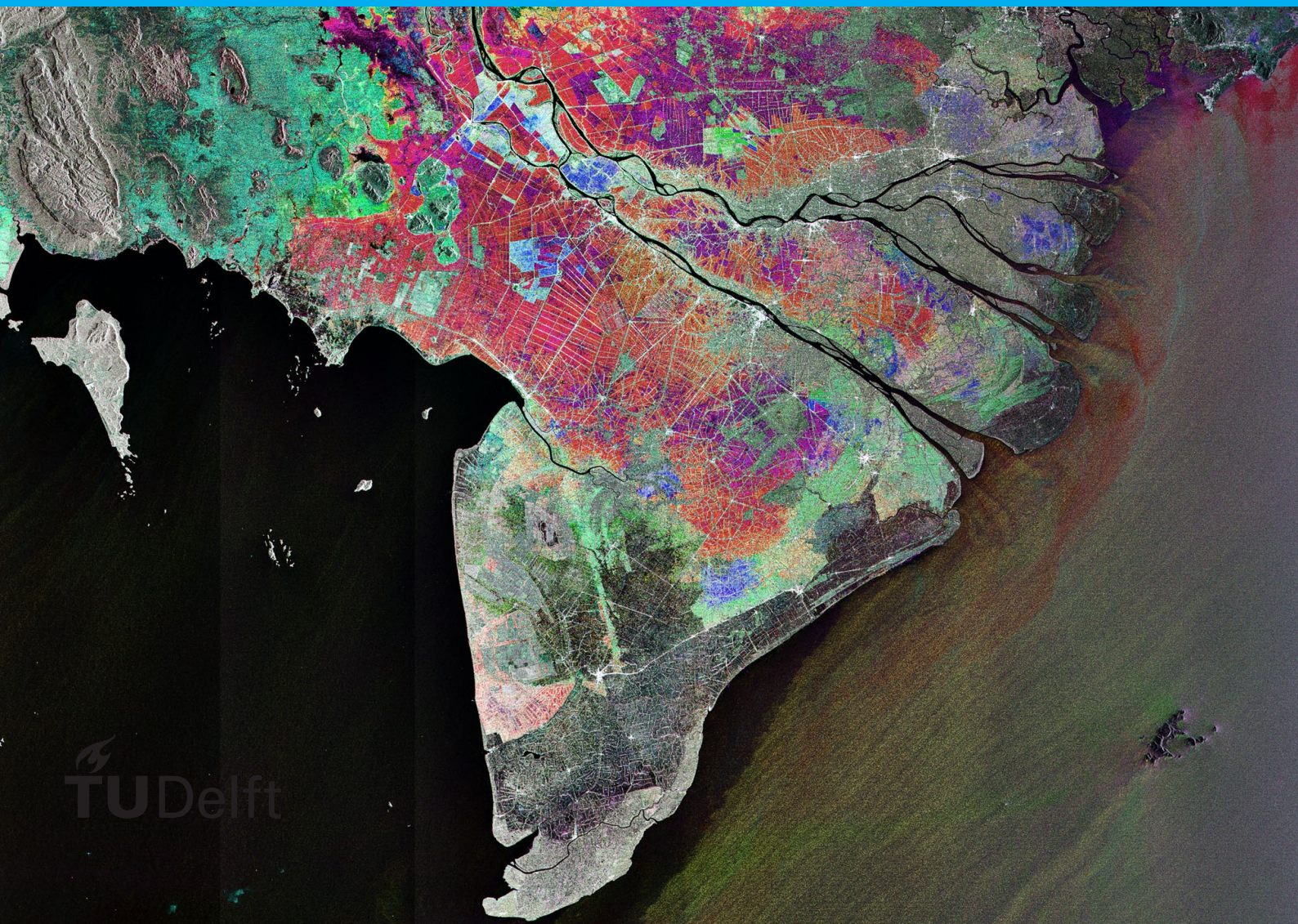


# Classifying Mangroves in Vietnam using Radar and Optical Satellite Remote Sensing

Processing Sentinel-1 and Sentinel-2  
Imagery in Google Earth Engine

E.G. Portengen





# Classifying Mangroves in Vietnam using Radar and Optical Satellite Remote Sensing

Processing Sentinel-1 and Sentinel-2 Imagery in  
Google Earth Engine

by

E.G. Portengen

to obtain the degree of Master of Science  
in Applied Earth Sciences (Geoscience & Remote Sensing)  
at the Delft University of Technology,  
to be defended publicly on 15th December 2017

Student number: 4156641  
Thesis committee: Dr. R. Lindenbergh, TU Delft, daily supervisor  
Prof. M. Menenti, TU Delft  
M.H. Phan, TU Delft  
R.A. Molijn, TU Delft  
Dr. V. Phan Hien, Ho Chi Minh City University of Technology

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.



# Abstract

Mangroves are forest ecosystems growing in (sub)tropical saline coastal environments. With their unique root structure they serve as important natural coastal protection and provide habitats with excellent conditions for cultivating fish, shrimp and crab species. Despite all benefits mangrove forests are disappearing at alarming rates around the world but especially in Asia such as the Mekong Delta coast. Therefore, this research focusses on the Ca Mau Province in Vietnam. The Ca Mau province is the southernmost province of Vietnam with mangroves present along the coastlines, the Mui Ca Mau National Park and in mixed mangrove aquaculture farms. Remote sensing has been widely proven to be essential in mapping mangrove ecosystems. Previous research used either expensive optical and radar data sources or free but lower resolution systems. This study is the first that uses the new Copernicus Sentinel-1 radar and Sentinel-2 multispectral satellite missions that provide free available data with high spatial (10-20 meter) and temporal (10-12 days) resolution. Since optical data is prone to cloud effects and radar data is hard to interpret, both data sets are combined to investigate improvements for classifying mangroves. The data is processed in the new online Google Earth Engine platform providing a powerful tool for big data applications such as land cover classification. Optical data is found to separate mangroves by their spectral reflectance mainly in the near-infrared wavelength domain. The dominant mangrove species in the Ca Mau province, *Rhizophora Apiculata* and *Avicennia Alba*, are found to be separable from comparing unsupervised clustering results with ground truth locations. The C-band radar signal is dominated by volume scattering, indicating the density of the canopy. Especially VV-polarization has good correlation with canopy parameters. To improve information from the radar signal a temporal analysis is executed. Seasonal variations are quantified and show an increase according to the spatial succession of mangroves. Pioneer species, such as *Avicennia* genus, show less seasonal variations than mature species, such as *Rhizophora* genus. With the previous information five classes are defined: urban area, water and three mangrove classes: *Rhizophora Apiculata* species in extensive shrimps, *Rhizophora Apiculata* species in natural environment and *Avicennia Alba* species. A classification method is set-up in the Google Earth Engine with a Random Forest classifier using the satellite data inputs and ground truth training input of the five classes. A combination of the optical data with the temporal information of the radar data is found to be the best data input for separating those five classes. Classification results are obtained for discriminating mangrove types up to an overall accuracy of 87%. The classification gets less reliable when mangrove species are mixed or at locations where the ground truth training input was scarce. With the resulting yearly land cover maps land cover changes can be detected. Comparing the land cover map of 2017 with a mangrove cover product of 2000 shows a regression along the southern coastline. No significant changes inside the shrimp farms are found between 2016 and 2017 but with the future availability of a long time series of Sentinel-1 and 2 data those can be detected with the method that is resulted from this study.



# Preface

In spring 2013 I came first in contact with remote sensing during the course of data analysis and geostatistics in the bachelor of Applied Earth Science. Since then it got my attention and I decided to contact Roderik Lindenbergh, resulting in a bachelor thesis about remote sensing of glaciers. After nearly 2 years of courses in the master track of Geoscience & Remote Sensing I came back to Roderik to discuss about a master thesis using satellite remote sensing data. Since I decided to quit my rowing career after six years of intensive training, I wanted to take my last opportunity of going abroad during my studies.

This resulted in a trip to Ho Chi Minh City, Vietnam, where I worked on the university and carried out a fieldwork campaign to gather the ground-truth data used for this thesis research. I would like to thank Vu Phan Hien for proposing the thesis topic to me and for helping me during my time in Vietnam. Introducing me to the many different kind of foods and not the least spending three days with me driving through the farms of Ca Mau looking for interesting mangrove locations.

Back in Delft I would like to thank my other supervisors, Ramses Molijn, Hung Phan and prof. Menenti, for the discussions about the processing of my data and an always critical look into my results. Of course I owe a big thanks to Roderik for the almost weekly meetings about my proceedings and your always cheerful mood to get the most out of my thesis.

Last but not least I would like to thank my family, friends and Swen for keeping me motivated. Swen, during all the hours we spend in the library, drinking coffee and complaining about our non-working scripts and bad results, you were my best company.

*Leonoor Portengen  
Delft, December 2017*





# Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Vietnamese Mekong Delta and Ca Mau province . . . . .	1
1.2 Mangrove ecosystems. . . . .	2
1.3 Sentinel missions . . . . .	3
1.4 Problem statement and project description . . . . .	3
1.5 Research questions . . . . .	3
1.6 Report outline. . . . .	4
<b>2 Remote sensing of mangroves</b>	<b>5</b>
2.1 Mangroves in Vietnam . . . . .	5
2.1.1 Coastal protection . . . . .	5
2.1.2 Fauna and aquaculture . . . . .	6
2.1.3 History. . . . .	6
2.1.4 Climate change . . . . .	6
2.1.5 Carbon storage. . . . .	6
2.2 Mangrove properties . . . . .	6
2.2.1 Rhizophora genus . . . . .	7
2.2.2 Avicennia genus . . . . .	7
2.2.3 Others . . . . .	7
2.3 Remote sensing data . . . . .	9
2.3.1 Optical remote sensing. . . . .	9
2.3.2 Radar remote sensing . . . . .	9
2.4 Remote sensing of mangrove properties . . . . .	11
2.4.1 Spectral properties of mangrove . . . . .	11
2.4.2 Spectral discrimination of mangrove species. . . . .	11
2.4.3 Mangrove properties in radar data . . . . .	12
2.4.4 Discrimination of mangrove species using radar. . . . .	14
2.5 Issues with mangrove remote sensing. . . . .	15
2.5.1 Data . . . . .	15
2.5.2 Data fusion . . . . .	15
2.5.3 Temporal analysis . . . . .	15
2.5.4 Cloud-based processing and classification. . . . .	15
2.6 Conclusions. . . . .	16
<b>3 Data and site description</b>	<b>17</b>
3.1 Region of interest . . . . .	17
3.1.1 Fieldwork campaign . . . . .	18
3.1.2 Training classes . . . . .	18
3.2 Sentinel-1 mission . . . . .	20
3.2.1 Radar data processing . . . . .	20
3.3 Sentinel-2 mission . . . . .	21
3.3.1 Optical data processing . . . . .	22
3.4 Exploratory data analysis . . . . .	23
3.4.1 Places of interest. . . . .	23
3.4.2 Spectral signatures. . . . .	28
3.4.3 NDVI. . . . .	30

3.5	Software . . . . .	31
3.5.1	Epicollect+. . . . .	31
3.5.2	Sen2Cor and SNAP. . . . .	32
3.5.3	Matlab . . . . .	32
3.5.4	Google Earth Engine . . . . .	32
3.6	Conclusions. . . . .	33
<b>4</b>	<b>Methods for mangrove classification</b>	<b>35</b>
4.1	Data input . . . . .	35
4.2	Unsupervised clustering . . . . .	36
4.2.1	K-means algorithm . . . . .	36
4.2.2	Spectral signatures of clusters . . . . .	36
4.2.3	Clustering of vegetated areas. . . . .	36
4.3	Supervised classification . . . . .	37
4.3.1	Classification inputs . . . . .	37
4.3.2	Random forest algorithm . . . . .	38
4.4	Temporal information . . . . .	39
4.5	Validation. . . . .	41
4.6	Conclusions. . . . .	42
<b>5</b>	<b>Results</b>	<b>43</b>
5.1	Classification results . . . . .	43
5.1.1	Yearly land cover map . . . . .	43
5.1.2	Accuracy and confusion matrix . . . . .	45
5.1.3	Classification confidence . . . . .	47
5.1.4	External validation. . . . .	48
5.1.5	Case studies . . . . .	49
5.2	Workflow . . . . .	52
5.3	Temporal analysis. . . . .	53
5.4	Unsupervised clustering . . . . .	57
5.5	Conclusions. . . . .	59
<b>6</b>	<b>Discussion</b>	<b>61</b>
6.1	Data processing. . . . .	61
6.1.1	Cloud masking. . . . .	61
6.1.2	Google Earth Engine . . . . .	61
6.1.3	Fieldwork campaign . . . . .	62
6.2	Classification results . . . . .	62
6.2.1	Single images . . . . .	62
6.2.2	Different data input . . . . .	63
6.3	Comparison with external research results . . . . .	64
6.4	Future applications . . . . .	66
6.5	Conclusions. . . . .	67
<b>7</b>	<b>Conclusions and recommendations</b>	<b>69</b>
7.1	Conclusions. . . . .	69
7.1.1	Research questions . . . . .	69
7.2	Recommendations . . . . .	71
7.2.1	Improvements . . . . .	71
7.2.2	Further research for future applications . . . . .	71
	<b>Bibliography</b>	<b>73</b>
<b>A</b>	<b>Results</b>	<b>77</b>
A.1	Land cover maps using different data input. . . . .	77
A.2	Detailed workflow. . . . .	80
A.3	Temporal analysis. . . . .	81
A.4	Clustering. . . . .	85

# List of Figures

1.1	Map of the Mekong Delta in Vietnam [33] . . . . .	2
1.2	Map of the Ca Mau province [9] . . . . .	2
1.3	Different types of shrimp farming in Ca Mau province . . . . .	2
2.1	Succession of mangrove in Ca Mau cape, showing differences typical of each zone [40] . . . . .	7
2.2	Different mangrove species in Mekong Delta . . . . .	8
2.3	Spectral signature plot. Examples of typical spectral signatures of green vegetation, water and soil are shown . . . . .	10
2.4	Cellular leaf structure and its interaction with electromagnetic energy. The top layer with most chlorophyll absorbs most visible light except for green light. The lower layers, consisting of irregularly shaped cells, highly reflect the near infrared wavelengths . . . . .	10
2.5	The electromagnetic spectrum showing the parts that are important in remote sensing. The Shortwave Infrared (SWIR) belongs in between the Near IR and Thermal IR ranging from 1 to 3 $\mu\text{m}$ [45] . . . . .	10
2.6	Influence of roughness on radar backscatter . . . . .	10
2.7	Influence of incidence angle on radar backscatter . . . . .	10
2.8	Surface and volume scattering of a radar beam for trees [19] . . . . .	10
2.9	Spectral characteristics and their influencing parameters of mangrove species <i>Avicennia marina</i> and <i>Rhizophora Conjugata</i> (synonym of <i>Rhizophora Apiculata</i> ) as measured with an field spectrometer in Ca Mau province, Vietnam [32] . . . . .	12
2.10	Dominating backscatter mechanisms at different stages of mangrove growth depending on bandwidth of radar beam [32]. . . . .	13
2.11	Schematic representation of the successive growth stages of the mangrove forests [37]. . . . .	13
2.12	Overall accuracy's for mangrove classification using PALSAR L-band and SPOT5 using two different classification schemes: Support Vector Machine (SVM) and Maximum Likelihood (MLC) . . . . .	14
3.1	Region of interest bounding box with locations of fieldwork campaign on Google Earth imagery. . . . .	17
3.2	Overview of training data classes in the ROI. The yellow numbers represent class 1: Mangrove vegetation: <i>Avicennia Alba</i> species. Bright green is class 2: Mangrove vegetation: <i>Rhizophora Apiculata</i> species in natural environment. Dark green is class 3: Mangrove vegetation: <i>Rhizophora Apiculata</i> in extensive shrimp farms. Blue is class 4: Water areas and red is class 5: Urban area and barren land . . . . .	19
3.3	Observation scenario of Sentinel-1 [16]. . . . .	20
3.4	Different modes of Sentinel-1 [16]. . . . .	20
3.5	Sentinel-1 color composite of bands of polarizations VV, VH and $\frac{VV}{VH}$ with the visible effect of temporal averaging . . . . .	21
3.6	Different bands of Sentinel-2 [12]. . . . .	21
3.7	Observation scenario of Sentinel-2 [17]. . . . .	22
3.8	Automatic ccene classification made during atmospheric correction for the Sentinel-2 image of 26-04-2016 . . . . .	22
3.9	Sentinel-2 images visualized in natural color composite. The yellow markers indicate the GPS locations of the fieldwork data. . . . .	23
3.10	Different POI in the natural color Sentinel-2 image of 26-04-2016. The green box contains the area inside the Mui Ca Mau National Park. The red boxes contains the urban communes of Dat Mui, Rach Goc and Nam Can. In the yellow box are the coastal areas with mangroves. The blue box contains area which consists of mixed mangrove and fish/shrimp farms. . . . .	23
3.11	Sentinel-1 color composite of bands of polarizations VV, VH and $\frac{VV}{VH}$ of temporal averaged April 2016. Zoomed into an area with extensive shrimp farms (up) and the Mui Ca Mau area (down). . . . .	24
3.12	Different zones in the Mui Ca Mau National Park . . . . .	24

3.13	Sentinel-2 images of 26th april 2016 zoomed into Mui Ca Mau National park . . . . .	25
3.14	Timeseries of Sentinel-2 imagery in Mui Ca Mau National Park . . . . .	26
3.15	Average precipitation (rain/snow) in Ho Chi Minh City, Vietnam [28] . . . . .	26
3.16	Time series of VV radar data on different mangrove places of interest . . . . .	26
3.17	Sentinel-2 images of 26th april 2016 zoomed in on the southern coast . . . . .	27
3.18	Time series of Sentinel-2 imagery along southern coast . . . . .	27
3.19	Sentinel-2 images of 26th April 2016 zoomed in on location of mangrove plantations . . . . .	28
3.20	Time series of Sentinel-2 imagery in extensive shrimp farms . . . . .	28
3.21	Spectral signatures for different places of interest . . . . .	29
3.22	NDVI values of a cloud free mosaic (the median) of all Sentinel-2 images. Color ranges from Blue (value 0) to White to Green (value 1) . . . . .	30
3.23	Time series of NDVI on different places of interest . . . . .	31
3.24	Epicollect+ project and question form on web server app and on mobile application . . . . .	31
3.25	Epicollect+ project and result table on web server app . . . . .	32
3.26	Screenshot of Google Earth Engine Javascript API . . . . .	33
4.1	Available images serving as input for the classification methods. The grey dots represent the available images from the Sentinel-2A satellite with the exact dates indicated. The blue and red dots represent images from the Sentinel-1A and 1B satellite respectively. . . . .	35
4.2	Spectral signatures for different places of interest . . . . .	37
4.3	Overall accuracy's for different classifiers in Sentinel-2 classification . . . . .	38
4.4	Random forest model. <b>A</b> represents the training process, where decision trees are built from random samples of the original data, which contains positive (green labels) and negative (red labels) examples. <b>B</b> represents the classification process where the majority of class from the individual trees is chosen. For each data in each tree the algorithm starts at the root node of a decision tree and traverses down the tree testing the variables values in each of the visited split nodes (pale pink nodes) until a leaf node is reached: green nodes predict for the positive class, red nodes predict for the negative class [34] . . . . .	39
4.5	Basics of Fourier analysis [8] . . . . .	39
4.6	HANTS Fourier fit at National Park . . . . .	40
4.7	Mangrove Forest of the World from 2000 [22] . . . . .	41
5.1	Yearly land cover maps of 2016 and 2017 and difference map. . . . .	44
5.2	Overall accuracy's for Random Forest classification . . . . .	46
5.3	Classification confidence for Sentinel-2 & Sentinel-1 temporal input classification of 2017. The images are coloured from white to the specific color corresponding to the minimum and maximum value. . . . .	47
5.4	Comparison of land cover map of 2017 with Mangrove Cover from 2000. Bright green, dark green and yellow pixels indicate the mangroves classified in this research and coincide with the Mangrove Cover from 2000. Black pixels indicate mangrove cover in 2000 that is not classified in 2017 as mangrove vegetation . . . . .	48
5.5	Different locations of the case studies . . . . .	49
5.6	Yearly 2016 and 2017 land cover maps with ground truth examples at farms near Nam Can bridge	50
5.7	Yearly 2016 and 2017 land cover maps with ground truth examples at farms along southern coast	50
5.8	Yearly 2016 and 2017 land cover maps with ground truth examples at north coast National Park location . . . . .	51
5.9	Workflow for classifying mangroves using Sentinel-1 & 2 satellite data . . . . .	52
5.10	Second amplitude term for VV backscatter time series. Black rectangles indicate POI . . . . .	53
5.11	VV backscatter time series at different POI . . . . .	54
5.12	VH backscatter time series at different POI . . . . .	54
5.13	Correlation of VV and VH backscatter for each different land cover class . . . . .	55
5.14	VV backscatter time series at POI 1 including different Fourier fits . . . . .	56
5.15	VV backscatter time series at POI 2 including different Fourier fits . . . . .	56
5.16	Clustering scheme showing the first step making four seperate clusters. The second step clusters only the (vegetation) pixels from cluster 3 into four new clusters to better differentiate vegetation . . . . .	57

---

5.17 Spectral signatures from cluster analysis for Sentinel-2 median image . . . . .	58
5.18 Cluster map from cluster analysis for Sentinel-2 median image . . . . .	58
6.1 Single Sentinel-2 image classification results . . . . .	62
6.2 Land cover result in 2017 from different data combinations zoomed into location with extensive shrimp farms . . . . .	64
6.3 Classification results from Vo et al [58] . . . . .	65
6.4 Classification results from Kuenzer et al [32] . . . . .	66
A.1 Land cover map of 2017 using Sentinel-2 data . . . . .	77
A.2 Land cover map of 2017 using Sentinel-1 data . . . . .	78
A.3 Land cover map of 2017 using Sentinel-1 & 2 data . . . . .	78
A.4 Land cover map of 2017 using Sentinel-1 temporal data . . . . .	79
A.5 Land cover map of 2017 using Sentinel-1 temporal and Sentinel-2 data . . . . .	79
A.6 Detailed workflow for classifying mangroves using Sentinel-1 and Sentinel-2 satellite imagery .	80
A.7 Second amplitude term for VH cackscatter time series. Black rectangles indicate POI. . . . .	81
A.8 Mean for VH backscatter time series. Black rectangles indicate POI . . . . .	81
A.9 Mean for VV backscatter time series. Black rectangles indicate POI . . . . .	82
A.10 First amplitude term for VV backscatter time series. Black rectangles indicate POI . . . . .	82
A.11 Third amplitude term for VV backscatter time series. Black rectangles indicate POI . . . . .	83
A.12 VV backscatter time series at POI 3 . . . . .	83
A.13 VV backscatter time series at POI 4 . . . . .	84
A.14 VV backscatter time series at POI 5 . . . . .	84
A.15 Cluster analysis for Sentinel-2 median image for $k=4$ . . . . .	85
A.16 Cluster analysis for combined Sentinel-1 & 2 median image for $k=4$ . . . . .	85
A.17 Cluster analysis for combined Sentinel-1 & 2 median image for $k=10$ in random colors . . . . .	86



# List of Tables

5.1	Overview of data input per method . . . . .	43
5.2	Land cover changes from 2016 to 2017. Percentages with respect to class in 2016. . . . .	45
5.3	Confusion matrix from Sentinel-2 and Sentinel-1 temporal data input classification results compared to validation ground truth data . . . . .	46
6.1	Overview of percentage of total pixels per class . . . . .	63





# Introduction

Mangroves are forest ecosystems that occur in saline coastal environments where a tropical or subtropical climate is present. With their salt tolerant roots mangrove ecosystems are able to survive in intertidal areas where a combination of salt sea water and fresh river water is present. Mangroves are valuable ecological and economic resource and they serve as a very important factor in coastal protection. However, mangroves are under big threat and many is already been lost in the last decades [2]. Remote sensing has been proven to be a valuable tool in analysing and monitoring mangroves. Many different methods have been used and this research will focus in using the new radar Sentinel-1 and optical Sentinel-2 missions for a better knowledge in mangrove mapping [32].

This first chapter includes some general background information. The research area will be described, some general information about mangrove ecosystems and the Sentinel satellite missions. In the end of this chapter the research objective is given by means of the problem statement and project subscription, followed by the research questions and a report outline.

## 1.1. Vietnamese Mekong Delta and Ca Mau province

The Mekong river is one of the world's greatest rivers, with a length of 4800 km and a basin area of 795 000  $km^2$  [47]. The Mekong river origins on the Tibetan plateau in China and flows through 5 other countries; Myanmar, Laos, Thailand, Cambodia and finally Vietnam.

The Mekong Delta is the region in south-western Vietnam where the Mekong river flows through a network of distributaries finally ending in the South China sea. The Mekong Delta region covers an area of 39 000  $km^2$  and is home to more than 17 million inhabitants. The delta is the principal rice-producing region in Vietnam and is therefore also called the 'rice bowl' of Vietnam. But, also fruit and vegetables are produced as well as a lot of aquaculture is present. During the wet season, from July to December, a large part of the Delta is flooded due to both the high discharges of the Mekong river and local precipitation. On the contrary, during the dry season there's a water shortage and salinity intrusion forms a serious threat. The Mekong Delta is the region with the smallest percentage forest area in Vietnam. The only provinces with large forests are Ca Mau Province and Kien Giang Province, together accounting for two thirds of the region's forest area.

Ca Mau Province is the most southern province of the 63 provinces in Vietnam. Because of the big coast line the area is most vulnerable for coastal erosion. An extensive network of canals provides a popular means of transport but since 1-2 years a new road to Dat Mui increases the accessibility of the area. The Mui Ca Mau National park serves as a important tourism destination in the Ca Mau province together with the U Minh biosphere reserve.

Ca Mau Province has the largest total area of mangrove forest in the Mekong Delta. However, the area of mangrove forest has declined 50% over the past few decades, primarily due to increasing population pressure and the expansion of shrimp farming [39, 53]. In the Mekong Delta there are multiple ways of shrimp farming: intensive, semi extensive, shrimp-rice and shrimp-mangrove. Intensive systems rely completely on concentrated feed and fertilizers. Intensive ponds do not have any mangrove and are recognizable by the paddle wheels (figure 1.3a). Semi extensive shrimp farms consist of natural feed with additional feed and fertilizer. In 2008 70% of Ca Mau's total land was occupied by semi-extensive shrimp farms. According to provincial regulations farmers must have 60% of their total area to be mangroves and 40% aquaculture [7, 21].

Shrimp ponds in mixed farms are often incorporated into the mangroves as long thin channels within the mangrove themselves (figure 1.3b).



Figure 1.1: Map of the Mekong Delta in Vietnam [33]

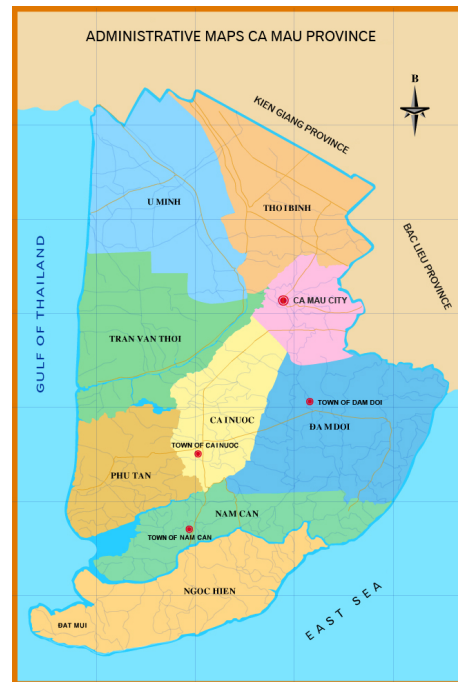


Figure 1.2: Map of the Ca Mau province [9]



(a) Intensive shrimp farming



(b) Extensive shrimp-mangrove farming

Figure 1.3: Different types of shrimp farming in Ca Mau province

## 1.2. Mangrove ecosystems

Mangroves refer to tree structures typically found in saline coastal environments in the tropics and subtropics, primarily between 25 degrees north and 25 degrees south. Mangrove forests are home to a large variety of fish, crab and shrimp species. The mangrove ecosystem has developed specialized adaptations to live in this tidal environment. The dense roots of mangrove forests help stabilize the coastline and prevents erosion from waves and storms. In Vietnam are 27 different mangrove species. Huge afforestation activities in Vietnam began in 1975, after the unification of the country. Many mangroves are lost due to the construction of shrimp farming. Since 2000 the government is trying to rehabilitate the mangroves [20]. Also climate change has a big effect on the mangrove area due to increasing sea-level, increasing CO<sub>2</sub> and higher air and water temperature. The average rate of mangrove area loss worldwide due to climate change is 1-2 % total area per year [3].

### 1.3. Sentinel missions

The European Commission in cooperation with the European Space Agency (ESA) direct world's largest single earth observation programme Copernicus. In the space component of the programme ESA develops a family of five different Sentinel missions. The goal of the Sentinel program is to replace the current older Earth observation missions. Many current missions are or will soon be nearing the end of the operational life time. The Sentinel program ensures continuation of ongoing studies with new and better equipment. Each mission will focus on a different aspect of Earth observation; Atmospheric, Oceanic, and Land monitoring. This study will focus on the Sentinel-1 and Sentinel-2 missions. The Sentinel-1A and 1B satellites carry a Synthetic Aperture Radar (SAR) in C-band that provides all weather, day and night imagery. The first imagery of Sentinel-1A became available on 6 October 2014 and his twin Sentinel-1B was launched 25 April 2016. Together they provide an image interval of every 6 days. The Sentinel-2A and 2B satellites carry a multispectral imager (MSI) with 13 spectral channels in the visible/near infrared(NIR) and short wave infrared (SWIR) spectral range with spatial resolutions of 10, 20 and 60 meters. The Sentinel-2A satellite delivers data since October 2015 and Sentinel-2B is launched in March 2017. Together they will reach an image interval of every 5 days.

### 1.4. Problem statement and project description

Mangrove ecosystems have been a widely studied subject around the globe for the last years. Remote sensing has been proven to be essential in monitoring and mapping this ecosystem. Many different remote sensing sensors have been used in research to monitor mangroves: from aerial photography to hyperspectral imagery and different radar data sources [32].

Medium-resolution optical data studies show promising results for multiple applications in different regions of interest. There has been studies on Vietnam specifically, most of them on change detection of the mangrove area. However, a investigation of the unique properties of different mangroves types in southern Vietnam hasn't been found. Also, the medium-resolution imagery that has been used (ASTER, IRS, SPOT, Landsat) have some drawbacks. Some of them have good resolution but are not freely available (e.g. SPOT), others are free but the resolution of 30 meter is not optimal (e.g. Landsat). The new Sentinel-2 mission is better on both sides: it is free available and has resolution of 10 and 20 meters, depending on the used band combinations.

