, the largest vegetation zone on Earth, cover large areas in Europe, Asia and North America. In these areas, the coniferous trees are the most dominating. These include pines and spruces. Throughout the year, the temperature changes significantly and in the winter the growing season has a pause when the ground is frozen and covered by snow. These seasonal variations have an effect on the total backscatter (Pulliainen et al. 1996). Variations are also present in optical imagery where, for example, winter scenes with snow have very different reflection properties than summer scenes (Warren 1982).

The use of optical satellite imagery in land cover classification has been researched broadly. The Landsat program is probably the most known mission for optical imagery. It is widely regarded as an important asset for land cover mapping. Its temporal, spatial and spectral resolutions make it ideal for mapping of large areas. (Wulder et al. 2008) As it can be seen in the studies presented next, Landsat data are still widely used and succeed in various tasks in different conditions. It has been proven to work in land cover classifications and in boreal areas. Therefore, it was a natural choice for this thesis.

The most recent satellite in the Landsat series is Landsat 8. Imagery from this satellite has been used in a wide variety of classification tasks. Different subsets of the data have been used and the images have been applied for different land cover types and their classification. Li et al. (2015) developed an automatic classification approach for urban areas with the help of Landsat 8 data. Test sites in China (Nanjing and Ordos) were used. The classification was based on different spectral indices, also some recently developed indices. After a preliminary classification which used the indices in a linear model, the results were improved with building a non-linear support vector machine. The method successfully extracted water and vegetation, and effectively extracted impervious surfaces and bare land.

Eisavi et al. (2015) used multi-temporal spectral and thermal Landsat 8 images to perform land cover classification in the area around Naghadeh, Azerbaijan. Seven images from spring, summer and autumn were used to better be able to separate different vegetation and agricultural classes, due to their changes in the growing season. Classifications were made with spectral and thermal data (separately and combined) and with up to 21 features, such as principal components and the best original data. The classification was done with a random forest classifier. All methods were successful, but choosing the best features gave the best results (overall accuracy of 91.8 %).

Landsat data have been used for land cover classification in most areas. Some research in the boreal forest areas is also available. Sasaki et al. (2001) used Landsat Thematic Mapper imagery from May 1998 and September 1995 for land cover mapping over an area by Fort McMurray, Alberta, Canada. The long time gap between the images was caused by cloudy images in between. The aim was to obtain classes of different forest types, such as open conifer forest or closed broad-leaved forest. The original radiance values of the different bands and spectral indices were used as features in a decision tree classifier. This resulted in an overall accuracy of 76 %. Potapov et al. (2011) created a method for boreal forest cover and change mapping which was tested in European Russia with composite Landsat Enhanced Thematic Mapper Plus (ETM+) images from 2000 and 2005. MODIS (Moderate Resolution Imaging Spectroradiometer) data were used for a radiometric normalization as large areas of the Russian boreal region are missing physically-based surface reflectance corrections. A training set was formed with visual interpretation with the help of several additional data sets. As an input for the classification, 50 features were created, including the original bands, spectral indices and cloud/shadow/water probabilities. An empirical model and a classification tree were used for the land cover mapping, resulting in a final accuracy of 89 %.

The long-running Landsat program has also allowed research on time series. In a typical time series analysis of EO data, a feature is extracted from each image and the temporal trajectory of the feature is then analyzed. Lhermitte et al. (2011) and McCloy (2010) give a good overview of state-of-the-art time series analysis methods for EO data. Other relevant techniques include Dynamic Time-Wrapping (Petitjean et al. 2012) which can analyze and compare time series with irregular sampling intervals or imagery contaminated by clouds, and Fourier time-frequency analysis (Brooks et al. 2012) which can handle missing data. Kennedy et al. (2010) and Cohen et al. (2010) contributed to the progress of time series analysis in forest monitoring and mapping. They used Landsat data over Oregon and Washington, USA, while developing a method for capturing short-term events and smoothing of long-term trends. The time series data allowed to form long and consistent global multiyear observations capturing both inter-annual and intra-annual variations of spectral reflectance with the help of temporal segmentation. The method successfully captured a wide range of phenomena and the segmentation method was regarded a feasible and robust way for increasing the information extraction from Landsat time series. Similarly to these methods using optical data, advances in acquisition times with imaging radars are now permitting studies with new and longer SAR time series. This opportunity will be exploited with Sentinel-1 data in this study.

A more recent trend in remote sensing has been the use of microwave remote sensing data for land cover classification. Previously, a big limitation for using spaceborne microwave radiometry (passive instruments) has been its moderate spatial resolution, leading to only a few studies of the boreal forest area (Kurvonen et al. 2002). Another limitation was the use of a single wavelength and only one polarization. Different types of polarizations can provide different kind of information, leading to better classification results. (Idol et al. 2015) More recently, newer satellites with synthetic aperture radars, such as Sentinel-1, have been introduced which have made it possible to monitor the Earth with better temporal and spatial resolution.

The high temporal resolution of newer SAR instruments make them optimal for time series approaches. Even if multispectral imagery is easier to interpret, SAR time series have

proven to be competitive. For example, Chust et al. (2004) found a SAR time series of seven images to contain more separable land cover information than Landsat Thematic Mapper (TM) data. Other studies using SAR time series also exist. Oyoshi et al. (2015) used RADARSAT-2 (C-band) time series over Japan to identify rice-planted areas. Optical data are not feasible for this study as the data need to be collected during the rainy season which would lead to clouds in the optical imagery. Both VV and VH polarized data were used. The algorithm used the temporal profile of the SAR backscatter. Optimal thresholds for the minimum value and the range (difference between minimum and maximum value) within the time series were created and used for classification. Finally, the VH polarized data resulted in a higher final accuracy (83 %) than the VV polarized data (76 %). Waske and Braun (2009) used three SAR time series over two agricultural regions in Germany for land cover classification. The time series were collected from ERS-2 (European Remote Sensing Satellite) and ENVISAT (Environmental Satellite) ASAR (Advanced Synthetic Aperture Radar) images. Either 15 or 17 dates were used, ranging from March to September in 2005 or 2007. The classification was performed with random forests and the results were compared to classifications made with the maximum likelihood classifier and a simple decision tree. Classifier ensembles, i.e. using different training samples for the same classification method and then combining the results, were also tested. The study showed that random forests is a very promising technique for classification of multi-temporal SAR data.

The Sentinel mission is a fairly new project, with the launch of Sentinel-1 in April 2014 (European Space Agency). Its characteristics, such as a high temporal resolution, are widely recognized, but limited research using this data is available at this point. However, it is expected to contribute a lot to an improvement in, for instance, land cover classification. Even though it might still be hard to create fully automatic multi-class classifiers for large areas, some classes seem promising for developing classifiers able to cover these classes on continental or even global scale on an annual basis. These include classes such as forest/non-forest, urban areas and crop classification. (Wagner et al. 2012)

Many land cover classification studies of boreal forest areas using SAR data are available. Most of these studies focus on the areas which are especially important in these regions: forests and wetlands. An example from Finnish territory was carried out by Kurvonen et al. (2002). They did a land and forest-type classification in southern Finland (Porvoo) and northern Finland (Sodankylä) with the use of multi-temporal SAR data from ERS-1, JERS-1 (Japanese Earth Resources Satellite) and RADARSAT, and radiometer data. Both texture and intensity values were used. With seven land cover classes, the texture measures led to a final classification accuracy of 65 % whereas the intensity values resulted in an accuracy of 50 %. In the forest-type classification with four classes, the texture measures gave an overall accuracy of 66 % and the intensity values resulted in an accuracy of 40 %. Engdahl and Hyyppä (2003) conducted a study over an area around Helsinki, in southern Finland, where the potential of ERS-1/2 Tandem InSAR data was studied in land cover mapping. A time series of 14 image pairs was processed, and backscattered intensity, Tandem coherence and long-time coherence information were used in the classification. The water class was first classified and the remaining five classes were classified with an unsupervised classifier, ISODATA (Iterative Self-Organizing Data Analysis Technique). The validation was done with the help of high-resolution aerial orthophotos, digital base maps and the Finnish National Forest Inventory. The resulting overall accuracy for the

classification was 90 %, leading to a conclusion that these data are very suitable for land cover mapping.

Studies in the boreal zone have also been conducted with the help of satellites providing fully polarimetric data. Antropov et al. (2014) used fully polarimetric SAR data in the Finnish boreal forest zone to perform land cover mapping. The used L-band data were collected with ALOS (Advanced Land Observing Satellite) PALSAR (Phased Array type L-band Synthetic Aperture Radar) in 2006-2007. Polarimetric features, such as entropy and anisotrophy, were extracted and supervised classifications with a probabilistic neural network (PNN) and the maximum likelihood method were performed. The PNN performed better and had accuracies up to 82.6 % in a land cover mapping of five classes and up to 90 % in a forest-nonforest mapping. The classification results were dependent on seasonal variations. Lönnqvist et al. (2010) performed land cover mapping with data from Kuortane in the Finnish boreal zone, using different data types and classification methods. Fully polarimetric and dual-polarized data from ALOS PALSAR were used and the classifications were performed with two supervised and two unsupervised methods. The data were used as fully polarimetric data, as features extracted from the fully polarimetric data or as intensity data from selected channels. The classifications were performed with three to six classes. Fully polarimetric data with three classes gave the best results (87.5 % - 84.7 %), whereas the intensity data reached lower accuracies (83.6 % - 78.6 %). It was also concluded that urban areas are hard to classify, and snow covered scenes also bring issues in a classification task.

With the rise of the improved SAR products, fusion with optical data has become more popular. Their complementary information has great potential within land cover classification. Much of the current research focuses on this data fusion and its capabilities of improving land cover classification results. Villa et al. (2015) used features derived from Landsat 8 and TerraSAR-X for urban land cover mapping in the Lombardy region in northern Italy. Four optical scenes from March to July and three SAR scenes from May to July were used. Two features were extracted from the Landsat 8 images: Urban Index and Soil and Vegetation Index. The backscatter in the X-band was used as the SAR feature. These features were fused with different combinations. Four different classification algorithms were used: classification and regression trees, random forest, support vector machines, and multilayer perceptron. The results showed that adding the SAR component improved the classification accuracy with 3 % on average. Idol et al. (2015) fused RADARSAT-2 and PALSAR data with optical ASTER (Advanced Spaceborne Therman Emission and Reflection Radiometer) data for a land cover classification in Sudan. Texture measures of pixels in the SAR data were also used. These give information about spatial arrangements of pixels, instead of only describing individual brightness values. The classification was performed with the maximum likelihood decision rule. The results showed higher accuracies than the results gained when using only optical or SAR data.

The maximum likelihood classifier has been used in similar applications already earlier. Huang et al. (2007) executed a land cover classification in St. Louis, Missouri, by merging three bands from Landsat 7 images with RADARSAT SAR imagery (C-band). Tests with radar imagery that had been processed with different filtering and texture techniques were done. The classification was performed with a maximum likelihood classifier. Results were compared to a classification done with only Landsat data and an improvement of up to 10 % in the overall classification accuracy was seen.

Research very closely related to the study of this master's thesis, i.e. fusing multi-temporal SAR imagery with one optical scene, is also available. Waske and van der Linden (2008) showed that adding even one multi-spectral image to a SAR time series helps the classification accuracy. The used SAR data consisted of nine ENVISAT ASAR and ERS-2 images from April to September 2005 and the optical image was one Landsat 5 Thematic Mapper image from May 2005. The study area was in Bonn, Germany. The classification was executed with decision fusion, based on a support vector machine (SVM). The SAR and optical data were first pre-classified separately with SVM. The fusion was then performed with two methods, a second SVM and with random forests. The fusion with random forests resulted in the best overall classification accuracy (84.9 %).

In another study, Michelson et al. (2000) used seven ERS SAR images and one Landsat TM scene, all from 1993, to evaluate the statistical separabilities of spectral and backscattering signatures in 16 Swedish land cover classes and to examine if data fusion improves the classification result. Three different classification algorithms were tested: maximum likelihood, sequential maximum a posteriori and neural network classification. The results showed that the SAR data had more separable land cover information, however, when combined with optical data, it was even more. The sequential maximum a posteriori classification proved to be the best method. The overall accuracy was the best when fusing the two data sets (57.1 %).

Finally, Chust et al. (2004) investigated the performance of two ERS SAR time series (1994/1995 and 1998) and optical imagery from SPOT (Satellite Pour l'Observation de la Terre) (1994), both separately and fused, for a classification task in Minorca, Spain. The classifications were done with the maximum likelihood classifier, based on the real Gauss-Wishart distribution. This is a contextual formulation of the Gauss distribution, which means that it takes into account the neighbourhood around the investigated pixel. Data fusion was performed with the Dempster-Shafer evidence theory as decision fusion. The results showed that the fusion improved the classification accuracies, compared to using only one sensor.

All of the studies presented above show that the field of study has been widely researched, and more recently more focus has been put on SAR time series and fusion with optical imagery. However, notable gaps in the research remain. First, both of the satellites used in this thesis are very new (launched in 2013 and 2014). This means that much research on their products is not yet available. Landsat 8, being part of a long existing mission, has already led to some research results, whereas Sentinel-1 imagery has not yet been used at all. With the advent of Sentinel-1, longer SAR time series are available, whereas previous studies had more limited data.

Second, the feature extraction and selection, and the different ways to combine features in data fusion create huge amounts of possibilities. None of the previous studies have used exactly the same strategy as this study does. In this thesis, a new way to combine multi-temporal SAR and optical data is proposed and evaluated.

Finally, the fusion of SAR time series with optical data is an important field of research in the boreal areas where cloud cover, shorter days and stronger seasonal changes make the lack of sufficient data a great limitation for studies. This study will contribute with a lot of

important information regarding SAR and optical satellite data fusion for land cover classification in boreal forest areas.

3 Data

The development of the desired methodology requires some data for the processing. A test site has to be chosen and different types of data from that area need to be acquired. This includes the data used for the classification, in this case SAR and optical satellite data, and some reference data. The next section will present all the data that are used for this project.

3.1 Test site

The area around the Hyytiälä Forestry Field Station was used as a test site for the developed classification methodology. The area extents and location can be seen in Figure 8. The Hyytiälä Forestry Field Station is a unit of the Department of Forest Sciences at the Faculty of Agriculture and Forestry at the University of Helsinki. The station is in active use for forestry teaching and research. This has resulted in a wide range of available in-situ data that can be used, for example, in validation of the classification results.

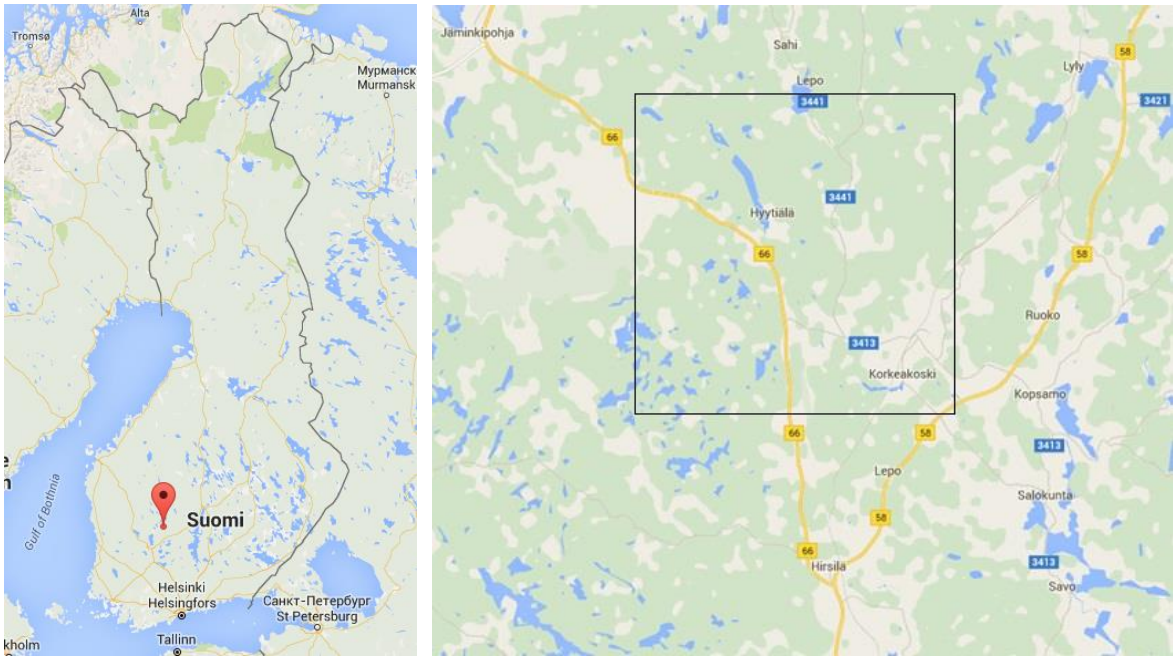

Figure 8 *Location of the test site (Map data ©2015 Google)*

The used area is 10 kilometers wide and 11 kilometers high. The exact coordinates can be found in Table 1. The area includes all the desired land cover classes. Most of the area is dominated by different types of forests. Smaller areas of lakes, agricultural land and wetlands are also present. Additionally, some urban areas can be found (Juupajoki).

Table 1 *Location and size of the test area*

Coordinate – top left	61°52'48.75" N, 24°12'13.66" E
Coordinate – bottom right	61°47'7.07" N, 24°24'7.67" E
Size	10 x 11 km

The classification methodology was developed with the help of this test area, but the aim is to be able to apply the same methodology also on a larger scale. However, this will require some additional work on the methodology.

3.2 Reference data

The reference data serve as a tool to create a training data set and to validate the results of the classification. Primarily, a very high resolution GeoEye image from the Hyytiälä area (Figure 9, left) was used for this purpose. The used image was pre-processed. An extract from the CORINE Land Cover (CLC) data base was also used as another way of validating the results (Figure 9, right).

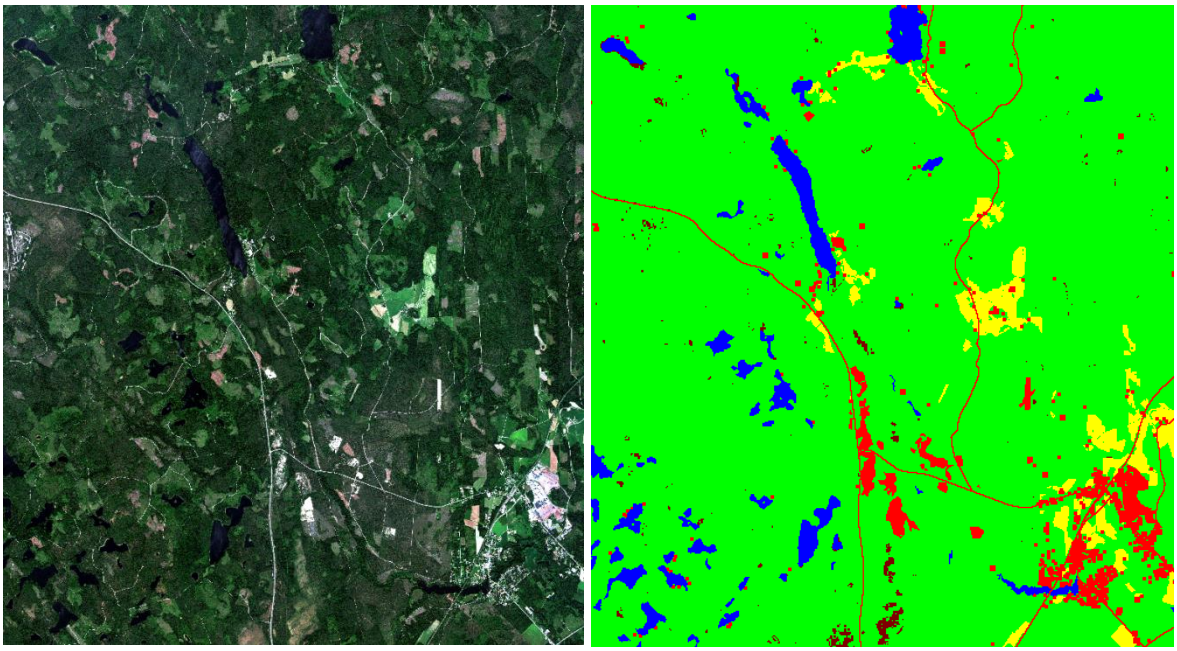

Figure 9 *GeoEye image (left) and CORINE Land Cover (right) extract used as reference images*

Some basic information about the GeoEye reference image can be seen in Table 2. The image was taken in 2010, so it has to be assumed that no change in the land cover has happened compared to the images that are used for classification. These images were acquired during 2014 and 2015. The panchromatic band of the GeoEye reference image can be used to get a very high spatial resolution.

Table 2 *Properties of the reference image*

Acquisition date	4.7.2010
Bands	Red, green, blue, NIR, panchromatic
Pixel size	2 m/0,5 m (panchromatic)
Area extents	5500 x 5000 pixels/22 000 x 20 000 pixels

As the GeoEye image does not include any information about the land cover classes, only a visual interpretation, an extract from the CORINE Land Cover (CLC) data base was used as another way of validating the results. CORINE (COoRdination of INformation on the Environment) is a data collection on land in Europe with common nomenclatures and methodologies. The used data are from 2012. CLC2012 has a pixel size of 20 meters which matches the data used for the classifications.

The standard CLC2012 data have three hierarchy levels. The first level has five classes: artificial surfaces, agricultural areas, forests and semi-natural areas, wetlands, and water bodies. The second level has 15 classes and the third level has 44 sub-classes. (Manakos

and Braun 2014, 55-74) The Finnish CLC2012 has omitted some of the classes that do not exist in Finland, and additionally, the Finnish CLC2012 has a national fourth level, resulting in a total of 48 classes.

The GeoEye image needed to be processed a bit further to be able to be used in validation of the classification results, including visual interpretation and a definition of training and test plots, whereas the CLC2012 was ready to be used. However, a visual interpretation and a comparison to more recent imagery show that the CLC2012 is not accurate. Faulty land cover classes can be found, especially in pixels around borders of areas. The GeoEye image has better spatial resolution which allows better interpretation. For these reasons, the CLC2012 was mainly used only as a help in the visual interpretation of the classification results. It was also tested for validation with some of the classification results. The GeoEye data were used as the primary source of validation as it is more reliable.

3.3 SAR and optical satellite data

The data used for the land cover classification in this project were two types of satellite data: SAR and optical satellite data. A time series of Sentinel-1 images was used as the SAR data and a Landsat 8 image was used as the optical data. All images were pre-processed, which is shortly presented in section 4.1. A smaller area, around the Hyytiälä Forestry Station, was extracted from all images for the classification.

The European Space Agency launched Sentinel-1A on 3 April 2014 on a Russian Soyuz rocket from Europe's Spaceport in French Guyana (European Space Agency). Sentinel-1, or the European Radar Observatory, is a polar orbiting two-satellite constellation. Sentinel-1B is expected to be launched in 2016. The goal of the mission is to provide free and open data access to the medium resolution data. Emphasis has been put on reliability and fast data flow. The satellites use the C-band and have a revisit time of 6 days. They work on an altitude of 693 km. Sentinel-1 has four operational imaging modes, where the main mode is the Interferometric Wide-swath mode. This mode supports operation in dual polarization. (Torres et al. 2012)

The used Sentinel-1 data consisted of 48 images with acquisition dates ranging from October 2014 to August 2015. Both VV and VH polarizations were available, resulting in 24 images per polarization. The images are multi-looking imagery and they are amplitude images. Table 3 summarizes the image properties of the Sentinel-1 data. The acquisition conditions were obtained from the open data provided by the Finnish Meteorological Institute (https://ilmatieteenlaitos.fi/avoin-data).

Table 3 *Properties of Sentinel-1 images*

Acquisition dates and conditions at the Juupajoki Hyytiälä weather station (Finnish Meteorological Institute)	**Date**	**Temperature (mean)**	**Precipitation**	**Snow depth**
	4.10.2014	10.2 °C	0.1 mm	
	9.11.2014	2.7 °C	0.1 mm	9 cm
	21.11.2014	-3.2 °C	0.6 mm	
	3.12.2014	1.7 °C	N/A	
	15.12.2014	1.1 °C	4.2 mm	5 cm
	27.12.2014	-7.6 °C	0.0 mm	14 cm
	8.1.2015	-0.8 °C	3.5 mm	22 cm
	13.2.2015	-5.2 °C	0.7 mm	40 cm
	25.2.2015	1.1 °C	0.0 mm	35 cm
	9.3.2015	3.0 °C	N/A	23 cm
	21.3.2015	-5.0 °C	N/A	15 cm
	2.4.2015	0.7 °C	0.2 mm	15 cm
	14.4.2015	1,4 °C	0.4 mm	
	26.4.2015	3.9 °C	7.3 mm	
	8.5.2015	8.9 °C	1.1 mm	
	20.5.2015	10.0 °C	0.0 mm	
	1.6.2015	10.9 °C	1.3 mm	
	13.6.2015	14.1 °C	0.0 mm	
	25.6.2015	13.3 °C	1.8 mm	
	7.7.2015	12.3 °C	9.3 mm	
	19.7.2015	13.3 °C	4.1 mm	
	31.7.2015	13.9 °C	0.3 mm	
	12.8.2015	16.7 °C	N/A	
Orbit number	57			
Orbit	Ascending			
Polarization	Dual (VV and VH)			
Band	C			
Pixel size	20 m			
Area extents	200 x 200 km 10 000 x 10 000 pixels			
Coordinate of top left corner	62°38'10.52" N, 22°15'30.04" E			

The National Aeronautics and Space Administration (NASA) launched its first satellite in the Landsat series in 1972. Afterwards, more satellites have been launched and until this day, the Landsat program has provided continuous data of the Earth's land areas. (Lauer et al. 1997) The newest satellite in the program is Landsat 8, launched in February 2013. This satellite was developed in cooperation with the Department of the Interior U.S. Geological Survey (USGS). The two sensors onboard Landsat 8 are the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). These produce more than 500 image scenes per day. The goal of Landsat 8 is to continue the data flow from the program's satellites, while maintaining consistent and comparable data to the previous systems. (Roy et al. 2014)

The used Landsat data were one Landsat 8 image, taken 23 July 2014. Eight bands were available. None of the pixels within the test area were affected by clouds. Table 4 summarizes the image properties of the Landsat 8 image.

Table 4 *Properties of Landsat 8 image*

Acquisition date	23.7.2014		
Bands	**1**	Coastal aerosol	0.43 – 0.45 µm
	2	Blue	0.45 – 0.51 µm
	3	Green	0.53 – 0.59 µm
	4	Red	0.64 – 0.67 µm
	5	Near Infrared (NIR)	0.85 – 0.88 µm
	6	SWIR 1	1.57 – 1.65 µm
	7	SWIR 2	2.11- 2.29 µm
	9	Cirrus	1.36 – 1.38 µm
Pixel size	20 m		
Area extents	200 x 200 km 10 000 x 10 000 pixels		
Coordinate of top left corner	62°38'10.52" N, 22°15'30.04" E		
Path	190		
Row	16-17		

These data sets were used to test the developed methodology, both separately and together. The results were mainly compared with information extracted from the GeoEye image. An additional validation, mainly visual, was performed with the CLC2012.

4 Methods

The main focus of this study was the development of a classification methodology for land cover mapping. The aim was to combine SAR and optical satellite data, and at the same time take advantage of the time series of the SAR data. The algorithm development was mostly done with MATLAB. Some other programs, such as ER Mapper, was used for smaller tasks, but the main work was done with MATLAB.

The land cover classification consisted of many different stages and experiments. These included the training phase, feature extraction and selection, data fusion and the actual classification. The next sections will introduce the different stages of the process.

4.1 Satellite data pre-processing

Satellite data need to be pre-processed before it can be used further. In this project, all satellite imagery was obtained already pre-processed. The pre-processing had been performed with in-house VTT software. Nevertheless, the next section gives a brief introduction to how the images had been pre-processed.

An atmospheric correction was performed on the GeoEye image. It was done with the SMAC (Simplified Method for Atmospheric Correction) package. This method is explained by Rahman and Dedieu (1994). The aerosol optical density (AOD) was estimated to be 0.1.

The Sentinel-1 data were downloaded from Sentinels Scientific Data Hub (scihub) of ESA. Ortho-rectification and a radiometric correction were performed. This was done with open source digital elevation model (DEM) data. Next, scenes with pixels possibly within the test site were searched for in each epoch in the stack. This was done by first listing the center points of the scenes and then choosing those scenes with center points less than 400 kilometers from the test site center. The unique orbit number modulo 175, the number of orbits per orbit repeat period for Sentinel-1, was used to identify the orbit configurations. Stacks were formed out of the chosen scenes for those orbit configurations that had at least one scene center within 100 kilometers from the stack center. Only scenes with the same orbit number modulo 175 were considered in the stacks, and this was also used as the stack identifier. The used stack has the number 57. All scenes were listed per date and a composite layer was made for each date. Dates with valid pixels covering over 50 percent of the whole test site were included in the final stack.

The Landsat data were downloaded from the Earth Explorer server of USGS. Two Landsat frames from the same orbit were chosen. These were pre-processed with the Envimon software, developed at VTT Technical Research Centre of Finland. The atmospheric correction was done with the SMAC method. To complete the calculations, the aerosol optical density needed to be estimated, and in this case, the calculations were iterative. AOD values were calculated separately for all pixels. The images were then resampled to a 20 meter pixel grid in the UTM zone 35. This was performed with the nearest neighbour method with ER Mapper 2013. Finally, the two frames were combined. The cloud masking was done manually with the help of band 9, the cirrus band.

4.2 Image analysis and data classification

The actual main part of this work was the development of a classification methodology for the multi-sensor and multi-temporal data in hand. All methods used supervised classification. This section will explain the different stages of this classification and the methods that were used.

Five classes were used for the classification: urban areas, agricultural areas, forest areas, wetlands and water bodies. These coincide with the five level 1 classes of CORINE Land Cover. These classes are actually a mix of land cover and land use classes (Manakos and Braun 2014, 55 - 74).

4.2.1 Training and test data

The training and test data sets for this study were created manually with the primary reference image, the GeoEye image. A grid of plots 500 meters apart was created. Each plot had the size of 60 meters which resulted in 3 x 3 pixels in the classification data. Plots were compared to the reference image and labeled with the right class using visual interpretation. Only plots completely covered with a single land cover class were used. However, to get a sufficient amount of plots for each class, the grid needed to be made denser at locations belonging to a certain class. In those areas, plots were spaced 100 meters apart. In total, this resulted in 618 plots (Figure 10), where 385 plots were completely in one land cover class. Most of these were forest areas.

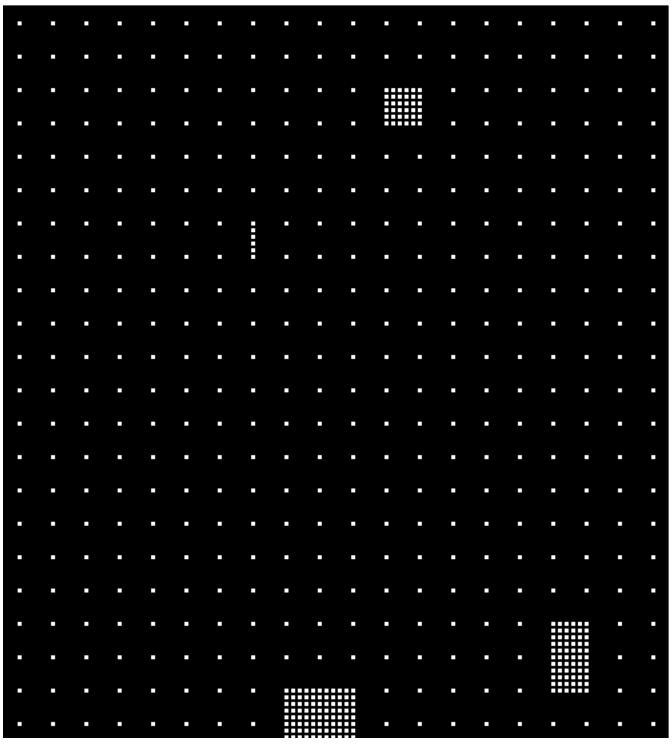

Figure 10 *Grid used for visual interpretation of plots. The denser areas are for agricultural areas (square on top), water bodies (line), urban areas (rectangle in right lower corner) and wetlands (rectangle on bottom).*

A training set and test set of five plots per class were formed, i.e. ten plots per class needed to be chosen. This can be done by ordering the plots according to some parameter and choosing every x^{th} plot to find ten plots. In this case, values from a Sentinel-1 image were

used. This image was chosen by taking the image with the biggest range in the amplitude values. In this case, that was the VV polarized image from 14.4.2015. The mean value for each plot was then calculated and used for comparison. The ten plots were then chosen from this ranked list.

4.2.2 Time series

The Sentinel-1 data used in this research have been collected during a time period of about a year. This time perspective gave a valuable insight to the dynamics of the area under investigation and helped to create better land cover classifications. It is important to examine the changes in the images over time to be able to explain the classification results.

To better understand the dynamics of the time series, the images were transformed to various figures. The created figures were line plots with markers representing different dates, i.e. different images. This means that one marker is a value calculated from one image, for example the mean value. Markers can be values calculated from the whole image or only using the training and test points. When statistics from the different classes were desired, only the training and test points were used.

4.2.3 Feature extraction

Feature extraction was performed on both the Sentinel-1 images and the Landsat 8 image. If the original images would be used directly for the classification, each Sentinel-1 polarization with 24 images would lead to a 24-dimensional feature vector, and the Landsat 8 image with 8 usable bands, would lead to an 8-dimensional feature vector. With feature extraction, these original values were instead used to calculate some new fewer features, leading to a smaller dimensionality for the feature vector. Hence, feature extraction helped to reduce the dimensionality while keeping much of the information.

The two data sets were handled separately and with varied techniques, leading to different kinds of features. Eight features were calculated for the Sentinel-1 data and four features were calculated for the Landsat 8 image. The features were calculated pixel-wise. For the Sentinel-1 data this meant that the values of a pixel in the time series were used to form one feature value. For the Landsat 8 image, the values for a pixel of the different bands were used to calculate the feature value.

Some statistical measures were used as features for the Sentinel-1 time series data. The freely available Time Series Analysis toolbox for Matlab was used for these calculations (http://pub.ist.ac.at/~schloegl/matlab/tsa/). The calculated measures included different basic statistics: minimum and maximum value, root mean square, mean value, standard deviation, skewness, kurtosis, and entropy. These measures characterize pixels and form feature vectors which are then used for the classification.

The first two features were the minimum and maximum values. The minimum and maximum value a pixel had in the time series was taken as the feature. This resulted in two values in the feature vector.

The root mean square (RMS) is, as the name implies, the square root of the arithmetic mean of the squares of a set of numbers. For a pixel p at (i, j) in a stack of m images with values $p_{ij1} \dots p_{ijm}$, the root mean square is as follows

$$\boldsymbol{RMS} = \sqrt{\frac{\sum_{k=1}^{m} \boldsymbol{p_k}^2}{\boldsymbol{m}}} \quad \textbf{(8)}$$

Moments are quantitative measures to describe the shape of a set of points. Four moments were used as features: mean (first moment), standard deviation (derived from the second moment), skewness (third moment), and kurtosis (fourth moment). The mean tells the average value of a set of values. For a pixel p at (i, j) in a stack of m images with values $p \ldots p_{ijm}$, the mean is as follows

$$\bar{\boldsymbol{p}} = \frac{\boldsymbol{1}}{\boldsymbol{m}} \sum_{\boldsymbol{k=1}}^{\boldsymbol{m}} \boldsymbol{p_k} \quad \textbf{(9)}$$

To characterize the variability, the standard deviation can be used. It tells how the values are spread out around the mean value. The standard deviation is the square root of the variance and is calculated as follows

$$\boldsymbol{\sigma} = \sqrt{\frac{\boldsymbol{1}}{\boldsymbol{m-1}} \sum_{\boldsymbol{k=1}}^{\boldsymbol{m}} (\boldsymbol{p_k} - \bar{\boldsymbol{p}})^2} \quad \textbf{(10)}$$

Skewness, the third moment, describes the asymmetry of a distribution around its mean. A positive value means the distribution has an asymmetric tail towards the positive side (from the mean value) whereas a negative value has a tail towards the negative side. The skewness is defined as

$$Skewess = \frac{1}{m} \sum_{k=1}^{m} \left[\frac{p_k - \bar{p}}{\sigma}\right]^3 \quad \textbf{(11)}$$

The kurtosis describes the relative peakedness or flatness of the distribution. The comparison is done with the normal distribution. A positive kurtosis has a sharper peak whereas a negative kurtosis has a wider peak. The kurtosis is calculated as follows

$$\boldsymbol{Kurtosis} = \left\{\frac{\boldsymbol{1}}{\boldsymbol{m}} \sum_{\boldsymbol{k=1}}^{\boldsymbol{m}} \left[\frac{\boldsymbol{p_k} - \bar{\boldsymbol{p}}}{\boldsymbol{\sigma}}\right]^{\boldsymbol{4}}\right\} - \boldsymbol{3}$$. (Press 1992) **(12)**

The eighth feature used for the Sentinel-1 data is entropy. Entropy is a statistical measure of randomness and is defined as follows

$$\boldsymbol{Entropy} = -\sum \boldsymbol{pdf} * \boldsymbol{log}(\boldsymbol{pdf}) \quad \textbf{(13)}$$

where pdf is the empirical probability distribution, calculated from the data (Shannon 1948).

The Landsat 8 image was described with the help of spectral indices which then could be used as features. These indices use the values from two bands to calculate an index that

describes the pixel. The indices are designed to accentuate some special land cover class. Four different indices were used as features.

The Normalized Difference Vegetation Index (NDVI) characterizes live green vegetation. It was first introduced by Rouse et al. (1973). The index uses the visible red band and the near-infrared band, and is calculated as follows

$$\boldsymbol{NDVI = \frac{NIR - RED}{NIR + RED}} \quad \textbf{(14)}$$

where RED corresponds to surface reflectance in Landsat 8 band 4 and NIR corresponds to surface reflectance in Landsat 8 band 5. This index works with the principle where vegetation results in a large difference in the two used bands. This leads to large NDVI values. On the contrary, soil results in values close to zero and water bodies result in negative values.

The Normalized Difference Water Index (NDWI) is used to find open water areas in remotely sensed digital imagery. The calculations are based on visible green light and the near-infrared radiation. The NDWI is expressed as follows

$$\boldsymbol{NDWI = \frac{GREEN - NIR}{GREEN + NIR}} \quad \textbf{(15)}$$

where $GREEN$ corresponds to surface reflectance in Landsat 8 band 3 and NIR corresponds to surface reflectance in Landsat 8 band 5. Water features will have positive values, whereas vegetation and soil will have zero or negative values. (McFeeters 1996) One drawback with this index is that built-up areas tend to result in similar values as water areas (Xu 2007).

The Normalized Burn Ratio (NBR) is most commonly used to find burned areas and burn severity. The bands where non-incendiary vegetation or incendiary vegetation are the most visible are used. These are near-infrared and short-wave infrared. The NBR is calculated as follows

$$\boldsymbol{NBR = \frac{NIR - SWIR\,2}{NIR + SWIR\,2}} \quad \textbf{(16)}$$

where NIR corresponds to surface reflectance in Landsat 8 band 5 and $SWIR\,2$ corresponds to surface reflectance in Landsat 8 band 7. (García and Caselles 1991)

The Normalized Difference Built-up Index (NDBI) is used to find built-up areas. It uses the knowledge of the reflectance of these areas, where the short-wave infrared (SWIR) values are much larger than the near-infrared (NIR) values. The NDBI is calculated as follows

$$\boldsymbol{NDBI = \frac{SWIR\,1 - NIR}{SWIR\,1 + NIR}} \quad \textbf{(17)}$$

where $SWIR\,1$ corresponds to surface reflectance in Landsat 8 band 6 and NIR corresponds to surface reflectance in Landsat 8 band 5. The advantage of this index is that the reflectance properties of built-up areas do not change during the year. However, some other areas might have similar reflectance properties during some times of the year. (Zha et al. 2003)

4.2.4 Feature selection

Feature selection was performed on both the SAR and optical data sets. Fisher's discriminant ratio and stepwise regression were applied on the Sentinel-1 data, whereas a knowledge-based way of choosing bands from the Landsat 8 image was used.

The FDR is the ratio of the between-class variance to the within-class variance. A large FDR, that is a high separability, is the result of a large between-class variance and a small within-class variance. The between-class variance is calculated by using the mean values of the data points in the classes. A large difference in the mean values means that the classes are separable. For the multi-class case, the between-class variance S_B is calculated as follows

$$\boldsymbol{S_B = \sum_{i=1}^{c} m_i(\mu_i - \mu)\,(\mu_i - \mu)^t} \qquad \textbf{(18)}$$

where m_i is the number of samples in class i, μ_i is the mean value of class i, and μ is total mean of all samples. The within-class variance S_W is calculated as follows

$$\boldsymbol{S_W = \sum_{i=1}^{c} \sum_{x_k \in class\ i} (p_i - \mu_i)(p_i - \mu_i)^t} \qquad \textbf{(19)}$$

where p_i are the samples of class i. Fisher's discriminant ratio is then calculated using these variances. This separability measure is defined as follows

$$FDR = \frac{S_B}{S_W}.$$ (Duda et al. 2012, 44-51) **(20)**

Feature selection of the SAR data was processed with the two different methods, but also applied on different SAR data. First, eight features were calculated from both VV and VH polarized data and a feature selection was done, reducing the features to four. Second, a scene selection was done with the feature selection methods. First, all 48 scenes were used as features and the best scene was chosen. Second, the 48 scenes (24 per polarization) were reduced to four scenes. Both polarization options for one date were regarded as one feature in this scene selection, but the classifications were later done with separate scenes.

After Fisher's discriminant ratio and stepwise regression were performed on the SAR data, the results were compared. The comparison was done with the classification results when using only SAR data and the features resulting in a better classification were used in the data fusion phase.

In addition to these methods, a feature (band) selection was performed on the Landsat 8 image. This selection was not based on any calculations, rather on prior knowledge of the band characteristics. Also in this case four features (bands) were selected.

4.2.5 Dimensionality reduction

A dimensionality reduction was performed by doing a principal component analysis. The principal components were calculated separately for the Sentinel-1 data and the Landsat 8 image. The following classifications were also done separately.

All 48 images (including both polarizations) from the Sentinel-1 data were used in the PCA calculations. The first three principal components were extracted which should give the most important information content these 48 images have. Thereafter, this 3-dimensional data set was used in the classification stage. A similar experiment was performed on the Landsat 8 data. The principal components were calculated from bands 2-7 and the first three components were chosen. Subsequently, this 3-dimensional data set was used to classify the image. The resulting classifications are presented in the next chapter.

4.2.6 Data fusion

One important aspect of this thesis is the data fusion and how it affects the classification results. In this study, the fused data was the Sentinel-1 data and the Landsat 8 data, i.e. SAR and optical data. The different characteristics of these data sets should complement each other to reach better results.

In this project, the data fusion was performed on both pixel and feature level. From both data sets, both the original pixel values and extracted features were used. These were fused together in different combinations. For example, four SAR features could be fused with four optical features, or with four reflection values from different bands. The different fusion combinations are presented in section 4.2.7. In all cases, four values/features per data set were used, making the feature vector 8-dimensional in all cases (Figure 11). The fusion was done by simply combining the four values from the different data sets to one feature vector. This means that every pixel used in the classification had eight values.