Radar data has a big advantage to optical data by not being affected to clouds, which are abundant in tropical regions. Different effects and relationships of canopy, stand structures etc have been examined. The different wavelengths of radar, L-, C- or X-band have their own benefits but each of them can give information on mangrove structures. However, most of the former radar missions are already ended (e.g. Envisat, Radarsat) so an investigation for present day mangrove properties discrimination is not possible with those. Sentinel-1 is scheduled to stay active for at least the next 4 years. Still, using radar data is more challenging than optical data. Therefore, in this study the possibilities to fuse both data to increase information of mangrove discrimination is investigated.

### 1.5. Research questions

The above mentioned objectives are transferred to the following research questions:

- What is the best method for discriminating mangrove types in Vietnam using radar and optical satellite remote sensing?
  1. Why do we need to study mangroves in Vietnam?
  2. What are the unique properties to discriminate mangroves?
  3. Can we extract those unique properties from satellite imagery? How?
  4. What differences in mangrove types can be extracted from satellite imagery?
  5. How can space-borne optical and radar data be used for the classification of mangroves?
  6. How to validate the accuracy of the classification results?
  7. How can the quality of the classification result be improved by combining optical and radar data?
  8. Can this method be used to reach nationwide or even worldwide coverage? How?

## **1.6. Report outline**

This report consists of seven chapters. Chapter 2 gives an introduction of how remote sensing can be used for mangrove monitoring. This chapter will give answers to research questions 1 till 4. Chapter 3 will go into detail on how Sentinel-1 and Sentinel-2 are processed to serve as input for the classification methods, also the area of interest is explained. Chapter 4 introduces the different methods for the classification of mangroves and is subdivided into unsupervised and supervised classification. This chapter gives answers to research questions 5 till 7. Chapter 5 describes the results from those methods in detail. The results are discussed in chapter 6 and a final conclusion with an answer on the primary research question is given in chapter 7.

# 2

## Remote sensing of mangroves

This chapter describes the needs to monitor mangroves and the ability of remote sensing for performing this task. Section 2.1 goes in detail about different applications where mangroves have unique and essential properties such as coastal protection and the conservation of flora and fauna. Section 2.2 describes the different species and properties of the mangroves that are present in the Ca Mau province. From section 2.3 the focus lies on remote sensing systems that will be used in this research. Specific mangrove properties that can be extracted from those systems are explained in section 2.4.

### 2.1. Mangroves in Vietnam

In 1.2 a small introduction is given in mangrove ecosystems. However, at first we need to state the reasons why there is a need to monitor mangroves along the Vietnamese Mekong Delta coast. Many reasons are stated in literature and are summarized in this section.

#### 2.1.1. Coastal protection

The main reason why mangroves in southern Vietnam need to be studied is their potential for coastal protection. Mangroves can significantly contribute to coastal protection through wave attenuation, storm surge reduction, erosion protection and sediment collection and stabilisation. The role of mangrove ecosystems are investigated for different coastal phenomena: waves, storm surges and sea level rise. Waves entering the mangrove forest lose energy as they pass through the network of trunks, branches and aerial roots. Wave height can be reduced 13% to 66% over 100 meter of mangroves. The highest rate of wave height reduction per unit distance occurs near the mangrove edge and when they pass a greater density of obstacles. Mangroves with aerial roots attenuate waves the most. Measured rates of storm surge reduction through mangroves range from 5 to 50 centimetres water level reduction per kilometre of mangrove width. Although mangrove belts of one kilometre are rare, a smaller reduction of water level also can have a great impact on the extent of flooding [35, 36]. Mangrove also has a role of keeping suspended sediments in order to protection erosion. For this, there is a need of supply of sediments from the rivers. If there are insufficient nutrients from this sediment source the mangroves can die. Therefore monitoring mangroves is important for integrated delta management.

Despite all these benefits at many locations along the Mekong Delta coast (and other mangrove coasts over the entire world) mangrove forests are disappearing at alarming rates. Therefore, there is a big need of maintaining existing mangrove forests and planting new mangrove where possible. In many cases the mangrove forest first needs to be brought back to the system, before the coast can rely on its protection services [50]. For this reason mangrove rehabilitation programmes are set up by, among others, the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). Stefan Groenewold is the Technical Adviser of this Integrated Coastal Management Programme (ICMP) and explained me about their mangrove rehabilitation programme. The objective of this programme is to protect and strengthen the coast in the Mekong Delta area. One of the methods is improving the stability of the mangrove belt along the coast [51]. Many steps are involved in those rehabilitation programmes: decision making for mangrove conservation and planting, site assessment, planting techniques and finally monitoring [49]. Mangrove monitoring contains different phases from data collection, storage, analysis to reporting. With this information the forest health and development,

the survival rate of newly planted seedlings, erosion and forest cover are needed to be found. Many methods are used for this monitoring and remote sensing is a very useful tool.

### 2.1.2. Fauna and aquaculture

As mentioned in section 1.2 mangroves are ecosystems unique in the tropics and subtropics which contain many special species. They provide habitats for species which are adapted to a saline tidal environment, for example aquatic organisms such as algae, crabs, worms, decapods (shrimps, squids, lobsters etc) and many kinds of fish [40]. Studies on fish indicate that the mangrove ecosystem in Vietnam consist of 258 different species belonging to 70 families. The mangrove ecosystem with many nutrients is also an excellent environment for cultivating fish and crab species. This generates the income of many people in Mekong Delta.

However, there is an enormous decrease of mangroves for the construction of those shrimp and fish ponds [6, 39, 52, 53, 55]. There have been set rules by the Vietnamese government to protect the mangroves. Following the law on rules for land use and natural resource utilisation for reforestation of mangrove forests (1999) shrimp farming-forestry enterprises (SFFE) needs to have 60% of the farm area to be protected mangrove. The mangroves provide useful help to the cultivation of shrimps in integrated farming systems [21]. The farmers don't see the added value of the mangroves, while having a higher percentage of mangrove is proven to be beneficial [58]. To monitor the 60/40 rule the mangroves need to be discriminated from other vegetation to monitor the (changes in) mangrove area.

### 2.1.3. History

During the Vietnam war (1955-1975) many mangrove forests have been destroyed due to toxic chemicals. This chemical, also known as Agent Orange, eroded the tree cover and seedling stock, making reforestation more difficult. Huge afforestation activities in Vietnam began in 1975, after the unification of the country. Since 2000 the government is trying to rehabilitate the mangroves [20]. The Can Gio biosphere reserve, for example, was almost completely destroyed after the war and is now still being restored. An area of 20.000 ha of mangroves is already been replanted. We need to study the mangroves to see how the rehabilitation programmes are functioning and how different species are developing.

### 2.1.4. Climate change

The average rate of mangrove area loss worldwide due to climate change is 1-2 % total area per year [3]. Climate change has a big effect on the mangrove area due to increasing sea-level, increasing CO<sub>2</sub> and higher air and water temperature. As mentioned in section 2.1.1 mangroves adapt to the sea level rise by increasing the soil surface elevation. By measuring this increase we can get an idea of how much sea level rise occurs along the coastal areas. Due to climate change the tropical regions can be expanding which means mangrove is able to grow on new locations. Monitoring (new) mangrove can indicate those locations.

### 2.1.5. Carbon storage

A final reason for studying mangrove is its potential for carbon storage. Mangrove forests are the most important sink of carbon (C) in the tropics. Conservation of mangrove forest is needed to increase ecosystem C storage and to offset carbon emission at the regional scale [54].

## 2.2. Mangrove properties

In the next section the properties of mangrove are described. There are many different species with different unique properties that are introduced first.

There are many different numbers available on the amount of different species of mangrove in the Mekong Delta, Vietnam. It differs from 27 species in whole Vietnam to 69 different species in only southern Vietnam [20, 40]. The correct number isn't really of most importance rather than the properties of the different species. We can subdivide mangroves in different categories associated with their location relative to the shore. The most important species in the Indonesian and West Pacific region belong to the genera *Rhizophora* (red mangroves), *Avicennia* (black/grey mangroves), *Sonneratia*, and *Laguncularia* (white mangroves). The most common species in Ca Mau Province are shown in figure 2.1. *Rhizophora apiculata* is the dominant species and is also the mangrove that is cultivated in the extensive shrimp farms. In this research the main focus is on differentiating the *Rhizophora apiculata* and *Avicennia alba*. This is because *Avicennia alba* is a pioneer species which occurs mainly on the coastline and is important for coastal protection. *Rhizophora apiculata*

is the dominant mature species. Figure 2.1 shows that the structure of the two mangrove types are quite different which makes it easier to differentiate them. More details on these two species are discussed in the next sections.

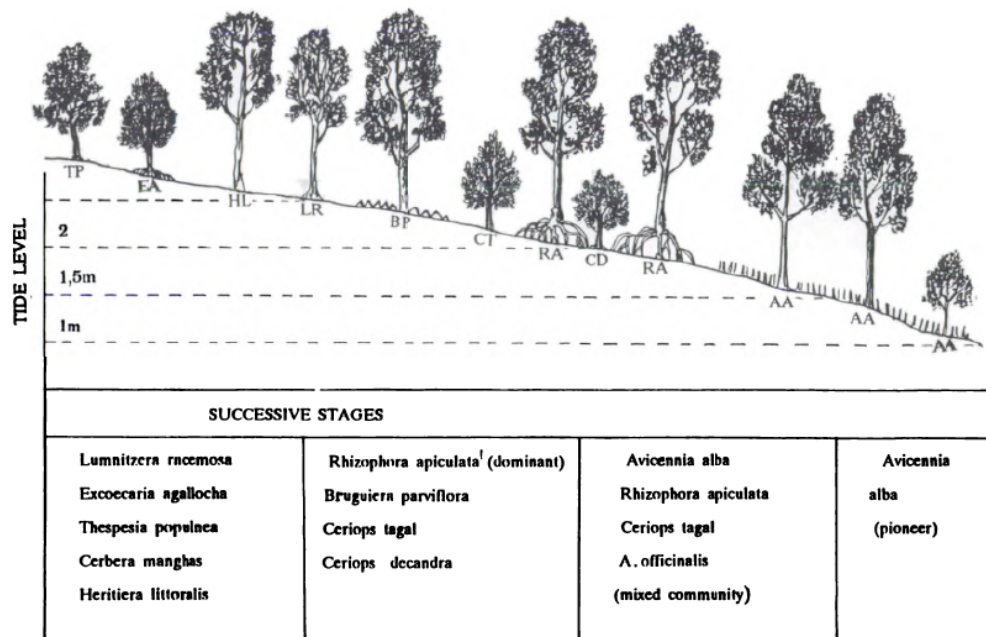


Figure 2.1: Succession of mangrove in Ca Mau cape, showing differences typical of each zone [40]

### 2.2.1. *Rhizophora* genus

The *Rhizophora* genus is called the 'true mangrove' due to its distinctive stilt roots. The most common *Rhizophora* in Vietnam is the *Rhizophora Apiculata*, Đước in Vietnamese. During low tide the stilt roots are often above the water and during high tide they are mostly covered with water. These aerial roots have a very strong function in wave attenuation, as described in section 2.1.1. *Rhizophora apiculata* can be distinguished by its colour and the shape of the plants generally. The leaves are large, dark-green and glossy. The developed stands of *Rhizophora apiculata* are mostly strong monotypic as observed typically in the middle estuarine reaches with muddy sediments [11]. Measures of *Rhizophora Apiculata* at different ages show that their growth rate is highest between the ages of 10-15. The mean height growth/year is the first ten years 0.99 meter per year. The maximum height in Ca Mau province is around 30 meter. Tidal flooding and duration are equally important factors which regulate growth and distribution of mangrove species. Experiments including *Rhizophora Apiculata* conducted in different coastal areas of Vietnam showed that it is unusual for those mangroves to grow in places lacking fresh water [40].

### 2.2.2. *Avicennia* genus

The *Avicennia* genus is known as the pioneer of the mangroves. They occupy a diversity of habitats within the tidal range and across salinity extremes of (sub)tropical sheltered areas. *Avicennia* genus is dominated by small trees or shrubs. The *Avicennia Alba*, Mắm trắng in Vietnamese, is the main pioneer in Vietnam together with the *Avicennia Officinalis*. The species is found along tidal river banks, on entrances of tidal inlets and extending along shoreline mudflats. It is the most common species in new formed mud banks, for example the accreting areas of Ca Mau Cape. Alluvium is accumulated here due to the shallow seabed and weak waves that die down before reaching the shore. There is no other place in Vietnam where mangroves so quickly and strongly occupy accreted land [40]. *Avicennia Alba* can be distinguished by its lighter green, pale, greyish color of the leaves.

### 2.2.3. Others

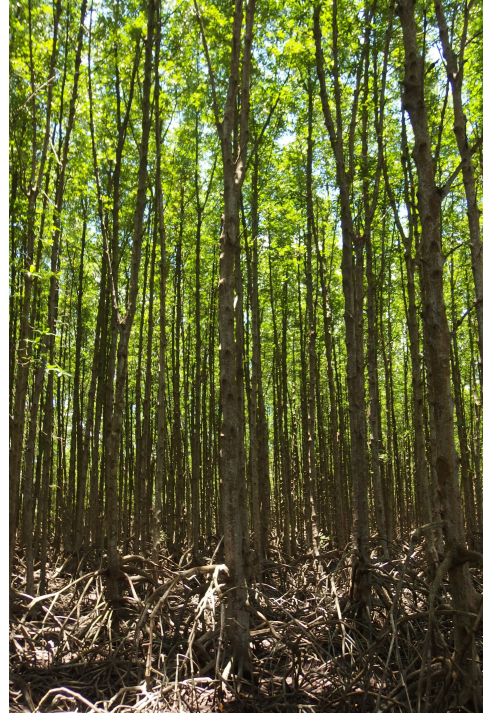
Other common species occurring in the Mekong Delta are *Sonneratia Alba*, *Bruguiera Parviflora*, *Nypa fruticans* and *Ceriops Tagal*. Those species are often present in mixed areas. Therefore it's hard to distinguish

them as separate species [40].

The *Bruguiera* genus is known as Orange mangroves and can be recognized from its distinctive 'knee roots'. The *Sonneratia* genus is known as the Apple mangroves and is thus recognizable from the growing fruits that look like apples. *Sonneratia* species have cone shaped roots which can grow in a radius of more than 10 meters around the trunk. *Nypa fruticans* is known as the Mangrove Palm and is easily recognizable from other mangroves. It is also known as the underwater coconut with its very big fruits on the stalk.



(a) *Avicennia Alba*



(b) *Rhizophora Apiculata*



(c) *Nypa Fruticans*



(d) *Sonneratia Alba*

Figure 2.2: Different mangrove species in Mekong Delta



## 2.3. Remote sensing data

The data used in this research is optical and radar satellite remote sensing data. Properties of this data will be explained in detail in sections 2.3.1 and 2.3.2. The next section, 2.4, goes in detail on how remote sensing can be used to extract mangrove properties. The last section, 2.5, discusses the issues that still remain in the field of mangrove remote sensing.

### 2.3.1. Optical remote sensing

Optical remote sensing is so-called passive remote sensing and makes images of the earth's surface detecting the solar radiation reflected by targets on the ground. The solar radiation is measured as electromagnetic waves in different wavelengths in the visible light, near infrared and short-wave infrared (fig 2.5). The intensity of the reflectance is dependent on the reflectivity of the material in the different wavelength domains. This **reflectivity** is dependent on many different factors of which colour, structure and surface texture are the most important. The differences in reflectance are well identified in a spectral signature plot with the reflectance as a function of wavelength. Different earth surfaces leads to unique spectral signature plots, as can be seen in figure 2.3.

Green vegetation is easily recognizable by low reflectance in the visible domain and high reflectance in the Near Infrared. The steep transition around  $0.7 \mu\text{m}$  is called the red edge. **Chlorophyll** strongly absorbs light at wavelengths around  $0.45 \mu\text{m}$  (blue) and  $0.67 \mu\text{m}$  (red) and reflects strongly in green light ( $0.51 \mu\text{m}$ ), therefore our eyes perceive healthy vegetation as green. The high reflectance between  $0.7$  and  $1.3 \mu\text{m}$  results primarily from the **internal structure of plant leaves** (fig 2.4). As this internal structure varies amongst different plant species, this wavelength range allows the distinction between plant species. Absorption minima found at longer wavelengths are caused by the **water content of the leaves**.

The spectral signature of bare soil is less variable. The brown color gives a higher reflectance in the visible domain and the absence of water increases the reflectance in the shortwave infrared. Other factors affecting the reflectance are **soil texture, surface roughness** and the presence of **minerals and organic matter**.

Clear water is recognisable due to the high absorption of wavelengths longer than  $0.8 \mu\text{m}$ . Therefore water is easily detected with optical remote sensing. Clear water has a lower reflectance in the visible domain than water with sediments or high chlorophyll concentrations due to algae for example.

### 2.3.2. Radar remote sensing

RADAR is an acronym that stands for Radio Detection and Ranging. Radar remote sensing are active remote sensing systems operating at microwave frequency. Active remote sensing means that the antenna actively transmit a signal/pulse, electromagnetic waves in the microwave domain, towards a target; in this case the earth's surface. The receiving antenna (often the same antenna for transmitting and receiving) measures the strength of the backscattered signal and the time delay between the transmitted and reflected signal to determine the distance to the target. The strength of the backscattered signal is dependent on many system parameters and on properties of the target.

One of those parameters is the **wavelength** of the radar. The most commonly used wavelengths in radar remote sensing are L-band  $\lambda = 235\text{mm}$  (1.3 GHz), C-band  $\lambda = 56\text{mm}$  (5.4GHz) and X-band  $\lambda = 30 \text{mm}$  (10GHz). Microwaves are able to penetrate clouds, canopies and even the upper layer of the soil. How far the waves penetrate depends on the exact frequencies. The longer the wavelength the further it can penetrate the canopy. X-band radar only detects the top of the canopy. The L-band radar is able to penetrate an entire canopy and can detect soil properties or water surface. This radar is often used to measure the soil moisture content. C-band does not penetrate the entire canopy (only during the emerging stages of the crops) and is during the mature stage affected by volume scattering in the canopy.

Another system parameter is the **polarization** of the wave signal. A wave is emitted in either a horizontal or vertical polarization. With a dual-receiver radar the received wave can be measured into a (relative) horizontal or vertical component, providing information on the target. In single polarization mode HH, horizontal emitted, horizontal received, and VV, vertical emitted, vertical received, are measured. For a dual-polarized mode a combination of either HH and HV or VV and VH is available. This polarisation scheme is dependent on the mode of the satellite and will be discussed later. Also the **incidence angle** influences the backscatter intensity. A bigger incidence angle causes a lower backscatter intensity. To make meaningful comparisons between images with a different look angle, the difference in angles should be as small as possible. In this research this objective is met with incidence angles varying between 35 and 40 degrees at the area of interest.

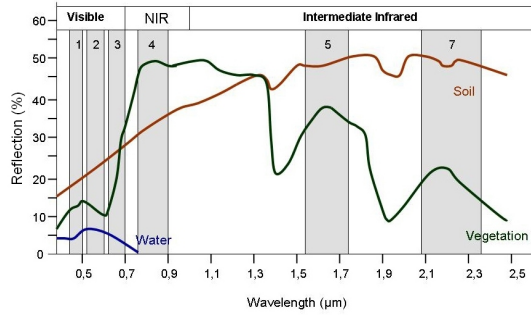


Figure 2.3: Spectral signature plot. Examples of typical spectral signatures of green vegetation, water and soil are shown

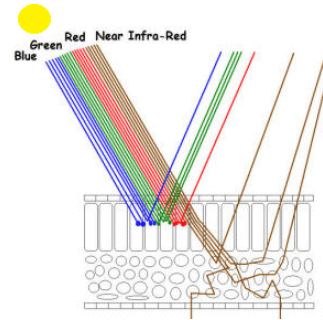


Figure 2.4: Cellular leaf structure and its interaction with electromagnetic energy. The top layer with most chlorophyll absorbs most visible light except for green light. The lower layers, consisting of irregularly shaped cells, highly reflect the near infrared wavelengths

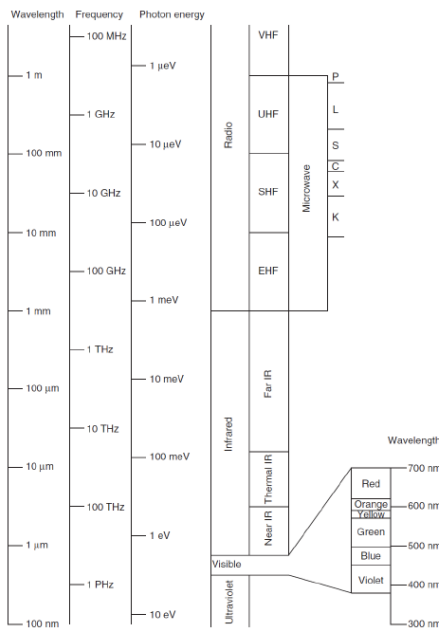


Figure 2.5: The electromagnetic spectrum showing the parts that are important in remote sensing. The Shortwave Infrared (SWIR) belongs in between the Near IR and Thermal IR ranging from 1 to 3 µm [45]

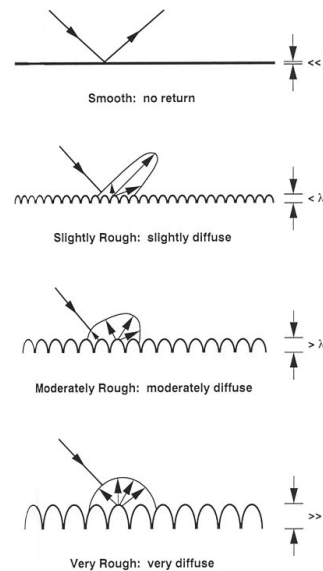


Figure 2.6: Influence of roughness on radar backscatter

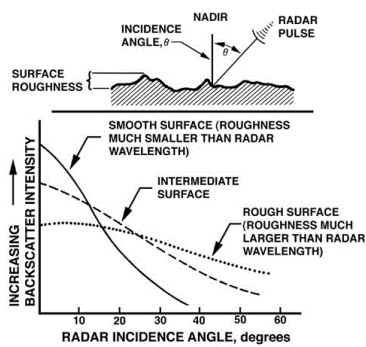
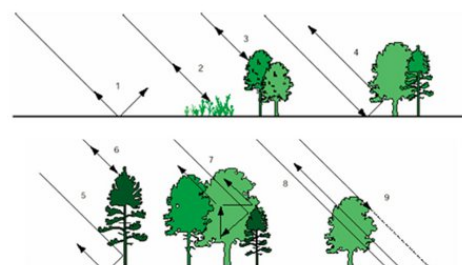


Figure 2.7: Influence of incidence angle on radar backscatter



1 Diffuse scattering from the ground; 2 and 3 Direct scattering from various vegetation components; 4 Double-bounce vegetation-ground interaction; 5 Corner reflector between tree trunks and ground; 6 Direct backscatter from the forest canopy; 7 Volume scattering from within the forest canopy; 8 Diffuse scattering from the ground; and 9 Shadowing by parts of the forest canopy of other parts of the canopy or the ground.

Figure 2.8: Surface and volume scattering of a radar beam for trees [19]

Important target properties that influence the radar's backscattered signal are geometry, roughness of the surface and dielectric constant. According to the **geometry** of the target, targets can be horizontal, vertical and random scatterers. Horizontal and vertical oriented scatterers have a dominantly horizontal or vertical structure which strongly reflects horizontal or vertical waves respectively (HH or VV). Cross-polarisations (VH and HV) are sensitive to both orientations and therefore are indicators of random scatterers. In case of trees many sorts of scattering are possible, for example direct canopy scattering, double-bounce scattering and volume. Eight different sources of scattering related to vegetation are shown in figure 2.8. The **roughness** of a surface influences the direction of the reflection. Smooth surfaces, with variations much smaller than the wavelength size, cause a reflection away from the incident wave and give no backscatter at all. Rough surfaces, variations much bigger than the wavelength size, reflect diffusively and therefore cause a big amount of backscatter. The **dielectric constant** is a ratio between electric permittivity of the material of the target and the permittivity of free space. The dielectric constant is dependent on water and a higher moisture content leads to a stronger reflection of electromagnetic waves because they penetrate less deep in the target. For example, also water in leaves makes the amount of reflection bigger and therefore causes a higher backscatter value.

## 2.4. Remote sensing of mangrove properties

Now the basics of optical and radar remote sensing are explained (sections 2.3.1 and 2.3.2) and the biological properties of mangroves are investigated (section 2.2), the next section investigates how those properties of mangroves can be extracted. Many research has been done on mangroves using remote sensing and a detailed review about this has been done by Kuenzer et al. [32] in 2011. However, a detailed description of the properties is not easily to be found. Therefore an overview is made for both spectral properties in optical data and properties in radar data. Also discriminating between different mangrove species using those properties is investigated.

### 2.4.1. Spectral properties of mangrove

Various kinds of green vegetation, including mangroves, produce a high peak in the NIR band and low in the red and green region of spectrum, as explained in section 2.3.1. A good indicator for vegetation is therefore the Normalized Difference Vegetation Index (NDVI) which is given by  $NDVI = \frac{NIR-RED}{NIR+RED}$ . In this index vegetation pixels will get a value close to 1 since the values of RED are low and the values for NIR are high. Other pixels such as bare land, urban area and water will get a value closer to 0.

Many studies use NDVI values for monitoring and mapping mangrove changes. Also false color composites are used to distinguish between very high NIR and lower NIR response. Smooth textures indicate mangrove from other land cover if they are densely distributed. Dense mangrove forests give higher NIR response than sparsely distributed and mixed mangrove areas [31, 42, 44]. Also visual interpretation is still a highly used method for mangrove mapping. However, confusion between mangroves and other vegetation is the most commonly reported source of classification error. Though, these traditional methods already give classification accuracies of mangrove classes ranging from 75% to 95% for producer's and user's accuracies [27]. More specific information on spectral properties of mangrove are based on the differences within the different species.

### 2.4.2. Spectral discrimination of mangrove species

Already in 1986 remote sensing studies on mangroves using photographic sensors were executed. Leaf optical signatures and canopy structures were used to classify mangrove communities and identify changes. Ramsey (1996) found that black leaf reflectance is higher than red leaf reflectance. Canopy height is correlated to canopy NIR reflectance, which is mainly dependent on the LAI. Also he found the complexity in describing the spectral and structural difference of mangroves [43]. Since then a lot of research has been done finding the origins of those differences in mangrove reflectances. Spectrometer data of those two mangrove species in Ca Mau, Vietnam (2010) as a stack of mangrove leaves showed that species differ because of the principal biophysical and chemical properties, such as water, cellulose, leaf pigments as chlorophyll and many more [32].

Figure 2.9 shows that discrimination in the 380-750 (visible) wavelength domain is weaker than in the near-infrared region. The near-infrared signal can better facilitate mangrove discrimination. However, other studies also have used the region of 400-800 nm to discriminate between different mangrove species. Those shorter wavelengths in the visible region are influenced by the pigments and chlorophyll content of the leaves

and therefore influence the color and thus reflectance in this wavelength range. Since the leaves of *Avicennia* are brighter green than *Rhizophora* the reflectance is slightly higher around 550 nm (green light). Longer wavelengths (800-1050nm) contain noise due to cloud cover, light source fluctuations etc. Differences between *Rhizophora Apiculata* and *Avicennia Alba* reflectance were biggest among 5 different mangrove species [29]. Some studies mention the near-infrared plateau at 750-800 nm together with the red edge, 690-750 nm, as important spectral regions for discrimination of vegetation species. With in-situ spectral measurements mean spectra of six studied mangroves in Ca Mau Province, Vietnam are obtained. In those spectra the near-infrared plateau and the red edge show the biggest differences [25]. Features that are also used to distinguish among mangrove communities with medium-resolution satellite data are textural and spectral characteristics of the canopy. The spectral characteristics within a species are defined by age, vitality, phenological and physiological parameters.

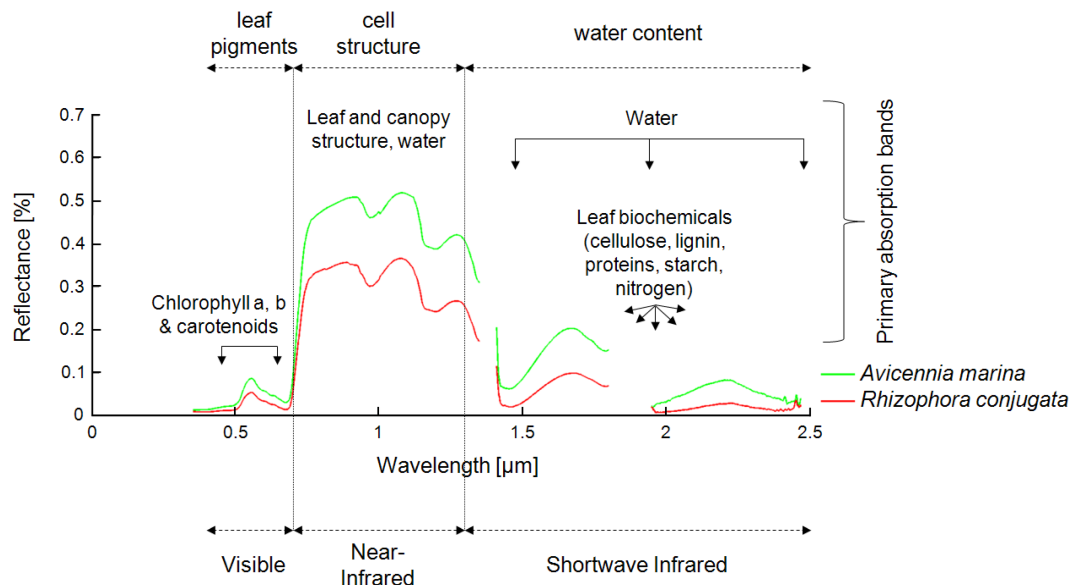


Figure 2.9: Spectral characteristics and their influencing parameters of mangrove species *Avicennia marina* and *Rhizophora Conjugata* (synonym of *Rhizophora Apiculata*) as measured with a field spectrometer in Ca Mau province, Vietnam [32]

### 2.4.3. Mangrove properties in radar data

Radar imagery from SAR systems are less intuitive to the common human perception than optical imagery. It is more challenging to make an interpretation than a natural color composite, for example. The intensity of the microwave signal that is measured back at the radar satellite is called the radar backscatter coefficient. This intensity is depending on a lot of different factors. Each system has a wavelength of the transmitted signal, the polarization of transmitted and received signals and the incidence angle of the signal. Finally, the specific received backscatter signal is dependent on the interactions with the received surface. For vegetation the factors that influence the backscatter are internal properties; moisture content, cell structure etc. and external properties; size, geometry, trunks, branches roots and leaves etc [32]. The different backscatter mechanisms due to the different wavelength of the radar system can be seen in figure 2.10. Shorter wavelengths are influenced mainly by the canopy because the signal does not penetrate through the leaves and trunks etc. When the wavelength increases there is more penetration and the backscatter can even reach the water or soil below the canopy. The presence of water results in a significant increase in ground-trunk/two-bounce scattering and thus in an increase in radar backscatter. This ground-trunk scattering is typically detected by the longer wavelength radar, L- and P-band. At shorter wavelengths, C- and X-band, this will only occur if few leaves are present since the forest canopy dominates the signal [30].