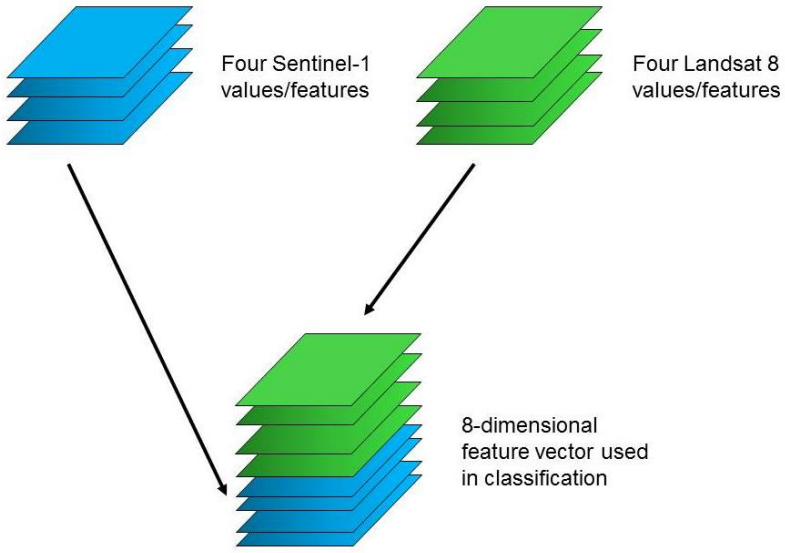

Figure 11 *Fusion of four values/features from each data set to an 8-dimensional feature vector*

Data fusion can refer to different ways of combining data. In this study, the data fusion refers to fusing SAR and optical data. Other options, such as fusing data from the same sensor but different acquisition times, can also be considered data fusion. However, it is the multi-sensor data fusion that is the focus of this work.

4.2.7 Classifications

Two classification algorithms were used in this study: the nearest neighbour rule and the maximum likelihood classifier. Three different variations of the nearest neighbour rule were used. All data sets were classified with these four different methods.

Three versions of the nearest neighbour rule were used in this study. The first two were k-NN with k=1 and k=9. This means that the first one used one neighbour for the classification (1-NN), and the second one used nine neighbours (9-NN). The third method used mean values of the training areas (later referred to as MDM). In this method, the mean value of all training pixels within a class was calculated and these values were compared to the pixels with the unknown classification to assign a class. The mean values were calculated separately for each image. The most similar mean value, compared to the pixel with the unknown class, decided the assigned class. This classifier is also called Minimum-Distance-to-Means (Lillesand et al. 2008, 551).

In combination to these algorithms and their variations, a majority voting was used to remove unnecessary "noise" in the results. This was done by moving a 3x3 window over the resulting classification image and the middle pixel was assigned to the class that the majority of the pixels within the window belong to. If two or more classes were equally represented, no change was made. Additionally, the border pixels were not changed.

These classification methods were applied on the different data sets to gain a land cover classification. A summary of the experiments is visualized in Figure 12. It can be seen that the different data sets were used both individually and together, as well as in their original form and with features extracted from them. Additionally, different classification algorithms were applied on these data sets and combinations.

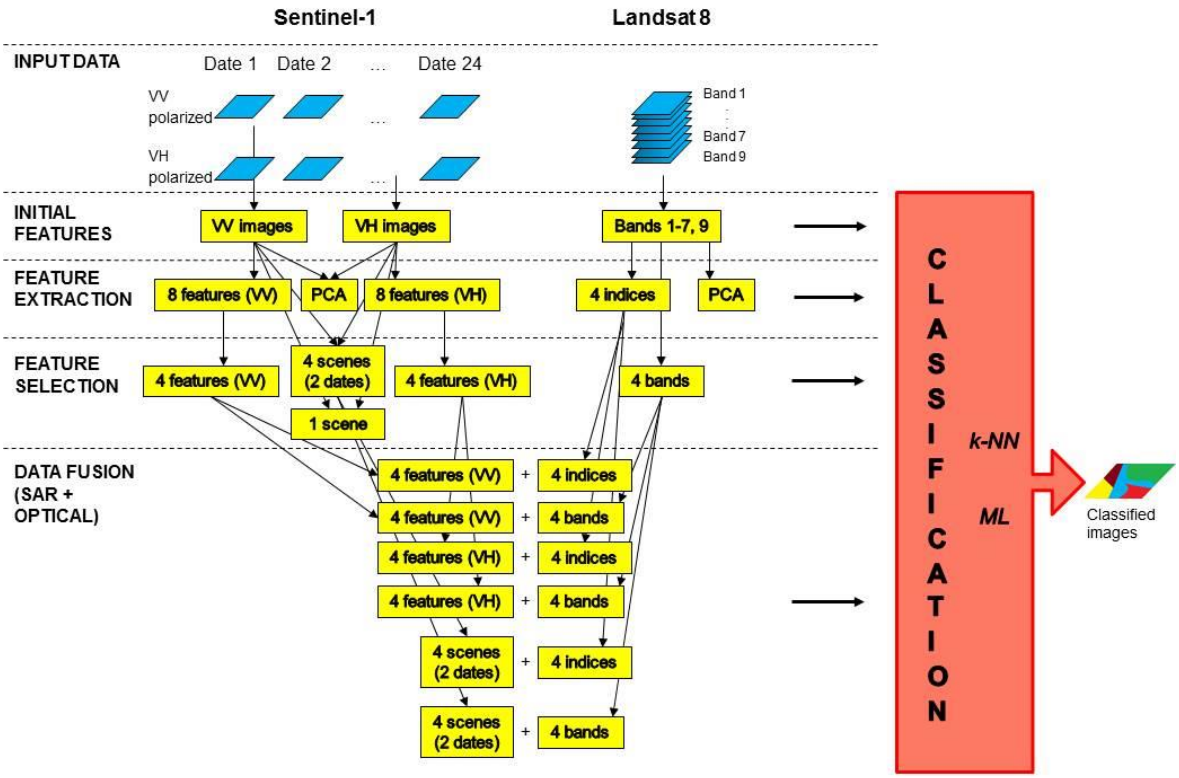

Figure 12 *The different classifications performed in this study*

All these different data combinations and classification methods resulted in big amount of classification results. The aim was to find the optimal method. The Results section will present the outcome and a final evaluation can be found in the Discussion section.

4.2.8 Validation methods

A validation of the classification results was done with confusion matrices and various accuracy calculations. A confusion matrix was calculated for each classification map. This also included values for the final accuracy, the producer's accuracy and the user's accuracy.

In this study, the confusion matrix used the test plots that were manually defined in an earlier stage. This kind of validation is called stratified validation. Each class had five test plots with nine pixels, leading to 45 test pixels per class. Using the same amount of test pixels per class helped to give an overview of the results for each class. However, the actual classes were not the same size in the test image and another way of defining the test plots could lead to different results. For example, if test plots were chosen by location rather than by class, the forest class would most probably dominate. This should be taken into consideration when doing a final validation of the results.

The test plots were the main validation data, but a few classifications were also compared to CLC2012. This gave a chance to get an estimate of the classification success for every pixel in the image and not only for the test areas. However, CLC2012 was found to not be very accurate, and for this reason, all validation was not performed with it. Even so, it gave good idea of the approximate accuracy of the whole image.

5 Results

The developed classification methodology for multi-sensor satellite data has resulted in various results. Different data sets and classification methods have been examined, as explained in chapter 4. This section will present the most significant results gained in this study. It first explains the training phase, and feature extraction and selection. It then continues with the actual classification results. Classification results from single-sensor classifications are explained, using either SAR or optical data. Additionally, the results of data fusion and classifications of those data sets are presented. Accuracy assessments for all classifications are also shown.

5.1 Training phase

The training and test sets were formed according to the method introduced in section 4.2.1. This resulted in five training plots and five test plots per class. Each plot had nine pixels, leading to 45 training pixels per class.

To ensure that the used training and test sets represent the data adequately, it is good to examine them a bit more. In this case, some figures to visualize the chosen training and test pixels were created (Figure 13). The figures were created with pixel-wise mean values of all Sentinel-1 VV polarized versus all Sentinel-1 VH polarized images. All pixels within the plots chosen for the training and test sets were used.

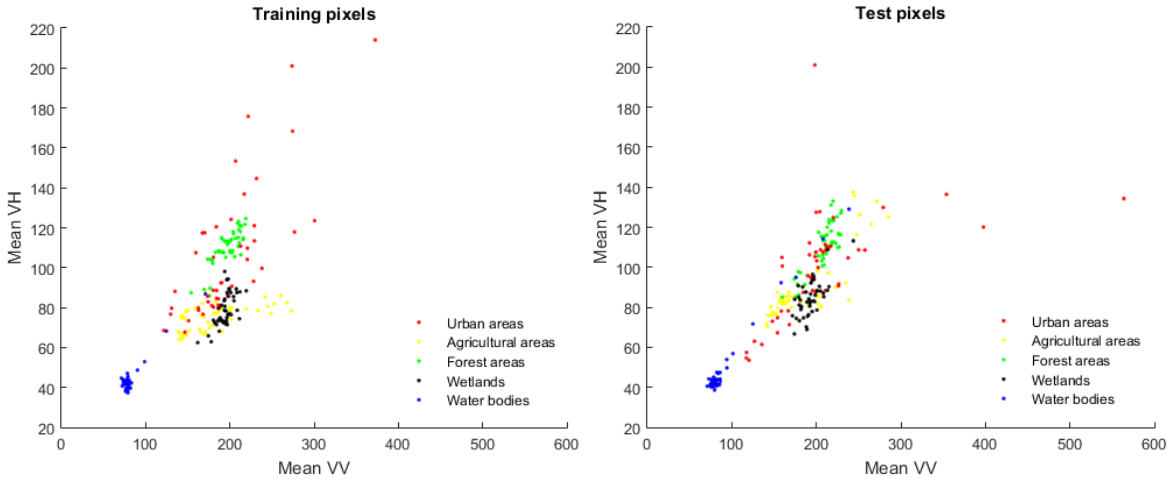

Figure 13 *Visualization of training (left) and test area pixels (right) with the mean value of the pixel from the VV polarized Sentinel-1 images (horizontal axis) and from the VH polarized Sentinel-1 images (vertical axis)*

The figures show the class separability for the training and test samples. It can be seen that the water class is well separated; only a few single pixels mix with the other classes. The urban areas are very spread out, both on top of the other three classes and outside. Out of these three classes, the forest class is mostly separated from the other two. The agricultural areas and wetlands overlap a bit and for the test pixels, and the agricultural class also overlaps with the forest class.

5.2 Time series

Time series can be used to characterize data and to give a better overview of the changes over time. The Sentinel-1 data used in this study was a stack of 48 images, 24 for each polarization, spanning over almost one year. This section presents some time series figures created from this data. The VV and VH polarized data were processed separately.

A general overview of the data can be seen by plotting the mean value and standard deviation for each image. Figure 14 shows the result of calculating the mean value and standard deviation for all pixels (amplitude values) in one image. These were calculated separately for the VV and VH polarized images, but it can be seen that both data sets follow a very similar trend. Only the magnitude is different. The mean values vary in the first half, whereas they stay more constant towards the end. They also increase on the second half. The standard deviation is quite constant during the whole time series, except for some variations in November to January.

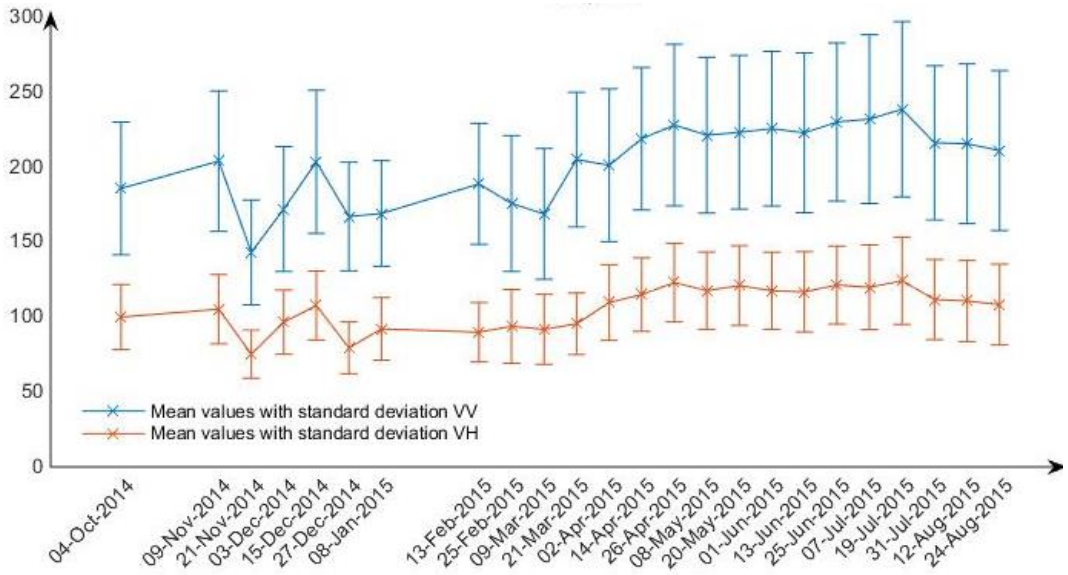

Figure 14 *Mean value and standard deviation for each image in the Sentinel-1 image stack*

Another way to examine the time series is to look at the different classes. The training and test pixels were used to create figures with mean values and standard deviation values for pixels in different classes. Each value was calculated from the 90 training and test pixels in a certain class from that image. The resulting time series figures for the VV polarized Sentinel-1 images are visualized in Figure 15.

Looking at the mean values (upper figure), it is apparent that the water bodies have the most differing values. The other classes follow a quite similar pattern with quite similar values. One noticeable disruption is the sudden drop in the wetlands' and agricultural areas' mean values on 25.2.2015 and 9.3.2015. Another class, the urban areas, stands out in the figure of standard deviations. It mostly has higher values than the other classes. The agricultural areas also show some higher values, especially from the beginning of April until the end of June. The other classes are very similar and do not have a great variability in the values throughout the time series.

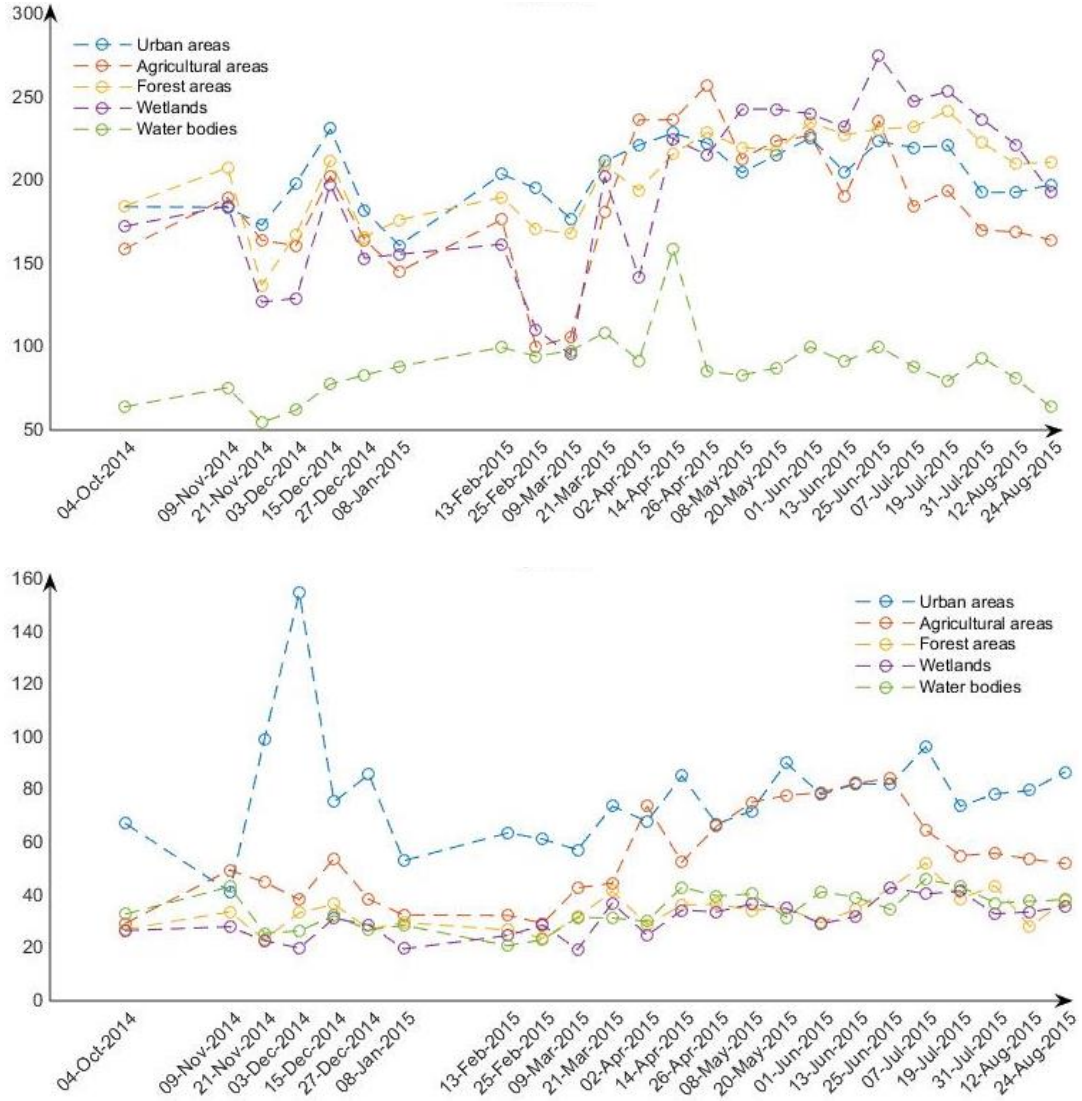

Figure 15 *Mean values (up) and standard deviation (down) for training and test pixels in each class for the VV polarized Sentinel-1 images*

The same graphs have been produced for the VH polarized Sentinel-1 data (Figure 16). Again the mean values of the water bodies are lower than the mean values of the other classes. The wetlands and agricultural areas have a very similar pattern, whereas the urban areas and forest areas follow this pattern only during certain times. They differ noticeably from the end of December to the beginning of March and the forest areas also show higher values during the rest of the time series. In the corresponding graph for the standard deviation values, the urban areas are again separated from the other classes, whereas the other classes are very similar. Only the agricultural areas have a higher peak during May. This coincides with a peak in the urban areas and a small peak in the forest areas.

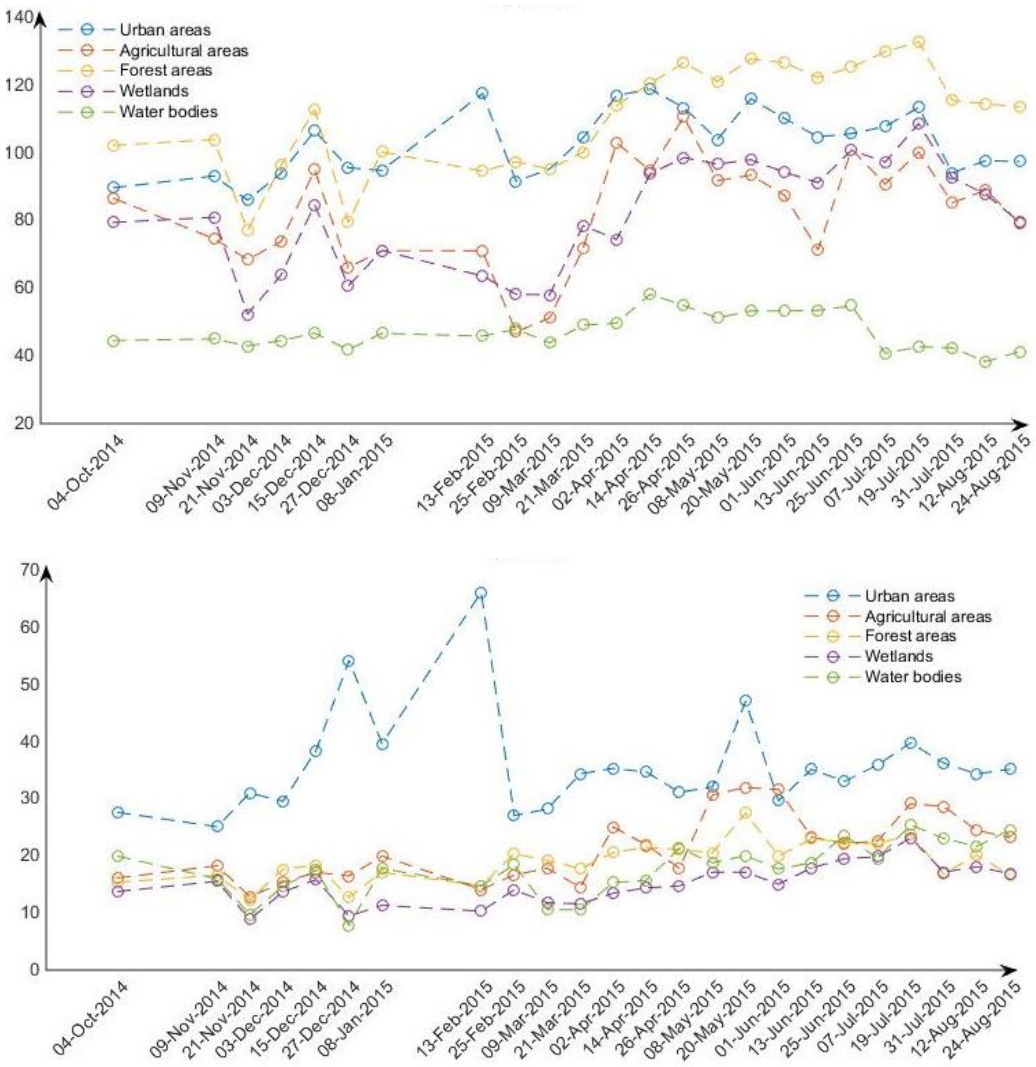

Figure 16 *Mean values (up) and standard deviation (down) for training and test pixels in each class for the VH polarized Sentinel-1 images*

5.3 Feature extraction and selection

In an attempt to reduce the dimensionality of the used data, a feature extraction was performed on both data sets (Sentinel-1 and Landsat 8). The features introduced in section 4.2.3 were calculated. This resulted in eight Sentinel-1 features per polarization and in four Landsat 8 features. These were then used in further steps of the classification.

Feature selection was performed on the extracted SAR features and as a scene selection for the SAR images. The selection was done with both Fisher's discriminant ratio and with stepwise regression. The calculations were done with features normalized to a zero mean and a standard deviation of one. Only training and test data were used for the feature selection as the class of the input data needed to be known. In the stepwise regression, the entrance tolerance was set to 0.05 and the exit tolerance was set to 0.10. Both polarizations' data from one date were combined to one feature during the feature selection, except when selecting only one scene.

The extracted SAR features were first ranked with Fisher's discriminant ratios (Table 5). The four features with the highest values were chosen for the classification. These features are the minimum value, the mean value, the root mean square and the entropy. A ranking was also performed with stepwise regression (Table 6). The number of iterations was limited to four to assure the right amount of features. Four features were added to the model at this stage. The p-values changed at each iteration so the four smallest values do not belong to these four features. The chosen features were the standard deviation, the maximum value, the minimum value and the kurtosis.

Table 5 *Fisher's discriminant ratios for SAR features calculated from both VV and VH polarized data*

Feature	Fisher's discriminant ratio
Minimum value	0.7451
Mean value	0.3925
Root mean square	0.3685
Entropy	0.2833
Maximum value	0.2390
Standard deviation	0.2270
Kurtosis	0.1138
Skewness	0.0986

Table 6 *p-values in the stepwise regression for SAR features after four iterations (polarizations combined)*

Feature	p-value
Mean value	0.5818
Standard deviation	0.00056436
Entropy	0.0585
Skewness	0.0037
Maximum value	0.000015315
Minimum value	0.0128
Root mean square	0.5242
Kurtosis	0.0090

A similar feature selection was performed on the unmodified SAR scenes. First, the best scene from the available 48 scenes was chosen. When using Fisher's discriminant ratio, this was the VV polarized scene from 19.7.2015 (Table 7). The results show that one of the polarizations does not seem to be significantly better. In the calculations for the stepwise regression, the iterations were limited to one. This led to choosing the VV polarized scene from 2.4.2015 (Table 8).

Table 7 *Ten best Fisher's discriminant ratios for dates of SAR scenes (polarizations not combined)*

Date	Polarization	Fisher's discriminant ratio
19.7.2015	VV	1.4542
7.7.2015	VH	1.4272
9.3.2015	VH	1.4269
9.11.2014	VV	1.4111
26.4.2015	VV	1.4046
4.10.2015	VV	1.3160
25.2.2015	VV	1.2889
26.4.2015	VH	1.2770
15.12.2014	VV	1.2521
2.4.2015	VH	1.2399

Table 8 *p-values in the stepwise regression for SAR scenes after one iteration (polarizations not combined)*

Date	p-value (VV polarized)	p-value (VH polarized)
4.10.2014	3.5251e-04	2.1269e-04
9.11.2014	3.4417e-04	1.3138e-05
21.11.2014	2.1875e-05	1.1045e-15
3.12.2014	0.0013	1.2873e-09
15.12.2014	6.0641e-07	2.4923e-09
27.12.2014	0.0018	4.8185e-06
8.1.2015	0.9141	0.0052
13.2.2015	6.9121e-09	1.9245e-12
25.2.2015	1.8366e-08	0.0018
9.3.2015	0.0035	1.7587e-07
21.3.2015	0.6940	9.8256e-08
2.4.2015	7.4595e-64	1.2032e-13
14.4.2015	5.7621e-05	2.6886e-07
26.4.2015	5.5596e-04	1.2440e-08
8.5.2015	0.0185	0.0778
20.5.2015	0.0133	0.0115
1.6.2015	0.0765	7.8414e-05
13.6.2015	0.0457	0.0211
25.6.2015	0.0353	4.0642e-04
7.7.2015	0.0343	2.3939e-05
19.7.2015	0.9422	0.0077
31.7.2015	0.0137	0.2815
12.8.2015	0.0395	0.0147
24.8.2015	0.6980	0.0885

A scene selection for finding the best four scenes was performed next. As polarizations were combined for one date, 24 scenes were used. The aim was to find the best two dates. The results from Fisher's discriminant ratio can be seen in Table 9, where the ten best ratios are visible. This method led to choosing scenes from 25.2.2015 and 9.3.2015. When

using the stepwise regression, the dates were 13.6.2015 and 19.7.2015. In this case the iterations were limited to two. The resulting p-values can be seen in Table 10.

Table 9 *Ten best Fisher's discriminant ratios for dates of SAR scenes (polarizations combined)*

Date	Fisher's discriminant ratio
25.2.2015	0.7996
9.3.2015	0.6148
19.7.2015	0.5996
3.12.2014	0.4996
24.8.2015	0.4925
4.10.2014	0.4852
2.4.2015	0.4442
7.7.2015	0.4363
12.8.2015	0.4328
26.4.2015	0.4312

Table 10 *p-values in the stepwise regression for SAR scenes after two iterations (polarizations combined)*

Date	p-value
4.10.2014	0.0418
9.11.2014	0.3010
21.11.2014	0.0311
3.12.2014	0.1941
15.12.2014	0.2299
27.12.2014	0.0833
8.1.2015	0.1393
13.2.2015	0.1176
25.2.2015	0.0182
9.3.2015	0.2129
21.3.2015	0.7172
2.4.2015	0.0189
14.4.2015	0.8357
26.4.2015	0.9315
8.5.2015	0.3486
20.5.2015	0.6299
1.6.2015	0.2764
13.6.2015	0.0143
25.6.2015	0.3863
7.7.2015	0.0015
19.7.2015	3.8593e-09
31.7.2015	0.3201
12.8.2015	0.2562
24.8.2015	0.2516

The choice of a feature selection method can have a big impact on the resulting features. In this case, the two methods lead to quite different results in the feature selection. The next sections will evaluate how well these features and scenes perform in the classification.

5.4 Single-sensor classification

The aim of this thesis was to perform multi-sensor classification. However, the two used sensors were first assessed separately to examine how well the classification worked using data from only one sensor. This makes it possible to later compare the results to the multi-sensor techniques. This section presents the different experiments performed on the Sentinel-1 imagery and the Landsat 8 image.

5.4.1 Sentinel-1

Different classifications were performed with the Sentinel-1 imagery. These included different algorithms, but also using different combinations of the Sentinel-1 images. This section will describe the most significant results of these experiments. The classification maps of 9-NN and ML will be shown as ML mostly has the highest final accuracy and as 9-NN mostly has the best visual result. Table 11 shows the different experiments and the resulting final accuracies of all classifications.

The Sentinel-1 data are both co- and cross-polarized (VV and VH). The first classifications were made separately for these polarizations, and with all 24 images per polarization. This means that the amplitude values were taken as the features, forming a 24-dimensional feature vector per pixel.

Two of the resulting classifications of the VV polarized images, 9-NN and ML, are shown below (Figure 17). The results are very noisy, even after the majority voting. Especially the maximum likelihood method has much extra urban areas. However, even with the noise, larger urban, agricultural and water areas are visible. The classification accuracies of these are 74.7 % and 80.4 %.

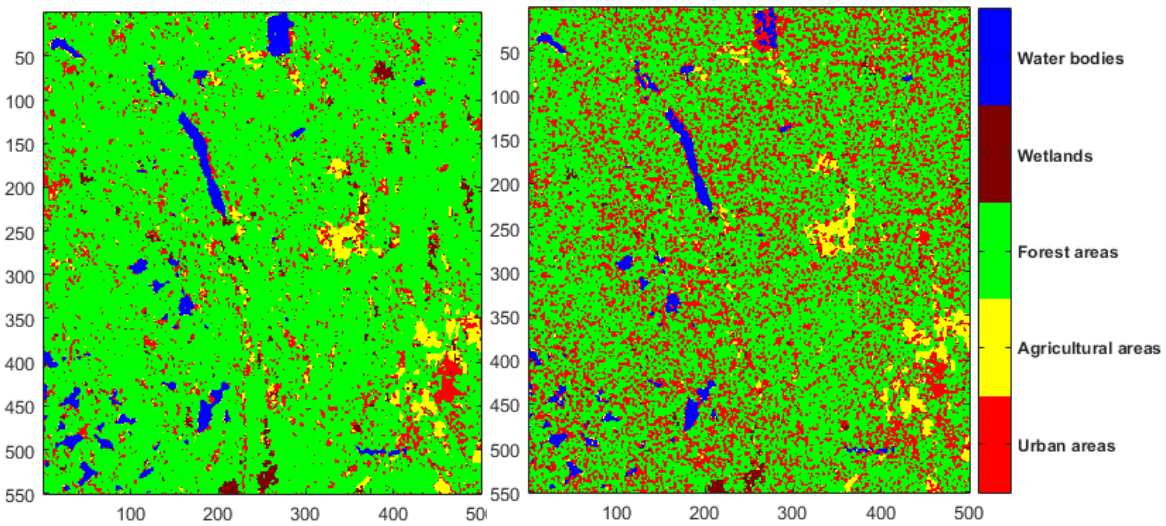

Figure 17 *Classification of Sentinel-1 VV polarized images with 9-NN (left) and ML (right)*

The same classifications were done with the VH polarized images (Figure 18). These results also display a lot of noise. The ML has much less noise than when using the VV polarized images. The characteristics of the noise have also changed. 9-NN has resulted in

extra noise of wetlands whereas the noise is urban areas with ML. The final accuracy of 9-NN is the same as with the VV polarized images, whereas for ML, it drops to 48.9 %.

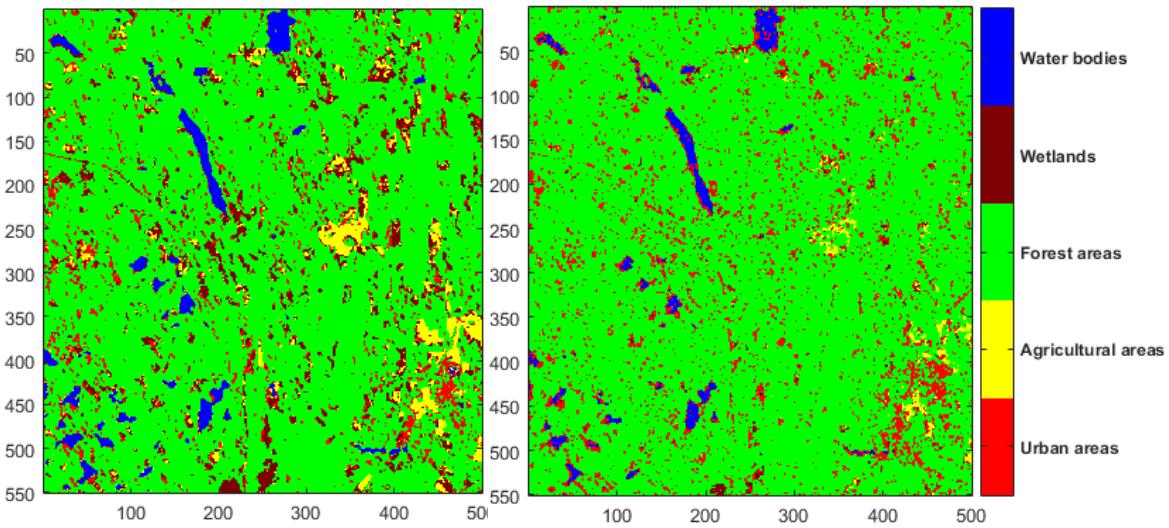

Figure 18 *Classification of Sentinel-1 VH polarized images with 9-NN (left) and ML (right)*

To reduce the dimensionality of the Sentinel-1 data, three principal components were calculated from the 48 VV and VH polarized images. This should result in images with the most important information. The three calculated components include 59.2 % of the variance in the original 48 images. These three components were then used for the classification. Figure 19 displays two of the results, with 9-NN and ML. Once again, there is much noise, especially of urban areas. However, compared to the previous results, the noise is now bigger areas. The corresponding classification accuracies are 70.2 % and 71.6 %.

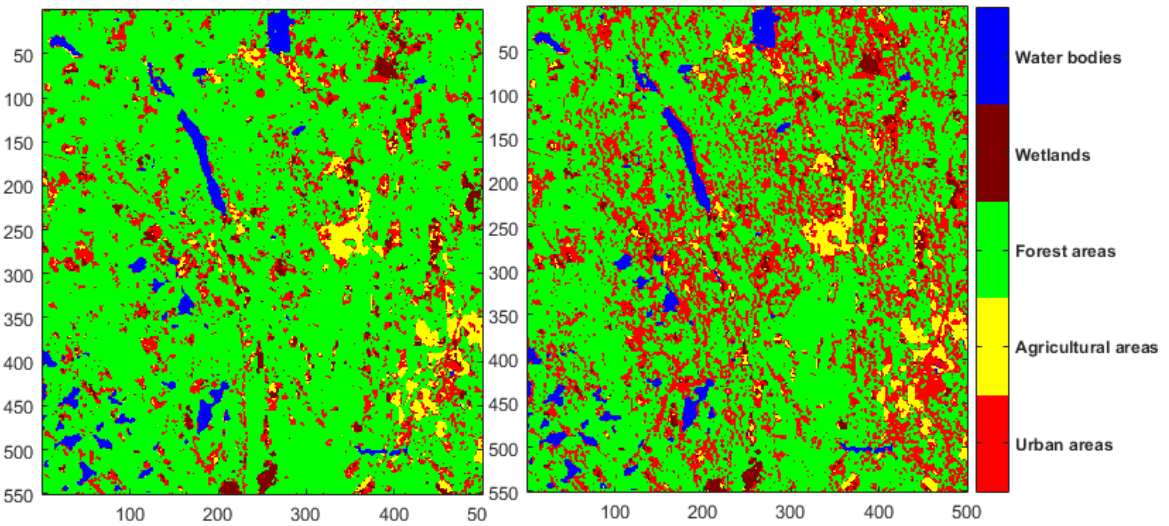

Figure 19 *Classification of three PCA components from the Sentinel-1 VV and VH polarized images with 9-NN (left) and ML (right)*

Another way to reduce the dimensionality is to do a scene selection. In this case, the most important image or images were chosen for the classification. In contrary to the principal component analysis, the scenes were not modified, but used just as they are. First, one scene was chosen with the help of the two presented feature selection methods and the

chosen scene was used for classification (Figure 20). Fisher's discriminant ratio led to choosing the VV polarized scene from 19.7.2015, whereas the stepwise regression led to choosing the VV polarized scene from 2.4.2015. The visual results are very noisy, and only the water bodies have some success in the classification. Classifications with the scene from 19.7.2015 led to a lot of extra wetlands, whereas the scene from 2.4.2015 led to more urban areas. Some of the bigger agricultural areas are also present. The final accuracies for the scenes chosen with stepwise regression are better (60.4 % and 57.3 % compared to 47.6 % and 47.6 %).

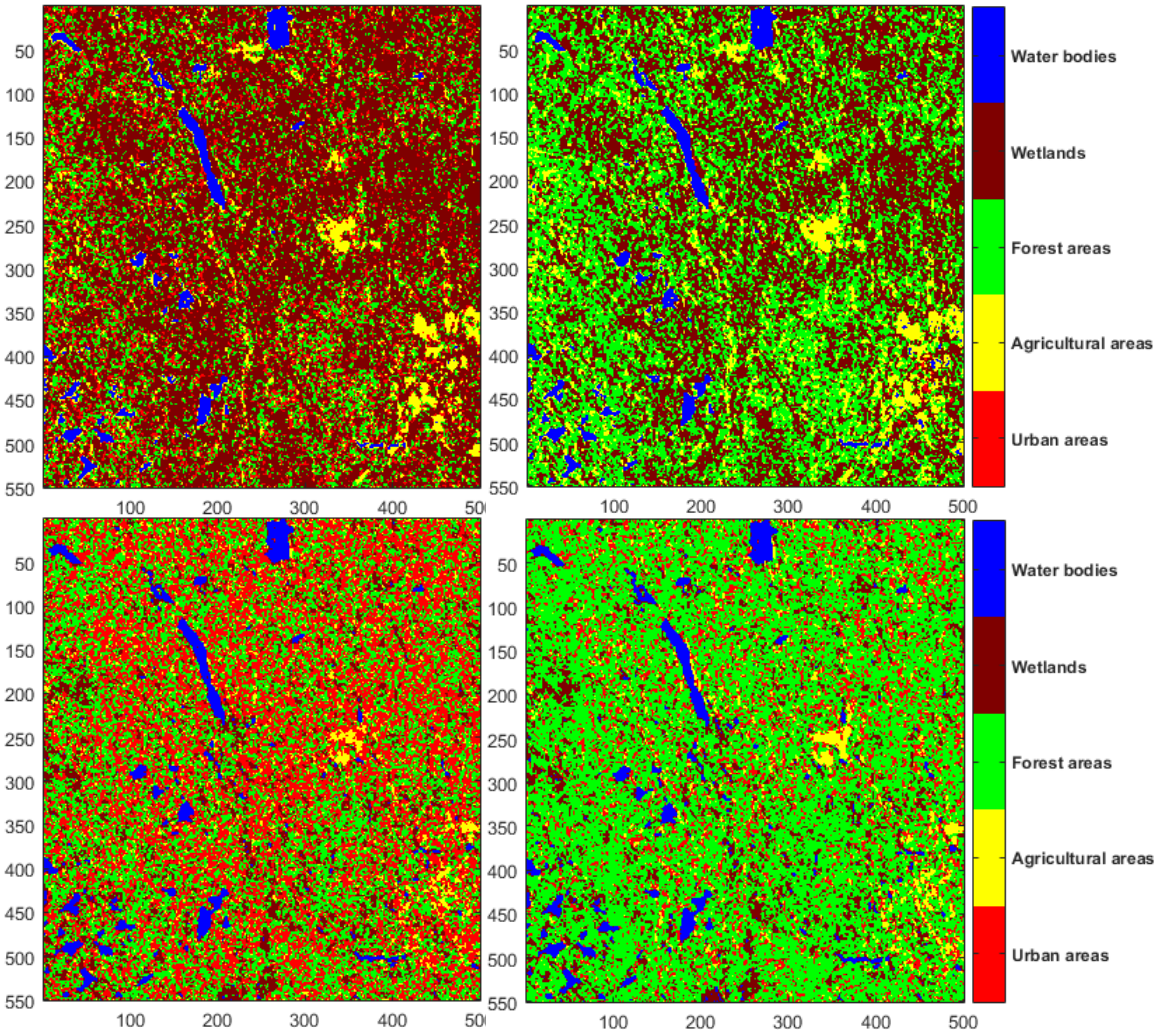

Figure 20 *Classification of one Sentinel-1 scene (one date, one polarization) with 9-NN (left) and ML (right). The scenes for the upper row (19.7.2015) are chosen with Fisher's discriminant ratio and the scenes for the lower row (2.4.2015) are chosen with stepwise regression. Both used scenes were VV polarized.*

A scene selection was also performed to find four images (two dates, both polarizations). Using Fisher's discriminant ratio, scenes from 25.2.2015 and 9.3.2015 were chosen, whereas the stepwise regression led to choosing the scenes from 13.6.2015 and 19.7.2015. The four classifications were then performed on both data sets. The resulting accuracies can be found in Table 11 and two classifications per data set are shown in Figure 21. Looking at the results, both data sets have resulted in very noisy classifications. However, scenes chosen with Fisher's discriminant ratio have even more areas misclassified as urban areas and quite a lot of extra areas classified as water bodies. The scenes chosen with

stepwise regression also resulted in better final accuracies (59.6 % and 58.2 % compared to 52.9 % and 51.1 %)

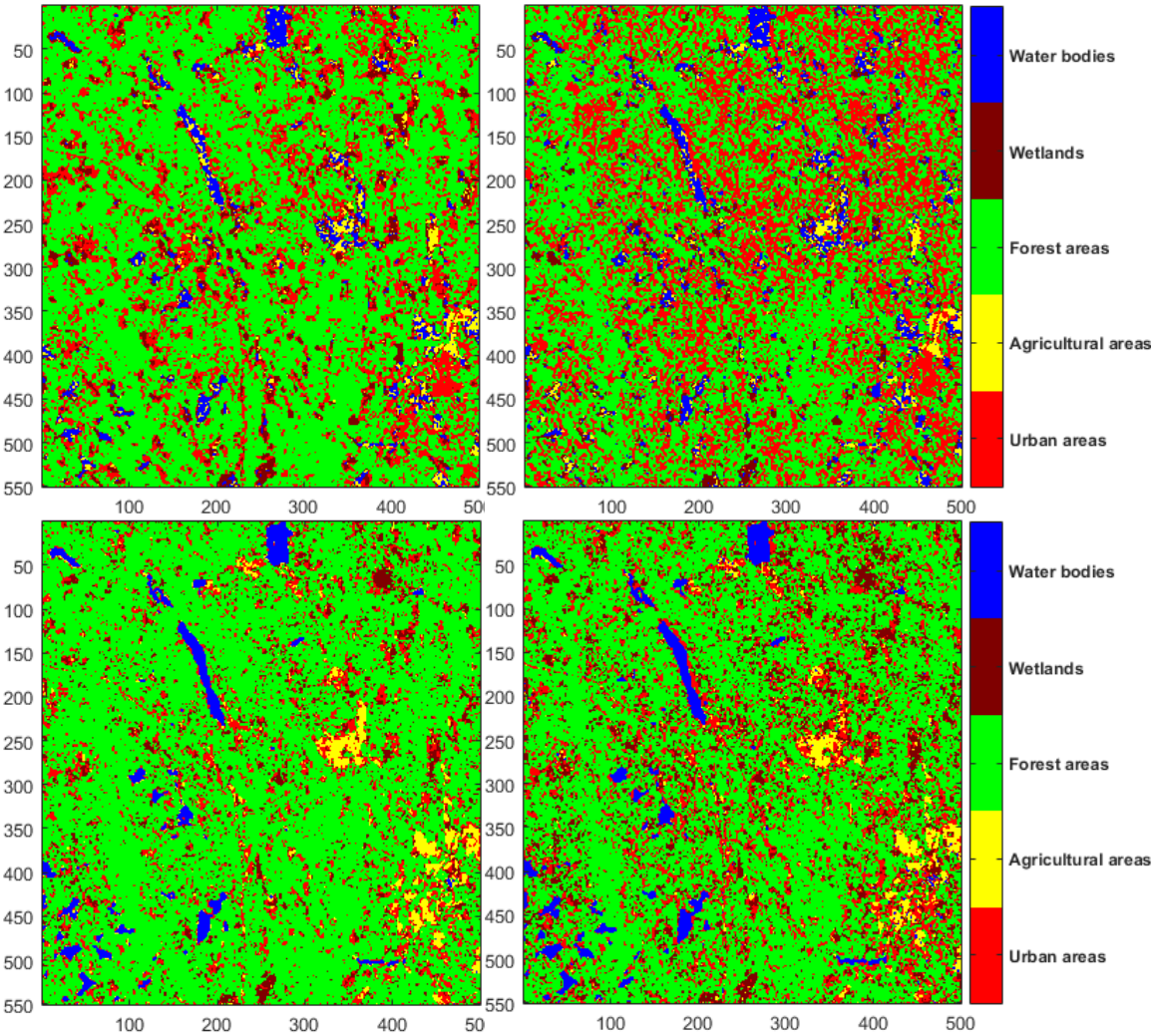

Figure 21 *Classification of four Sentinel-1 scenes (two dates, both polarizations) with 9-NN (left) and ML (right). The scenes for the upper row (25.2.2015 & 9.3.2015) are chosen with Fisher's discriminant ratio and the scenes for the lower row (13.6.2015 & 19.7.2015) are chosen with stepwise regression.*

The amount of features can also be reduced by feature extraction, followed by feature selection. By calculating new features from the amplitude values, each pixel got a reduced amount of features. The features were then reduced again with features selection. In this experiment, eight features were calculated from each polarization and four of these were selected for the classification. The feature selection was done with both Fisher's discriminant ratio and stepwise regression, leading to different classification results. The features selected with Fisher's discriminant ratio were the mean value, the minimum value, the root mean square and the entropy, whereas stepwise regression led to using the standard deviation, the maximum value, the minimum value and the kurtosis.