Measurements from simultaneous P-, L- and C-band (AIRSAR) show that at **C-band** volume scattering is dominated by the branches for all different polarizations HH, VV and HV. The penetration depth of C-band is in the order of a few meters within the crown. The radar wave interacts with the upper part of the canopy, mainly with leaves and twigs. This makes C-band sensitive to crown characteristics like the density, size, permittivity, and orientation of leaves, but also to canopy structure, crown architecture (leaves and branches)

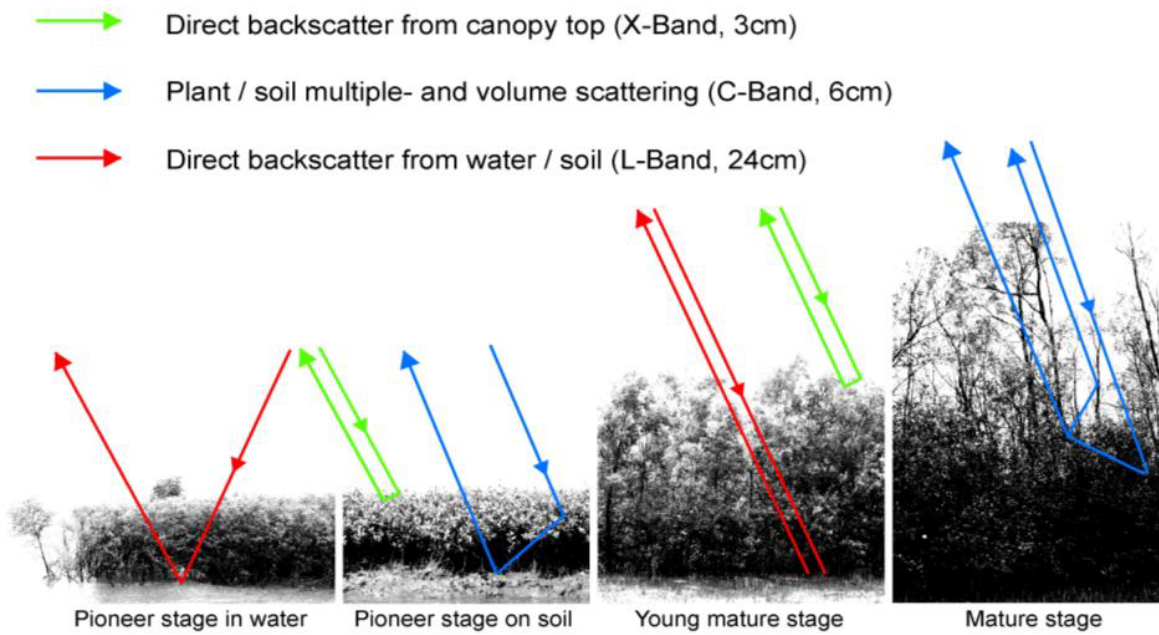


Figure 2.10: Dominating backscatter mechanisms at different stages of mangrove growth depending on bandwidth of radar beam [32].

and canopy heterogeneity. The Leaf Area Index (LAI) plays a dominant role in the observed backscatter since LAI is a measure of density of forest canopy. Dense canopies, high LAI, are expected to have more backscatter than sparser mangrove areas which contain fewer or no leaves ( $LAI \approx 0$ ). As leaf biomass increases, the backscatter coefficient usually saturates rapidly for the C-band SAR. This saturation in C-band is mainly the case for HV-polarization, which is the result of multiple scattering in the crown of the mangrove tree. HH- and VV-polarizations behave differently and continue to increase to a higher biomass saturation threshold. This increase in backscatter coefficient may be caused by the variations in the canopy structure due to the different growth stages (Fig 2.11). More mature stages have in general a higher leaf biomass. Overall, at C-band, the best correlations with canopy parameters are obtained with VV-polarization [37]. Another application of C-band SAR is DEM generation by using interferometry. Assuming the ground is flat the DEM provides information of the height of the vegetation. This can be valuable input for classification algorithms or in decision tree [26].

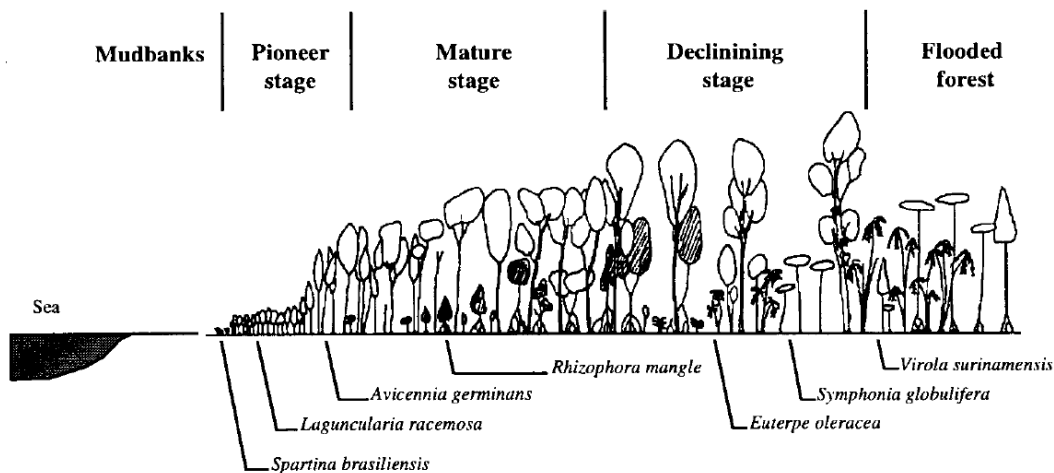


Figure 2.11: Schematic representation of the successive growth stages of the mangrove forests [37].

**L-band** penetrates deeper and can pass through the total canopy in pioneer stages. Scattering is influenced by the trunk and/or ground surface. Multiple scattering can occur between trunk and ground surface. More mature stages cause volume scattering and for increasing biomass VV and HH show similar values and stay constant. Cross-polarization HV shows increasing values with increasing biomass. High correlations at HV-polarization are found with basal area and tree height [37].

Finally, **P-band** has the longest wavelength ( $\sim 74$  cm) and penetrates most to the underlying surface. At VV polarization the increasing backscatter is clearly visible for water surfaces. A high sensitivity to biomass is found with the HV-polarization [37].

During this research The Vietnam Southern Satellite Technology Application Center (STAC), part of the Vietnam National Satellite Center (VNSC), was visited. This institute has projects about coastal zone environment management, impacts of climate change on environment in the Mekong Delta, rice & mangrove monitoring in Southern Vietnam and about applying active microwave remote sensing for estimating mangrove forest biomass. Although they used Envisat ASAR C-band radar for rice and mangrove classification they suggested the PALSAR L-band radar is better for discriminating mangrove properties. With classifications of 6 classes (>70%, 50-70%, 30-50% canopy cover, mixed shrimp-mangrove, urban and water) and 4 classes (>60%, 30-60% canopy cover, urban and water) the global accuracies for L-band were around 60% for classification with 6 classes and around 80% with 4 classes. The global accuracies for C-band were around 40% with 6 classes and around 55% for 4 classes.

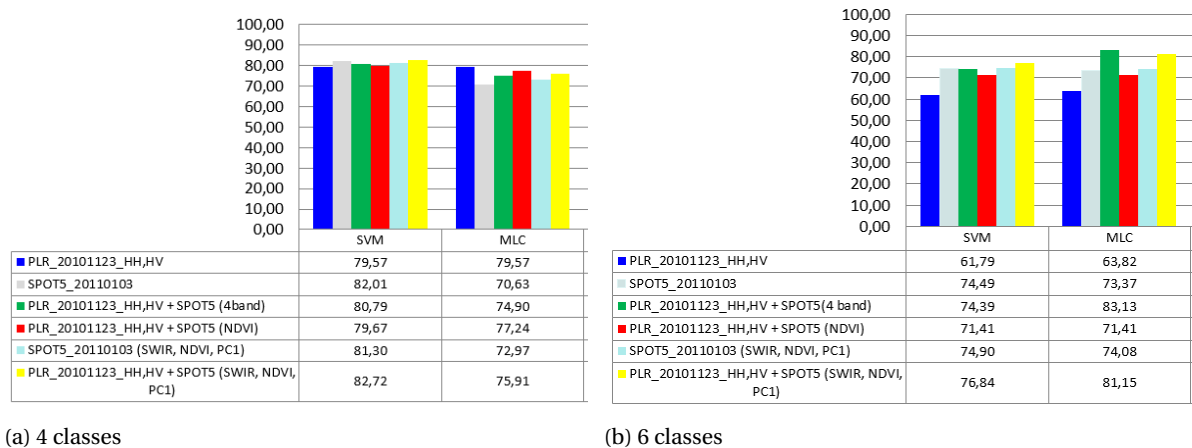


Figure 2.12: Overall accuracy's for mangrove classification using PALSAR L-band and SPOT5 using two different classification schemes: Support Vector Machine (SVM) and Maximum Likelihood (MLC)

#### 2.4.4. Discrimination of mangrove species using radar

Discriminating mangrove from non-mangrove has proven to be possible with most different radar systems. However, discriminating different mangrove species using radar is hard. In this research, the main discrimination in mangrove species is between *Rhizophora Apiculata* and *Avicennia Alba*. Since those two species have a different canopy cover and C-band SAR is dominated by volume scattering of the canopy it can be expected that the backscatter coefficient of those two species differ. *Rhizophora* genus consists of dense closed canopy while *Avicennia* genus occurs in more open areas with shrubs and less dense canopy. They are found in different growth stages as can be seen in figure 2.11. However, the best results are obtained when using a combination of optical and radar input [26]. Extra knowledge, based on local surveys, additional optical data, maps of zonation patterns and species communities are essential for good discrimination [32].

## 2.5. Issues with mangrove remote sensing

As can be noticed in the prior sections there has been lots of research on mangrove ecosystems using remote sensing. During this project there has also been some investigation with other researchers in the field on the issues still remaining with mangrove remote sensing.

### 2.5.1. Data

As can be noticed in section 1.4 and in the previous sections the used data is of big impact on the research outcomes. Not only the technical specifications as well as the availability of unpaid data. Many research has been done using expensive high (1 to 4m) resolution optical data like SPOT, Quickbird, IKONOS etc. For example by Dr. Tuan Quoc Vo, one of the authors of Kuenzer et al. [32], Vo et al. [57, 58] who has done a PhD research in the field of economic evaluation of mangrove ecosystem services using combined approaches of remote sensing data and socioeconomic data. The same holds for the paid data of the Vietnamese VNRedSat-1 images of 10m resolution. On the other hand many research has been done using the free Landsat optical data that has 30m resolution data with a 16 day revisit time. The same holds for radar imagery which is often expensive to obtain. Especially long wavelength SAR which are the only waves penetrating through the mangroves, like PALSAR with L-band is only available as a paid service. But also RADARSAT-1 and 2 with C-band and Cosmo-SkyMed with X-band are expensive data sources. Some free available data, like ASAR on Envisat, is not operational any more so a new free alternative is highly recommended to use.

As already mentioned in the introduction sections the Sentinel-1 and Sentinel-2 missions are both good improvements for monitoring mangrove ecosystems. They are both free to use data with good spatial (10-20 meter) and temporal resolution (10-12 days). All details about the Sentinel missions will be covered in the next chapter.

### 2.5.2. Data fusion

Another issue that is still largely missing in the field of mangrove remote sensing is the use of data fusion to improve mangrove monitoring. Some research has already been done, for example by the Vietnam National Satellite Center (VNSC). They fused PALSAR L-band together with SPOT5 data, the results can be found in figure 2.12. Optical data showed better results but the combination with radar can improve the overall accuracy, especially using Maximum Likelihood Classification (MLC). However, both of these data sources are still paid services. Combining satellite data for better forest monitoring has high potential [26], also with free available data [46]. The potential of Sentinel-1's dense SAR time-series is noted in combination with the global available optical Landsat-8 and/or Sentinel-2 data. Many earth observation research groups either optical or SAR remote sensing instead of both. Multi-sensor development are helped by the use of large storage and processing capabilities like the Google Earth Engine. This is already used by Hansen et al but also here only Landsat optical data is used [24]. In this research the data fusion between C-band SAR (Sentinel-1) and optical (Sentinel-2) is investigated.

### 2.5.3. Temporal analysis

The last issue is the use of temporal analysis for mangrove properties extraction. Mangrove properties are now mainly characterized by certain patterns in wavelength reflectance for example the high peak in the Near Infrared but the behaviour of this pattern over time has not been discussed. The fruit and flowering of the mangrove, the propagule maturation but also the seasonality of the weather can influence those patterns. The same holds for radar backscatter which is known to be dependent on water, roughness and volume scattering that can differ through time. A temporal analysis will be done to investigate mangrove characteristics in time series applications.

### 2.5.4. Cloud-based processing and classification

As mentioned in previous section 2.5.2 Google Earth Engine is a powerful tool for storage of big data sets but also has big processing capabilities. This makes it possible to execute classifications algorithm on a big scale or on multiple images at the same time. Google Earth Engine is equipped with different classifiers for different classification algorithms most of them in the machine learning category, for example support vector machine (SVM), classification and regression tree (CART) and random forest (RF). The latter two are found to generate good classification results with spectral imagery from Landsat-8 in the Google Earth Engine [48].

## 2.6. Conclusions

This chapter investigates the functions of mangroves and the main types of mangroves in Ca Mau province in Vietnam. The basics of remote sensing are described and how this data can be used to extract mangrove properties. Conclusions from this chapter bring some answers to the research questions described in section 1.5.

- The main reason for studying mangroves in the Mekong Delta is the need for coastal protection in this highly vulnerable area where mangroves contribute by reducing wave height and storm surges and holding sediments to protect erosion. Also conservation of flora and fauna is an important driver in monitoring mangroves. Finally, a big part of Ca Mau province's area is covered with extensive shrimp farms which consists of water ponds where fish and shrimps are cultivated surrounded by mangroves. Monitoring these mangroves is important to check governmental regulations and changes towards intensive shrimp farming where no mangrove are present.
- Mangroves are found in saline coastal environments and are recognized by its unique roots, which are stilt, pencil or knee roots depending on the species. The two main species in the Ca Mau Province are *Avicennia* and *Rhizophora* genus. Close to the shoreline pioneer species with pencil roots are common, such as *Avicennia* with its light green and greyish color of the leaves. More inland grows the *Rhizophora* genus that has darker green leaves, can grow much taller and has distinctive stilt roots.
- In optical satellite imagery vegetation is easily discriminated from other land covers by its high reflectance in the near infrared (NIR) and its low reflectance in the red and blue wavelengths. Dense mangrove forests give higher NIR reflectance than sparse forests. Within different mangrove species *Rhizophora* and *Avicennia* genus are easiest to discriminate by their spectral signatures.
- In radar imagery the signal wavelength determines the penetration depth in the canopy. Shorter wavelengths are dominated by volume scattering while longer wavelength are dominated by the trunk or even the ground surface. At C-band the best correlations with canopy parameters are obtained with VV-polarization. Denser mangrove forests cause more backscatter than sparse forests. Mangrove species are not yet discriminated using radar backscatter and therefore a temporal analysis will be investigated to improve in extracting mangrove properties from radar data.
- This research focuses on investigating mangrove properties using Sentinel-1 and Sentinel-2 imagery since this data is free to use and has both a good spatial and temporal resolution. Methods for combining those data are investigated and temporal analysis to improve in mangrove classification. Google Earth Engine is used for cloud-based processing with online available remote sensing data.



# 3

## Data and site description

This study uses the data of the Sentinel-1 radar mission and the Sentinel-2 optical mission. They are chosen because they are free to use, still have long future operation time and both the temporal and spatial resolution are good for the function of mangrove monitoring. Both missions will be explained in full detail in sections 3.2 and 3.3. In section 3.1 a site description is given on the Region of Interest (ROI). It gives a detailed explanation of the area inside the Ca Mau province in Vietnam. A field campaign is executed to obtain ground truth data in this ROI. Information is gathered from which training classes are described. A first exploratory data analysis is done using the data from both Sentinel-1 as Sentinel-2 mission, giving information about those ground truth places. In the last section 3.5 a description is given on the software that is used during this research. Concluding, this chapter gives all information about the data and how it is used in this research.

### 3.1. Region of interest

Figure 3.1 shows the Region Of Interest (ROI). It ranges from the southern lobe of Phu Tan District in the north to the city of Nam Can in the north-east, the city of Rach Goc in the south-east and to the end of the cape of Mui Ca Mau National park. The Google Earth image shows that the ROI is an area consisting of many green areas and few urban development except for some small villages. The polygons inside the red bounding box show the locations on which ground truth data is obtained. The area is chosen because Ca Mau province is the province with the highest percentage of mangrove left. Inside the red bounding box both natural mangroves are present in the Mui Ca Mau National Park and a small strip along the southern coastline and also cultivated mangrove in extensive shrimp farms.

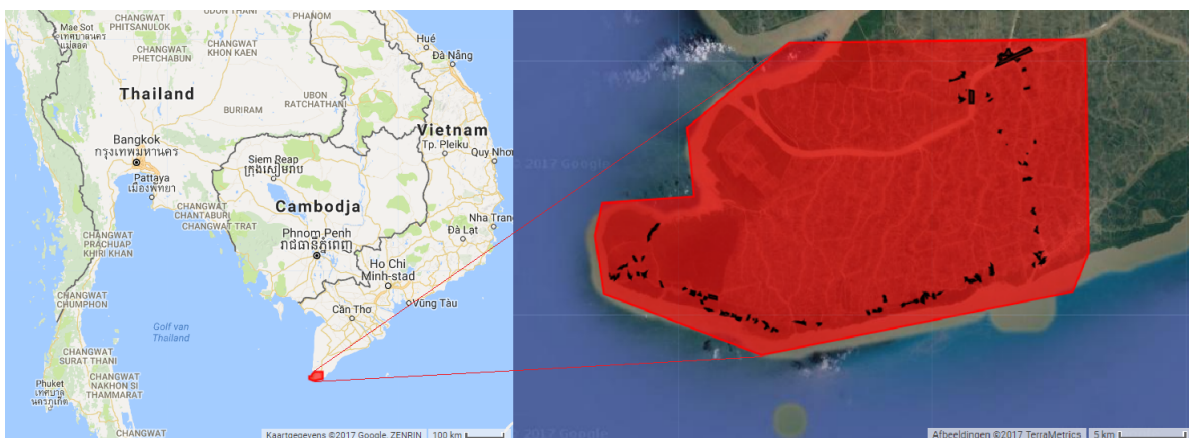


Figure 3.1: Region of interest bounding box with locations of fieldwork campaign on Google Earth imagery.

### 3.1.1. Fieldwork campaign

To get insight on the different land cover, different mangrove types and surroundings of the region of interest a fieldwork campaign is executed. During a two-day fieldwork many locations are observed within the Region of Interest. During those days representative sample sites of known land cover type are studied. Those sample sites are called training sites or areas and are used for the supervised classification algorithm to serve as ground truth training data. It is noticeable that most locations follow the pattern of the main road, which was just completed. However, it is tried to spatially divide the observation locations but since special admission is needed to access the inland areas, especially the Mui Ca Mau National Park, this was not always possible. Mangroves are poorly accessible areas so sometimes a boat is the only means of transportation. For the Mui Ca Mau National Park a boat trip is used to get insight in the mangrove type. This was the only place where access into the National Park was allowed for non-Vietnamese researchers. During the fieldwork GPS locations are taken from 140 different locations. On these locations the surroundings and landcover are observed, pictures taken, the distribution of water and vegetation estimated, species of mangrove observed and the height measured or estimated. Sometimes also the stem diameter is measured or a fish eye view of the forest canopy is made.

### 3.1.2. Training classes

During the fieldwork many insight is gained in the land cover inside the ROI. With the GPS locations and corresponding photos training polygons are created using Google Earth to be able to cover the mangrove canopy pixels instead of the GPS locations that are often taken from the side. Those training polygons are divided into 5 different classes:

1. Mangrove vegetation: *Avicennia Alba* species
2. Mangrove vegetation: *Rhizophora Apiculata* species in natural environment
3. Mangrove vegetation: *Rhizophora Apiculata* species in extensive shrimp farms
4. Water (rivers, sea and shrimp ponds)
5. Urban area and barren land

Since this research focusses on monitoring mangrove types the majority of classes are mangrove vegetation. The other two classes are useful since water covers a big part of the area in the sea and rivers and between the mangrove vegetation and in intensive shrimp ponds. Urban area is very distinctive and different than water and vegetation and also includes parts with barren land and roads. Within the mangrove vegetation it is chosen to make classes for the two main species that are found within the region of interest. Those *Avicennia Alba* and *Rhizophora Apiculata* are easy to separate as found in the literature [32]. The *Rhizophora Apiculata* species is separated into two classes. The mangrove that is located in the National Park is very old and tall and very different from the *Rhizophora* mangrove that is cultivated in the extensive shrimp farms. Inside this cultivated mangrove class the difference in size is big because the cultivated mangrove do not get older than 18 years, after that they are cut down and replanted.

The differences in the training classes in the field are shown in figure 3.2. The *Avicennia* has clearly lighter leaves and is more a shrub than a tree. The natural *Rhizophora* in the park differs from the cultivated *Rhizophora* by the size and height and the surroundings. The cultivated *Rhizophora* is planted in straight lines with always a lot of water around. Water and urban areas are easy distinguishable.

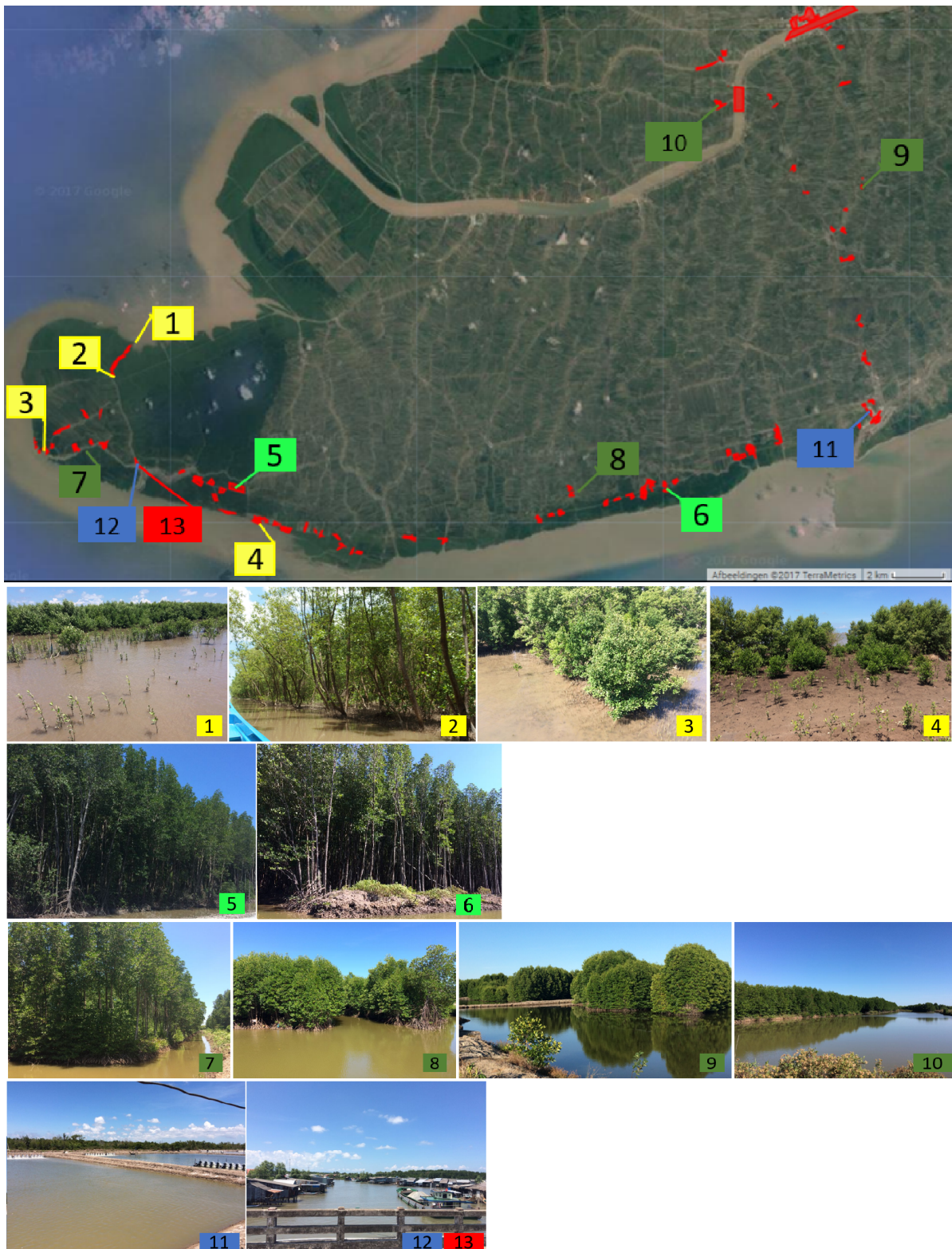


Figure 3.2: Overview of training data classes in the ROI. The yellow numbers represent class 1: Mangrove vegetation: *Avicennia Alba* species. Bright green is class 2: Mangrove vegetation: *Rhizophora Apiculata* species in natural environment. Dark green is class 3: Mangrove vegetation: *Rhizophora Apiculata* in extensive shrimp farms. Blue is class 4: Water areas and red is class 5: Urban area and barren land

### 3.2. Sentinel-1 mission

The Sentinel-1 mission consists of a constellation of two satellites, Sentinel-1A and Sentinel-1B. Sentinel-1A was launched on 3 April 2014 and Sentinel-1B on 23 April 2016. The first data was available in October 2014 and 2016 respectively. Both satellites carry a C-band Synthetic Aperture Radar (SAR) that provides all-weather, day and night imagery. C-band SAR are antennas of a wavelength between 3.75 to 7.5 cm (4.0 to 8.0 gigahertz (GHz))(figure 2.5). As the constellation orbits 180 ° apart, the mission images Europe every six days and every twelve days the rest of the world (figure 3.3).

Sentinel-1's radar can operate in four modes: Interferometric Wide Swath (IW), Extra Wide Swath (EW), Wave (WV) and Stripmap (SM) as shown in figure 3.4. These modes can operate in several polarization schemes. By varying the polarization of the transmitted signal, SAR systems provides information on the polarimetric properties of the observed surface.

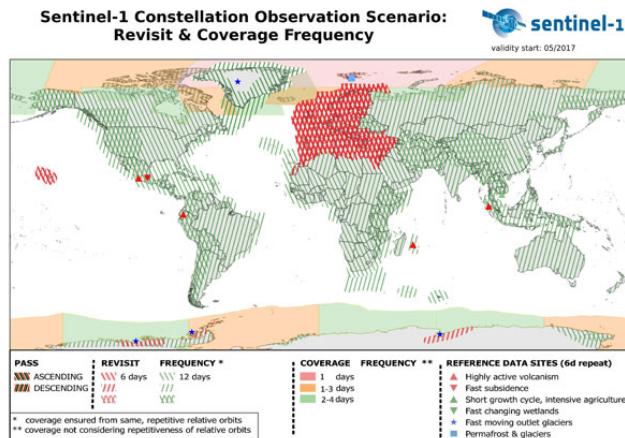


Figure 3.3: Observation scenario of Sentinel-1 [16].

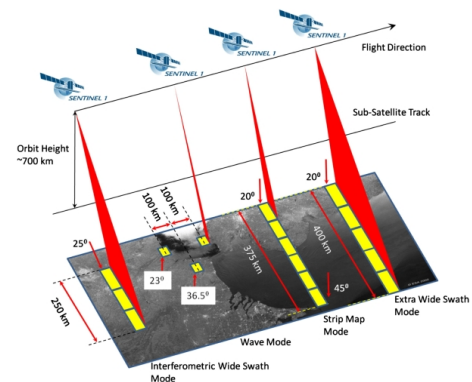


Figure 3.4: Different modes of Sentinel-1 [16].

#### 3.2.1. Radar data processing

The primary mode over land is Interferometric Wide Swath (IW) with VV+VH polarization which will be used in this study [16]. In IW mode the radar acquires data with a 250 km swath at 5m by 20m spatial resolution for a single look. The data is processed from Level-1 Single Look Complex (SLC) to Level-1 Ground Range Detected (GRD) scenes. For each scene 5x1 looks are needed and are averaged over a grid with a pixel size of ten by ten meters. Those GRD scenes are available in the repository of the Google Earth Engine. This collection of images is updated weekly. Each available scene is pre-processed with the Sentinel-1 Toolbox using the following steps:

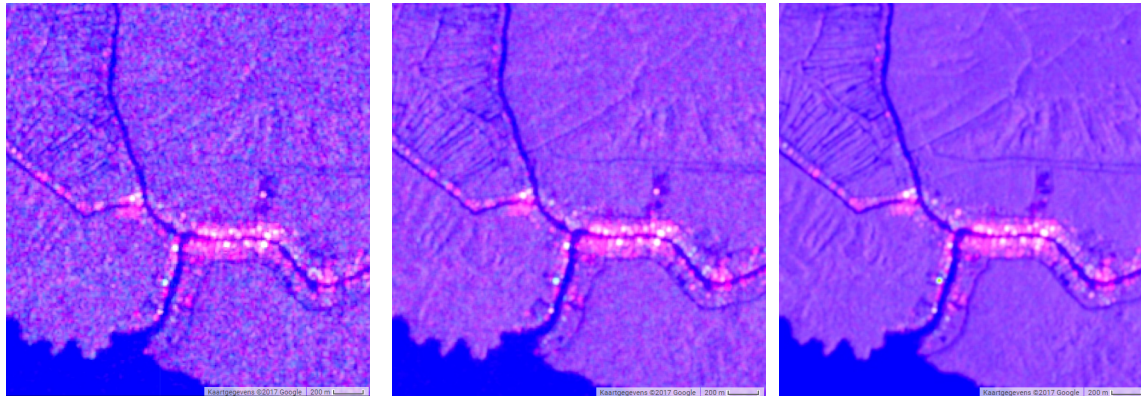
1. Apply orbit file (using restituted orbits)
2. Thermal noise removal
3. Radiometric calibration
4. Terrain correction (orthorectification)

During thermal noise removal dark strips near scene edges with invalid data are removed. This operation cannot be applied to some earlier images and causes error strips along the scene edges for data from 2015 until the beginning of 2016. To make time series comparable, radiometric calibration computes backscatter intensity using sensor calibration parameters in the GRD metadata. Orthorectification is applied to account for the terrain differences in the image. The height data of the SRTM 30 meter DEM is used for this.