The selected features were used for the classification, performed with the four methods. The results when using VV polarized images can be seen in Figure 22 and the corresponding final accuracies are found in Table 11. The results show again quite a lot of noise. The maximum likelihood classifications are noisier, with a lot of misclassified urban areas. Overall, the results when using the features chosen by stepwise regression look

better. However, the accuracies do not completely agree as the final accuracy for the maximum likelihood classification of the features from the Fisher's discriminant ratio is better.

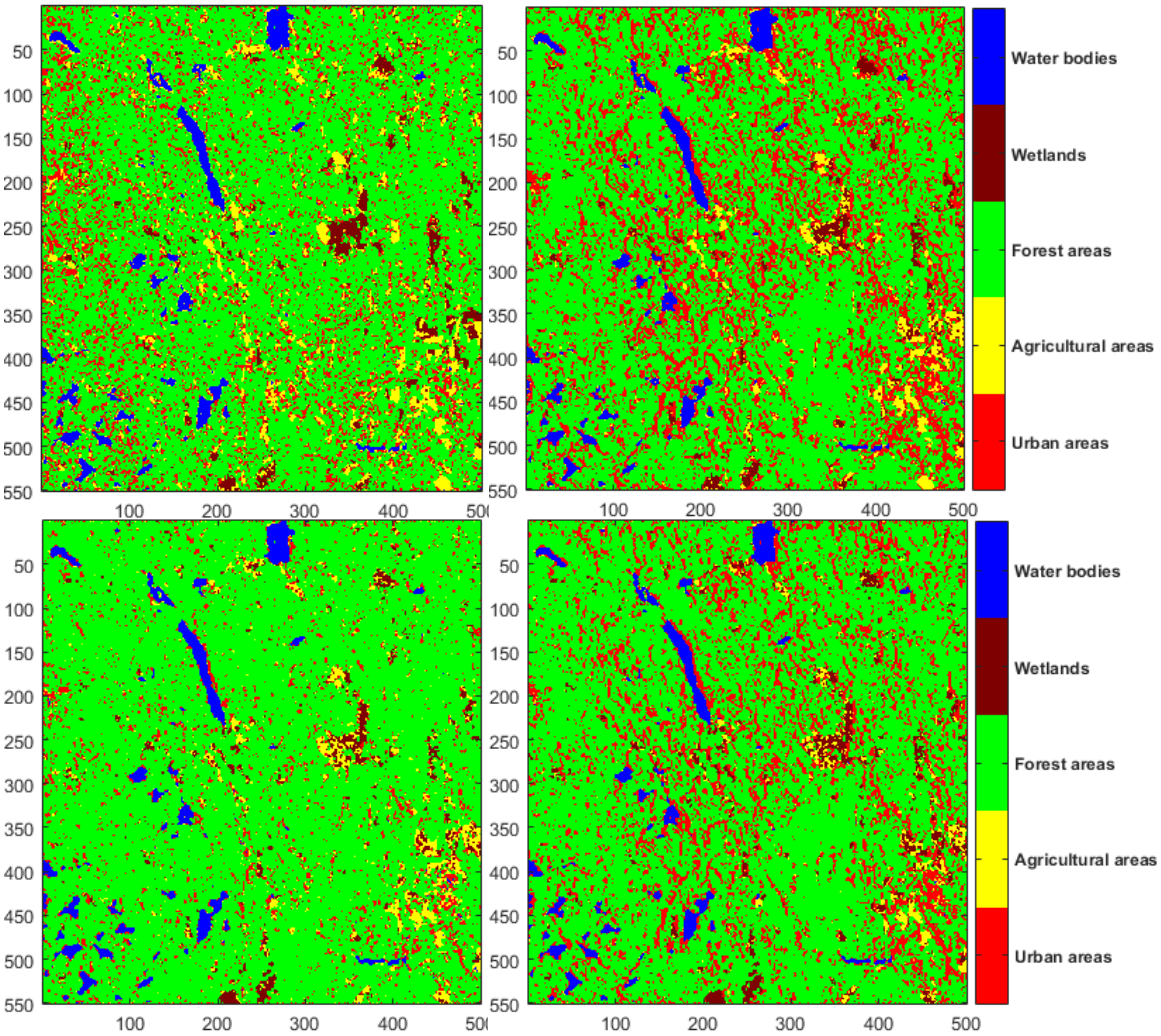

Figure 22 *Classification of Sentinel-1 VV polarized images after feature extraction and selection with 9-NN (left) and ML (right). The features for the upper row (mean, min, RMS, entropy) are chosen with Fisher's discriminant ratio and the features for the lower row (std, max, min kurtosis) are chosen with stepwise regression.*

The four classification methods were performed also on the features extracted and selected from the VH polarized data (Figure 23). Compared to the features from VV polarized data, the amount of noise decreases, but is still present. The noise changes as more wetlands are now present. Again, the features from stepwise regression give a better visual result. However, the final accuracies (Table 11) do not give a straightforward solution to which method works better as the better accuracy varies between classification methods.

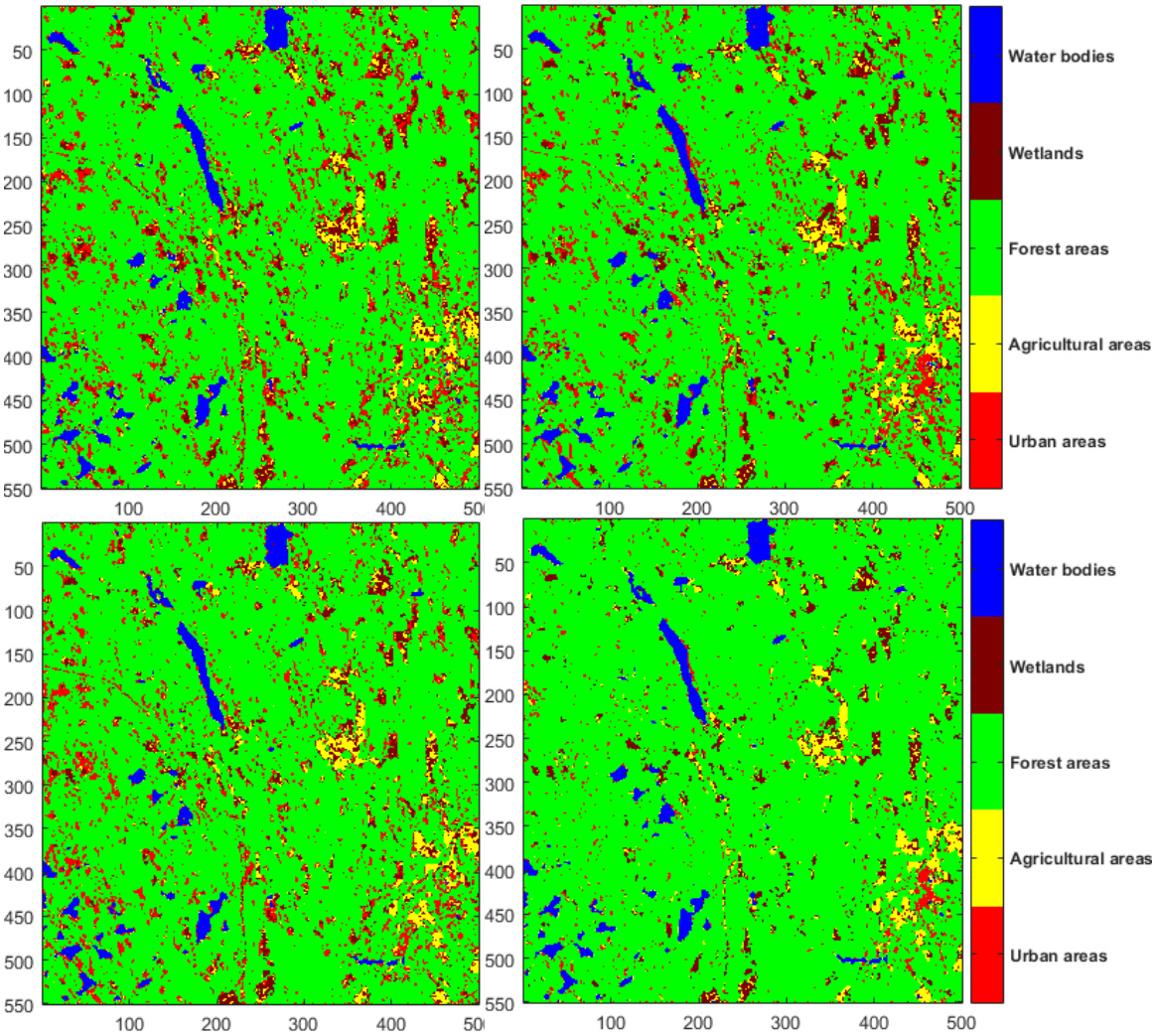

Figure 23 *Classification of Sentinel-1 VH polarized images after feature extraction and selection with 9-NN (left) and ML (right). The features for the upper row (mean, min, RMS, entropy) are chosen with Fisher's discriminant ratio and the features for the lower row (std, max, min kurtosis) are chosen with stepwise regression.*

The visual interpretation gives a good idea of the classification result, but the final accuracies should also be examined. Table 11 shows the final accuracies of all conducted experiments, also those not shown visually. Out of the four classification methods, none stands out for all experiments. 9-NN mostly performs as well as the best method for a certain data set, whereas maximum likelihood ranges from being clearly the best to being clearly the worst. It can be noted that a dimensionality reduction mostly leads to a worse accuracy (except for PCA). However, even in those cases good results have been reached for some classifications.

Table 11 *Accuracies for classifications of Sentinel-1 data*

Sensor	Data	MDM	1-NN	9-NN	ML
Sentinel-1	VV polarized images	76.0 %	79.1 %	74.7 %	80.4 %
Sentinel-1	VH polarized images	71.1 %	67.6 %	74.7 %	48.9 %
Sentinel-1	3 PCA images calculated from all VV and VH polarized images	81.3 %	79.1 %	70.2 %	71.6 %
Sentinel-1	Scene selection (1 date, one polarization): - using Fisher's ratio (19.7.2015, VV polarized) - using stepwise regression (2.4.2015, VV polarized)	 50.7 % 56.9 %	 36.9 % 59.6 %	 47.6 % 60.4 %	 47.6 % 57.3 %
Sentinel-1	Scene selection (2 dates, both polarizations): - using Fisher's ratio (25.2.2015 & 9.3.2015) - using stepwise regression (13.6.2015 & 19.7.2015)	 51.1 % 60.9 %	 44.0 % 59.1 %	 52.9 % 59.6 %	 51.1 % 58.2 %
Sentinel-1	Feature extraction and selection from VV polarized images: - using Fisher's ratio (mean, min, RMS, entropy) - using stepwise regression (std, max, min, kurtosis)	 44.4 % 52.9 %	 62.7 % 65.8 %	 61.3 % 64.9 %	 70.2 % 65.3 %
Sentinel-1	Feature extraction and selection from VH polarized images: - using Fisher's ratio (mean, min, RMS, entropy) - using stepwise regression (std, max, min, kurtosis)	 48.0 % 61.3 %	 64.0 % 61.3 %	 64.0 % 65.3 %	 69.8 % 68.4 %

It has been shown that the different ways of using the Sentinel-1 data lead to varying results. A common result is the presence of noise. This can clearly be seen in the visual presentations. The varying characteristics of the noise could be a reason to differences in the accuracies. Later this data will be merged with optical satellite data to improve the results.

5.4.2 Landsat 8

The single Landsat 8 image was used for different classifications. The original bands or some extracted features were used, and the four different methods were tested. This section will describe the results. The experiments and the final accuracies can be seen in Table 12.

The used Landsat image has eight channels. The first classifications were done by choosing six of these (2-7). The reason to leave out band 1 (coastal aerosol) was that it was originally designed for coastal water and aerosol monitoring which does not match the study's purpose. The reason to leave out band 9 (cirrus) was that it is designed to identify clouds, which is not needed at the classification stage when the cloud masking was already performed. Each band was regarded as one feature, leading to a 6-dimensional feature

vector per pixel. Two classifications, 9-NN and ML, are visualized below (Figure 24). The results have some noise, but compared to the Sentinel-1 results, it is bigger and clearer. In 9-NN, the noise is mostly agricultural areas, whereas ML resulted in too much urban areas. The corresponding final accuracies are 78.2 % and 91.6 %.

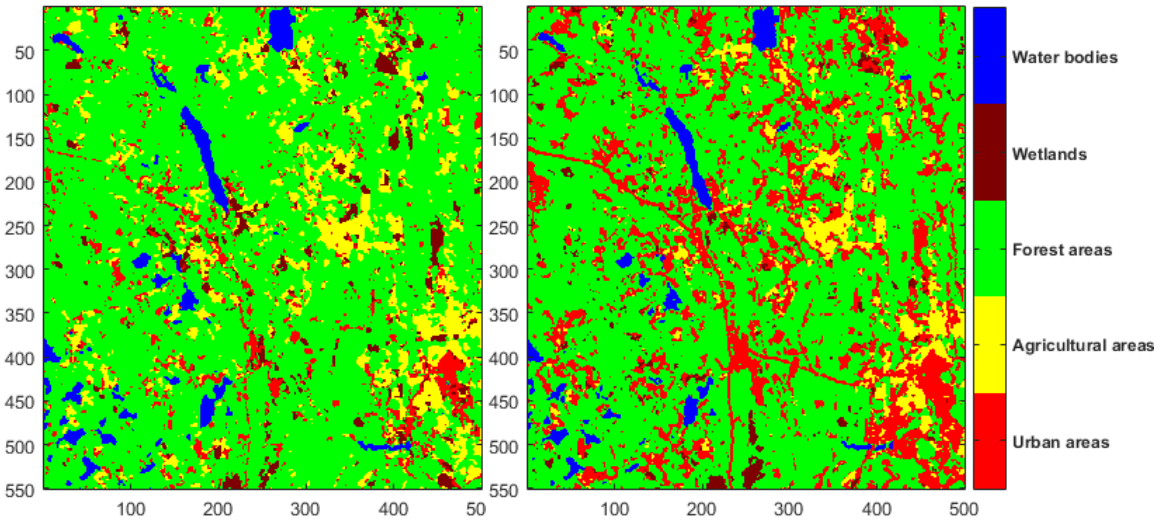

Figure 24 *Classification of the Landsat 8 image bands 2-7 with 9-NN (left) and ML (right)*

Three principal components were calculated from the Landsat 8 image, using bands 2-7. These three components included 99.7 % of the variance in the original images. Two of the resulting classifications can be seen in Figure 25. The results do not change much from classifications calculated with the original pixel values. Some noise is again present, agricultural and urban areas, but to a bit of a smaller extent. The final accuracies for these two classifications are 80.9 % and 83.1 %.

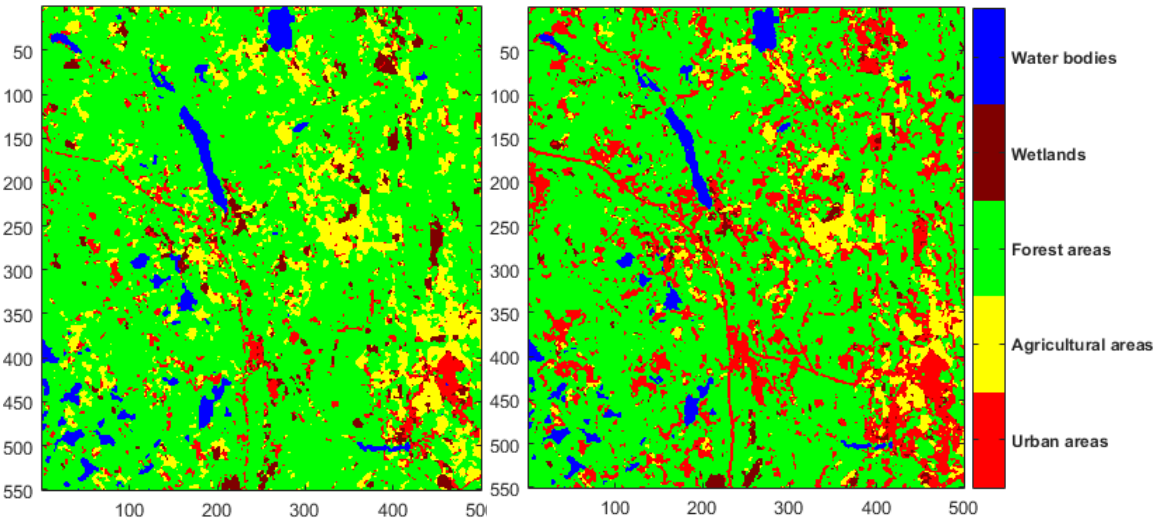

Figure 25 *Classification of three PCA components from the Landsat 8 image with 9-NN (left) and ML (right)*

A band selection was performed for classifications with four bands. For this purpose, bands 2-5 were chosen. Again, bands 1 and 9 were left out due to their characteristics. Bands 6 and 7 were left out as they were regarded less suitable for this kind of land cover mapping task. Two of the resulting classifications can be seen in Figure 26. The visual result when using 9-NN improved a lot, whereas the maximum likelihood classification

remains very similar. Much noise of urban areas is present. The final accuracy for the 9-NN classification is 81.3 % and for the maximum likelihood classification, it is 90.7 %.

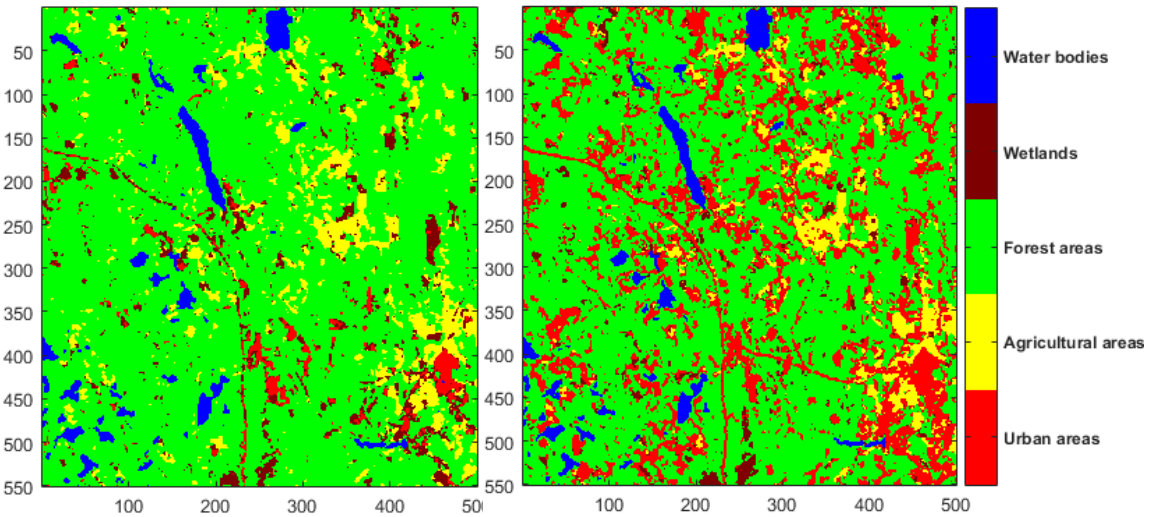

Figure 26 *Classification of the Landsat 8 image bands 2-5 with 9-NN (left) and ML (right)*

A feature extraction was performed by calculating some indices from the Landsat image. These were the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Water Index (NDWI), Normalized Burn Ratio (NBR), and the Normalized Difference Built-Up Index (NDBI). This reduced the dimensionality of the feature vectors to four. The resulting classifications change in character (Figure 27). Compared to directly using the channel values, these results also have noise but the noise is agricultural areas instead of urban areas. The classification accuracies change to 72.4 % % and 76.0 %.

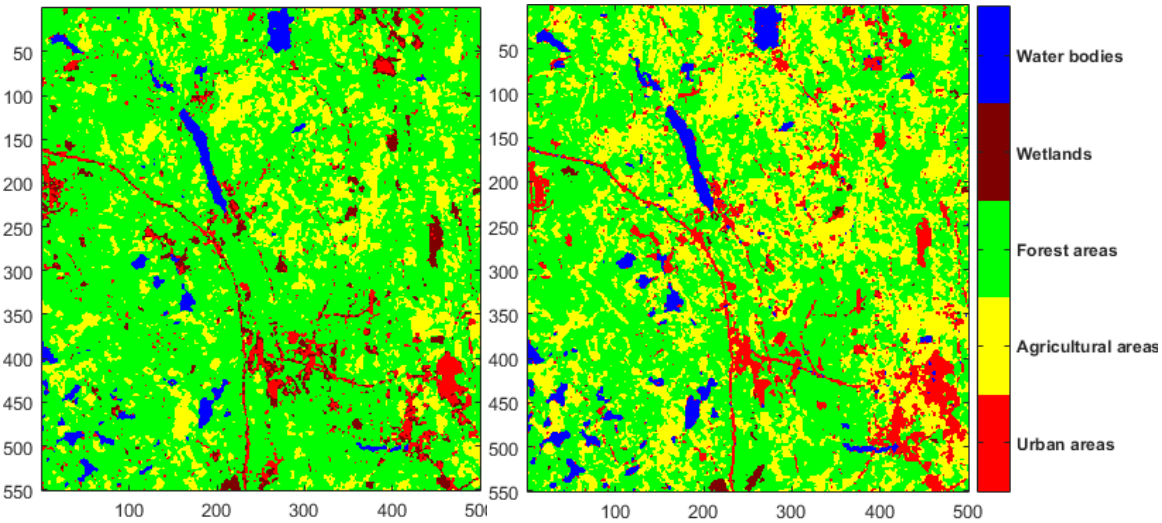

Figure 27 *Classification of the Landsat 8 image with four indices with 9-NN (left) and ML (right)*

Visually, the classification results of the Landsat 8 data look quite clear, especially when using 9-NN. The final accuracies of all experiments (Table 12) also show good accuracies. There is not a very big difference in most accuracies and no clear pattern can be seen, except lower accuracies when using the extracted features. The maximum likelihood classifier tends to result in the best final accuracy.