The pixel values represent the backscatter coefficient which is the target backscattering area (radar cross-section) per unit ground area. The backscatter coefficient can vary by several orders of magnitude and is therefore converted to decibels (dB) as  $10 * \log_{10}(\sigma^0)$ . Values are clipped to the 1st and 99th percentile to preserve the dynamic range against anomalous outliers, and quantized to 16 bits.

All of the above mentioned methodology is pre-processed in the Google Earth Engine. With the available GRD scenes the users can enhance the data further with other filters and averaging methods. For Sentinel-1 data there has been discussion to apply a speckle filter. In the end, no speckle filter is used since it reduces the spatial resolution of the images. To reduce other outliers and noise a temporal averaging is done for every full month of images. This is done by selecting the available imagery from the first to the last day of

a specific month and calculating a pixel-based mean. From October 2014 until October 2016 there is a new image available every 12 days in Vietnam. From October 2016 onwards there is a new image every 6 days. This means that the data is averaged over  $\pm 3$  images until October 16 and  $\pm 5$  or 6 images after this time. The result of this temporal averaging is shown in figure 3.5. Noise is reduced when averaging over multiple images. This effect is largest when averaging over a complete year but is already useful for a monthly average. The final pre-processing step for Sentinel-1 imagery is clipping the monthly average images to the Region of Interest. The data is now ready to execute the methods described in the next chapter.



(a) Single image 18 December 2016 (b) Monthly average December 2016 (c) Yearly average 2016

Figure 3.5: Sentinel-1 color composite of bands of polarizations VV, VH and  $\frac{VV}{VH}$  with the visible effect of temporal averaging

### 3.3. Sentinel-2 mission

Sentinel-2 is an earth observation mission consisting of two satellites carrying a multi-spectral imager (MSI) with 13 spectral channels in the visible/near infrared (NIR) and short wave infrared spectral range (SWIR) with different resolutions, shown in figure 3.6. Those channels are strictly chosen in the atmospheric windows with enough transmission to observe earth's surface. The bands with 60 meter resolution will not be used in this research since they are less detailed and the wavelengths are of lower interest. Sentinel-2A was launched 23 June 2015 and Sentinel-2B on 7 March 2017. The first data data of Sentinel-2A was available in January 2016 and the data of Sentinel-2B in September 2017. Together they will have a revisiting time of 5 days for Europe, Africa and Greenland and some other locations. The rest of the world has a revisiting time of 10 days when both satellites are operating together (figure 3.7). It can be seen that Vietnam is also a special location with a higher revisit time. However, the most southern point is just missing out on this, probably because of an early switch off over sea, and therefore the Ca Mau province is having only 20 days revisit time within the time span of this research since data from the Sentinel-2B satellite was not available yet.

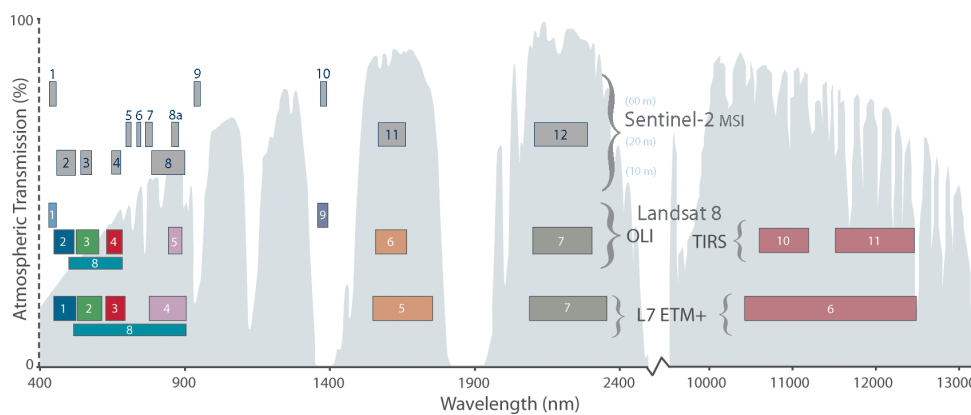


Figure 3.6: Different bands of Sentinel-2 [12].

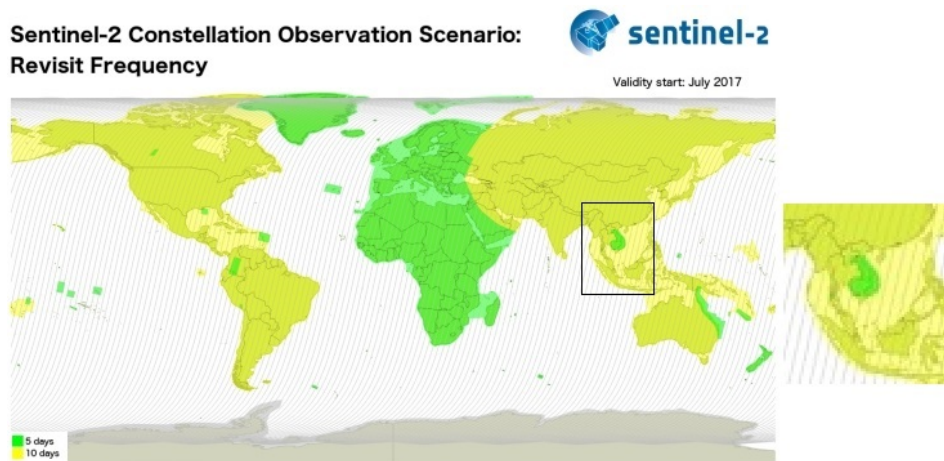


Figure 3.7: Observation scenario of Sentinel-2 [17].

### 3.3.1. Optical data processing

The multi-spectral imager on board of the Sentinel-2 satellites undertakes systematic acquisitions in a single observation mode, from which different levels of data are available. Level-1 consists of the Top of Atmosphere radiances (TOA), Level-1B being in sensor geometry and Level-1C being in fixed cartography geometry (UTM projection and WGS84 ellipsoid). The Level-2A product is Bottom of Atmosphere (BOA) reflectance in cartographic geometry. Sentinel-2 Level-1C data is downloaded from the Copernicus repository because the atmospheric correction which is needed to convert to Level-2A Bottom of Atmosphere (BOA) reflectance is not (yet) available in the Google Earth Engine. This correction is executed using the Sen2Cor processor that will be explained in more detail in section 3.5.2. After this correction a subset is made of the bands that were recorded with 10 or 20 meter resolution, see figure 3.6. Those bands are upsampled to the same pixel size of 10 meter. Extra bands of the cloud confidence and the scene classification are included in the file that is uploaded to Google Earth Engine for the next steps. An example of this scene classification is shown in figure 3.8. In Google Earth Engine the pixels with a cloud confidence value of more than 5% are masked. Also pixels that are classified as shadow, high or medium probability cloud in the scene classification are masked. Sometimes masking causes loss of data of interest but the amount of error caused by cloudy pixels is worse than the loss of some data pixels. Just like Sentinel-1 data the final pre-processing step is to clip the data to the Region of Interest. The data is now ready to execute the methods described in the next chapter.

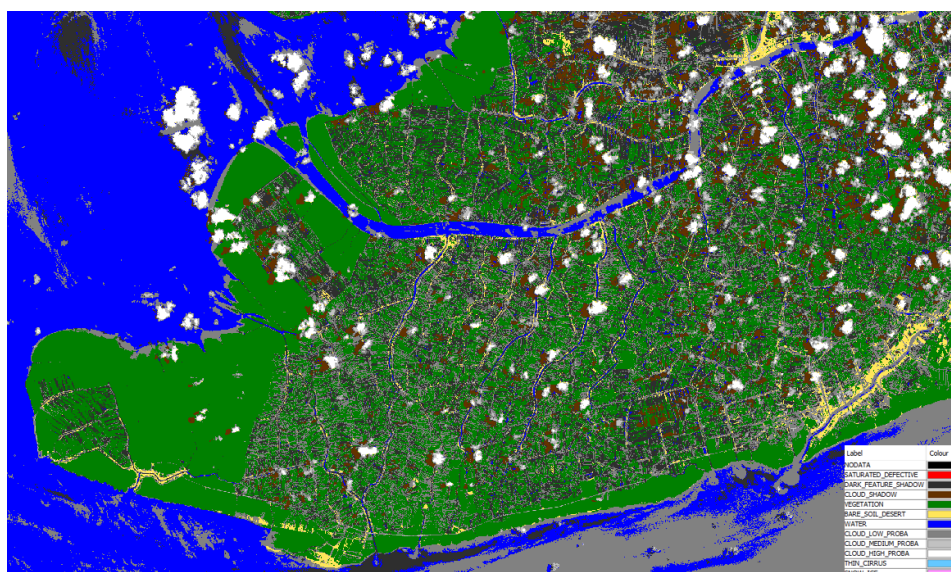


Figure 3.8: Automatic scene classification made during atmospheric correction for the Sentinel-2 image of 26-04-2016

### 3.4. Exploratory data analysis

To get a first insight in the information that is available from the above mentioned sources of data, an exploratory data analysis is done. In figure 3.9 Sentinel-2 images are shown, figure 3.11 shows images from Sentinel-1.

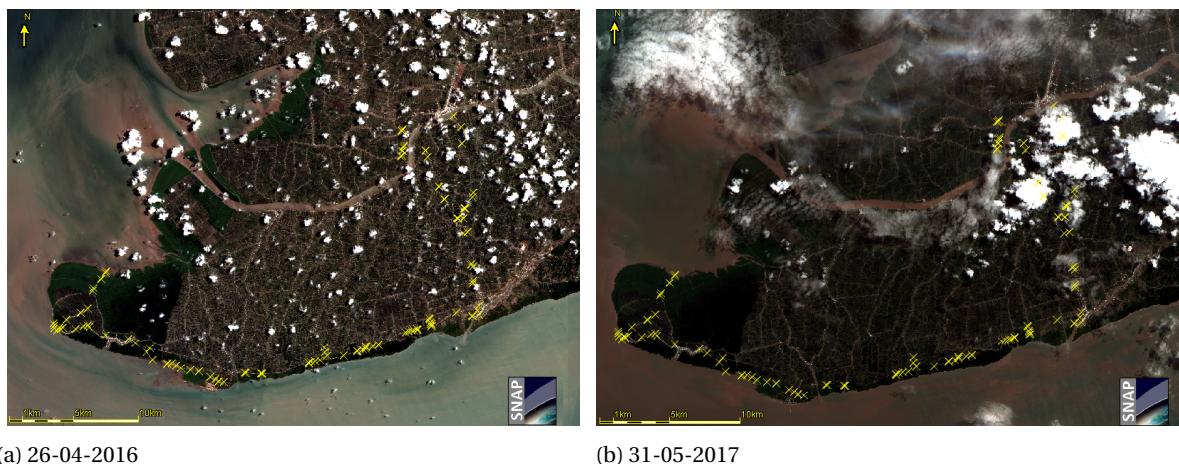


Figure 3.9: Sentinel-2 images visualized in natural color composite. The yellow markers indicate the GPS locations of the fieldwork data.

#### 3.4.1. Places of interest

There are different features visible in the ROI which are of interest. The first is the Mui Ca Mau National Park on the south-western tip which is an area totally covered by mangrove forest and is a highly restricted area. In the area there are also some urban areas from the villages of Dat Mui, Rach Goc and Nam Can. Along the southern coast also mangrove appears but this area is less restricted than the National Park. Finally, the most abundant feature represents the area where mangrove and fish/shrimp farms are mixed. This is the main part of the area but also the hardest since there's a lot of mixed pixels on water and vegetation. The different places are marked on figure 3.10 and will be discussed in more detail using natural and false color composites. When zooming into those areas with false color images this can improve on the visibility of (healthy) vegetation. Using the vegetation false color composite, consisting of bands of near-infrared, red and green, vegetation appears bright red and water areas appear grey.



Figure 3.10: Different POI in the natural color Sentinel-2 image of 26-04-2016. The green box contains the area inside the Mui Ca Mau National Park. The red boxes contains the urban communes of Dat Mui, Rach Goc and Nam Can. In the yellow box are the coastal areas with mangroves. The blue box contains area which consists of mixed mangrove and fish/shrimp farms.

Also radar data gives some first information on the places of interest. Figure 3.11 is a false color composite with bands VV, VH and  $\frac{VV}{VH}$  from the radar backscatter. It can be seen that inside the National Park there's a more homogeneous signal than in the surroundings. The mixed area of mangrove/shrimp farms consists of more yellow together with purple which indicates more mixed pixels with water.

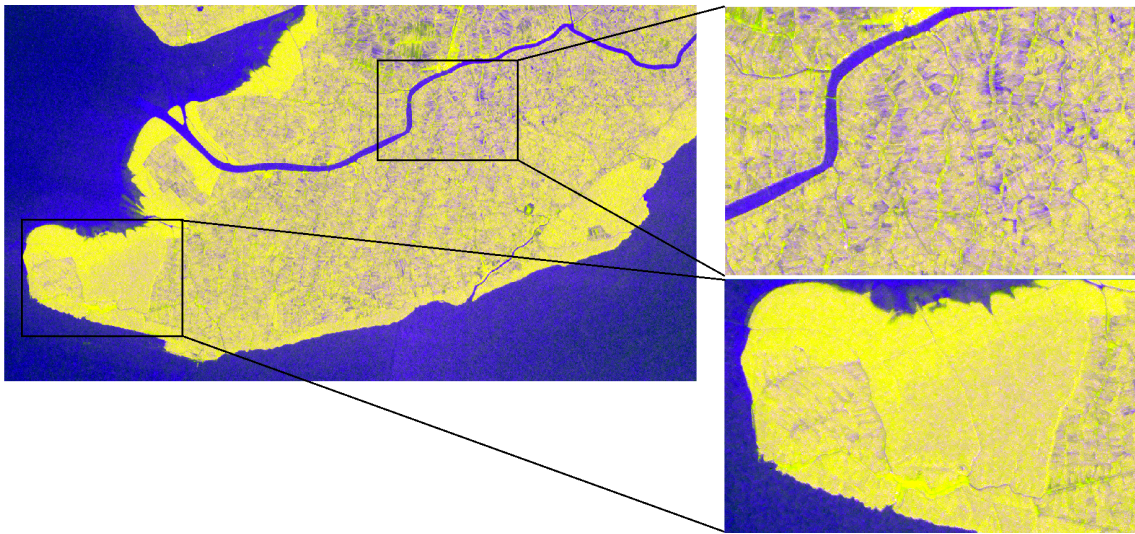


Figure 3.11: Sentinel-1 color composite of bands of polarizations VV, VH and  $\frac{VV}{VH}$  of temporal averaged April 2016. Zoomed into an area with extensive shrimp farms (up) and the Mui Ca Mau area (down).

#### 3.4.1.1. Mui Ca Mau National Park

The Mui Ca Mau National Park, or National Park of Cape Ca Mau, is the most southern point of Vietnam. The National Park is listed on Unesco's list of biosphere reserve and consists of three different zones. The core area, a strictly protected ecosystem that contributes to the conservation of landscapes, ecosystems, species and genetic variation. The buffer zone surrounds the core areas, is slightly less strict and is used for scientific research, monitoring etc. In the transition area the greatest activity is allowed, fostering economic and human development that is socio-culturally and ecologically sustainable.

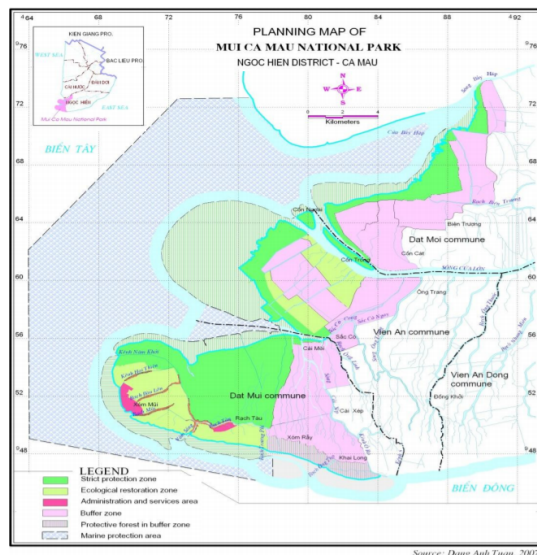


Figure 3.12: Different zones in the Mui Ca Mau National Park

An optical natural and false color composite shows the distinction between vegetation and non-vegetation, water and houses more clearly in figure 3.13. The western coast shows the dyke that is been built from North to South. Behind this dyke is some vegetation in different tones of green. Along the canal are small houses



with mixed mangrove and shrimp/fish farms in clear straight lines. The second picture zoomed into Mui Ca Mau National park shows the village of Dat Mui with the boundary of the core zone at the northern side and the restoration zone on the southern side. Here the brightest red and most green reflection is seen.



(a) Natural color composite at western coast



(b) False color composite at western coast



(c) Natural color composite near Dat Mui



(d) False color composite near Dat Mui

Figure 3.13: Sentinel-2 images of 26th april 2016 zoomed into Mui Ca Mau National park

Cloud-free pixels are selected in the National Park and a time series is made of the available Sentinel-2 imagery, which runs from January 2016 to August 2017. In figure 3.14 reflectance values of the different spectral bands are visible as well as the NDVI that is calculated using band 8 (NIR) and band 3 (Red). It can be seen that Bands 6,7,8 and 8A, all in the near infrared, have much higher reflectance than the other bands, around 0.35. NDVI values inside the National Park are mostly around 0.9 except for August 2016 and since June 2017. Figure 3.15 shows that the rainy season starts in April which is indicated in the time series. Within one season a small curved shape can be seen for the bands in near infrared (B6,7,8 and 8A). Reflectance is higher when there is more rainfall. Since the near-infrared are influenced by the leaf and canopy structure (figure 2.9) it might be related that during wet season the canopy is more closed, more leaves are present causing the higher reflectance.

A time series of available Sentinel-1 imagery has a longer time interval from January 2015. This time series shows a sinusoidal trend for all the places of interest consisting of mangrove. The trough of the wave is from December until April which happens to be the dry season in Southern Vietnam (figure 3.15). The crest of the wave is similar to the wet season. During a dry period there is less backscatter than average while during a wet period there is more. The National Park has the highest radar backscatter coefficient compared to the other places with mangrove vegetation, indicating highest amount of biomass in this area (section 2.4.3).

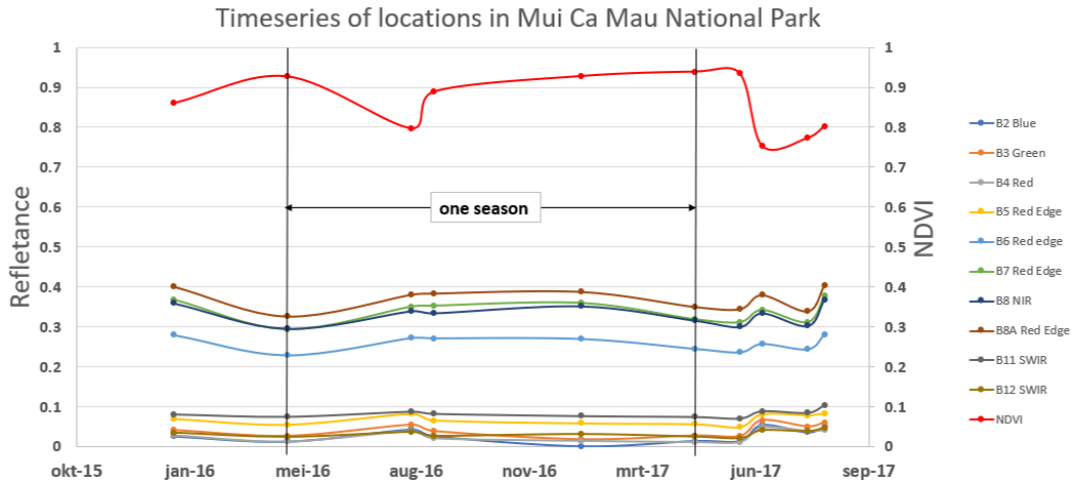


Figure 3.14: Timeseries of Sentinel-2 imagery in Mui Ca Mau National Park

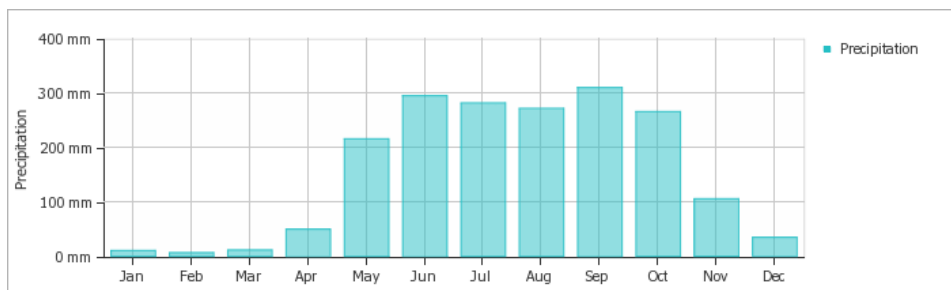


Figure 3.15: Average precipitation (rain/snow) in Ho Chi Minh City, Vietnam [28]

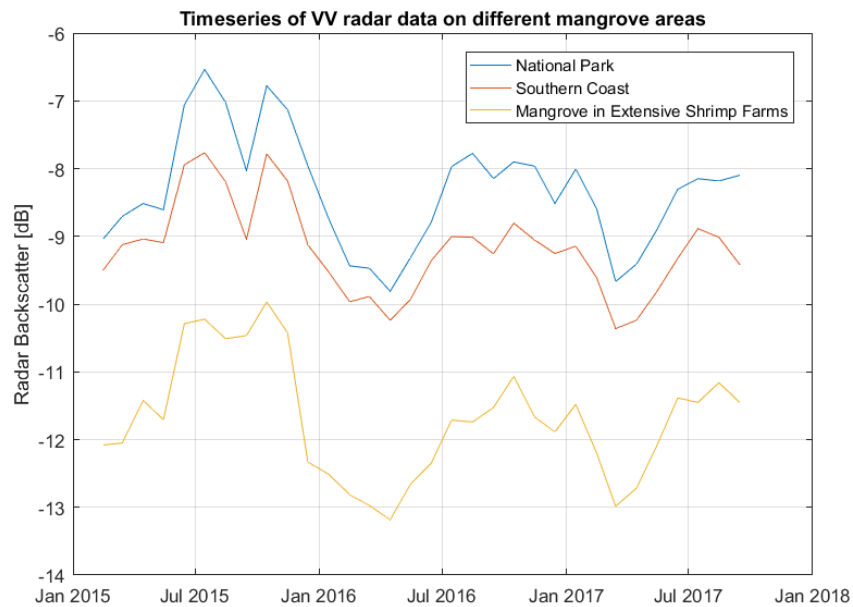


Figure 3.16: Time series of VV radar data on different mangrove places of interest

**3.4.1.2. Southern mangrove coast**

Along the southern coast mangroves grow without being cultivated in the extensive shrimp farms. The area is marked by the new road which runs from Rach Goc village (east of the ROI) to Dat Mui village (west of the ROI). On the northern side the area consists of many shrimp and fish farms with partly mangroves and on the southern side mangrove vegetation dominates until the sea starts. This is exactly the area where mangrove has its function to protect the coastal areas and the farms located behind, as described in section 2.1.1. Conform the succession of mangroves from the sea shore to more land inward locations (figure 2.1) the species of mangroves is slightly different here compared to the national park. The *Avicennia Alba*, a pioneer species, is more common here especially directly at the coastline. Further inland also *Rhizophora Apiculata* grows but is expected younger than inside the National Park. An investigation of the natural and false color Sentinel-2 imagery here shows similar vegetation/non-vegetation patterns. The natural color along the southern coastline looks lighter green than the very dark green around Dat Mui (figure 3.13c).

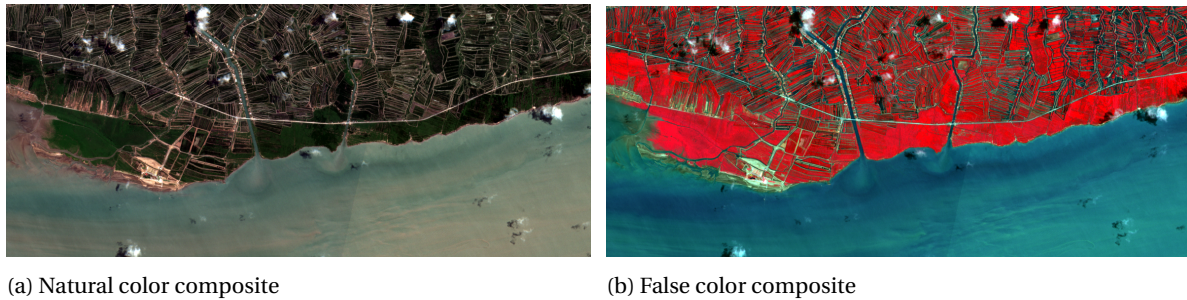


Figure 3.17: Sentinel-2 images of 26th April 2016 zoomed in on the southern coast

A time series of the spectral bands and the NDVI shows that the reflectance in the NIR bands is quite similar but the red band (as well as the green and blue bands) is higher. This causes the NDVI values along the southern coast being lower on average than in the National Park. This differences in the visible wavelengths can be caused by the different species of *Avicennia Alba* that has lighter color leaves than the *Rhizophora Apiculata*. Another option is that the older mangroves inside the National Park result in a darker forest compared to the same species mangroves that are younger along this southern coastline. Again, in the time series the values of August 2016 and especially June 2017 are not in line with the other dates. The high reflectance in the wet season is again seen in the bands of the infrared (B6,7,8,8A) but the differences are smaller than inside the National Park. It might be that younger or more pioneer mangroves show smaller increase in leaf of canopy closure in the wet season.

The time series of the Sentinel-1 radar imagery (fig 3.16) shows exactly the same seasonal pattern as the National park, with high backscatter in the wet season and low in the dry season, but than with lower backscatter in general.

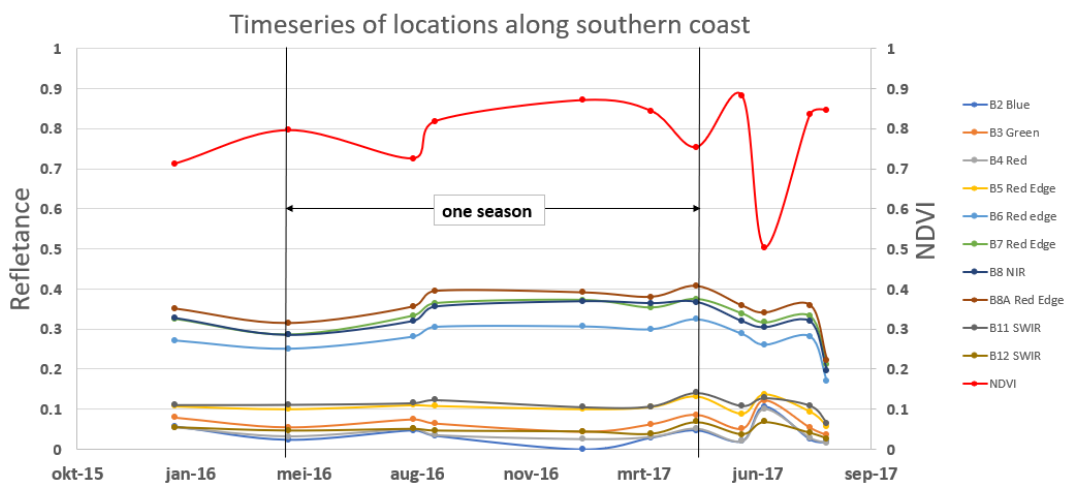
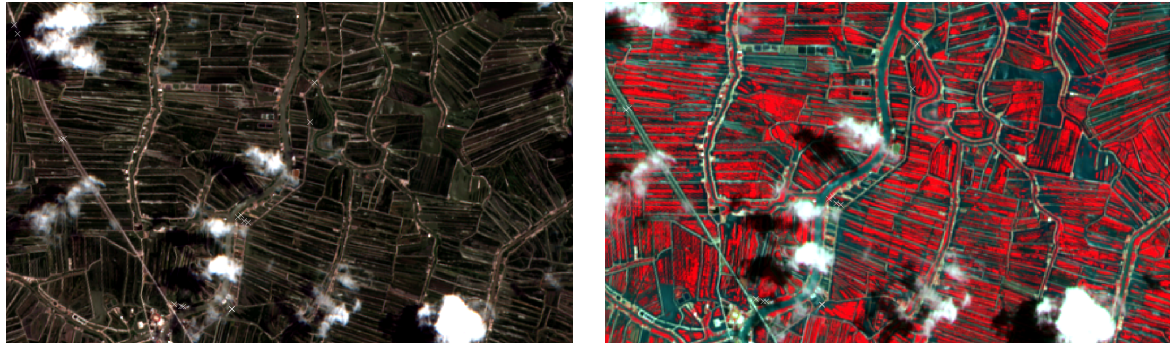


Figure 3.18: Time series of Sentinel-2 imagery along southern coast

### 3.4.1.3. Mangrove plantations and shrimp farms

The figures in 3.19 are located in the blue box of figure 3.10. This place of interest consists of farmers that grow mangrove trees in water ponds where fish and shrimps are cultivated. As mentioned in section 2.1.2 those shrimp farming-forestry enterprises (SFEs) need to have at least 60% of the area within their farm to consist of mangrove. To monitor this 60/40 rule the mangroves need to be recognized. A problem with those locations is the pattern of the mangroves in combination with the surroundings of all the water. Often the reflectance within a pixel of 10 by 10 meter is not only influenced by the mangrove but also by the surrounding waters.