Table 12 *Accuracies for classifications of Landsat 8 data*

Sensor	Data	MDM	1-NN	9-NN	ML
Landsat 8	Bands 2-7	85.8 %	81.3 %	78.2 %	91.6 %
Landsat 8	3 PCA images calculated from bands 2-7	81.3 %	84.4 %	80.9 %	83.1 %
Landsat 8	Feature (band) selection: Bands 2-5	78.7 %	81.3 %	81.3 %	90.7 %
Landsat 8	Feature extraction: NDVI, NDWI, NBR, NDBI	58.2 %	76.4 %	72.4 %	76.0 %

Classifications of the Landsat 8 data have high final accuracies, whereas the visual results still lack in precision. Much noise is present, mostly agricultural or urban areas. In an attempt to improve the classifications, the data will be fused with SAR data in the next section.

5.5 Multi-sensor classification

The multi-sensor classifications were done by fusing the data from Sentinel-1 and Landsat 8. Six different data fusion combinations were tested. These combinations used either extracted features or directly the pixel values in the images. This section presents the classification results. The classification maps for all experiments are shown to be able to assess both the classification methods and the fused data sets. All experiments and final accuracies can be seen in Table 13.

Some of the single-sensor classifications of the Sentinel-1 data were performed with features and scenes selected from a larger set of available features/scenes. Visually and statistically, the features/scenes selected with stepwise regression led to better results. Therefore, these features/scenes were used in the data fusion.

First, a data fusion using the four features extracted from the VV polarized Sentinel-1 data and the four extracted indices from the Landsat 8 image was performed and the classifications were run. This led to four classification maps (Figure 28). It can be noted that MDM produced too much agricultural areas, whereas ML has a lot of extra urban areas. However, the amount of this urban noise reduces significantly compared to most of the single-sensor experiments. 1-NN and 9-NN look quite “clean”, only a bit of extra agricultural areas and some urban areas. All classifications, except the MDM classification, show all the bigger land cover areas that can be seen in the CORINE image, only lacking in precision of the area extents. The ML classification has the best final accuracy (93.8 %).

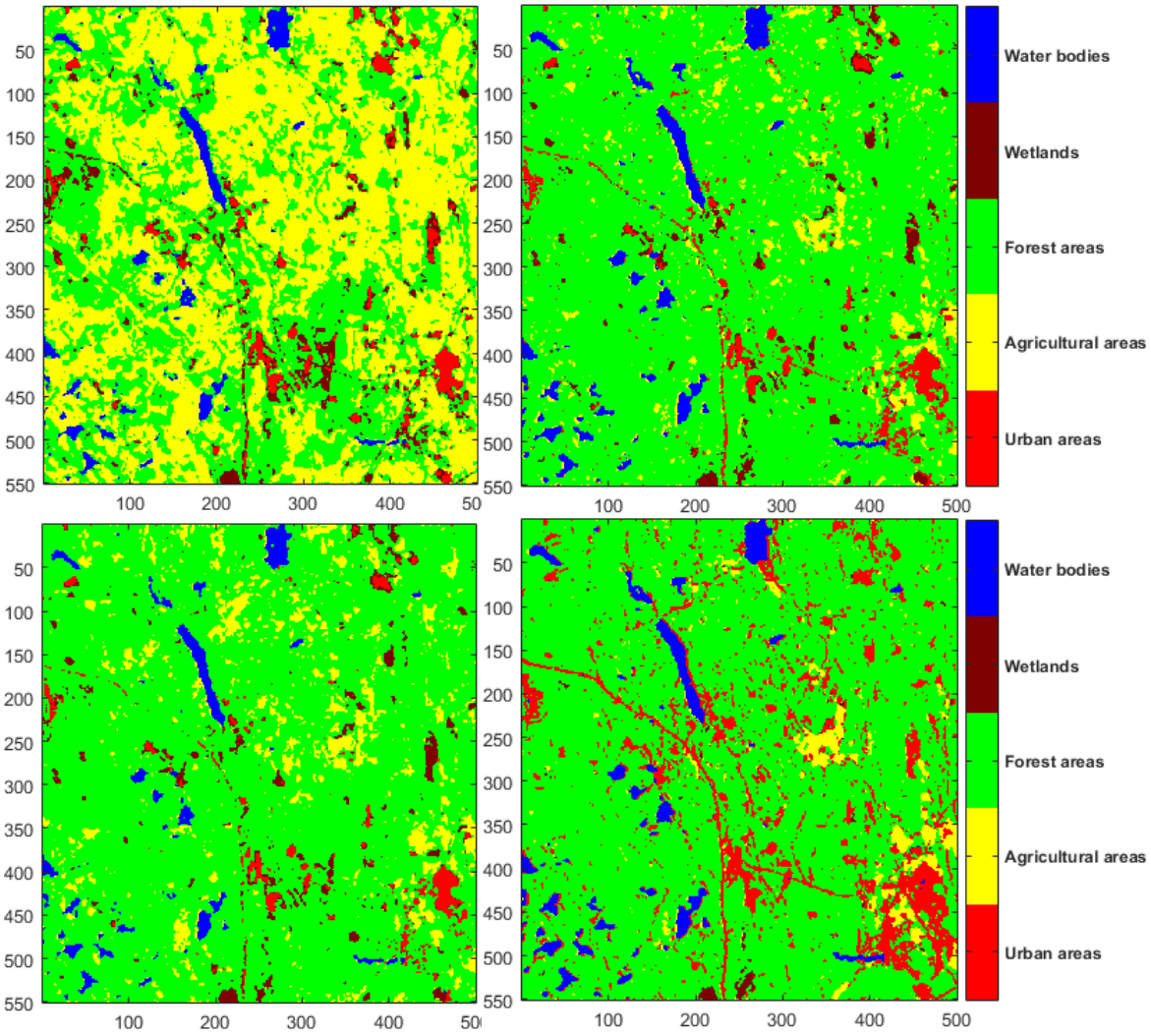

Figure 28 *Classification of fused data (4 features from the VV polarized Sentinel-1 images and 4 indices from the Landsat 8 image) with MDM (upper left), 1-NN (upper right), 9-NN (lower left) and ML (lower right)*

The four features from the VV polarized images were also fused with four Landsat 8 image bands, as can be seen in Figure 29. The MDM classification still produced much noise, but this time wetlands. However, these wetland areas are considerably smaller than the agricultural areas in the previous case. The 1-NN and 9-NN classifications remain visually similar to the previous experiment. The ML produced again too much urban areas, this time even more. Out of these four classifications, the 1-NN classification has the best classification accuracy (90.2 %).

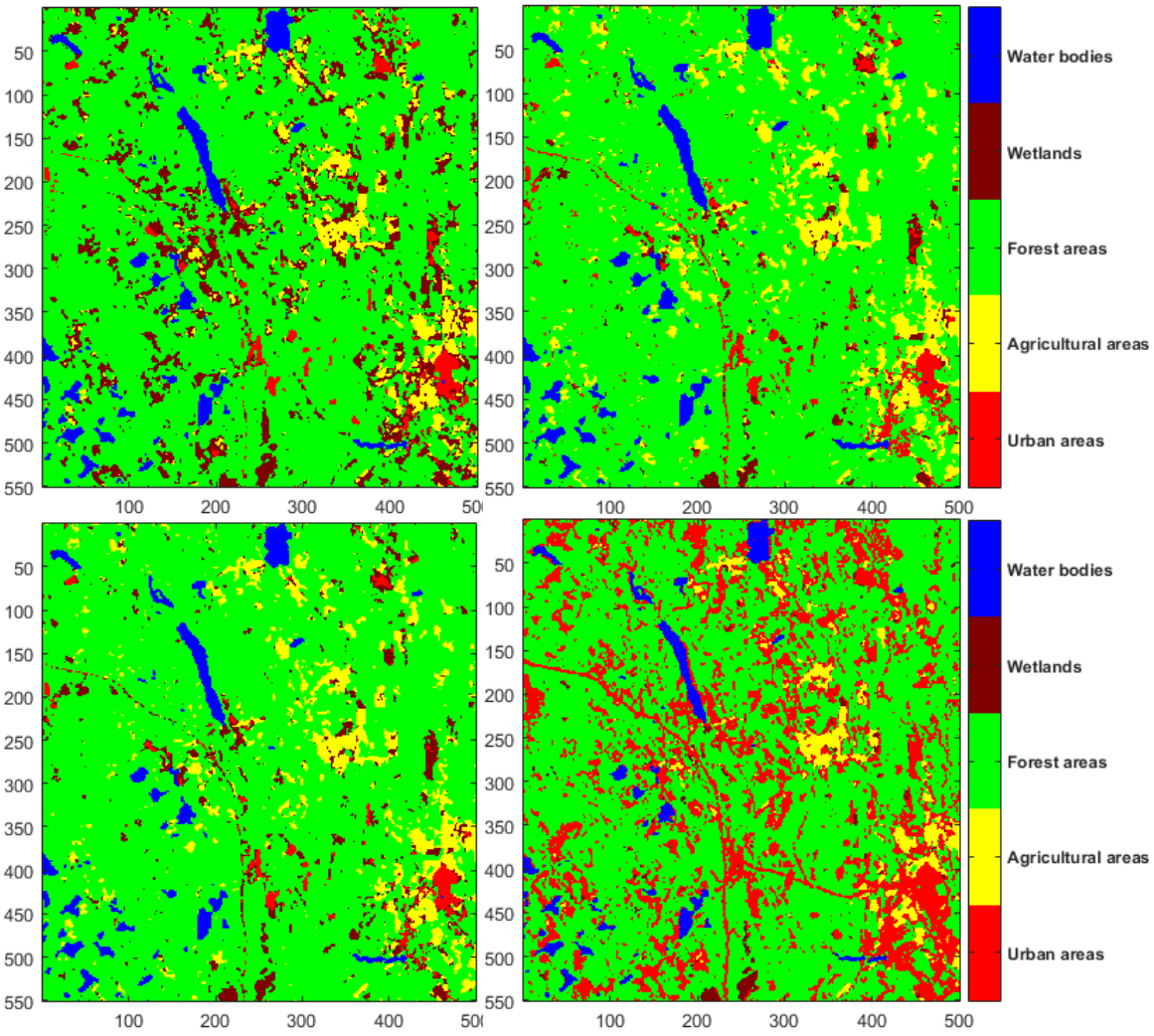

Figure 29 *Classification of fused data (4 features from the VV polarized Sentinel-1 images and 4 Landsat 8 bands) with MDM (upper left), 1-NN (upper right), 9-NN (lower left) and ML (lower right)*

The next two experiments used the four features extracted from the VH polarized Sentinel-1 imagery. The first one fused those with four indices calculated from the Landsat 8 image. The four classification methods were performed, leading to the results in Figure 30. Visually, the classifications with MDM, 1-NN and 9-NN are very similar to the ones made with VV polarized data and four Landsat indices. The final accuracies do not differ much either. A significant improvement can be seen in the ML classification. Much of the extra urban areas have disappeared, leading to a result close to the CORINE image (Figure 9, right). This classification also has the best final accuracy of these four (89.8 %).

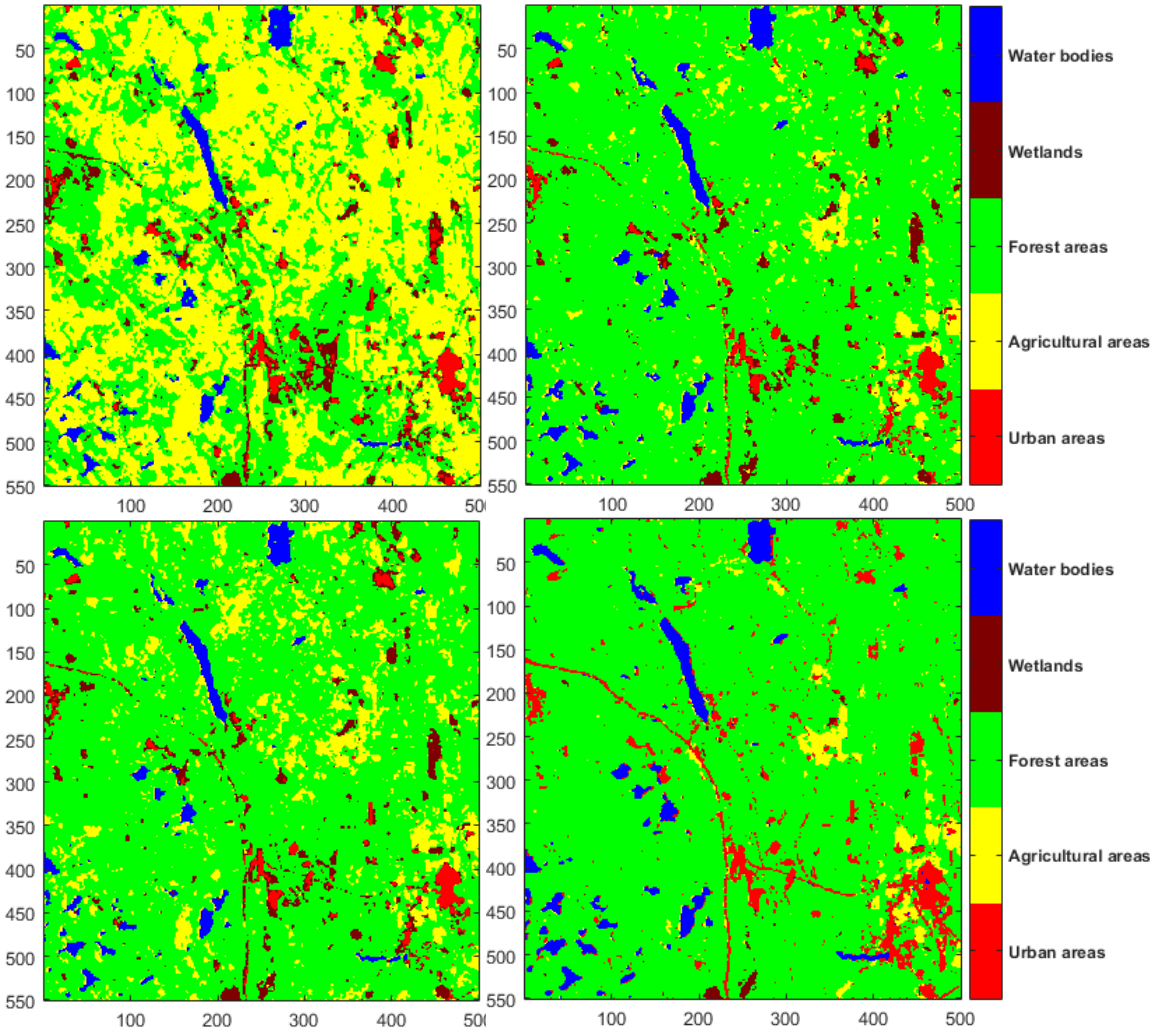

Figure 30 *Classification of fused data (4 features from the VH polarized Sentinel-1 images and 4 indices from the Landsat 8 image) with MDM (upper left), 1-NN (upper right), 9-NN (lower left) and ML (lower right)*

The features extracted from the VH polarized Sentinel-1 data were next fused with four bands from the Landsat 8 image, and the four classification methods were tested (Figure 31). Again, the first three classifications are visually very similar to when using extracted features from VV polarized Sentinel-1 data and four Landsat bands. The accuracies did decrease though. On the contrary, the ML classification's accuracy improved. Much extra urban areas are still present, but to a smaller extent. Even if the visual interpretation shows some flaws, this classification actually has the best classification accuracy of all experiments (97.3 %).

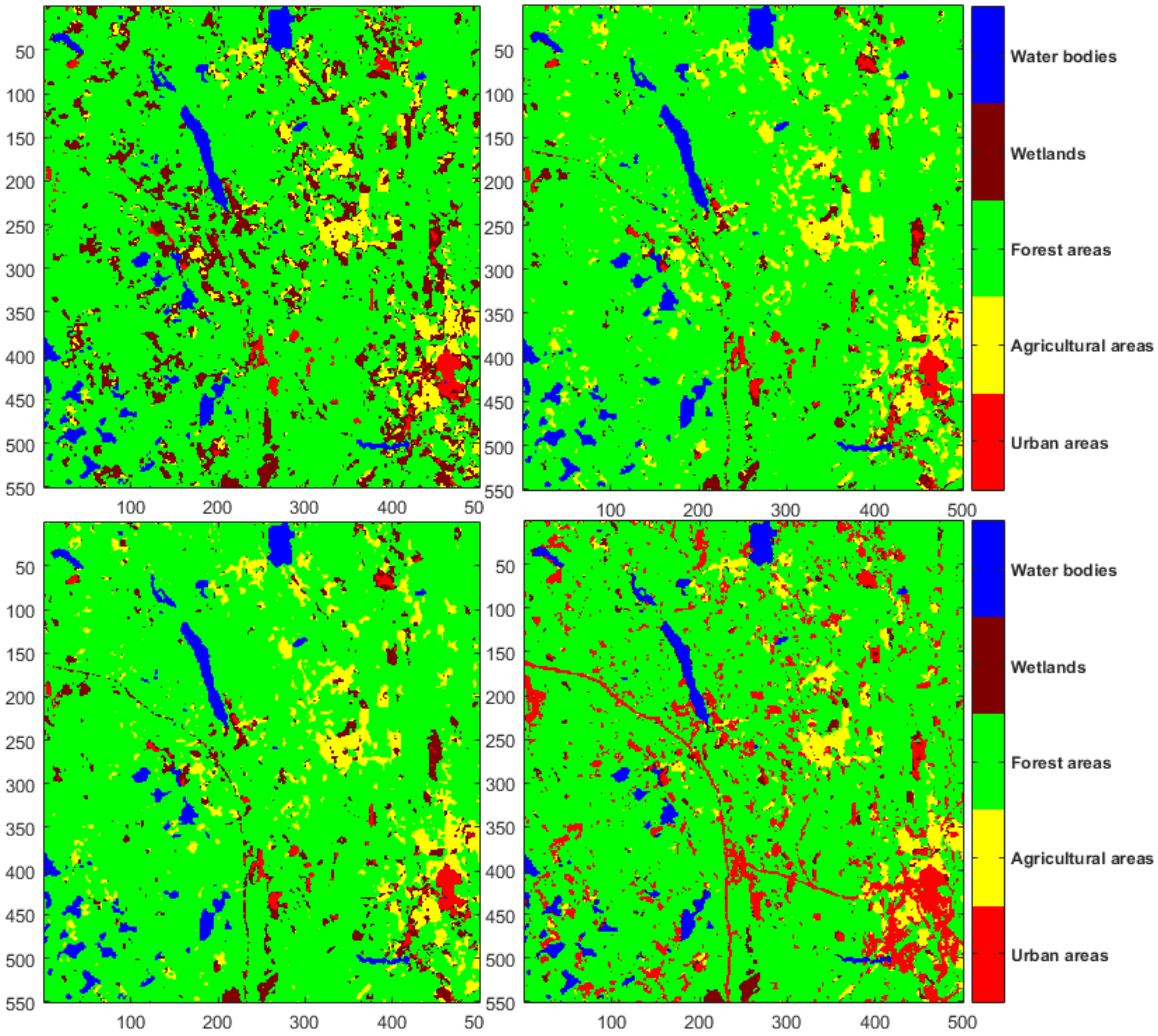

Figure 31 *Classification of fused data (4 features from the VH polarized Sentinel-1 images and 4 Landsat 8 bands) with MDM (upper left), 1-NN (upper right), 9-NN (lower left) and ML (lower right)*

The last two fusions used four Sentinel-1 images, two scenes per polarization, and combined those with different Landsat data. These scenes were first fused with the four indices calculated from the Landsat 8 image (Figure 32). The results do not differ much from the other classifications when the four indices were fused with Sentinel-1 data. The accuracies also do not change significantly. One remark would be the extra agricultural areas that appear on the western side of water bodies and the bigger road in the 1-NN and 9-NN classifications. The 1-NN classification resulted in the best final accuracy (87.1 %).