(a) Natural color composite

(b) False color composite

Figure 3.19: Sentinel-2 images of 26th April 2016 zoomed in on location of mangrove plantations

In the time series of locations inside extensive shrimp farms (figure 3.20) the mixing of mangrove vegetation with water causes the bands in the NIR range to have much lower reflectance. However, the curved shape of the near-infrared bands (bands 6,7,8 and 8A) is again visible within one season that begins when the rainy season starts, from April to April. The NDVI values are much lower than the other mangrove vegetation areas, around 0.55 compared to 0.9 for the National Park and 0.8 for the Southern Coast. The same effect holds for the time series of the VV radar backscatter (figure 3.16). Also here the backscatter coefficient is lower than for the other two places of interest but the seasonal pattern is still visible.

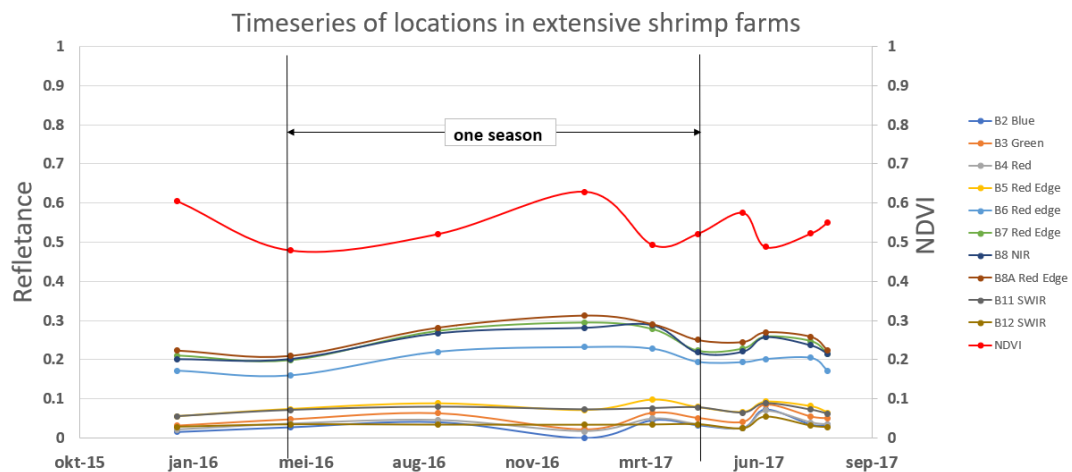


Figure 3.20: Time series of Sentinel-2 imagery in extensive shrimp farms

### 3.4.2. Spectral signatures

In the previous section some interesting places of interest with mangrove vegetation are discussed. The time series already showed differences in NDVI values and in specific wavelengths. A spectral signature is the variation of reflectance or emittance with respect to wavelength for a specific material. In figure 2.9 spectral characteristics of a former study were shown where parameters of two different mangrove species are measured with a field spectrometer. With the bands that are available on Sentinel-2 (figure 3.6) such spectral signature plot is created. A spectral signature plot is often a single record of one image on a certain time

stamp but can also be made from a mean or median of a certain time period. In this case a spectral signature plot is made for the mean of the reflectances over the whole available time series: figure 3.21. Together with the three places of interest with mangrove vegetation also urban area and water areas are included.

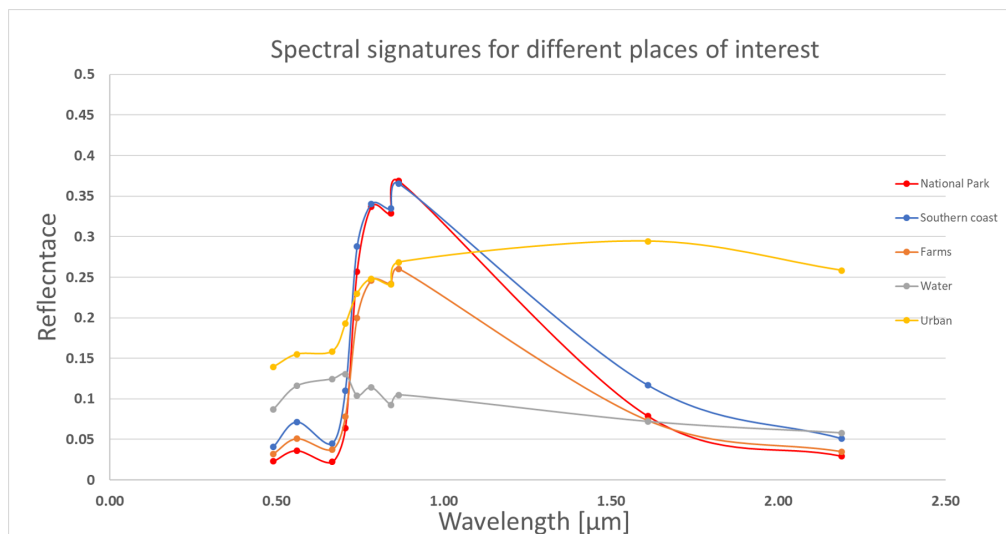


Figure 3.21: Spectral signatures for different places of interest

### 3.4.2.1. Different classes

The spectral signature plot of figure 3.21 shows that both urban and water area are very different than the vegetated areas. Urban area can easily be distinguished by the very high reflectance in the shortwave infrared domain and water area by the very low reflectance in the near infrared. Both urban and water have a much higher reflectance in the visible wavelength range than the vegetated areas.

For the three different places of interest with mangrove vegetation, that were already discussed in the previous section, the differences are smaller. The shapes of the points on the national park and the southern coast are very similar to the shape of the reference spectral characteristics of figure 2.9. The mangroves inside the extensive shrimp farms have a much lower reflectance in the near infrared, which might be caused by the mixing with the water pixels. Comparing the signature of the points in the National Park and the southern coast shows biggest differences in the visible wavelengths. The mangrove type at the southern coast show higher reflectance on those wavelengths. According to figure 2.9 the southern coast is likely to be dominated by *Avicennia* mangrove types and the National Park by *Rhizophora* mangrove types. This coincides with the training data (section 3.1.2).

The time series show also differences inside the classes. The bands in the infrared show more variability in the reflectance between wet and dry season inside the National Park than at the southern coastline. This might be related to the different phenological properties of the mangrove species. Also the time series of the radar backscatter will be further investigated as a measure of seasonal variability of the different classes.

### 3.4.2.2. Wavelength ranges

Figure 2.9 showed that the visible wavelength range is most influenced by leaf pigments, e.g. chlorophyll. The Near-Infrared range is most influenced by the cell structure of the observed vegetation, e.g. leaf and canopy structure. The shortwave Infrared range is influenced by the water content.

Differences in the **visible wavelengths** occur between mangrove and non-mangroves and within the mangroves. Water and urban areas reflect much more than vegetated areas. Between the types of mangrove in the National park and along the coast also differences occur. This means that the chlorophyll content of those two mangrove types are different. This is related to the different colors of the *Rhizophora* and *Avicennia* species.

Differences in the **near infrared wavelengths** for vegetation are influenced by cell structure such as leaf and canopy structure as well as water content. For the mangrove types different reflectance is expected. Significant differences are visible between the pure mangroves and the water areas. Areas of extensive shrimp farms have a value between those two. The spectral reflectance in the NIR range within the farms are significantly lower than both the NP and the coastline. Urban areas are similar to the farms in the NIR range but

the water areas are clearly less reflective and thus easy to distinguish. In the time series it was seen that the near infrared wavelengths also significantly change according to the seasonality. Phenological properties of the different mangroves might become visible in those wavelengths.

The last range, the **shortwave infrared wavelengths**, highly influenced by water content in vegetated areas, shows bigger differences between the mangrove species. The *Avicennia* along the southern coast might consist of higher water content than the *Rhizophora* species that is present in both the National Park as the shrimp/fish farms. The behaviour of water areas are similar to mangrove areas in this wavelength regime. Urban areas show very different patterns in the shortwave infrared compared to all other classes and is therefore easy to discriminate.

### 3.4.3. NDVI

As mentioned in section 2.4.1 the Normalized Difference Vegetation Index (NDVI) is the best indicator for vegetation detection. Figure 3.22 shows the NDVI for a cloud free composite of the ROI using a median of the complete Sentinel-2 imagery collection. The western side of the area and the southern coastline appears to be the most green and consist of the highest amount of vegetation. Low NDVI values, coloured blue, are visible in the water canals from the river and sea together with houses along those canals. Those houses are located next to the farms with mixed water and mangroves, mostly coloured white and light green.

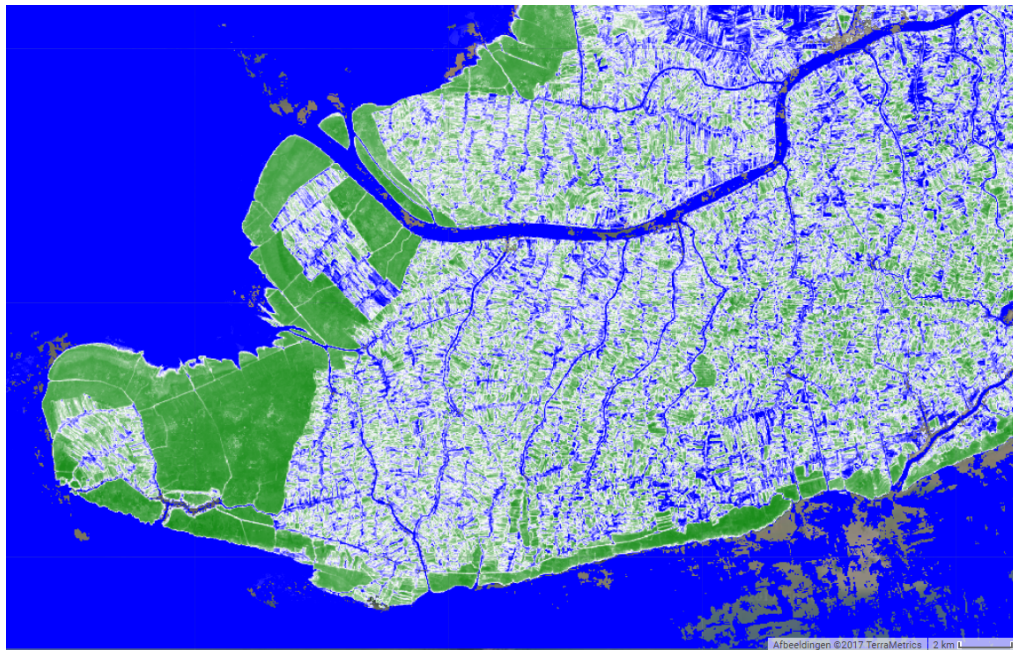


Figure 3.22: NDVI values of a cloud free mosaic (the median) of all Sentinel-2 images. Color ranges from Blue (value 0) to White to Green (value 1)

A combination of the time series of NDVI values along the places of interest including urban and water areas is shown in figure 3.23. The differences in NDVI values for the mangrove places are clearly visible. The National Park has highest values, followed by the southern coast and finally the shrimp farms. Urban area has a NDVI value of around 0.2 and water areas often have a negative NDVI. The value in the image from June 2017 is an outlier because of the much lower NDVI values except for urban area that is extraordinary high. The most likely reason for this is the high cloud cover in this acquisition, especially vague clouds which are hard for the cloud detection algorithm to find. The clouds do not only cause masked pixels but also shadows which provide unreliable data.

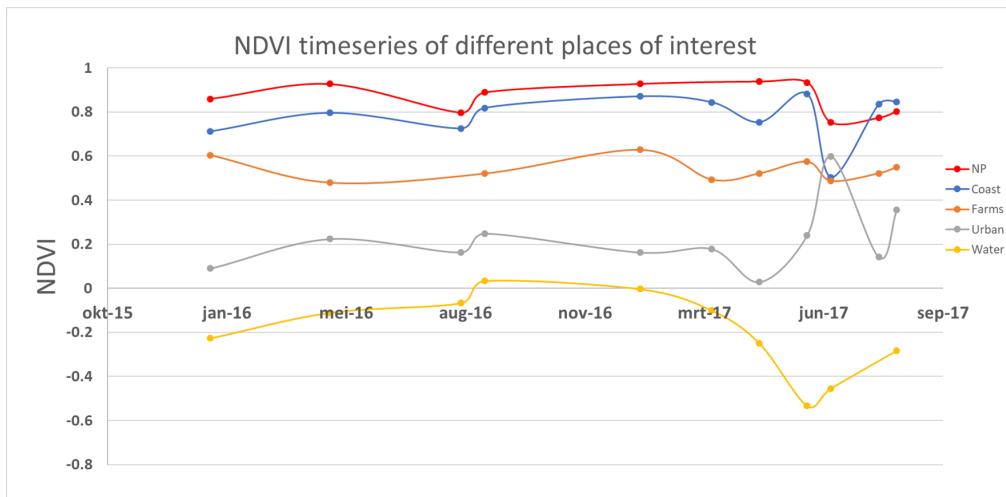


Figure 3.23: Time series of NDVI on different places of interest

### 3.5. Software

All the processing of those data cannot be done without the use of different software. The first is the Epicollect+ mobile application and website. This has been used for acquiring the GPS locations and information during the fieldwork campaign. Image processing of Sentinel-2 data is done using the Sen2Cor and SNAP software. Most of the work in this thesis has been executed in the Google Earth Engine. The processing of Sentinel-1 data, the generation of time series and the complete classification is done using the Google Earth Engine Javascript API. Finally, some extra calculations are done in Matlab for the Fourier coefficients extraction using HANTS (Harmonic Analysis of Time Series) algorithm, which will be discussed in more detail in section 4.4.

#### 3.5.1. Epicollect+

Epicollect+ is a free software developed by Imperial College London created for mobile data gathering. Users can make their own project using the web server app (figure 3.24). Using this web server app a form can be created which is used in the field to fill in the data that needs to be collected. Therefore it also consists of a mobile app that provides an interface for users to gather the data that is specified in the form (figure 3.24). This can be media fields like GPS, photos, videos as well as plain text, checkboxes, dropdown lists etc. After collecting the data in the field the web server app visualizes the data in tables and maps and it can be downloaded for further analysis (figure 3.25).

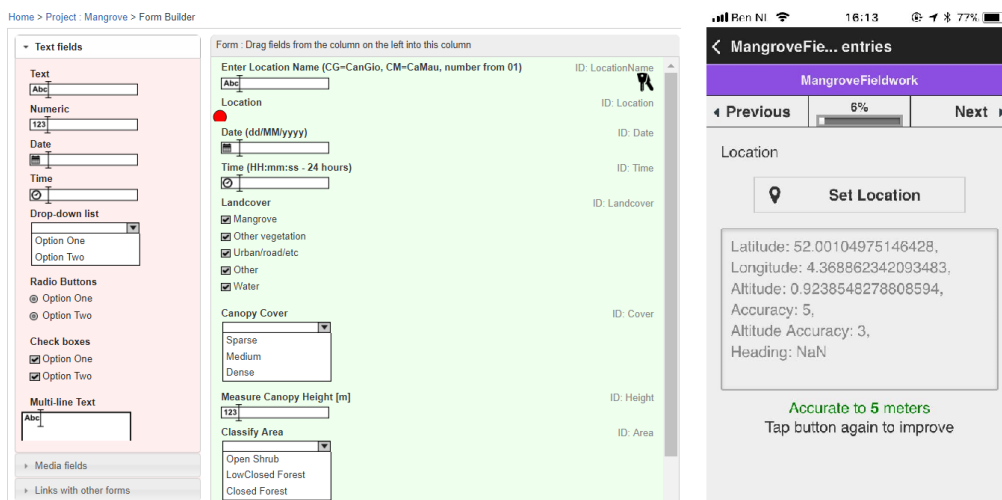
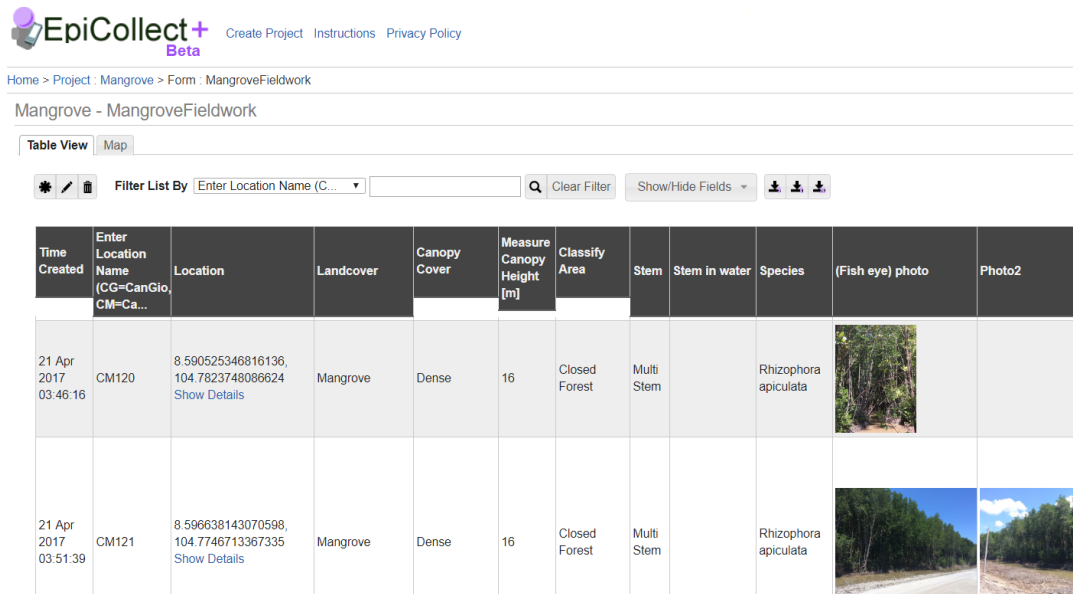


Figure 3.24: Epicollect+ project and question form on web server app and on mobile application



Home > Project : Mangrove > Form : MangroveFieldwork

Mangrove - MangroveFieldwork

Table View | Map

Filter List By: Enter Location Name (C... | Clear Filter | Show/Hide Fields |




Time Created	Enter Location Name (CG=CanGio, CM=Ca...)	Location	Landcover	Canopy Cover	Measure Canopy Height [m]	Classify Area	Stem	Stem in water	Species	(Fish eye) photo	Photo2
21 Apr 2017 03:46:16	CM120	8.590525346816136, 104.7823748086624 <a href="#">Show Details</a>	Mangrove	Dense	16	Closed Forest	Multi Stem		Rhizophora apiculata		
21 Apr 2017 03:51:39	CM121	8.596638143070598, 104.7746713367335 <a href="#">Show Details</a>	Mangrove	Dense	16	Closed Forest	Multi Stem		Rhizophora apiculata		

Figure 3.25: EpiCollect+ project and result table on web server app

### 3.5.2. Sen2Cor and SNAP

As mentioned in section 3.3.1 the correction that is needed for Sentinel-2 data from Top of Atmosphere (TOA) reflectance to Bottom of Atmosphere (BOA) reflectance, known as atmospheric correction, is not yet available in Google Earth Engine. Therefore this is done with the Sen2Cor processor, a third party plug-in for SNAP. SNAP, the Sentinel Application Platform, is software from ESA with a very fast image display and navigation even of giga-pixel images [18]. The Sen2Cor processor can be integrated in SNAP but can also run in an Anaconda command line editor from which version 2.3.1. is used for this research. Sen2Cor converts Sentinel-2 Level 1C products to Level 2A. This includes atmospheric correction to Bottom of Atmosphere (BOA) reflectance but also automatic scene classification. Sen2Cor creates also additional Aerosol Optical Thickness-, Water Vapor-, Scene Classification Maps and Quality Indicators for cloud and snow probabilities. Its output product format is equivalent to the TOA input: JPEG 2000 images with bands of three different resolutions, 60, 20 and 10 m [15]. In the SNAP software the resampling and subsetting as described in 3.3.1 is executed.

### 3.5.3. Matlab

Matlab is used for the extraction of Fourier coefficients from harmonic time series. A Matlab implementation for the HANTS algorithm is provided by Mohammad Abouali [1].

### 3.5.4. Google Earth Engine

Google Earth Engine is an online geospatial processing platform with a very big catalogue of satellite imagery including Sentinel-1, but also multi year archive data from Landsat and MODIS. The functionality of Earth Engine is exposed through an API available in both JavaScript and Python. For the javascript API tutorials are available and a web-based IDE is available known as the Earth Engine code-editor on `code.earthengine.google.com`. An example of this code-editor, used in this research, is shown in figure 3.26. Figure 3.26 shows that on the left side of the code-editor a script manager ("scripts"), documentation of the complete javascript API algorithm ("docs") and asset manager ("assets") is available. In the script manager some example scripts are available to show some of the available operations. In the middle is the javascript API that is created by the user. On the right side is a click tool that gives values from the maps below ("inspector"), the "console" where the print commands are shown and a task manager ("task"). A 'Get Started Guide' is available on <https://developers.google.com/earth-engine/getstarted> covering basic operations as importing data, filtering, sorting, masking and visualizing image collections.



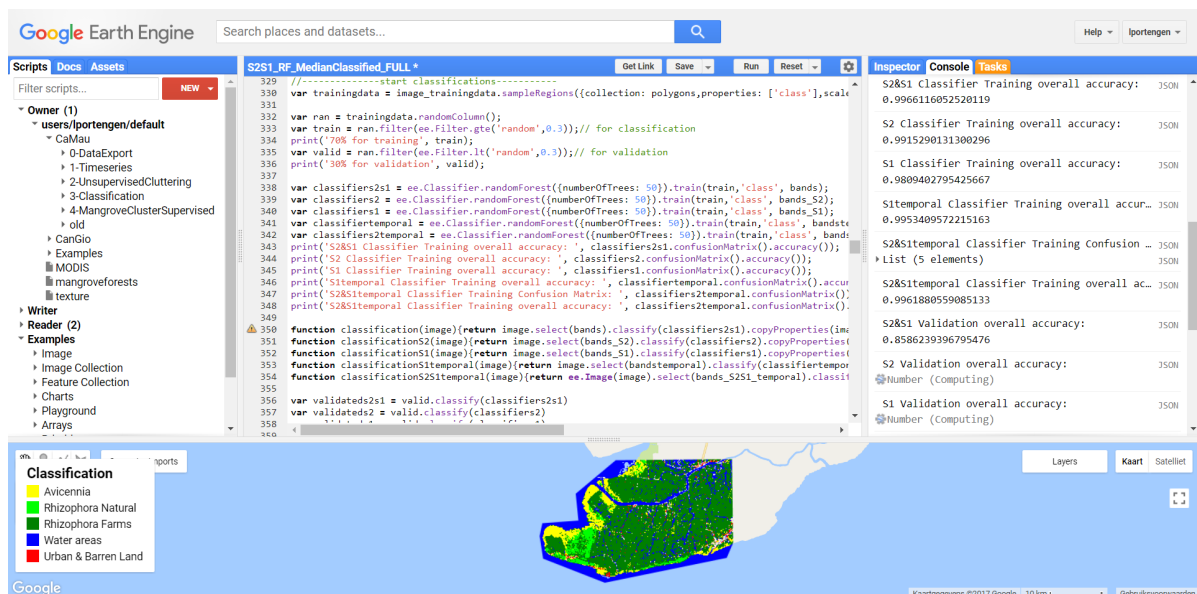


Figure 3.26: Screenshot of Google Earth Engine Javascript API

## 3.6. Conclusions

This chapter extensively described the area of interest and the data that is used in this research.

- The **region of interest (ROI)** is located in the most southern part of Vietnam, Ca Mau province which consists of the highest percentage of mangroves left in the Vietnamese Mekong Delta. Inside this ROI locations are visited during a fieldwork campaign where ground truth data is gathered. This data is divided into 5 training classes: Mangrove vegetation of *Avicennia Alba* species, Mangrove vegetation of *Rhizophora Apiculata* species in natural environment, Mangrove vegetation of *Rhizophora Apiculata* species in extensive shrimp farms, Water (rivers, sea and intensive shrimp ponds) and Urban area and barren land (figure 3.2).
- Radar data is used from Sentinel-1 satellite mission satellites, that carry a C-band Synthetic Aperture Radar. Acquired data is processed to Ground Range Detected (GRD) scenes that represent the backscatter coefficient  $\sigma^0$ . C-band backscatter is dominated by volume scattering in forest canopies and therefore useful for mangrove discrimination.
- Optical data is used from Sentinel-2 mission satellites, that carry a multi-spectral imager (MSI) with 13 spectral channels. The selected channels have 10/20m pixel resolution and represent the surface reflectance in different wavelength domains, from visible, near infrared to shortwave infrared.
- An **exploratory data analysis** investigated places of interest (POI) consisting of mangrove vegetation. The Mui Ca Mau National Park shows highest NDVI values compared to the mangroves along the southern coast and the mangroves inside shrimp farms. Mangroves along the southern coast have a higher reflectance in the red band, causing a lower NDVI value. Mangroves inside the extensive shrimp farms have a much lower reflectance in the NIR, causing a lower NDVI value. The latter is the effect of mixing of the pixel reflectance with the water surrounding the mangroves.
- All of the POI with mangrove vegetation show a seasonal pattern for the VV-polarized backscatter coefficient which seems to be related to the variations between the wet and dry season. This **temporal variability** is also seen in the optical data and most significant in the near infrared wavelengths. Variations between wet and dry season are larger inside the National Park compared to the southern coast line. This might be caused by the different phenological properties of the mangrove species that are present at the different locations. *Rhizophora Apiculata* species shows more variation than the *Avicennia Alba*. Mature mangroves inside the National Park increase more in canopy structure or leaf water content in the wet season compared to the younger trees and shrubs along the southern coastline.



# 4

## Methods for mangrove classification

An exploratory data analysis (section 3.4) investigated the information that can be extracted from the Sentinel-1 and 2 satellite imagery. With this satellite imagery clustering algorithms are used to find properties of homogeneous pixels. Analysing those clusters and comparing them with information of the exploratory data analysis gives insight on the meaning of the clusters. Adding ground truth training classes as input, supervised classification methods will be used for making land cover maps. This supervised classification has many options and different classifiers and data inputs will be investigated. Verification and validation is needed to see which gives the most accurate result.

### 4.1. Data input

In chapter 3 different data sources and the region of interest are extensively described. After all the processing steps from the different satellite sensors (sections 3.2.1 and 3.3.1) the final available images are summarized in figure 4.1. Those available images will be used as input for the following unsupervised and supervised classification methods. Sentinel-1A and 1B images together are averaged into monthly average radar images. For each available Sentinel-2 image the corresponding monthly average from Sentinel-1 data is merged into a combined image.

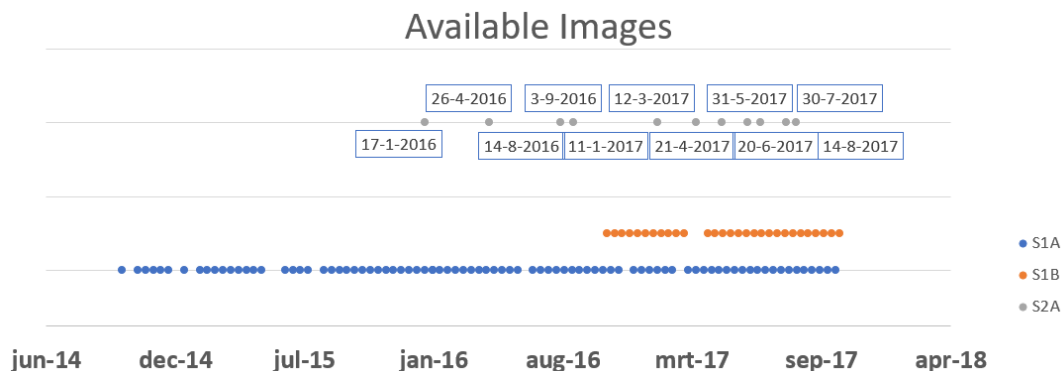


Figure 4.1: Available images serving as input for the classification methods. The grey dots represent the available images from the Sentinel-2A satellite with the exact dates indicated. The blue and red dots represent images from the Sentinel-1A and 1B satellite respectively.

## 4.2. Unsupervised clustering

Unsupervised classification, commonly referred to as clustering, is an effective method of partitioning remotely sensed data in groups of pixels (clusters) [45]. Without any prior knowledge on the area of interest the computer calculates clusters that have similar characteristics. To create a representative image that covers the whole area of interest the median of the available images is calculated. With this averaging the temporal properties vanish but this way one clustering event covers the whole area of interest without having too much effect of outliers by clouds, error strips etc. The median is calculated for the available time series of Sentinel-1 radar imagery, for the Sentinel-2 optical imagery time series and the combined time series (figure 4.1).

### 4.2.1. K-means algorithm

In this research the K-means clustering algorithm is used. K-means clustering aims to partition  $n$  observations into  $k$  number of clusters  $S$  in which each observation belongs to the cluster with the nearest mean. This also means the variance within the cluster  $S$  is minimized:

$$\min \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\| = \min \sum_{i=1}^k |S_i| \text{Var} S_i \quad (4.1)$$

where  $\mu_i$  is the mean of the points in  $S_i$ .

The algorithm is an iterative procedure where each observation  $x_{1..n}$  is assigned to the cluster which mean has the least squared Euclidean distance. With these new clusters the new mean of those clusters is calculated. The algorithm is finished when the observations no longer change cluster [4].

The name K-means indicates the  $k$  number of clusters which have to be specified by the user. Since no preliminary information about the data is available, also no information about the best number of clusters is available. For the time-averaged image collections the K-means clustering is executed for  $k=4$  and  $k=10$ . It is expected that there are at least two different species of mangrove and also different compositions of water and mangrove. The clustering algorithm is executed for three data inputs: the median image of the Sentinel-1 time series, the median of the Sentinel-2 times series and a combination of those two images. The data output is an image consisting of  $k$  different clusters.

### 4.2.2. Spectral signatures of clusters

To explore the information that is in the resulting clusters the mean value of all the pixels within a cluster is calculated for each input band. This gives an average reflectance per wavelength and/or an average backscatter for VV and VH polarization for every cluster. Those average reflectance values per wavelength regime are compared with the spectral signatures from the exploratory data analysis (section 3.4.2) to see what the clusters are representing in terms of classes and land cover.

### 4.2.3. Clustering of vegetated areas

K-means clustering divides the input data such that the input pixels are closest to the mean of cluster  $k$ . The bigger the differences in input pixels, the further away the values of pixels are from the mean while still belonging to that cluster  $k$ . Since variations between mangrove species are small compared to variations between mangrove and water the K-means clustering algorithm will not give information between different mangroves. Therefore, first the K-means algorithm for  $k=4$  is executed over the whole Region of Interest (ROI). Secondly, the cluster which is most likely to completely consist of vegetation (according to reference signature such as 2.3) is used as a new input for the clustering algorithm. Within this 'vegetation cluster' the algorithm is more likely to find differences within the (mangrove) vegetation. Again, those cluster properties can be compared with the spectral signatures from the exploratory data analysis (section 3.4.2).

### 4.3. Supervised classification

In supervised classification the user 'supervises' the pixel classification process. The user specifies the various pixels values or spectral signatures that should be associated with each class. This is done by selecting representative sample sites of known land cover type and are called training sites or areas. The training classes of the ROI are described in detail in section 3.1.2. The result of such a supervised classification is a map of the land covers which represent the different classes of the input data. A land cover map indicates the different land covers at a specific moment in time or over longer period of time. From these land covers the total area of the different classes can be estimated. Multiple land cover maps can detect changes in land cover and area. There are many things that influence the supervised classification process. Those include the classification inputs as well as the the classification algorithm. Some factors will be discussed separately in the next sections.

#### 4.3.1. Classification inputs

Executing the supervised classification algorithm is done with many different data inputs. This means that different amount of information is put in the classification schema, also known as different features. The ground truth data and resulting training classes are discussed in section 3.1.2. This ground truth data is randomly subdivided in a set of training (70%) and validation (30%) polygons. Since this separation is random, this division slightly changes every time the algorithm is executed but never more than 1%. The polygons contain the locations properties with the assigned class. For the training polygons to serve as input for the classification algorithm they are needed to be sampled from a training image. This way the training polygons get properties that are specific for the class and a feature space is created. The training image in this research is the median of all images from 2017 since the data is gathered in 2017. Using only the image from april 2017, when the fieldwork is exactly executed, is not possible since a lot of input polygons cannot be used because they are contaminated by clouds. The resulting spectral signatures from the training polygons are visible in figure 4.2. Compared to figure 2.9 the differences in the near infrared regime are slightly smaller between the three mangrove types. The training polygons will be used to learn the classifier what is the corresponding class for a specific pixel with the properties of the satellite data. The validation polygons will be used after the classification to check if they coincide with the resulting classes after the classification procedure.

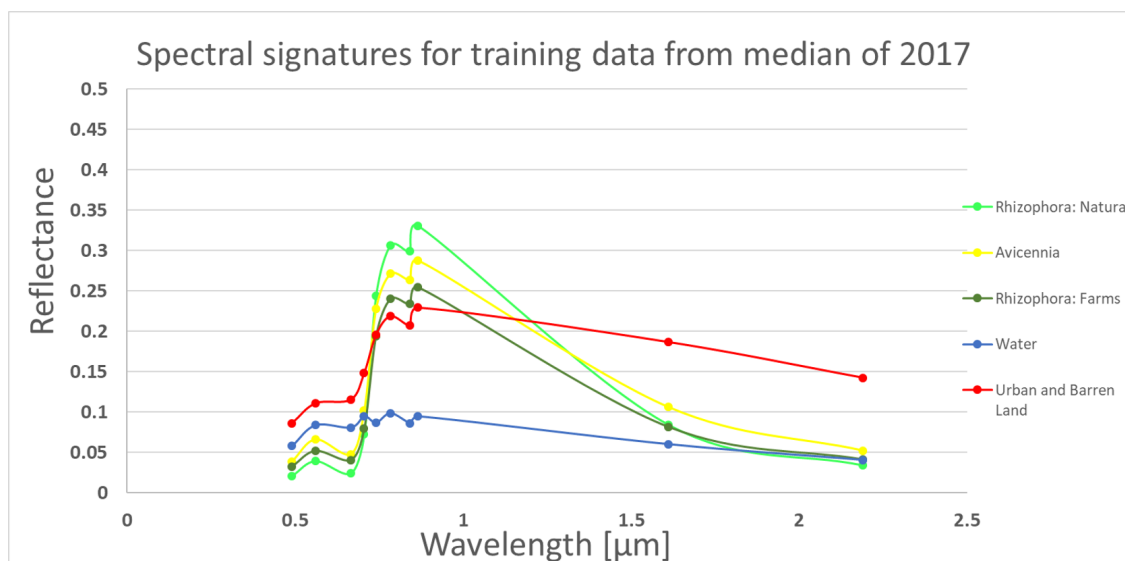


Figure 4.2: Spectral signatures for different places of interest

The training polygons contain properties per class for every feature that is available. However, classifications are executed for five different set of features. Both Sentinel-1 radar data and Sentinel-2 optical data as well as the data fusion are already discussed in section 4.1. Later, also the temporal information is added as classification feature. This will be discussed in detail the next section. The five different input sets are:

1. Sentinel-2 optical data: Bands 2,3,4,5,6,7,8,8A,11,12 (figure 3.6) and NDVI
2. Sentinel-1 radar data: VV and VH backscatter
3. Sentinel-1& Sentinel-2 fusion: all bands 1. and 2.

4. Sentinel-1 temporal information: Mean  $a_0$  and yearly harmonics term  $A_2$  for VV and VH backscatter at 10m pixel resolution and 50m pixel resolution (8 bands total)
5. Sentinel-2& Sentinel-1 temporal information: all bands 1. and 4. (total 19 features)

The classification is executed for every month that an image from Sentinel-2 is available, which was shown in figure 4.1. Yearly land cover maps are obtained by calculating the most common classification result for all available satellite images within the year. This way the complete Region of Interest (ROI) is covered and the effect of the outliers, such as clouds, is reduced.

### 4.3.2. Random forest algorithm

The supervised classification is executed in the Google Earth Engine (section 3.5.4) where some built-in classifications algorithms are available. Some of them require a manual input decision tree or classifier schema but others are more straightforward to use. The four most simple examples are the Minimum Distance, Classification and Regression Trees (CART), Support Vector Machine (SVM) and Random Forest classifiers. Those four classifiers are tested with the first input set, only Sentinel-2 optical data. The results for overall accuracy's calculated at the training sites and at the validation sites for the four different classifiers are shown in figure 4.3. It is shown that the Random Forest and CART classifier show the best results. Random Forest has the highest overall accuracy values of around 99% using the training data and around 85% for the validation data. The Random Forest classifier is an ensemble classifier that uses a set of CARTs to make a prediction and that makes it a little bit better than CART. Therefore, the Random Forest is the chosen classifier in this research.

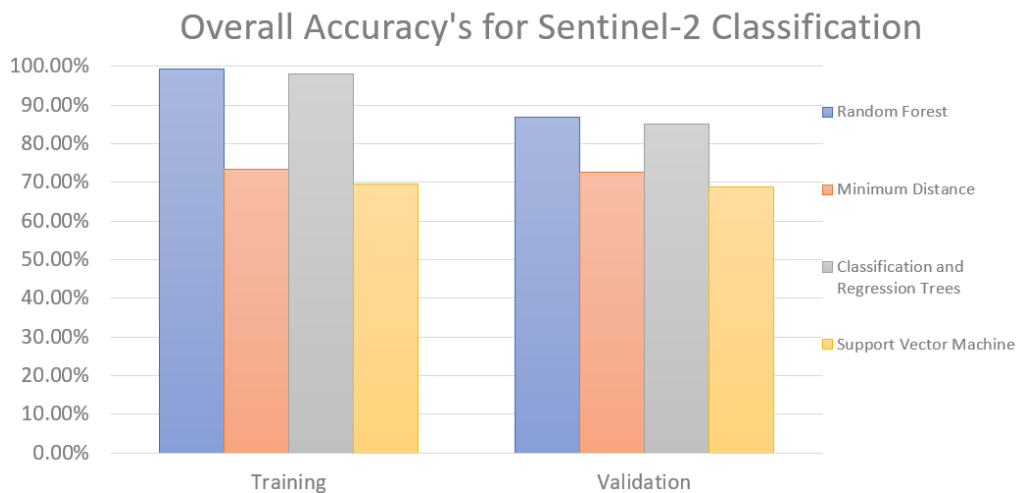


Figure 4.3: Overall accuracy's for different classifiers in Sentinel-2 classification

The Random Forest algorithm operates by executing a multitude of decision trees. The trees are created by drawing a subset of training samples through replacement. This means that the same sample can be selected several times, while others may not be selected at all. Each decision tree is independently produced and each node is split using a chosen number of features. In this case the number of features per split defaults to the square root of the total number of features available, depending on the data input set between 2 and 19 features. Using all variables for splitting the nodes causes a big increase in computation time and is therefore not desirable. The user also defines the number of trees within the forest. The majority of studies use 500 trees since this is the default setting in the R package which is often used for Random Forest applications [5]. An investigation on the sensitivity of the number of trees has shown that a much smaller number also gives good results with much shorter computation times [5]. In this research 50 trees are chosen because this gives good results with a computation time that can be handled by the Google Earth Engine. Each decision tree gives a class result from which the majority is chosen as the final class.

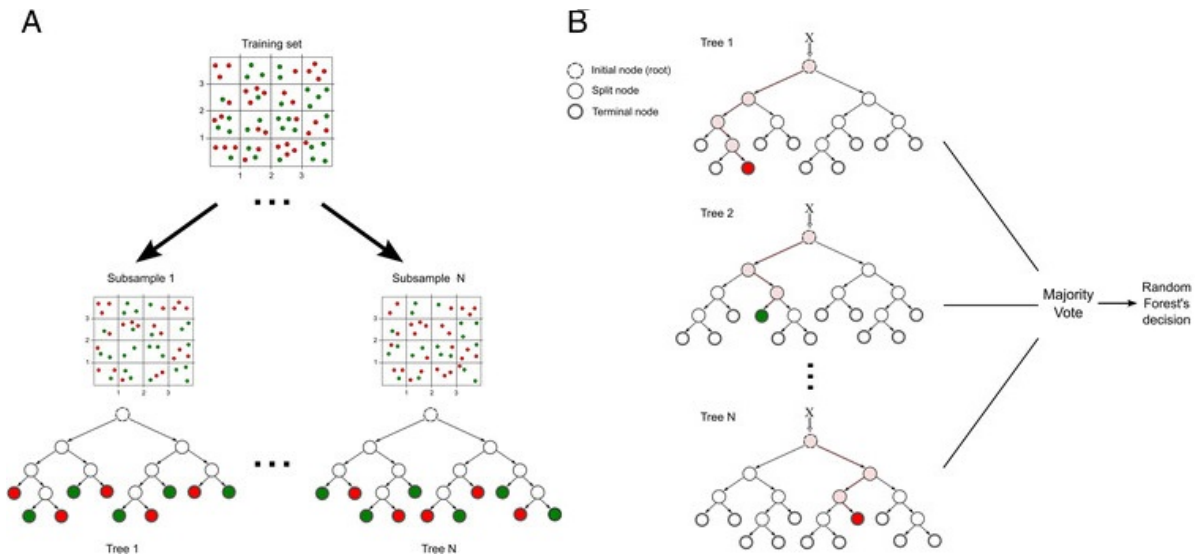


Figure 4.4: Random forest model. **A** represents the training process, where decision trees are built from random samples of the original data, which contains positive (green labels) and negative (red labels) examples. **B** represents the classification process where the majority of class from the individual trees is chosen. For each data in each tree the algorithm starts at the root node of a decision tree and traverses down the tree testing the variables values in each of the visited split nodes (pale pink nodes) until a leaf node is reached: green nodes predict for the positive class, red nodes predict for the negative class [34]

### 4.4. Temporal information

In section 2.5.3 temporal analysis has already been noted as possible improvement for mangrove properties extraction. In section 3.4 the first time series gave insight in the periodic changes. Especially the Sentinel-1 VV backscatter (figure 3.16) showed a clear periodic pattern interfering with the wet and dry season in southern Vietnam. Including this information in the classification algorithm is done by using harmonic analysis using Fourier series which appears to be an ideal way to facilitate multitemporal analyses using remote sensing data, particularly for phenological studies, health and land development [8]. Fourier series is a way to represent a time-series as the sum of a set of oscillating functions, namely sines and cosines (figure 4.5):

$$\begin{aligned}
 f(x) &\approx a_0 + a_1 \cos(x) + a_2 \cos(2x) + a_3 \cos(3x) + \dots \\
 &\quad + b_1 \sin(x) + b_2 \sin(2x) + b_3 \sin(3x) + \dots \\
 &= a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx))
 \end{aligned}
 \tag{4.2}$$

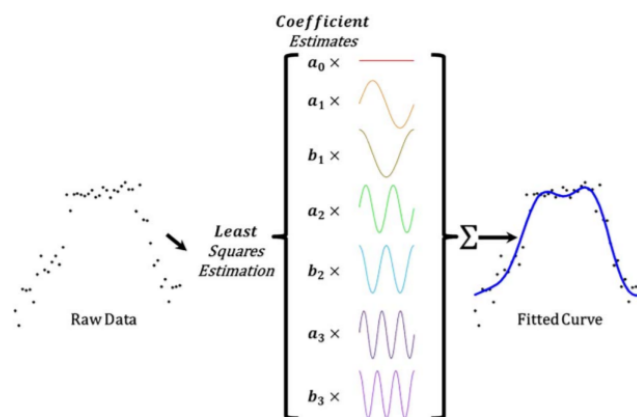
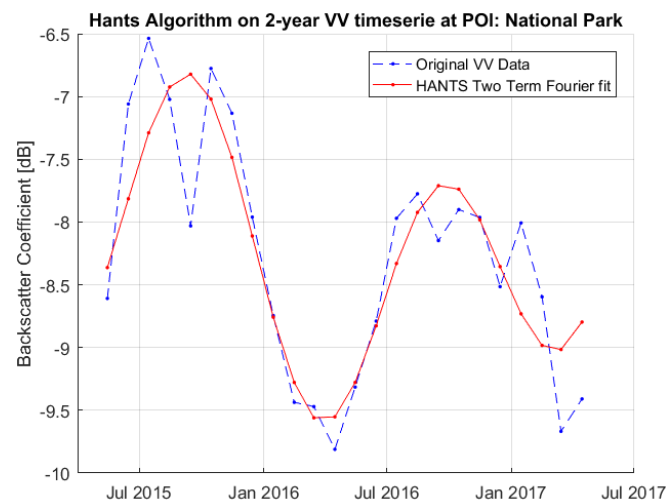


Figure 4.5: Basics of Fourier analysis [8]

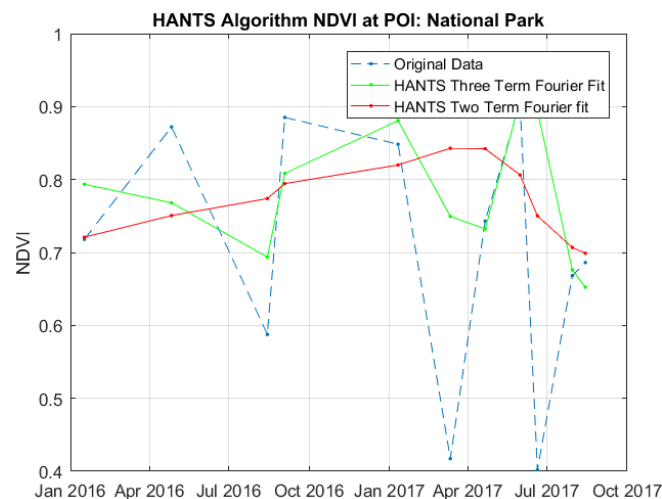
The constant term  $a_0$  is called the mean of the series and the pairings of the cosines ( $a_1, a_2, a_3$ ) and sines ( $b_1, b_2, b_3$ ) are the harmonics of the series. The amplitude of the pairings is calculated by:

$$A_n = \sqrt{a_n^2 + b_n^2} \quad (4.3)$$

In this research the HANTS (Harmonic Analysis of Time Series) procedure is used which decomposes a periodic time-dependent data set into sinusoidal functions but also smooths the data, removes outlier and fills the gaps of missing data [59]. The result is a smoothed dataset and the phases and amplitudes of the sinusoidal functions, which serves as temporal information for each pixel [1]. The HANTS procedure is executed for a two-year time series of Sentinel-1 VV and VH backscatter and for the total available time-series of Sentinel-2 NDVI values. The fitted datasets at the location of the National Park are shown in figure 4.6. The VV time series shows a more clear pattern than the NDVI and the Fourier fit is much better. The NDVI time series is not as long as the backscatter and also contains outliers, making the time series even sparser since those are left out. With fewer data points it is harder to obtain a proper fit to the data and the Fourier coefficients provide no reliable temporal information. The resulting amplitudes from the VV and VH backscatter time series are calculated for each pixel on both 10m resolution scale as well as 50m resolution. This is done to reduce the spatial variability in time. Especially with radar backscatter values those can be slightly different over time for example due to a slightly different viewing angle on the GRD scenes. Using superpixels is a way of reducing noise effects of the radar signal differences.



(a) VV Backscatter time series



(b) NDVI time series

Figure 4.6: HANTS Fourier fit at National Park



## 4.5. Validation

Accuracy assessment is an important part of image classification. In this study, a total of 195 reference areas were surveyed in the field to serve as input samples for both training and validation. 70 percent of those areas are used in the training procedure and the remaining 30 percent is for validation. The most common accuracy assessment method for supervised learning algorithm is using an error matrix, also confusion matrix [45]. A confusion matrix compares two columns within an image collection: one containing the actual values, and one containing predicted values. The rows of the matrix correspond to the actual values, and the columns to the predicted values. Descriptive statistics (user's accuracy, producer's accuracy and overall accuracy) are computed from the confusion matrix.

Another way of validating the classification is by quantifying the stability of the classification outcome. As mentioned in the previous section 5.2 the yearly classification is calculated as the most used class outcome from the available classified images. The stability of the classification is calculated by a ratio between the number that the final class is chosen and the total number of available classification.

The last validation method is comparing the classification result with an existing mangrove cover product. In Google Earth Engine the Mangrove Forest of the World (2000) is a globally available raster dataset, visualized in figure 4.7. This database is made using Landsat satellite data from the year 2000. More than 1,000 Landsat scenes were classified using hybrid supervised and unsupervised digital image classification techniques [22]. The database is being used for identifying priority areas for mangrove conservation, studying the role of mangrove forests in saving lives and properties from natural disasters, carbon accounting and biodiversity conservation. The database serves as a baseline for mangrove monitoring.

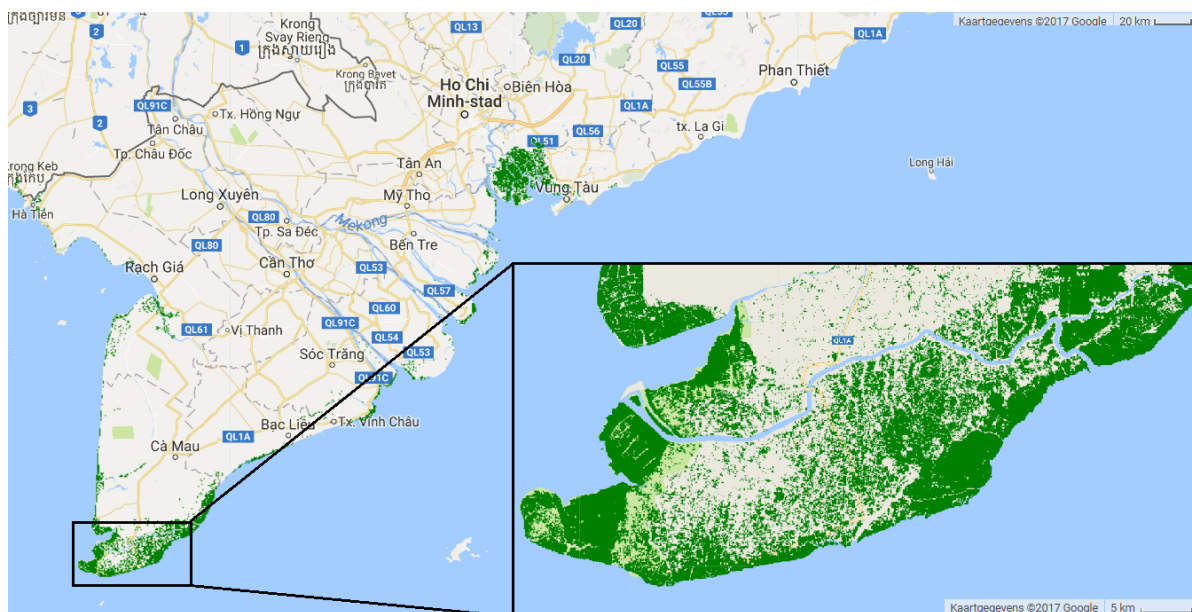


Figure 4.7: Mangrove Forest of the World from 2000 [22]

## 4.6. Conclusions

This chapter describes the different methods used in this study to find the best classification discriminating mangrove types in Vietnam.

- A simple, fast and unbiased method that provides information on similar pixels within an image is **K-means clustering**. Users need to specify the amount of the clusters made. By calculating spectral signatures from the resulting clusters the corresponding land covers can be estimated from literature or ground-truth information.
- **Random Forest supervised classification** is a machine learning method where ground-truth data is used as training input for assigning each pixel to a class. The Random Forest algorithm operates by executing multiple decision trees. Each class represents a land cover assigned from the training input data. For each (monthly) available image of Sentinel-1 or 2 the classification is executed for five different data input combinations. Yearly land cover maps are obtained by calculating for the pixel class that occurs most in all classified images in a year.
- To improve the classification extra features are introduced. Classic features such as NDVI is calculated using the Sentinel-2 optical data. Novel features are introduced by executing a **Fourier analysis** on two year time series to extract the mean and the amplitudes of the seasonal variation. Results are significant for the Sentinel-1 radar backscatter time series but not for the NDVI time series. The temporal radar information is calculated for both 10 and 50 meter resolution to serve as extra input features for the supervised classification.
- Validation is done by calculating the confusion matrix for both training as well as validation data. Overall accuracy's are compared for the different classification input data: Sentinel-2 (S2) optical data, Sentinel-1 (S1) radar data, S1 and S2 combined, Sentinel-1 temporal information and finally S2 and S1 temporal information. Also the consistency of the classification outcome is validated by a confidence map showing the stability of the classification for the different images within a year. Finally, the resulting land cover map will be compared with existing mangrove cover dataset to detect outliers and land cover changes.

# 5

## Results

The results of the different methods are presented in this chapter. First, the results of the supervised classification are shown. This includes the resulting land cover maps and the land cover changes that can be conducted from the land cover maps from 2016 and 2017. Sections 5.1.2 to 5.1.5 go deeper into the classification results with an accuracy assessment, the classifications confidence, validation with external mangrove cover map and case studies that examine the reliability of the classification results at different locations. Another result of this study is the workflow that is set-up for the best mangrove classification, shown in section 5.2. The classification results are obtained by first executing the methods described in sections 4.2 and 4.4, the unsupervised clustering and temporal analysis. A more technical approach to those results are shown in sections 5.3 and 5.4.

### 5.1. Classification results

The method of Random Forest supervised classification is explained in section 4.3. According to the input training data (section 3.1.2) the resulting classes are subdivided into urban area and barren land, water areas (sea, rivers and shrimp ponds) and three different mangrove vegetation: *Avicennia Alba* species, *Rhizophora Apiculata* species in natural environment and *Rhizophora Apiculata* in extensive shrimp farms. Those classes result in a land cover map representing the different physical coverage of the Earth's surface.

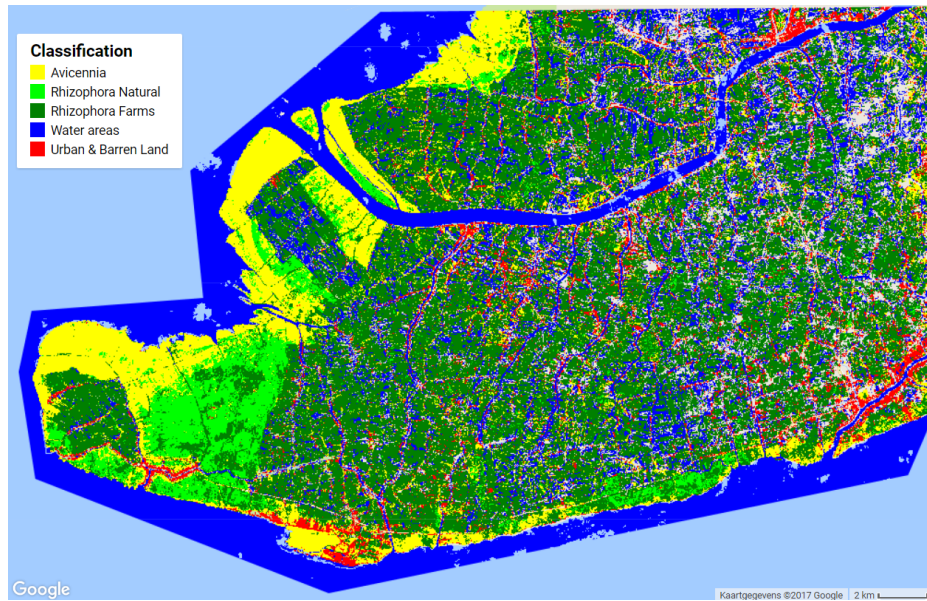
#### 5.1.1. Yearly land cover map

The final yearly land cover maps are shown in figure 5.1. This result is obtained by using the data input of the available images of Sentinel-2 optical data combined with the temporal information of the Sentinel-1 temporal analysis. An overview of the final data inputs per method is shown in table 5.1.

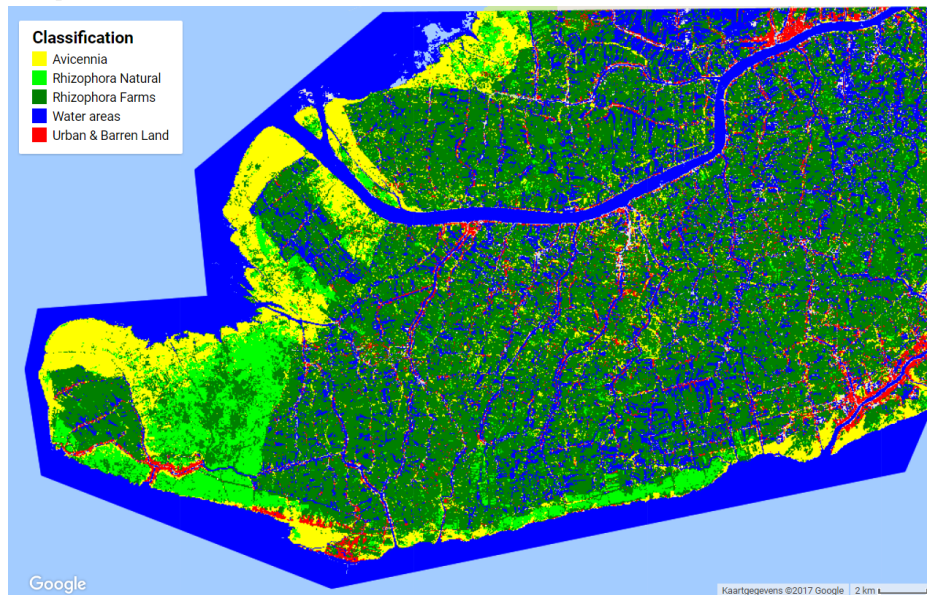
Input \ Method	Unsupervised Clustering	Temporal Analysis	Yearly Land Cover Map	
Sentinel-2 (optical)	Median of all available (11x) images	x	2016	S2 bands of 4 images
			2017	S2 bands of 7 images
Sentinel-1 (radar)	Median of all monthly (32x) images	2 year monthly images	S1 temporal result a0's and A2's	

Table 5.1: Overview of data input per method

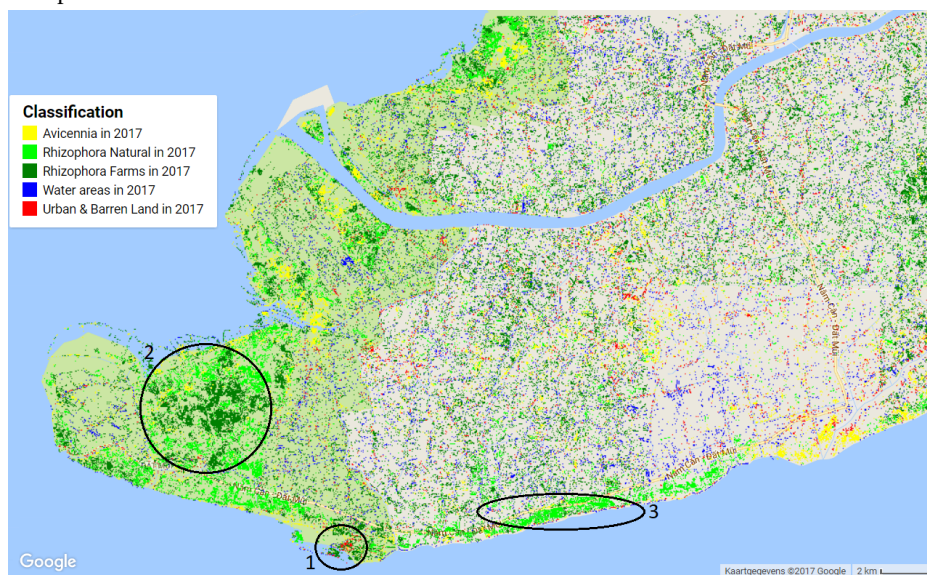
The two maps of figure 5.1 show an almost completely filled ROI. The map of 2016 has some empty pixels since there were only four images of Sentinel-2 available in 2016 and the input is masked according to the Sentinel-2 images. Some pixels are masked in all four images and don't have any value and remain empty. Since the 2017 input consists of seven images the amount of empty pixels is much smaller.



(a) Land cover map of 2016



(b) Land cover map of 2017



(c) Difference map between 2017 and 2016 image. Colors correspond to 2017 classes.

Figure 5.1: Yearly land cover maps of 2016 and 2017 and difference map.

**Land cover changes** By comparing the two maps of 2016 and 2017 land cover changes are observed. The changes in class are covered in the difference map of figure 5.1. In this map the colors represent the 'new' class of 2017 that has changed from the class in 2016. In table 5.2 the percentage of pixels that have changes from 2016 to 2017 are shown.

2016 \ 2017	Rhizophora Farms	Avicennia	Water	Urban	Rhizophora Natural
Rhizophora Farms	76%	4%	12%	3%	5%
Avicennia	29%	57%	2%	2%	10%
Water	12%	1%	87%	1%	0%
Urban	46%	10%	8%	35%	1%
Rhizophora Natural	31%	8%	0%	0%	60%

Table 5.2: Land cover changes from 2016 to 2017. Percentages with respect to class in 2016.