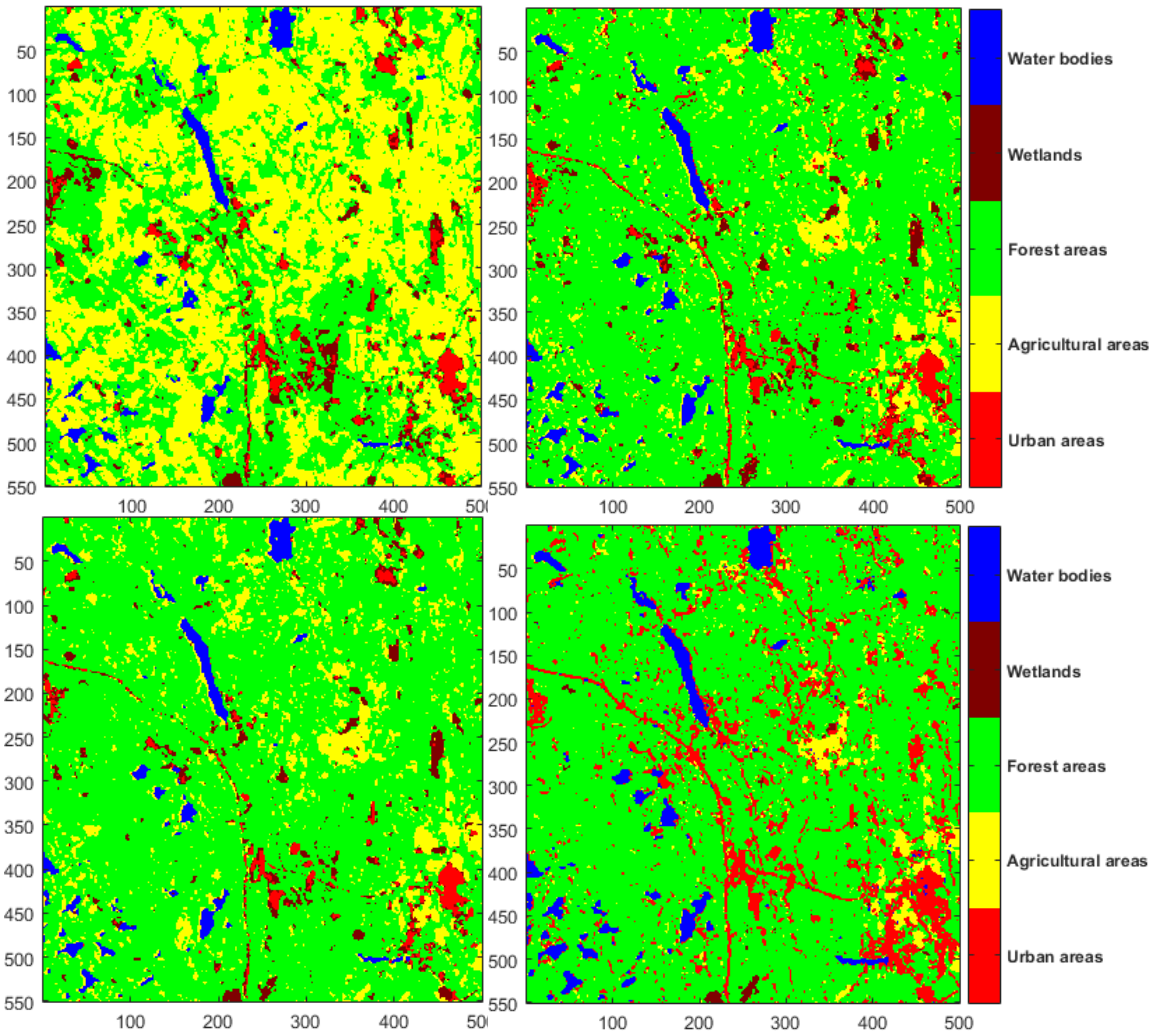

Figure 32 *Classification of fused data (4 Sentinel-1 images and 4 indices from the Landsat 8 image) with MDM (upper left), 1-NN (upper right), 9-NN (lower left) and ML (lower right)*

Finally, the four chosen Sentinel-1 scenes were fused with four bands selected from the Landsat 8 image. These classifications are shown in Figure 33. As in the previous case, these results resemble the other results when using the four Landsat 8 bands with a lot of extra wetlands in MDM, a lot of extra urban areas in ML, and quite good visual results in 1-NN and 9-NN. The similarity is also true for the final accuracies. The ML classification in this case has the highest accuracy when comparing the four results (85.8 %).

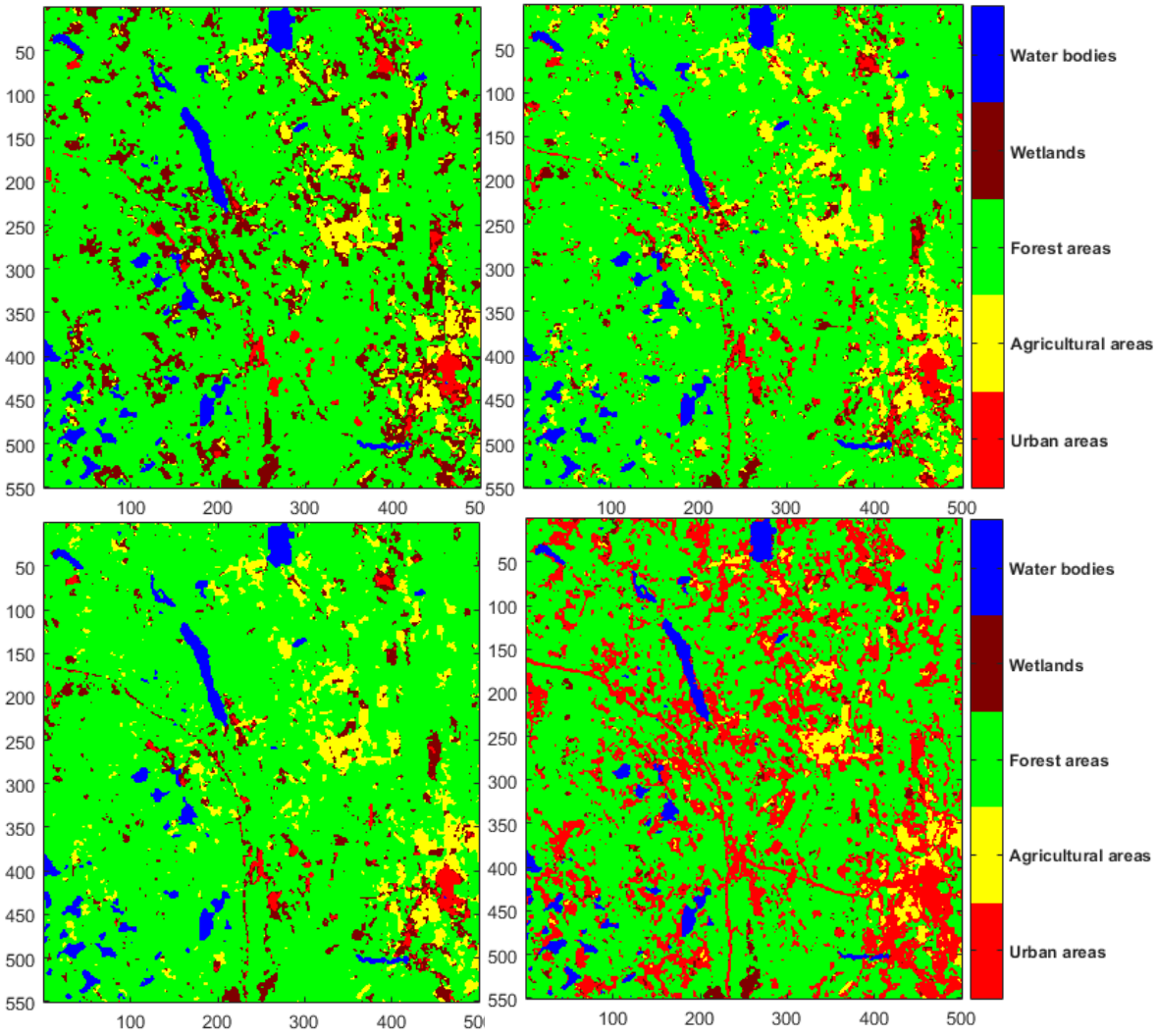

Figure 33 *Classification of fused data (4 Sentinel-1 images and 4 Landsat 8 bands) with MDM (upper left), 1-NN (upper right), 9-NN (lower left) and ML (lower right)*

The final accuracies in Table 13 support the visual results for the most part. The MDM classifications look the worst and also have the lowest accuracies, especially when using the four indices from the Landsat 8 image. The other three methods consistently show high accuracies. Most of the 1-NN and 9-NN classifications visually look very similar (when using the same data), but the accuracies still vary in some cases. The ML classifications mostly have the best final accuracy. However, in some cases the visual results are worse for this algorithm.

Table 13 *Accuracies for classifications of fused data*

Sensor	Data	MDM	1-NN	9-NN	ML
Sentinel-1 Landsat 8	4 features from VV polarized images (std, max, min, kurtosis) 4 indices (NDVI, NDWI, NBR, NDBI)	58.2 %	88.9 %	79.6 %	93.8 %
Sentinel-1 Landsat 8	4 features from VV polarized images (std, max, min, kurtosis) 4 bands (2-5)	81.3 %	90.2 %	88.4 %	89.3 %
Sentinel-1 Landsat 8	4 features from VH polarized images (std, max, min, kurtosis) 4 indices (NDVI, NDWI, NBR, NDBI)	57.8 %	88.0 %	76.0 %	89.8 %
Sentinel-1 Landsat 8	4 features from VH polarized images (std, max, min, kurtosis) 4 bands (2-5)	80.0 %	84.0 %	81.8 %	97.3 %
Sentinel-1 Landsat 8	4 scenes (13.6.2015 & 19.7.2015) 4 indices (NDVI, NDWI, NBR, NDBI)	58.7 %	87.1 %	79.6 %	86.2 %
Sentinel-1 Landsat 8	4 scenes (13.6.2015 & 19.7.2015) 4 bands (2-5)	79.1 %	81.3 %	82.2 %	85.8 %

In all the previous experiments, the validation was performed with the help of the test plots that were defined from the reference image. However, these cover only a small part of the whole image area. Therefore, a validation was also performed with the help of the CLC2012. Even though this image has some flaws, it should give an approximate evaluation of the classifications.

Two of the created classifications were chosen for this experiment. These were chosen both visually and with the help of the calculated accuracies. The first used image was the result that was visually the best: the ML classification using four extracted features from the VH polarized Sentinel-1 imagery and four indices calculated from the Landsat 8 image (lower right in Figure 30). The second used image was the classification that resulted in the best final accuracy when comparing to the test plots. This was the ML classification that used four extracted features from the VH polarized Sentinel-1 data and four bands from the Landsat 8 image (lower right in Figure 31).

Confusion matrices were calculated for these two best classifications, comparing to the test plots and to CLC2012. For the visually best classification, these can be seen in Figure 34. The green diagonal presents the amount and percentage of pixels that are classified to the right class. The blue box in the lower right corner is the final accuracy (green number), whereas the gray boxes represent the producer's accuracies (on the bottom) and the user's accuracies (on the right). In this experiment, the test plot validation resulted in a final accuracy of 89.8 % and the CLC2012 validation resulted in a final accuracy of 91.3 %. The characteristics of these matrices are quite different though. Using the test plots, the correctly classified pixels dominate, and the producer's and user's accuracies are mostly high. When using CLC2012, the amount of pixels is much higher, but also the different classes differ a lot in pixel amounts. Only the forests and water bodies have high

producer's and user's accuracies, but as these already constitute a high percentage of the total pixels, then the final accuracy remains high.

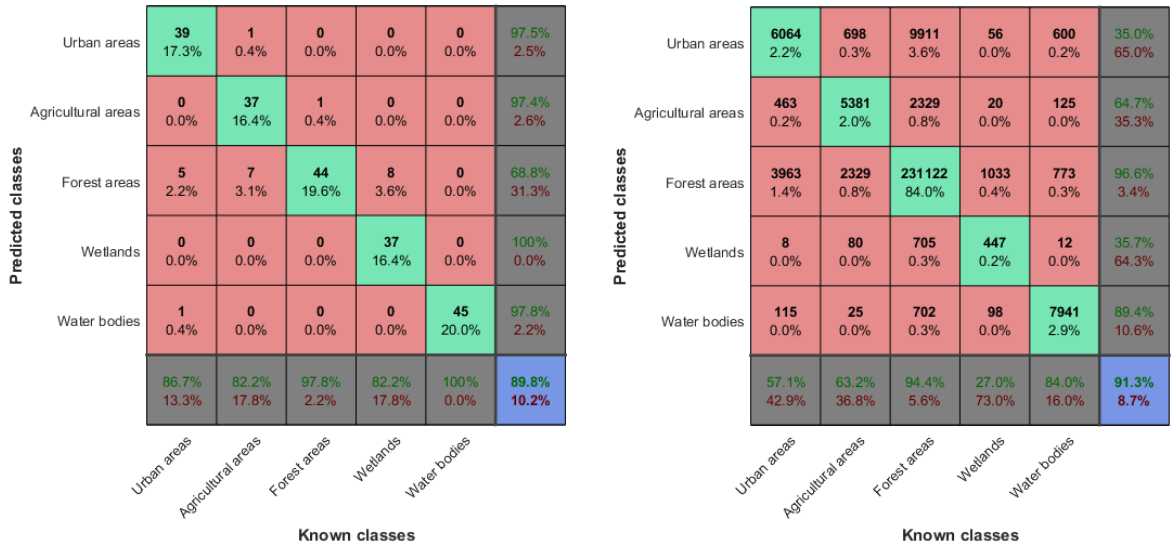

Figure 34 *Confusion matrices for the visually best classification (4 features from VH polarized Sentinel-images and 4 indices from Landsat 8 image) when comparing to test plots (left) and to CORINE (right)*

The same confusion matrices were created for the statistically best result (Figure 35). These matrices show bigger differences. The final accuracies were 97.3 % when using the test plots and 84.1 % when using CLC2012. When comparing to the test plots, almost all test pixels have been correctly classified, but the other matrix shows that there are still many misclassified pixels remaining. This is also confirmed by the visual interpretation. The user's accuracies remain high for the forests and water bodies, whereas the producer's accuracies are quite high for all but the wetlands.

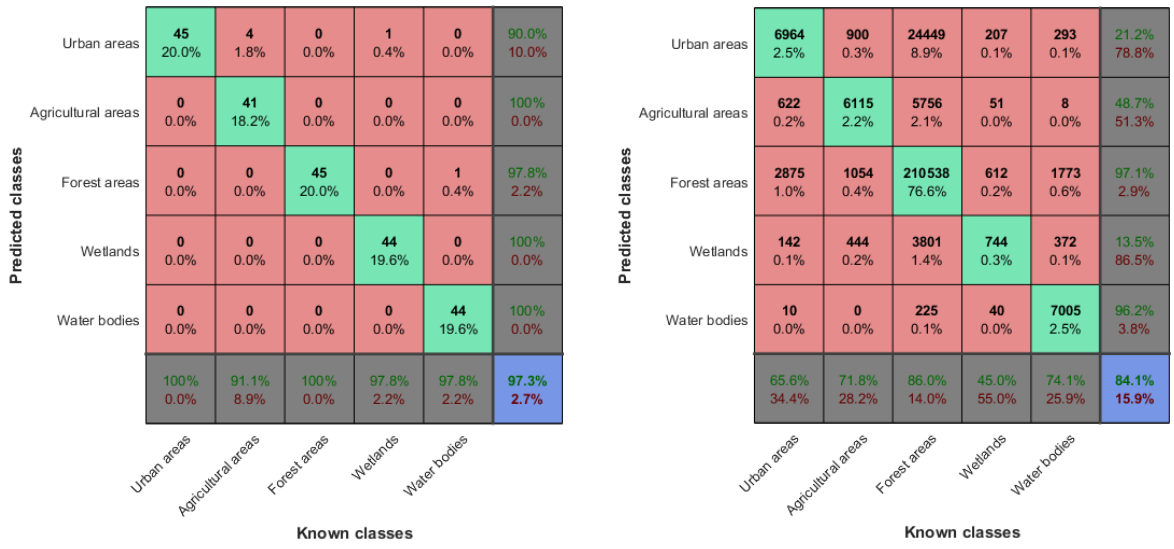

Figure 35 *Confusion matrices for the statistically best classification (4 features from VH polarized Sentinel-images and 4 Landsat 8 bands) when comparing to test plots (left) and to CORINE (right)*

All the demonstrated results have been presented both visually and statistically. By combining these two representations, a final evaluation of the success of a classification can be assessed. The following section will give a deeper explanation of the results.

6 Discussion

The aim of this study has been to develop an efficient land cover classification methodology when fusing multi-temporal SAR and optical data. Different data sets and classification methods have been tested. This section will assess their performance and give recommendations for future research. Some of the results will later be communicated at an international remote sensing conference (Sandberg et al. 2015).

6.1 Interpretation of results

The used data were a Sentinel-1 time series and one Landsat 8 image. As the results imply, a single SAR image typically does not contain enough information for a successful classification task and, therefore, the choice of a time series was well justified. This also gave the chance to examine the variation in time of the study area. The boreal forest areas typically have strong seasonal variations which should be seen in the time series and can help in the classification. However, the earlier presented temporal profiles show that the mean values and standard deviations for the whole images do not change too much. The mean values vary more in the beginning of the time series compared to the end of it. These variations in the winter season might be due to more distinct changes in the weather conditions, such as snow cover. The mean values rise a bit as the growing season starts and then remain quite similar. The profiles created separately for the different classes show greater variability. All classes show bigger changes in the winter, whereas the growing season seems to affect the VV polarized images more. The mean values rise more than in the VH polarized images. In the VH polarized images, the agricultural areas show this rise as well. These variations imply that these long time series contain information that separates the classes.

6.1.1 Training and test data

The training and test samples were formed with visual interpretation from a very high resolution image over the study area. Five training and five test plots per class were extracted (each containing nine pixels in the test data). This amount of data was sufficient for performing the classifications and the validation, but it might have been better to have a larger amount of samples. More training samples could most probably have describe the statistics of the different classes better, and more test samples could have given a more reliable estimate of the classification accuracy.

Looking at the figures created from the training and test samples (Figure 13), it is evident that some classes are not very well separated. The figures were created using mean values from the different polarization images and using other values could of course affect the outcome of the figure. The issues with class separability can lead to problems in the classification stage. The most separable class, water bodies, is well classified in most cases, whereas the least separable class, urban areas, has the most issues in the classifications. The difficulties of classifying the urban class has been noted in many previous studies (e.g. Engdahl and Hyyppä 2003, Lönnqvist et al. 2010), and therefore, these results were not surprising. Poor training samples also affect the classification accuracy calculations. The accuracies indicate how well the training samples are able to represent the variability in the image and can, therefore, be misleading.

Badly chosen test samples can affect the classification accuracy. In this case, each class had the same amount of test samples even though the actual classes were of very varying size. If the test sample amounts per class were closer to the actual pixel amount per class, the results might have been different. This can be seen when comparing the validation with the test samples and the one with CLC2012. A classification map with visually clear flaws can give high final accuracies as long as the test plot pixels are classified correctly.

6.1.2 Features

The choice of features has a significant impact on the classifications. In this study, four features were chosen for both the SAR and optical data in the data fusion stage. For the SAR data, these features consisted of statistical measures. These measures work on the single pixel level and are easy to calculate. They also use all the information the time series provide which gives a bigger variability and should help in describing the pixel. Nevertheless, it could be argued that some more refined measures that take into account the neighbourhood of a pixel, so called texture measures, would have been more suitable for the SAR data.

The use of different spectral indices as features for the Landsat 8 image is justified by their common use in many other studies. Their advantage is that each index is developed to find some certain land cover class and they typically use at least two spectral bands in the calculations. The same statistical measures that were applied to the SAR data could have been used, but they might not have been as effective as only one Landsat image was used instead of a time series.

The feature selection is an important stage of the classification process. Choosing the best features is very crucial when looking for the best classification results. Two methods were chosen for this purpose and applied on the SAR data: Fisher's discriminant ratio which ranks the features' discriminatory power and stepwise regression which looks for the features that work the best together. As classes needed to be known for the data used in the feature selection, only the training and test data could be used. This data set is quite small with only 90 observations per class and could affect the success of the feature selection. If a more comprehensive data set would have been available, a more reliable choice of features could have been made. However, using two different selection methods gave a chance to evaluate the success and choose the better result. It was expected that the stepwise regression would give better results as it takes into account how features interact, but features chosen with Fisher's discriminant ratio also gave some competitive final accuracies. Nevertheless, it is evident that the visual classification results of the features chosen with stepwise regression have less noise and, therefore, look better.

An interesting result is the scene selection of four scenes performed with the two different feature selection methods. One resulted in two late winter scenes whereas the other one resulted in two summer scenes. The time series could partly give an explanation to this. The Fisher's discriminant ratio looks for the most discriminant scenes and a sudden change can be seen in some time series. The mean values of the VV polarized images have a sudden drop during the two winter dates, and this drop is also visible for the wetlands and agricultural areas for the VV polarized images. This could lead to higher FDR values. When looking at the two summer scenes, they are from the growing season which could affect the choice. Stepwise regression looks for the most significant model of scenes for the given classes and using summer scenes might result in the most information. For

example, being in the middle of the growing season improves class separability compared to a snow covered images. One of the chosen summer images also has a high FDR value.

With these ideas in mind, the selection of only one scene and one polarization acts differently. One summer and one spring scene were proposed, both VV polarized. As the time series already showed, the VV polarized images have a larger range in the pixel values, which could lead to better class separability. The spring scene is finally chosen, but looking at the acquisition conditions, (Table 3) it actually is very similar to the winter scenes with some snow cover and a similar temperature.

When working with SAR time series, it was possible to do a feature selection or scene selection. Looking at the classification results, it seems like a feature extraction was the better choice. When performing the classification with extracted features from all Sentinel-1 data, the visual results have less noise compared to classifications with selected scenes and the final accuracies are better. When doing a scene selection, all the valuable information that the time series provides is lost which can lead to a worse classification. Additionally, using only one or four scenes seems to not be enough to contain a sufficient amount of information for a land cover classification.

6.1.3 Classification methods

The classifications were performed with the nearest neighbour rule (MDM and k-NN with k=1 and k=9) and the maximum likelihood classifier. These methods are easy to implement and widely used and, therefore, are a good way of analyzing the success of data fusion. Previous studies show that especially ML has been popular in similar tasks (e.g. Huang et al. 2007, Michelson et al. 2000). More sophisticated methods could be beneficial, especially when dealing with high dimensional data. Overall, MDM was the worst method and also had very varying results. In some classifications it had even the best final accuracy, but in others it could be significantly worse than the other methods. Visually, it mainly does not succeed well. This method uses the mean values of the training samples in one class when comparing to an unknown pixel which gives so few training samples that the results are easily wrong. Therefore, it is not recommended to use this method.

The nearest neighbour methods that use all the training samples when comparing to a new pixel (1-NN and 9-NN) are effective. The results are quite similar, varying from one data set to another. Overall, these two methods produced the least noise, but it is still present. The noise varies between agricultural areas, wetlands and urban areas. The final accuracies remain more consistent than for the other two methods. They also always stay close to the best accuracy reached with each data set.