**Vegetation loss** The most important changes are where vegetation has changed into water or urban area. For example the resort which is located on the most southern location of Vietnam, Khai Long, is rapidly expanding, indicated with number 1 in the difference map of figure 5.1. At this location a 100MW wind power plant will be developed on earth and in the sea between 2016 and 2018 [56]. Changes from vegetation into water might be caused by the harvesting of the mangrove but can also be a change from extensive shrimp farming (including mangroves) to intensive shrimp farming (no mangroves present). In the first scenario the mangroves should change back in later years but in the second scenario the area changed land use. Table 5.2 shows that the amount of change from vegetation to water and urban is relatively small, only the class farms in 2016 has changed for 12% to the water class in 2017 and 3% to the urban class in 2017.

**Vegetation changes** Table 5.2 shows that most land cover changes occur within the vegetation class. The Natural Rhizophora class has changed 31% from 2016 to Rhizophora Farms in 2017. This is most likely caused by the differences in the National Park that are visible in the difference map (indicated with number 2). Here, most changes from Rhizophora Farms to Rhizophora Natural and vice versa are found. Many changes are also found from Avicennia (2016) to Rhizophora Farms (2017). This is related to coastal Avicennia areas that are in 2017 sometimes classified as Rhizophora Farms. The case studies will show that Avicennia in 2016 was sometimes misclassified as 'other vegetation' that was not mangrove at all. Finally, the variation of the 2016 land cover map along the southern coast line is classified as Rhizophora Natural in 2017 (indicated with number 3 in the difference map). The latter coincides with the ground-truth data from the fieldwork campaign, which will be discussed in the case studies.

**False differences** Some less important differences seem to occur which are the result of a bad classification result: something seems to have changed but in fact there was no change but the yearly classification itself was not accurate. In figure 5.1 can be seen that the 2016 land cover map has much more urban area compared to 2017, while it is expected that urban areas grow instead of shrink. Table 5.2 shows 46% of the Urban pixels in 2016 that has changed to Rhizophora Farms. Clouds had a big effect on the 2016 land cover map and both clouds as urban land cover have a distinctive high amount of reflectance in the short wave infrared that other classes do not have. Therefore, the amount of urban class was much overestimated. The classification of 2017 had a smaller effect of these clouds and thus also less falsely classified Urban pixels.

### 5.1.2. Accuracy and confusion matrix

The Random Forest classification is executed for five different data input combinations. They were set up in section 4.3.1 and are:

1. Sentinel-2 optical data: Bands 2,3,4,5,6,7,8,8A,11,12 (figure 3.6) and NDVI
2. Sentinel-1 radar data: VV and VH backscatter
3. Sentinel-1& Sentinel-2 fusion: all bands 1. and 2.
4. Sentinel-1 temporal information: Mean  $a_0$  and yearly harmonics term  $A_2$  for VV and VH backscatter at 10m pixel resolution and 50m pixel resolution (8 bands total)
5. Sentinel-2& Sentinel-1 temporal information: all bands 1. and 4., in total 19 features

The overall accuracy's of the different scenarios are visible in figure 5.2. The accuracy's are calculated with the training data input and with a separate set of ground truth validation data. The training accuracy's are very high since the Random Forest is fit to those data inputs. The validation accuracy's are more representative. It

is shown that a combination of optical Sentinel-2 data fused with radar backscatter Sentinel-1 temporal information shows highest accuracy. Therefore, this data input scenario is used for the processing of land cover maps and investigating land cover changes. Resulting land cover maps of 2017 from data input combinations 1-4 are visible in Appendix A.1.

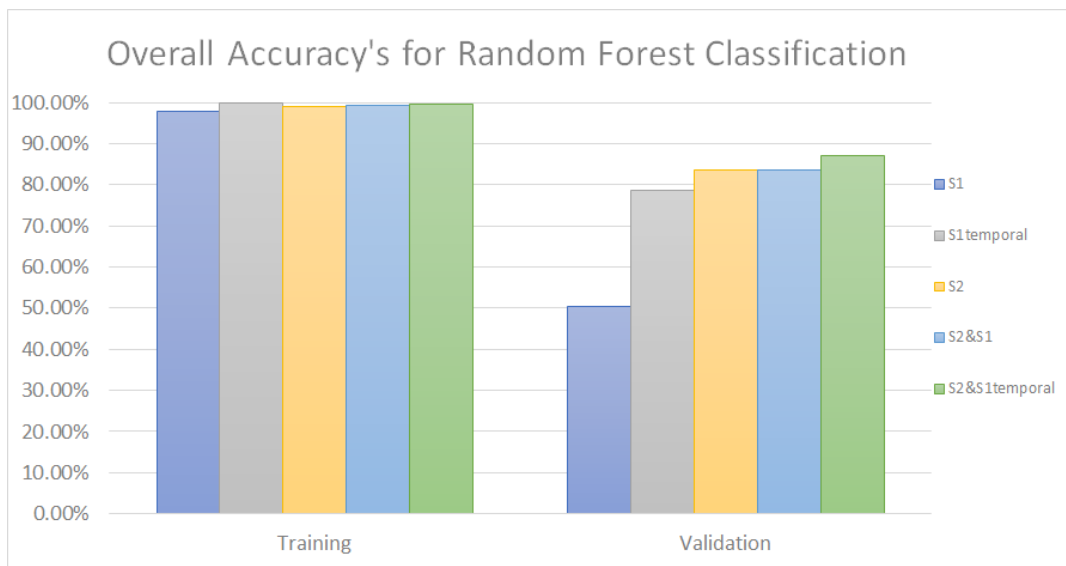


Figure 5.2: Overall accuracy's for Random Forest classification

The full confusion matrix for the Sentinel-2 and Sentinel-1 temporal input from the validation data is shown in table 5.3. The confusion matrix includes values for the producer's and user's accuracy. The producer's accuracy is the fraction of correctly classified pixels with regard to all pixels of that input training class. The user's accuracy is the fraction of correctly classified pixels with regard to all pixels classified as this class in the classified image. Both of these accuracy's are corrected for the available amount of validation pixels that is different per class. This ensures that classes with much more available pixels in the confusion matrix automatically get higher accuracy values.

Water and urban class both score highest in the user's and producer's accuracy. This was expected because those classes are very different from the vegetation. Biggest confusion of the water class is with the class Rhizophora Farms. This is caused by the fact that the water ground-truth input is sometimes small and very close to the neighbouring mangroves. From the different mangrove vegetation classes the Rhizophora Natural scores highest user's accuracy and Avicennia highest producer's accuracy. Rhizophora Farms scores worst in both accuracy's. Most confusions are between Rhizophora Farms and Natural Rhizophora. This is likely to occur if the Rhizophora in the farms is very old and almost ready to harvest or if the water ponds are relatively small and don't mix the pixel reflectance. Rhizophora Natural has only one confusion with a non-vegetation class, making it the best distinguishable mangrove class.

Input \ Predicted	Rhizophora Farms	Avicennia	Water	Urban	Rhizophora Natural	Producer's Accuracy
Rhizophora Farms	194	12	7	1	19	83,3%
Avicennia	6	118	3	8	3	85,5%
Water	19	1	195	2	1	89,4%
Urban	6	4	6	210	0	92,9%
Rhizophora Natural	24	11	0	0	189	84,4%
User's Accuracy	75,9%	87,4%	91,9%	92,9%	88,7%	

Table 5.3: Confusion matrix from Sentinel-2 and Sentinel-1 temporal data input classification results compared to validation ground truth data

### 5.1.3. Classification confidence

The yearly land cover map is the result of selecting, for each pixel, the class that came out most of the separate classifications per image within one year. This most common class is calculated by counting the total number of available classifications (cloud masked pixels are not classified) and the class which is chosen most is the winning class. The ratio between the total number of available classifications and the number of times the winning class came out is called the classification stability. This process is visualized in figures 5.3a to 5.3d. Figure 5.3a indicates the total available separate classifications in the year 2017 for each pixel. Figure 5.3b is the amount of times the final class came out of those separate classifications. Figure 5.3c shows the ratio between the two, indicating the stability of the classification. Figure 5.3d shows the same classification stability but coloured per class.

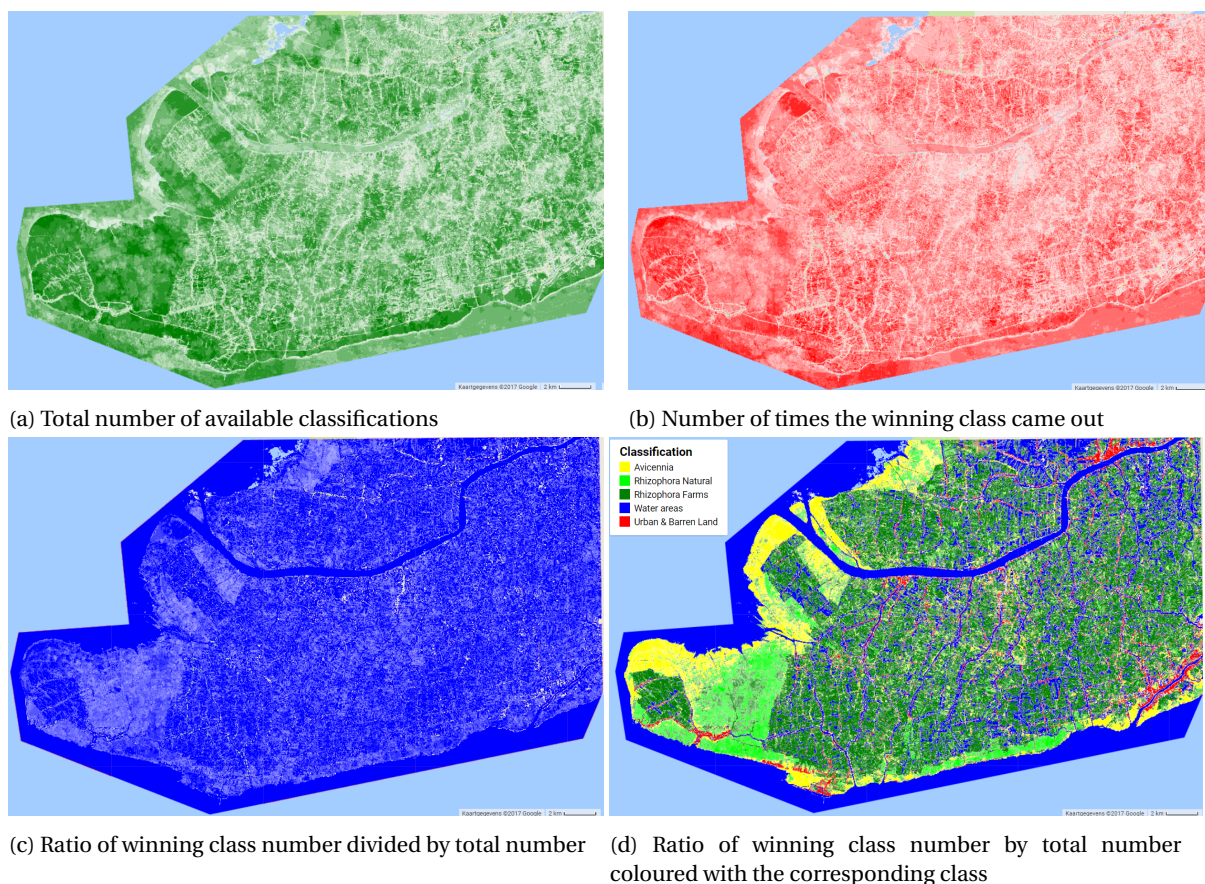


Figure 5.3: Classification confidence for Sentinel-2 & Sentinel-1 temporal input classification of 2017. The images are coloured from white to the specific color corresponding to the minimum and maximum value.

From the figures above the National Park shows lowest classification stability. This is not caused by the available classifications but by the low number of the winning class. In other words, the classification result is here less reliable. On the other hand, the ratio of classification confidence is very high in the area with many extensive shrimp farms. This is caused by the low total available classifications. Some locations only have one or two available classifications within the year 2017 and if these one or two show the same class outcome the ratio is very high. From figure 5.3d, coloured per class, can be seen that water and urban show high confidence which is in line with the low amount of confusions that were present in the confusion matrix. The high amount of confusions between Rhizophora Natural and Rhizophora Farms is related to the low confidence in the National Park area where the classification result is not very reliable.

#### 5.1.4. External validation

The previous validation methods looked into the accuracy and confidence of the classification made in this research. External validation is done with the Mangroves of the World (2000) map from Giri et al [22]. The resulting yearly land cover map of 2017 is compared with this mangrove cover map. The result is visualized in figure 5.4.

It can be seen that almost all mangrove cover from 2000 is covered in the classification of 2017 from this research. Some differences occur at the north-western lobe where in 2017 farms are present with water in between and in 2000 there was still only mangroves. This is a shift from mangrove cover to aquaculture. The clearest difference is the regression in mangrove cover at the southern coastline. At the eastern side a strip of about 500 meter shows the loss of mangrove cover between 2000 and 2017. More towards the western cape this strip is about 100 meters wide.



Figure 5.4: Comparison of land cover map of 2017 with Mangrove Cover from 2000. Bright green, dark green and yellow pixels indicate the mangroves classified in this research and coincide with the Mangrove Cover from 2000. Black pixels indicate mangrove cover in 2000 that is not classified in 2017 as mangrove vegetation



### 5.1.5. Case studies

All previous results were visualized for the entire Region of Interest (ROI). Figure 5.5 shows locations that have been visited during the fieldwork campaign and are investigated in more detail in a case studies. The case studies compare the classification results of both 2016 and 2017 with Google Earth Imagery including ground truth information. The reliability of the classification results and the land cover changes will be described.

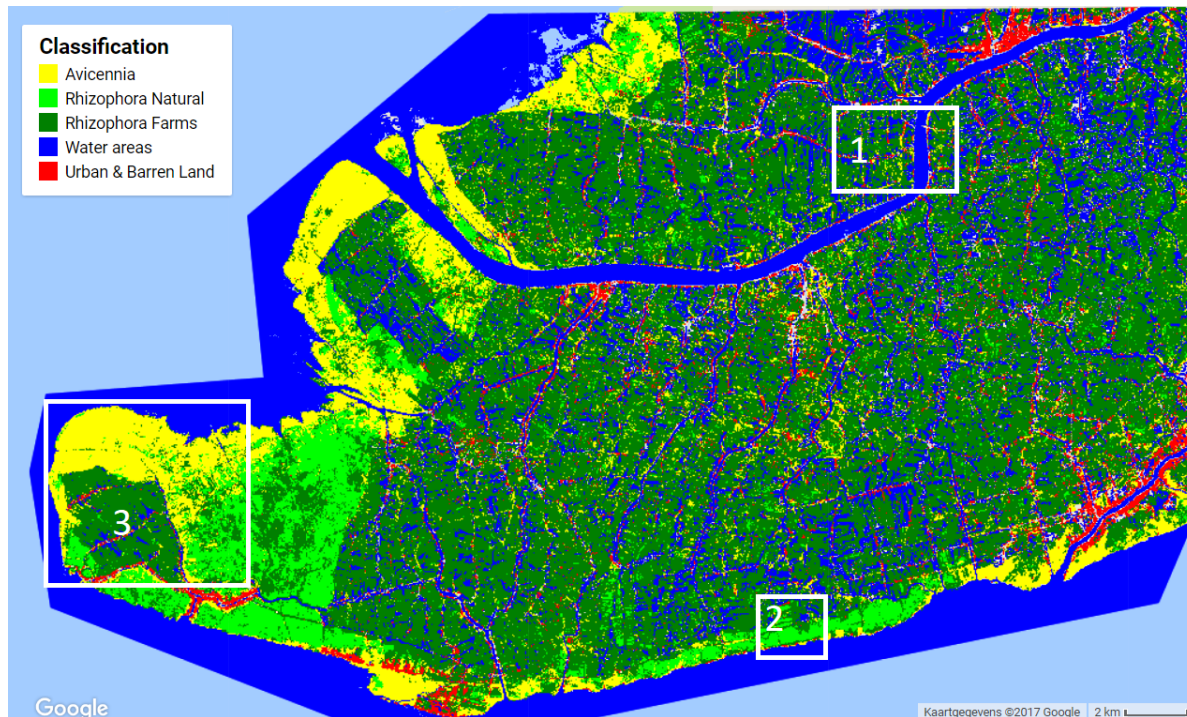


Figure 5.5: Different locations of the case studies

**1. Extensive shrimp farms near Nam Can bridge** Figure 5.6 shows the land cover zoomed into location one from figure 5.5. The result of 2016 shows some masked pixels caused by the clouds in the input images. Number one of the ground truth locations shows a dirt road with vegetation around. This vegetation is dry grass and bushes but it is classified as urban & barren land pixels together with Avicennia. There is no class where other vegetation than mangrove is covered. Apparently Avicennia is most similar to grass and bushes. This is explicable since Avicennia is the mangrove vegetation class which looks most like shrubs. Small rivers with the houses alongside are well classified just as the extensive shrimp farms. This class, *Rhizophora Apiculata* mangrove species in extensive shrimp farms, is dominant in this area. Also *Rhizophora* in natural environment is classified at some pixels but this is not likely in those surroundings. These pixels can be mangroves inside farms that are relatively old, causing the reflectance to be higher and more similar to *Rhizophora* in natural environment. The 2017 land cover maps shows less false classifications of Avicennia and natural *Rhizophora* but besides that no significant changes are detected from 2016 to 2017.

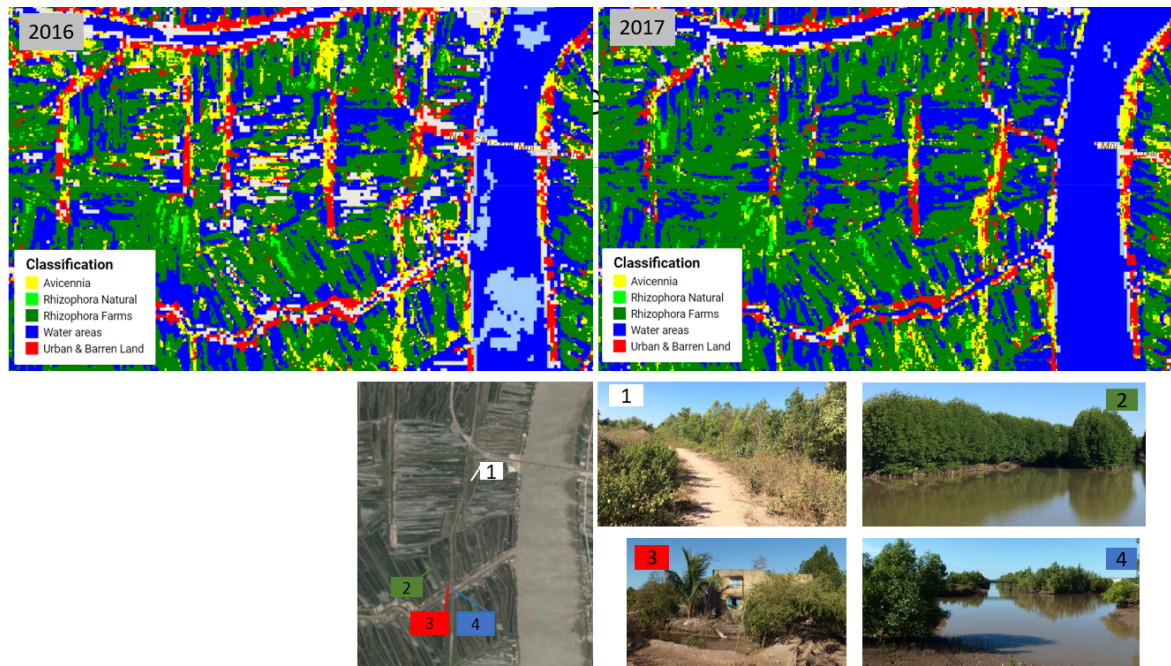


Figure 5.6: Yearly 2016 and 2017 land cover maps with ground truth examples at farms near Nam Can bridge

**2. Extensive shrimp farms near southern coast line** The second case study is located close to the southern coast line with also many farms (figure 5.8). This area has a wide strip of Natural Rhizophora along the coast and a very small strip of Avicennia closest to the sea. Also some pixels are classified as urban close to the sea but those are not correct. During the fieldwork campaign at the locations indicated at number 2 and 3 the mangrove trees were just cut down and water ponds were left. In the image of 2016 Rhizophora Farms is classified here and in 2017 water is classified, indicated with the black circles. Furthermore, there are less alternations in the 2017 result and the separation between the two Rhizophora classes is more clear than 2016. Again some Avicennia is classified between the farms but in this case ground truth location number four showed that this was indeed Avicennia present on a very small scale.

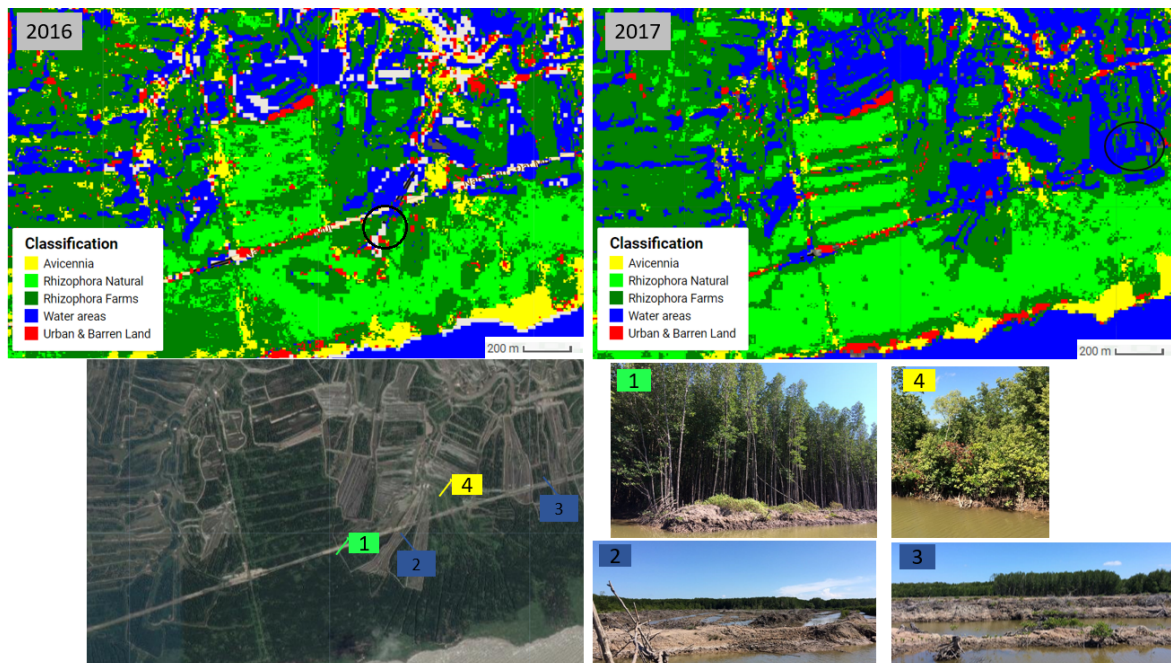


Figure 5.7: Yearly 2016 and 2017 land cover maps with ground truth examples at farms along southern coast

**3. North-west cape in Mui Ca Mau National Park** The last case study is located at the most western tip of the Vietnamese mainland. The Mui Ca Mau National Park's strict protection zone (figure 3.12) consists completely of Natural Rhizophora as well as Avicennia Alba. Separating those two classes is found to be challenging because there are a lot of confusions along the transition between those species. On this transition the data of 2016 makes a better distinction, but there are some other confusions visible, for example some pixels inside the Rhizophora Natural that are classified as Urban & Barren Land. As explained before this is related to the similar reflectance of urban land cover and clouds. Obtaining ground truth data around this area inside the National Park's protection zone was not possible. Only some locations of Avicennia Alba on the edges were observed by using a boat. Avicennia Alba mangroves were present in different sizes and ages but are all merged in one class. The closer to the northern coastline the smaller the shrubs, as visible in ground truth locations three, two and one. The transition between the two species is in reality more smooth than the classification results with a combination of both species growing. It is hard to find the best resulting class in the classification algorithm. As mentioned in section 5.1.3, this area also has the lowest confidence of the classification result.

The ecological restoration zone (figure 3.12) is where Rhizophora mangroves are cultivated around shrimp ponds. The mangroves look more 'natural' than the rest of the ROI but still consists of straight rows of mangroves alternated with water, as can be seen in the number four ground truth location. When 2016 is compared to 2017 there seem to be no significant changes here.

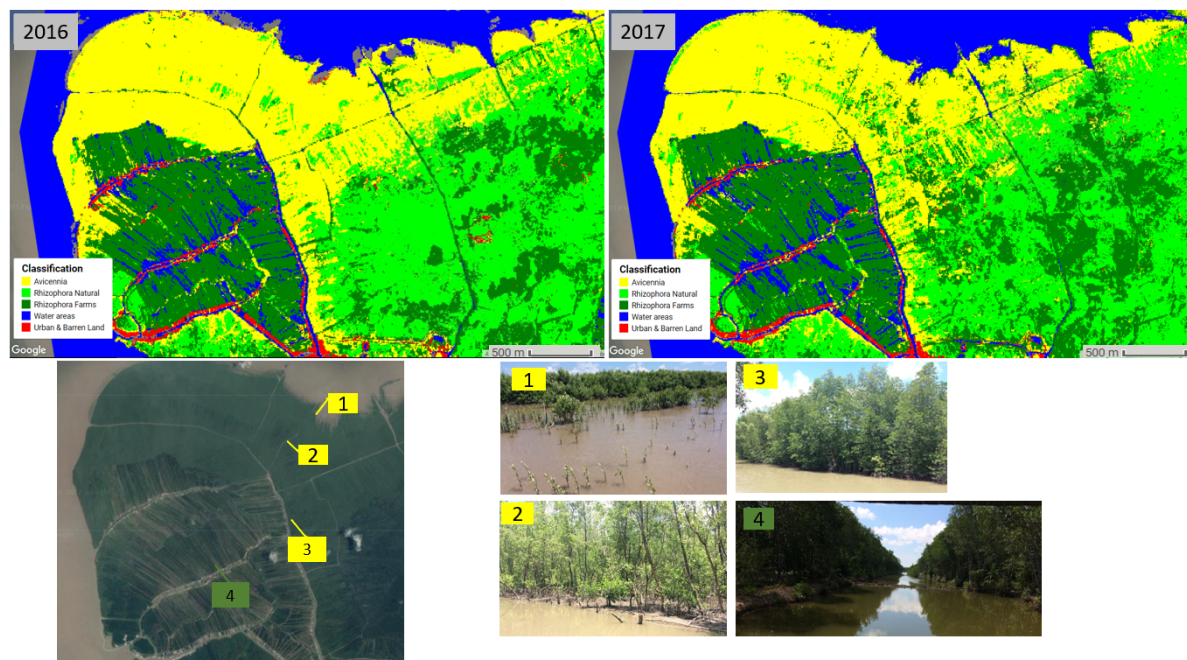


Figure 5.8: Yearly 2016 and 2017 land cover maps with ground truth examples at north coast National Park location

## 5.2. Workflow

In the previous section it has been described how Sentinel-1 and Sentinel-2 are used for the classification of mangroves. Sentinel-2's optical bands, including an extra NDVI band, gives a highly accurate land cover classification of mangroves, as seen in figure 5.2. The highest accuracy is obtained when fusing optical data with the temporal information obtained from the radar backscatter from Sentinel-1 satellite mission. Since this research investigates the best method for discriminating mangroves using remote sensing one of the results is a workflow describing such method. The workflow for classifying mangroves using Sentinel-1 and Sentinel-2 satellite imagery is shown in figure 5.9. A more detailed workflow, including the many different (pre-)processing steps involved, can be found in the appendix (figure A.6).

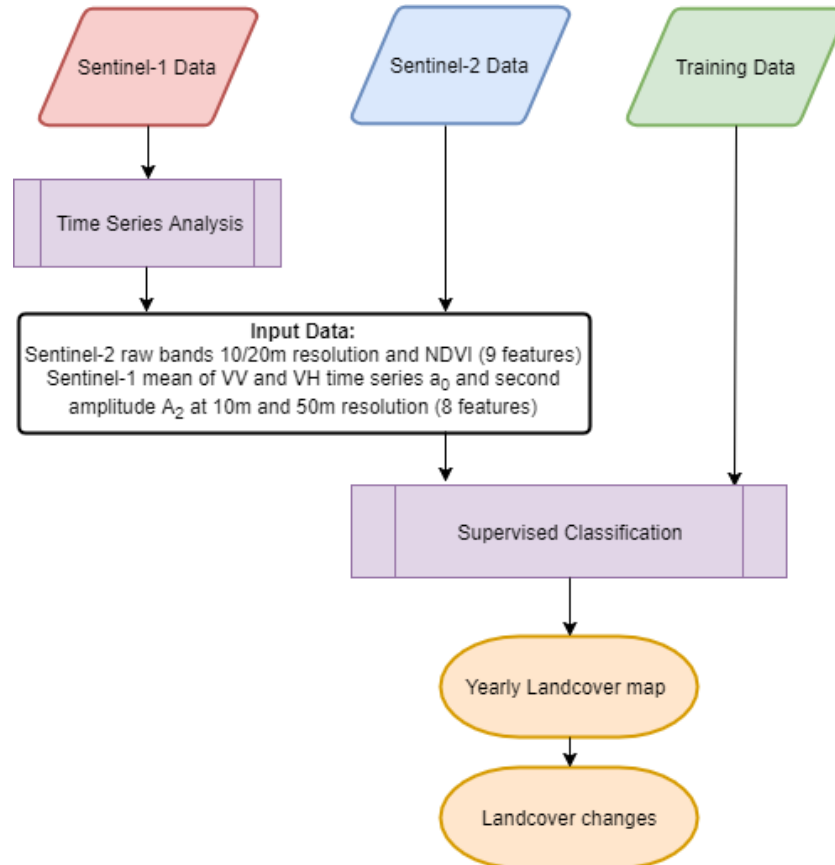


Figure 5.9: Workflow for classifying mangroves using Sentinel-1 & 2 satellite data

### 5.3. Temporal analysis

In the previous sections it has been shown that adding temporal information from the Sentinel-1 radar backscatter data to Sentinel-2 optical data increases the accuracy of the classification results. In this section the results of the temporal analysis, as described in section 4.4, are discussed in more detail.

The temporal analysis is executed for both 2 year time series of VV and VH backscatter. Also both 10 and 50 meter pixel size are used as input because with the extra 50 meter input the noise effect is reduced and more spatial information is gathered. A pixel based three-term Fourier Fit is made resulting in a representation of the mean,  $a_0$ , the first,  $A_1$ , second,  $A_2$ , and third,  $A_3$ , amplitude over the whole Region of Interest (ROI). The second term  $A_2$  shows the most variation, especially for the VV time series and is visible in figure 5.10. All other resulting maps can be found in Appendix A.3. All results show a straight line which is an error strip caused by the stitching of multiple GRD scenes.