The maximum likelihood classifier resulted in both good and bad results. Overall, it had some of the best final accuracies, but in some cases the accuracy dropped very low. The visual interpretation also shows the same trend, however, the visually worse results and those with a lower final accuracy do not always match. The classification maps imply that the maximum likelihood classifier has a tendency to classify too much area as urban areas. This is evident in most of the visual results. However, this is also one reason why the final accuracies are so high. The method seems to classify the other classes as well as the other methods, but as it also has a lot of urban areas, which is the biggest issue for the other methods, it reaches high final accuracies even if the class extends too far.

If one of the four methods would have to be chosen as the best, the choice would go to either 9-NN or ML. The advantage of 9-NN is that it is very consistent throughout all data sets. Additionally, it does have some of the best results, both visually and statistically. ML has much more varying results. Some are very good, but some are worse. It has some of the best final accuracies, but mostly due to the use of stratified sampling for the test data. Therefore, ML could be the best choice for a data set known to work well with this algorithm, whereas 9-NN is a valid option when the classification method needs to work on many different data sets.

6.1.4 Classification results

The classifications were performed both separately on the SAR and optical data and with fused data. The single-sensor results mainly have worse results. The classifications made with the Sentinel-1 data are characterized by large amounts of noise. This could largely be caused by the existence of speckle in SAR imagery due to mixing of backscatter from uneven surfaces. When the original images used for the classification have some kind of noise, it is quite likely that the classifications also have this. The noise in the classification images is quite small, but covers large areas and many classes. Only the water areas are well classified, whereas the larger urban and agricultural areas also can be identified but have a lack in precision. The VH polarized images seem to result in a bit less noise. As this polarization uses both vertical and horizontal modes, one for transmitting and one for receiving, it has a bigger chance of capturing more information. Cross-polarization also gives a better contrast than a co-polarization. Other studies have also reached better classification results with VH polarized data compared to VV polarized data (e.g. Oyoshi et al. 2015). Looking at the final accuracies, it seems like using all images' information gives better results than using only one or a few scenes or some extracted features. This supports the idea of SAR images not containing enough information by themselves, but using longer time series can solve this issue.

The classification results of the Landsat 8 image are better than those using SAR data. However, much noise is still present. When using optical imagery, some areas belonging to different classes might have very similar reflectance values, which lead to misclassified areas. For example, agricultural areas and grasslands look very similar. This misinterpretation of the reflectance values is more likely the reason for the noise in this case, compared to SAR imagery where the actual images already contain some noise. An interesting phenomenon is that the noise changes from urban to agricultural areas in the ML classification when using spectral indices. This might be caused by the characteristics of the indices which focus more on vegetation than on urban areas.

One main goal of this study was to investigate data fusion of multi-temporal SAR and optical data and its classification. It has been shown that the fusion significantly improves the results. Six different fusions were made, all resulting in good classifications (mainly over 80 % final accuracies). One important improvement when using fused data in this study was the decrease of the amount of noise in the results. Compared to single-sensor classifications, these results mostly contain much less noise. This might be due to the different noise sources in the two data sets which can be eliminated when the data are used together. Even the large areas of misclassified urban areas have decreased significantly and the actual urban areas stand out better.

The visually best result used the four extracted features from the VH polarized Sentinel-1 data and four indices from the Landsat 8 image, and the classification was performed with ML. One key feature here is that the Landsat indices do not create so much urban area noise, as can be seen in the single-sensor classification. Adding the SAR data then helps with removing the agricultural noise. This classification has a final accuracy of 89.8 % which is a very good result. Even when using CLC2012 for the comparison, it reaches a 91.3 % accuracy.

The highest final accuracy that was gained was 97.3 % (four features from the VH polarized Sentinel-1 images and four Landsat 8 bands). It is apparent in the visual result that is still has some issues. When comparing the result to CLC2012, the accuracy drops to 84.1 % which is much more reasonable. This shows that it is important to examine both the visual result and the final accuracy.

The data fusion is highly dependent on the used data sets, and the fusion would not be successful if the used data were of poor quality. These two data sets already give quite good results by themselves, but still have some flaws. The fusion manages to decrease the issues. As the optical imagery give very good results by itself, the SAR imagery basically has to fill in the problematic areas that the optical images cannot classify correctly.

6.2 Comparison of results to previous studies

Comparing these results to previous studies, similarities are found, but also improvements. When comparing single-sensor classifications, it is important to note that this study was more focused on developing an efficient methodology for fused data and single-sensor classifications might not have been optimized. Additionally, previous studies used more researched data sources whereas the used SAR data in this study was from a new source, Sentinel-1, and no previous studies existed as reference. Previous studies also did not have access to as long time series as this study. Nevertheless, the results of the classifications with SAR data do not reach as high as some of the previous studies performed with SAR data. Research conducted in similar conditions, such as the study by Engdahl and Hyyppä (2003) who also worked with C-band SAR and used 14 images pairs over southern Finland, gained a final classification accuracy of 90 %. Also other similar studies have reached final accuracies between 80 % and 90 %, whereas this study had some accuracies even below 40 % and only a few of the best around 80 %.

The classification of the Landsat 8 image performed well compared to previous studies. Unlike SAR data, Landsat images do not need as much processing and are easier to use for land cover classification and other remote sensing tasks. This could be a reason why final accuracies similar to those in previous studies could be reached. Most of the reached final accuracies are around 80 %, some even over 90 %, which corresponds to previous studies' results. For example, Potapov et al. (2011) reached a final accuracy of 89 % in a study in boreal Russia.

The most important outcome of this study was the development of a new land cover classification methodology with fused data. The same data and methods have not been used in any previous study, however, similar studies which also use multi-temporal SAR and optical data exist. Like all the previously presented studies, this thesis has proven that multi-sensor classifications perform better than single-sensor classifications. Furthermore, the gained accuracies in this study are very competitive compared to earlier research and

even surpass some previous results. Previous studies, such as the one by Michelson et al. (2000) which reached a final accuracy of 57.1 %, have not been able to gain as good results as this one where most final accuracies were above 80 %. Others, such as the study by Waske and van der Linden (2008), gained a final accuracy of 84.9 %, which coincides with the results in this thesis. Overall, it can be concluded that the developed methodology has been able to reach as good and even better results than other similar tasks within the field, while constructing a method using new available data and new ways to extract features and combine them in a data fusion stage. These new strategies have improved the current methodologies.

6.3 Recommendations for future work

This study has shown that fusing multi-temporal SAR and optical data is a successful strategy for land cover classification. Many different methods and data sets have been examined, but some improvements can still be made. First, a more extensive training and test set would be recommendable. A larger training set could better characterize the different classes and a larger test set could help to give a more reliable validation result. Second, another set of features could improve the results. A wide variety of features are available and another set might characterize the data better. For example, features that are able to better characterize the variations in a time series could be beneficial. The amount of features could also affect the results. Third, the used classification methods play a great role in the success of the classifications. Some more sophisticated methods could be tested as they might be able to better separate the classes.

On a larger scale, multi-temporal SAR and optical data fusion for land cover mapping can be developed much further. When even longer SAR time series will be available, the seasonal changes will be even more visible which can be used for creating better features. Especially the development of SAR data and its processing will enable better ways to interpret the data. Regarding different classification methods and algorithms, the need for new ones is not very urgent. However, new and improved algorithms could be useful and help to this could be searched from other fields, such as machine learning. Another aspect would be to create more automatic and near real-time methods. With a larger and faster data acquisition, processing of this data will not be possible if larger parts of the system are not automatized. This would also require up-to-date and reliable reference data. To conclude, the emergence of new data and better methods provide endless possibilities for multi-sensor and multi-temporal data fusion for land cover classification. As new techniques are tested, the results will surely improve even more.

Bibliography

Aghabozorgi, S., Seyed Shirkhorshidi, A. & Ying Wah, T. 2015. "Time-series clustering – A decade review". *Information Systems*, vol. 53, no. C, pp. 16-38. ISSN 0306-4379.

Antropov, O., Rauste, Y., Astola, H., Praks, J., Hame, T. & Hallikainen, M.T. 2014. "Land Cover and Soil Type Mapping From Spaceborne PolSAR Data at L-Band With Probabilistic Neural Network". *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 9, pp. 5256-5270. ISSN 0196-2892.

Asmala, A. 2012. "Analysis of maximum likelihood classification on multispectral data". *Applied Mathematical Sciences*, vol. 6, no. 129-132, pp. 6425-6436. ISSN 1314-7552.

Atzberger, C. & Eilers, P.H. 2011. "Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements". *International Journal of Remote Sensing*, vol. 32, no. 13, pp. 3689-3709. ISSN 1366-5901.

Brooks, E.B., Thomas, V.A., Wynne, R.H. & Coulston, J.W. 2012. "Fitting the Multitemporal Curve: A Fourier Series Approach to the Missing Data Problem in Remote Sensing Analysis". *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 9, pp. 3340-3353. ISSN 0196-2892.

Chust, G., Ducrot, D. & Pretus, J.L. 2004. "Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape". *International Journal of Remote Sensing*, vol. 25, no. 17, pp. 3513-3528. ISSN 0143-1161.

Cohen, W.B., Yang, Z. & Kennedy, R. 2010. "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync — Tools for calibration and validation". *Remote Sensing of Environment*, vol. 114, no. 12, pp. 2911-2924. ISSN 0034-4257.

Draper, N.R. & Smith, H. 1998. *Applied Regression Analysis*. 3rd ed. New York, NY: John Wiley & Sons. ISBN 9780471170822.

Duda, R.O., Hart, P.E. & Stork, D.G. 2012. *Pattern Classification*. 2nd ed. New York, NY: John Wiley & Sons. ISBN 9781118586006.

Eisavi, V., Homayouni, S., Yazdi, A.M. & Alimohammadi, A. 2015. "Land cover mapping based on random forest classification of multitemporal spectral and thermal images". *Environmental monitoring and assessment*, vol. 187, no. 5, pp. 1-14. ISSN 1573-2959.

Engdahl, M.E. & Hyyppä, J.M. 2003. "Land-cover classification using multitemporal ERS-1/2 InSAR data". *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 41, no. 7, pp. 1620-1628. ISSN 0196-2892.

European Space Agency, *Introducing Sentinel-1* [Homepage of European Space Agency], [Online]. Available:

http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-1/Introducing_Sentinel-1 [2015, 09/25].

Finnish Meteorological Institute, *Weather Observations, Juupajoki Hyytiälä weather station* [Homepage of Finnish Meteorological Institute], [Online]. Available: https://ilmatieteenlaitos.fi/avoin-data [2016, 01/20].

Foody, G.M. 2002. "Status of land cover classification accuracy assessment". *Remote Sensing of Environment*, vol. 80, no. 1, pp. 185-201. ISSN 0034-4257.

García, M.J.L. & Caselles, V. 1991. "Mapping burns and natural reforestation using thematic Mapper data". *Geocarto International*, vol. 6, no. 1, pp. 31-37. ISSN 1010-6049.

Gupta, R.P., Tiwari, R.K., Saini, V. & Srivastava, N. 2013. "A simplified approach for interpreting principal component images". *Advances in Remote Sensing*, vol. 2, no. 2, pp. 111-119. ISSN 2169-2688.

Guyon, I. 2003. "An introduction to variable and feature selection". *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182. ISSN 1533-7928.

Hall, D.L. & McMullen, S.A.H. 2004. *Mathematical Techniques in Multisensor Data Fusion*. 2nd ed. Norwood, MA: Artech House. ISBN 9781580533355.

Henderson, F.M., Chasan, R., Portolese, J. & Hart Jr, T. 2002. "Evaluation of SAR-optical imagery synthesis techniques in a complex coastal ecosystem". *Photogrammetric Engineering and Remote Sensing*, vol. 68, no. 8, pp. 839-846. ISSN 0099-1112.

Huang, H., Legarsky, J. & Othman, M. 2007. "Land-cover classification using Radarsat and Landsat imagery for St. Louis, Missouri". *Photogrammetric Engineering & Remote Sensing*, vol. 73, no. 1, pp. 37-43. ISSN 0099-1112.

Idol, T., Haack, B. & Mahabir, R. 2015. "Comparison and integration of spaceborne optical and radar data for mapping in Sudan". *International Journal of Remote Sensing*, vol. 36, no. 6, pp. 1551-1569. ISSN 1366-5901.

Japan Association on Remote Sensing 1993. *Remote Sensing Note*. Japan Association on Remote Sensing.

Jin, H., Stehman, S.V. & Mountrakis, G. 2014. "Assessing the impact of training sample selection on accuracy of an urban classification: a case study in Denver, Colorado". *International Journal of Remote Sensing*, vol. 35, no. 6, pp. 2067-2081. ISSN 0143-1161.

Johnson, R.A. & Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. 6th ed. Upper Saddle River, NJ: Pearson Prentice Hall. ISBN 9780131877153.

Kennedy, R.E., Yang, Z. & Cohen, W.B. 2010. "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation

algorithms". *Remote Sensing of Environment,* vol. 114, no. 12, pp. 2897-2910. ISSN 0034-4257.

Kurvonen, L., Pulliainen, J. & Hallikainen, M. 2002. "Active and passive microwave remote sensing of boreal forests". *Acta Astronautica,* vol. 51, no. 10, pp. 707-713. ISSN 0094-5765.

Landgrebe, D.A. 2003. *Signal theory methods in multispectral remote sensing*. Hoboken, NJ: John Wiley & Sons. ISBN 9780471420286.

Lauer, D.T., Morain, S.A. & Salomonson, V.V. 1997. "The Landsat program: Its origins, evolution, and impacts". *Photogrammetric Engineering and Remote Sensing*, vol. 63, no. 7, pp. 831-838. ISSN 0099-1112.

Lhermitte, S., Verbesselt, J., Verstraeten, W.W. & Coppin, P. 2011. "A comparison of time series similarity measures for classification and change detection of ecosystem dynamics". *Remote Sensing of Environment,* vol. 115, no. 12, pp. 3129-3152. ISSN 0034-4257.

Li, E., Du, P., Samat, A., Xia, J. & Che, M. 2015. "An automatic approach for urban land-cover classification from Landsat-8 OLI data". *International Journal of Remote Sensing*, vol. 36, no. 24, pp. 5983-6007. ISSN 0143-1161.

Lillesand, T.M., Kiefer, R.W. & Chipman, J.W. 2008. *Remote sensing and image interpretation*. 6th ed. New York, NY: John Wiley & Sons. ISBN 9780470052457.

Lönnqvist, A., Rauste, Y., Molinier, M. & Häme, T. 2010. "Polarimetric SAR Data in Land Cover Mapping in Boreal Zone". *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 10, pp. 3652-3662. ISSN 0196-2892.

Lu, D. & Weng, Q. 2007. "A survey of image classification methods and techniques for improving classification performance". *International Journal of Remote Sensing*, vol. 28, no. 5, pp. 823-870. ISSN 0143-1161.

Manakos, I. & Braun, M. 2014. *Land use and land cover mapping in Europe*. Dordrecht, Heidelberg, New York, London: Springer. ISBN 9789400779693.

McCloy, K.R. 2010. "Development and Evaluation of Phenological Change Indices Derived from Time Series of Image Data". *Remote Sensing,* vol. 2, no. 11, pp. 2442-2473. ISSN 2072-4292.

McFeeters, S.K. 1996. "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features". *International Journal of Remote Sensing*, vol. 17, no. 7, pp. 1425-1432. ISSN 0143-1161.

Michelson, D.B., Liljeberg, B.M. & Pilesjö, P. 2000. "Comparison of Algorithms for Classifying Swedish Landcover Using Landsat TM and ERS-1 SAR Data". *Remote Sensing of Environment*, vol. 71, no. 1, pp. 1-15. ISSN 0034-4257.

Mountrakis, G., Im, J. & Ogole, C. 2011. "Support vector machines in remote sensing: A review". *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247-259. ISSN 0924-2716.

Oyoshi, K., Tomiyama, N., Okumura, T., Sobue, S. & Sato, J. 2015. "Mapping rice-planted areas using time-series synthetic aperture radar data for the Asia-RiCE activity". *Paddy and Water Environment*, pp. 1-10. ISSN 1611-2490.

Petitjean, F., Inglada, J. & Gançarski, P. 2012. "Satellite image time series analysis under time warping". *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 8, pp. 3081-3095. ISSN 0196-2892.

Petitjean, F., Gançarski, P., Masseglia, F. & Forestier, G. 2010. "Analysing Satellite Image Time Series by Means of Pattern Mining". *In: Intelligent Data Engineering and Automated Learning - IDEAL 2010*, eds. Fyfe, C., Tino, P., Charles, D., Garcia-Osorio, C. & Yin, H., Paisley, UK, 1-3 September 2010, Berlin, Heidelberg: Springer, pp. 45-52. ISBN 9783642153808.

Pohl, C. & Van Genderen, J.L. 1998. "Multisensor image fusion in remote sensing: Concepts, methods and applications". *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823-854. ISSN 0143-1161.

Pölönen, I. 2013. *Discovering knowledge in various applications with a novel hyperspectral imager*. Dissertation. University of Jyväskylä. Jyväskylä.

Potapov, P., Turubanova, S. & Hansen, M.C. 2011. "Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia". *Remote Sensing of Environment*, vol. 115, no. 2, pp. 548-561. ISSN 0034-4257.

Powell, S.L., Pflugmacher, D., Kirschbaum, A.A., Kim, Y. & Cohen, W.B. 2007. "Moderate resolution remote sensing alternatives: A review of Landsat-like sensors and their applications". *Journal of Applied Remote Sensing*, vol. 1, no. 1, pp. 1-16. ISSN 1931-3195.

Press, W.H. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. New York, NY: Cambridge University Press. ISBN 9780521431088.

Pulliainen, J.T., Mikhela, P.J., Hallikainen, M.T. & Ikonen, J.-. 1996. "Seasonal dynamics of C-band backscatter of boreal forests with applications to biomass and soil moisture estimation". *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 34, no. 3, pp. 758-770. ISSN 0196-2892.

Rahman, H. & Dedieu, G. 1994. "SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum". *International Journal of Remote Sensing*, vol. 15, no. 1, pp. 123-143. ISSN 0143-1161.

Reiche, J. 2015. *Combining SAR and optical satellite image time series for tropical forest monitoring*. Dissertation. Wageningen University. Wageningen.

Richards, J.A. 2005. "Analysis of remotely sensed data: the formative decades and the future". *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 3, pp. 422-432. ISSN 0196-2892.

Rouse, J.W., Haas, R.H., Schell, J.A. & Deering, D.W. 1973. "Monitoring Vegetation Systems in the Great Plains with ERTS". *In: Third Earth Resources Technology Satellite-1 Symposium- Volume I: Technical Presentations. NASA SP-351*, eds. Freden, S.C., Mercanti, E.P. & Becker, M.A., Washington, DC, USA, December, 10-14, 1973, Washington, D.C.: NASA, pp. 309-17.

Roy, D.P., Wulder, M.A., Loveland, T.R., C.E., W., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H. & Zhu, Z. 2014. "Landsat-8: Science and product vision for terrestrial global change research". *Remote Sensing of Environment*, vol. 145, pp. 154-172. ISSN 0034-4257.

Sandberg, M., Rauste, Y., Häme, T. & Antropov, O. 2015. "Land cover mapping with multitemporal C-band SAR and optical satellite data in the boreal forest zone". *In: International Geoscience and Remote Sensing Symposium (IGARSS)*, submitted, 2015.

Sasaki, N., Takejima, K., Kusakabe, T. & Sweda, T. 2001. "Forest cover classification using Landsat Thematic Mapper data for areal expansion of line LAI estimate generated through airborne laser profiler". *Polar bioscience*, vol. 14, pp. 110-121. ISSN 1344-6231.

Shannon, C.E. 1948. "A mathematical theory of communication". *Bell System Technical Journal, The*, vol. 27, no. 3, pp. 379-423. ISSN 0005-8580.

Theodoridis, S. & Koutroumbas, K. 2009. *Pattern recognition*. 4th ed. Amsterdam, Boston, Paris: Academic Press. ISBN 9781597492720.

Theodoridis, S., Pikrakis, A., Koutroumbas, K. & Cavouras, D. 2010. *Introduction to Pattern Recognition: A Matlab Approach*. Burlington, MA: Academic Press. ISBN 9780123744869.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I.N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M. & Rostan, F. 2012. "GMES Sentinel-1 mission". *Remote Sensing of Environment*, vol. 120, pp. 9-24. ISSN 0034-4257.

Villa, P., Fontanelli, G. & Crema, A. 2015. "Integration of multi-seasonal Landsat 8 and TerraSAR-X data for urban mapping: An assessment". *In: 2015 Joint Urban Remote Sensing Event (JURSE)*, Lausanne, Switzerland, 30 March - 1 April 2015, IEEE, pp. 1-4. ISBN 9781479966530.

Wagner, W., Sabel, D., Doubkova, M., Hornacek, M., Schlaffer, S. & Bartsch, A. 2012. "Prospects of Sentinel-1 for land applications". *In: International Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, Germany, 22-27 July 2012, IEEE, pp. 1741-1744. ISBN 9781467311595.

Warren, S.G. 1982. "Optical properties of snow". *Reviews of Geophysics and Space Physics*, vol. 20, no. 1, pp. 67-89. ISSN 1944-9208.

Waske, B. & Braun, M. 2009. "Classifier ensembles for land cover mapping using multitemporal SAR imagery". *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 5, pp. 450-457. ISSN 0924-2716.

Waske, B. & van der Linden, S. 2008. "Classifying Multilevel Imagery From SAR and Optical Sensors by Decision Fusion". *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 5, pp. 1457-1466. ISSN 0196-2892.

Wulder, M.A., White, J.C., Goward, S.N., Masek, J.G., Irons, J.R., Herold, M., Cohen, W.B., Loveland, T.R. & Woodcock, C.E. 2008. "Landsat continuity: Issues and opportunities for land cover monitoring". *Remote Sensing of Environment*, vol. 112, no. 3, pp. 955-969. ISSN 0034-4257.

Xu, H. 2007. "Extraction of Urban Built-up Land Features from Landsat Imagery Using a Thematicoriented Index Combination Technique". *Photogrammetric Engineering & Remote Sensing*, vol. 73, no. 12, pp. 1381-1391. ISSN 0099-1112.

Zha, Y., Gao, J. & Ni, S. 2003. "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery". *International Journal of Remote Sensing*, vol. 24, no. 3, pp. 583-594. ISSN 0143-1161.