In figure 5.10 the canals are visible just as the north-western coastline that consists of high amplitude values along the shore (square 3). Biggest differences in mangrove vegetation are visible within the National Park's inner core (square 2) and the coastline (square 1). Within the National Park the temporal variation is high and decreases towards the coastline. This coincides with the succession of mangroves in Ca Mau Cape that has been found in figure 2.1. Mangroves that are located more inland are more mature and are dominated by *Rhizophora Apiculata* species. Inside the National Park those *Rhizophora Apiculata* mangroves are very old and thus very tall and cause a lot of variation in backscatter between the wet and dry season. One reason might be caused by the water intake of the leaves that is influenced by the amount of rainfall. Vegetation that has a greater moisture content will return more energy than drier vegetation because of the increased dielectric constant (section 2.3.2). Also, mangroves that are located inland are only influenced by high tide levels and thus not always covered with sea water. Therefore those mangroves might be more dependent on the availability of rain water. The difference in water content is larger between dry and wet season than mangroves which have abundant water supply. Along the shorelines where *Avicennia Alba* is the dominant mangrove species the roots are covered in water most of the time. This pioneer mangrove species is much smaller, especially towards the northern shoreline where new mangroves are growing on the accreting mudflats.

**Second Amplitude Term at 50m pixel size for VV backscatter time series**

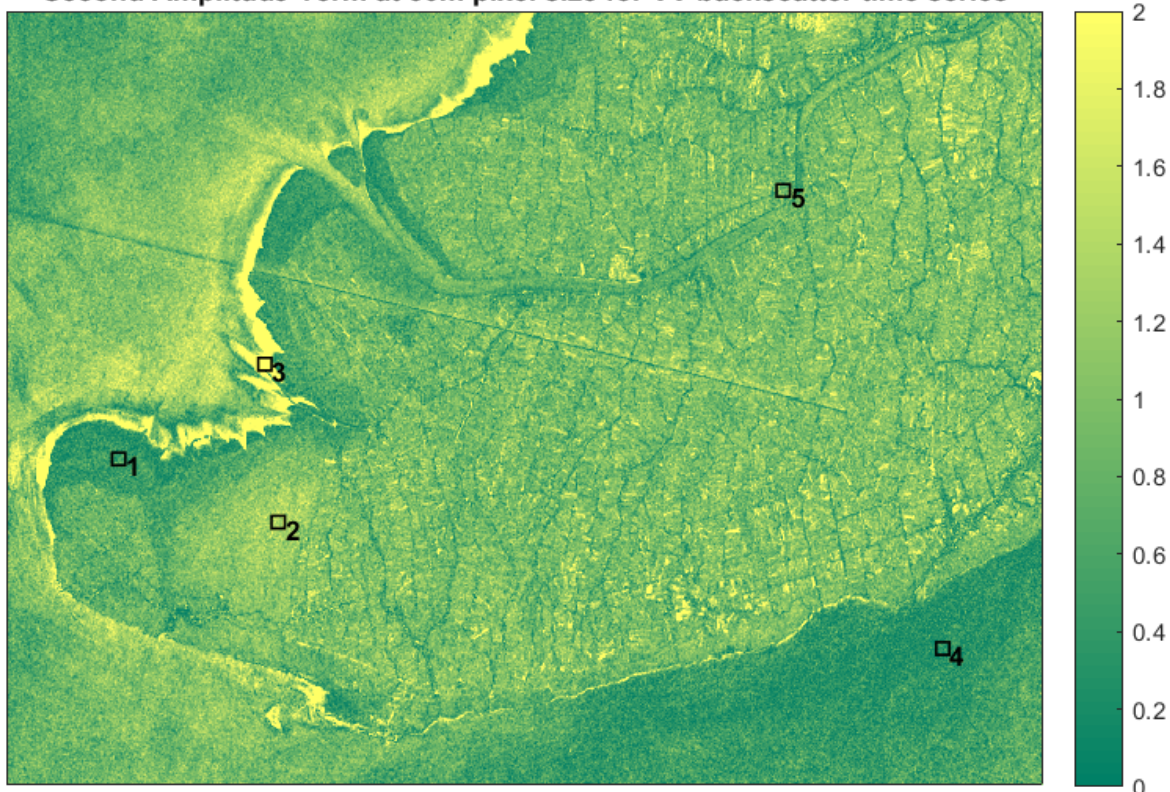


Figure 5.10: Second amplitude term for VV backscatter time series. Black rectangles indicate POI

In figure 5.10 some Places of Interest (POI) are indicated. The time series of both VV and VH backscatter on those places are shown in figures 5.11 and 5.12. The time series, made from the 50 meter pixels data, show the peaks that occur in the sea in yellow and purple. At the POI Sea South the peaks are quite random. Along the northern shore the time series follows a more sinusoidal pattern. The peaks in this time series, visible in both VV and VH, are caused by a high outflow of sediment from the canals towards the sea which is also influenced by the seasonality. Since the backscatter coefficient over sea is much higher than over vegetation the sinusoidal fit causes a very high second amplitude term  $A_2$ . The more random peaks along the southern sea shore are likely to be caused by the roughness of the sea surface waves, causing more or less backscatter (figure 2.6). The time series of the mangrove locations show less variation along the coastline and much more variation in the National Park and the farms, especially for the VV backscatter. From figures 5.10 and A.7, both VV and VH backscatter look to contain the same land cover information. However, figure 5.13 shows that the correlation between the two different polarizations is not very high for the different classes. Therefore, both VV and VH backscatter are used as input for the classification.

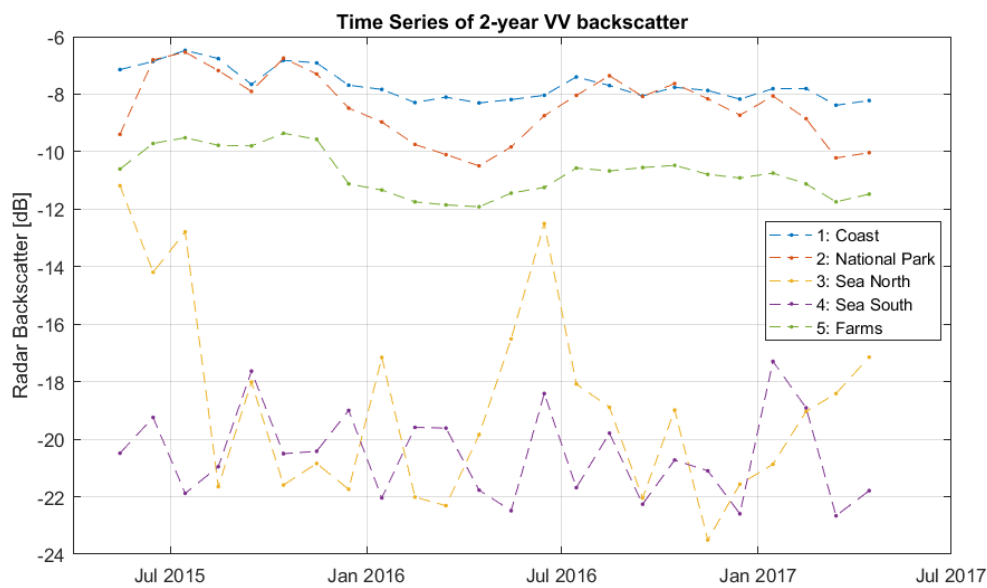


Figure 5.11: VV backscatter time series at different POI

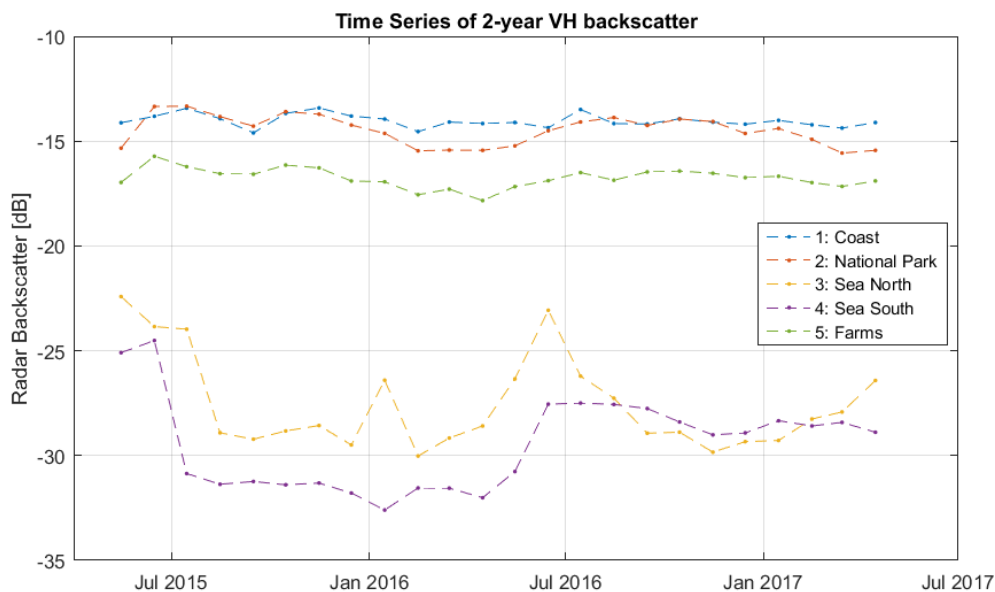


Figure 5.12: VH backscatter time series at different POI

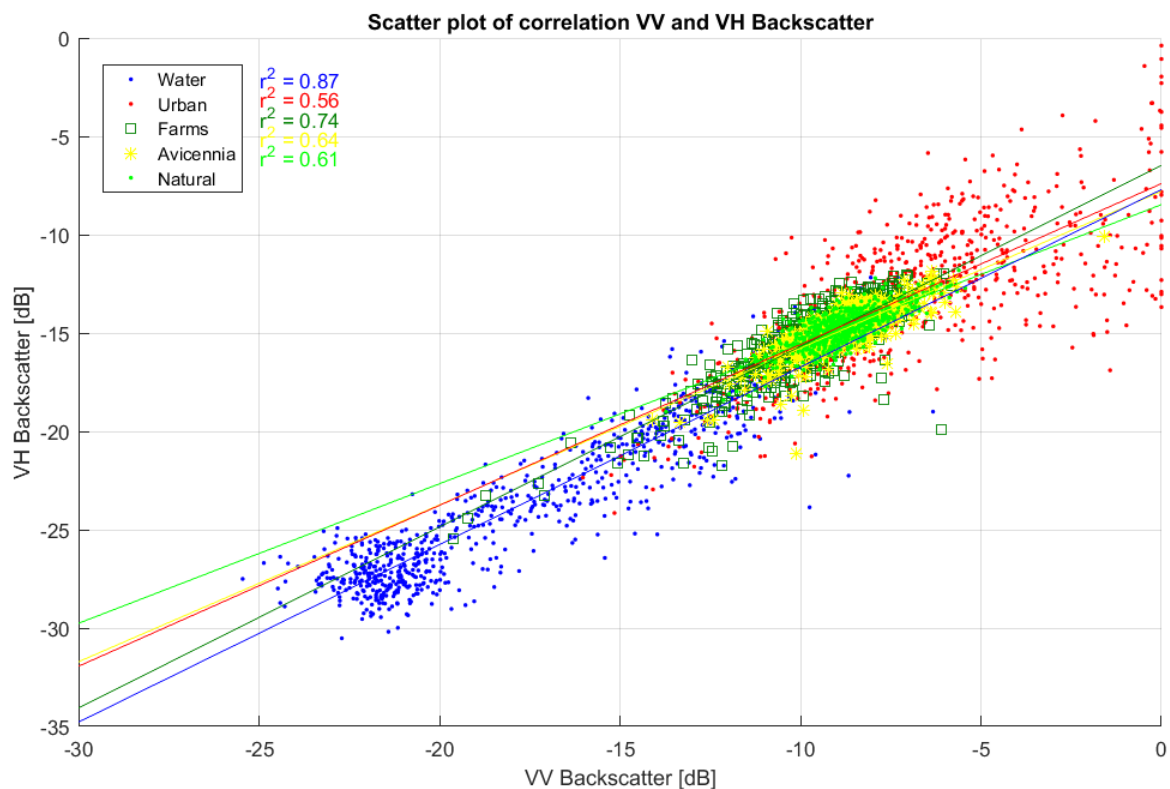


Figure 5.13: Correlation of VV and VH backscatter for each different land cover class

The temporal behaviour on mangrove locations 1 and 2 are shown separately in figures 5.14 and 5.15 including the different Fourier fit results. The other three locations are shown in appendix A.3. The second amplitude term is dominant since the one term Fourier Fit does not cover the pattern yet and the two and tree term Fourier Fit are so similar indicating that the third term does not add any extra information. The seasonal pattern inside the farms and the National Park looks very similar. The seasonal pattern along the coastline looks to decrease in amplitude over the years and is much smaller.

The amplitude of the second term of the Fourier fit contains most information about the different mangrove classes. The differences in amplitude are related to the succession of the mangrove species from the coast towards more inland areas. More mature species show greater differences in backscatter between wet and dry season than pioneer species. This is related to the canopy structure of the mangroves and/or the availability of water during the dry season.

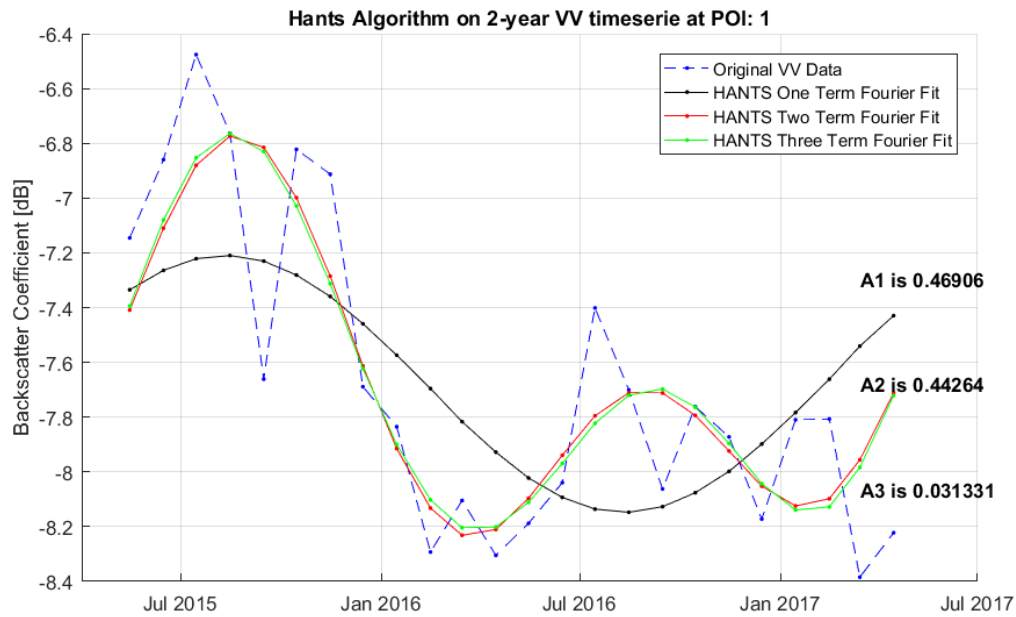


Figure 5.14: VV backscatter time series at POI 1 including different Fourier fits

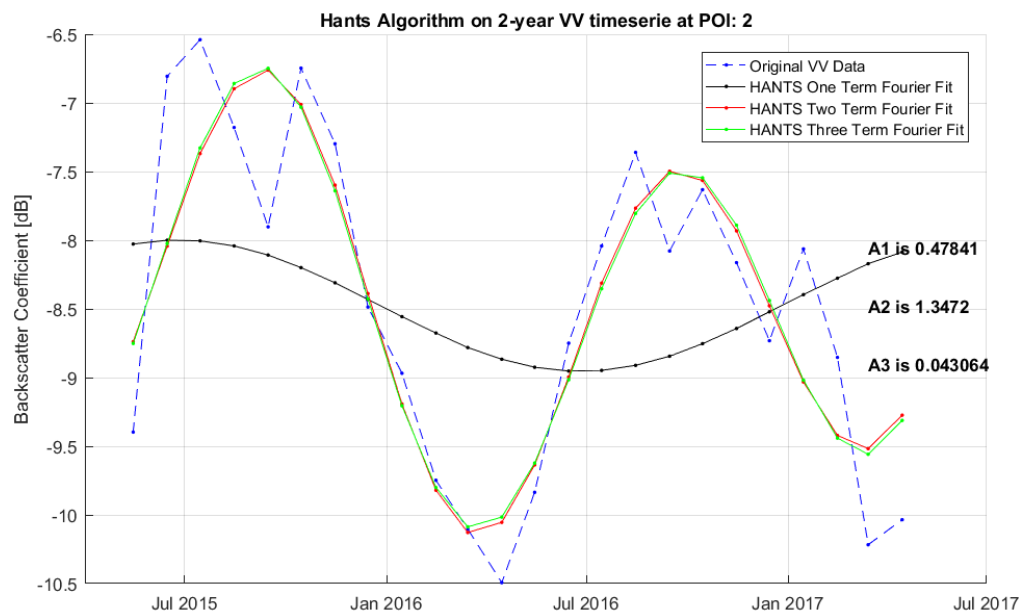


Figure 5.15: VV backscatter time series at POI 2 including different Fourier fits



## 5.4. Unsupervised clustering

The results of the unsupervised clustering indicate the separability of the different classes that were chosen as input for the supervised classification. The clustering algorithm, as described in section 4.2, is executed on the median image of the available Sentinel-2 images and on the median image of both Sentinel-1 and 2 images, as visualized in table 5.1. Both  $k=4$  different clusters and  $k=10$  different clusters are made. The resulting maps can be found in appendix A.4.

The results for four clusters ( $k=4$ ) show a clear distinction between water and densely vegetated area and two mixed classes in between as can be seen in figure 5.17. When comparing the spectral signatures of those clusters to the spectral signatures of the places of interest (POI) from the exploratory data analysis (figure 3.21) and with the spectral signatures of the training classes (figure 4.2) similarities are found. Cluster 0 is exactly the same as the training class water. Cluster 3 is very much similar to the training class of the *Rhizophora* mangrove in natural environment and thus the POI from the National Park. The other two clusters are not exactly the same as one of the training classes. Cluster 1 looks most similar to *Rhizophora* mangroves in the extensive shrimp farms but has less reflectance in the near infrared and more reflectance in the visible wavelength range compared to figure 4.2. The same holds for cluster 2 which has relatively high reflectance in the visible wavelength and low reflectance in the near infrared compared to the training classes and the POI. It seems that the clustering algorithm is dominated by the bands in near infrared wavelengths because here the biggest differences occur compared to other wavelengths. To improve the differentiation between the mangrove vegetation a new clustering is made. This is done by selecting the pixels within cluster 3, the most likely to be mangrove vegetation, from the initial clustering result and cluster those pixels again, see figure 5.16 This way, more detailed variability can be found.

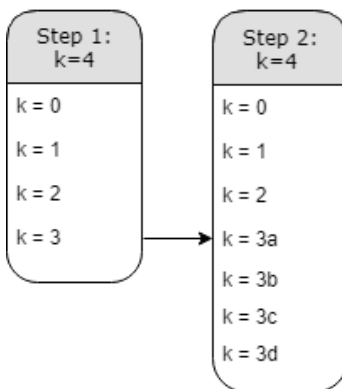


Figure 5.16: Clustering scheme showing the first step making four separate clusters. The second step clusters only the (vegetation) pixels from cluster 3 into four new clusters to better differentiate vegetation

The results of the clustering scheme are visible in figure 5.17 (spectral signatures) and 5.18 (cluster map). Within cluster 3 three very different clusters are found in the near infrared wavelengths. Cluster 3c, brown, has highest reflectance and occurs in the cluster map as random separate patches inside the ROI. It is not well known why those locations have such higher reflectance in the near infrared. Cluster 3a, turquoise, has highest reflectance in the other wavelengths, the visible and the shortwave infrared wavelengths. This coincides with the *Avicennia* mangrove signature from the training input data in figure 4.2. The clustering map shows that cluster 3a is often located along the coastlines. Mostly along the north-western coastline but also along the coastline in the south-east part of the ROI. Cluster 3b covers a very big part of the core zone of the National Park which is found to be similar as some parts of the southern coastline since they are in the same cluster. This coastline is less homogeneous as has been thought, since the variability inside cluster 3 here is quite large. Cluster 3d covers a negligible small area of pixels.

The unsupervised clustering showed that vegetation can easily be separated from other land covers. From step 2 in the clustering scheme, vegetation clustering, it is found that three main differences are separable. At least two are related to species and the third might be related to water content, age or canopy structure.

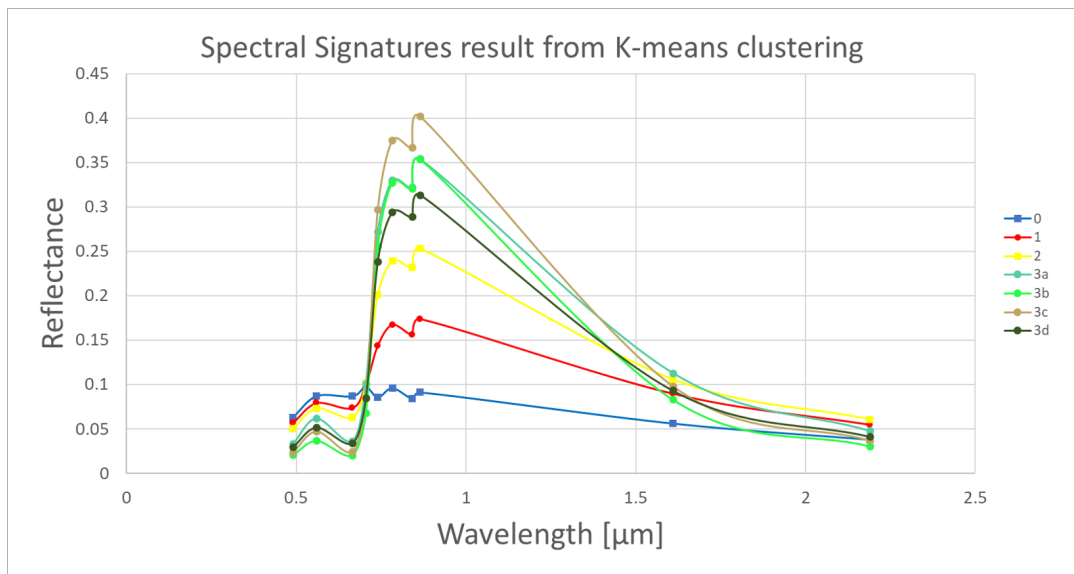


Figure 5.17: Spectral signatures from cluster analysis for Sentinel-2 median image

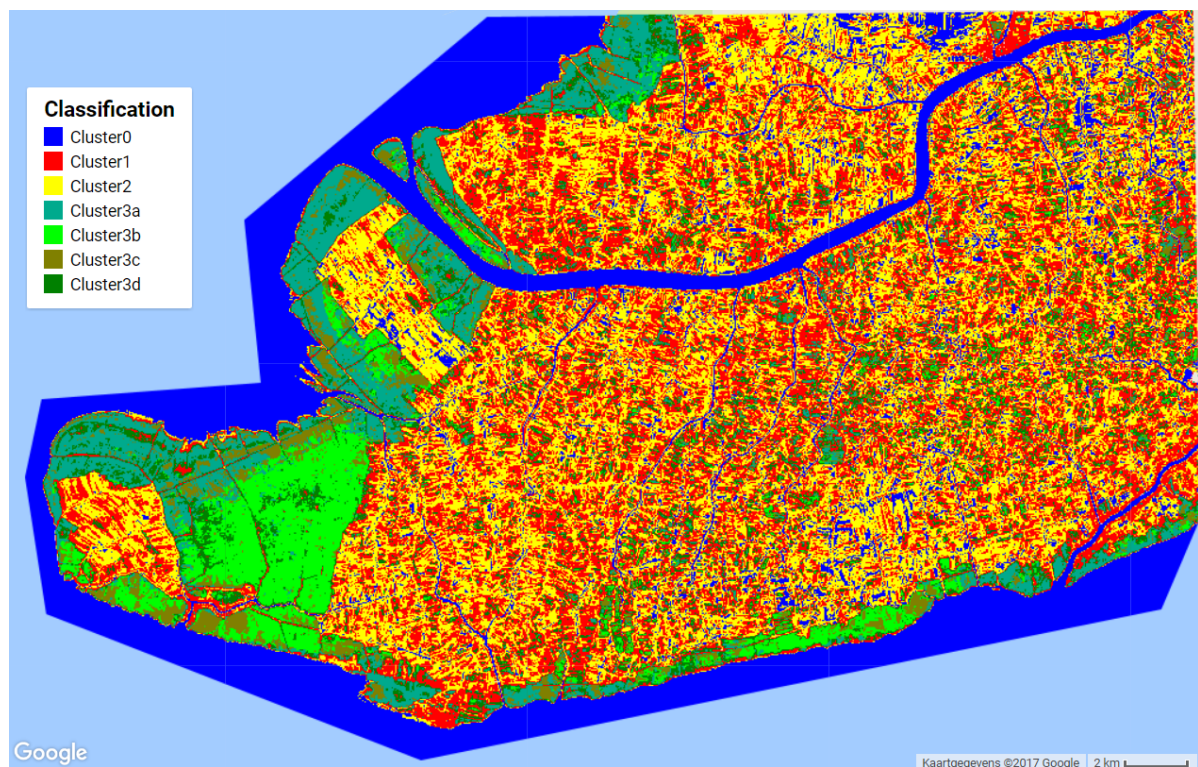


Figure 5.18: Cluster map from cluster analysis for Sentinel-2 median image

## 5.5. Conclusions

This chapter shows the results of the methods that were described in the previous chapter. The results give different insights to obtain the best classification for monitoring mangroves using Sentinel optical and radar satellite data.

- **Unsupervised clustering** results show that vegetation can easily be separated from other land covers. Specific vegetation clustering results show that three main differences are separable: two related to species and the third to water content, age or canopy structure.
- The **temporal analysis** results found that the amplitude of the second term of the Fourier fit contains most information about the different mangrove classes. Differences in amplitude are related to the succession of the mangrove species from the coast towards more inland areas. Mature species show greater differences in backscatter between wet and dry season than pioneer species. This is related to the canopy structure of the mangroves and/or the availability of water during the dry season.
- The final **classification results** consist of yearly land cover maps made from the most common classification result for all single image classifications within one year.
- Using Sentinel-2 optical data combined with temporal properties from Sentinel-1 radar data obtains the highest accuracy for the classification results. An overall accuracy of 87% is reached compared to validation ground-truth data.
- Validation with mangrove cover from 2000 shows a regression of mangrove along the southern coastline of Ca Mau province from 2000 to 2017.
- Different case studies show that the classification works well in areas with extensive shrimp farms. Differences inside shrimp farms between mangroves and water are detected. Since there is no class of other vegetation than the three mangrove types, *Avicennia* is often misclassified as other non-mangrove vegetation. The classification is less reliable inside the National Park when mangrove species are mixed and when there is no ground truth training input available.
- The land cover map of 2017 shows more precise results than 2016. It is expected that this trend continues to improve with the availability of more reliable data from the Sentinel-2 mission.
- A workflow is set up for the classification of mangroves using Sentinel-1 & 2 satellite data and is visualized in figure 5.9.



# 6

## Discussion

In this discussion the different findings of this research are reviewed. This includes difficulties and improvements with data processing and explanations that led to the classification results of the previous chapter. This results will be compared with external mangrove maps and future applications will be discussed.

### 6.1. Data processing

Satellite data processing has been an important part of this research. Several aspects that are addressed during this processing will be discussed.

#### 6.1.1. Cloud masking

From the different methodology and classification results it is concluded that clouds have a big effect on optical satellite imagery. Clouds cause outliers, misclassifications and decrease the availability of data. In this research the automatic Sen2Cor software (section 3.5.2) is used to detect clouds and shadows. Although this is one of the best fully automatic cloud detection techniques [41], the distortive effect of clouds is still not fully removed. Improving cloud detection techniques will also improve the classification results when using optical data. Since radar data is not affected by clouds, processing techniques to obtain reliable land cover information from the backscatter signal are very useful. The temporal analysis results from section 5.3 can serve as a start in this process showing significant differences in mangrove species from radar data.

#### 6.1.2. Google Earth Engine

Google Earth Engine is found to be a fast online platform with increasing amount of functionalities. Also the availability of a catalog with satellite imagery and other geospatial datasets makes that you do not need to download a lot of separate data. Planetary-scale analysis capabilities can be used for a world-wide coverage of land cover classification [23]. However, there are also some drawbacks of using this online platform.

In the beginning of this research the atmospheric correction for Sentinel-2 optical imagery was not yet available in the Google Earth Engine. Therefore this is executed separately via the Sen2Cor software and afterwards imported to the Google Earth Engine. From June 2017 a Python application is published that can be used in the Google Earth Engine API executing atmospheric correction [38].

Executing a classification procedure is possible with a limited amount of classifiers. Some simple, but often used, classifiers such as Maximum Likelihood and Nearest Neighbours are missing. Also, new state-of-the-art classification algorithms such as convolutional neural networks would be a great addition to the available classifiers.

Besides the classification algorithms many other algorithms are available in the Google Earth Engine but the code of those algorithms are hidden. This makes it hard to understand what the exact procedure has been. The classifiers, for example, are not explained with any code except for the corresponding literature from which they are derived. In this research, when using the Random Forest Classifier, the confidence of the classification result was not available, indicating how often the resulting class is noted by the different decision trees. Those hidden algorithms make it hard to investigate what is really happening in the process, to quantify the reliability of the results, to find error sources and make adaptations in the processes.

### 6.1.3. Fieldwork campaign

In figure 3.2 was seen that the available ground-truth data that is used as training input and for validation was not equally displaced over the ROI. This has caused the classification to perform well on locations with available training input (case study 2) and worse on locations with few training input (case study 3). Also the distribution of ground truth locations was not equal over the different classes. Therefore the confusion matrix needed to be corrected for the amount of available validation pixels. In this confusion matrix was seen that especially *Avicennia Alba* mangrove species had the least ground-truth. For a new fieldwork campaign in this area this species should be investigated more and also a better spatial distribution must be obtained.

## 6.2. Classification results

To obtain the best classification results that were shown in section 5.1 many decisions and some assumptions are made. Single images that are used for the yearly land cover maps are discussed and the different data input combinations.

### 6.2.1. Single images

As described in the methodology section the classification algorithm is executed for each available monthly image. To get insight on the result of these 'single image classifications' some examples are shown in figure 6.1. The input of these classifications is one Sentinel-2 image of the specified date.

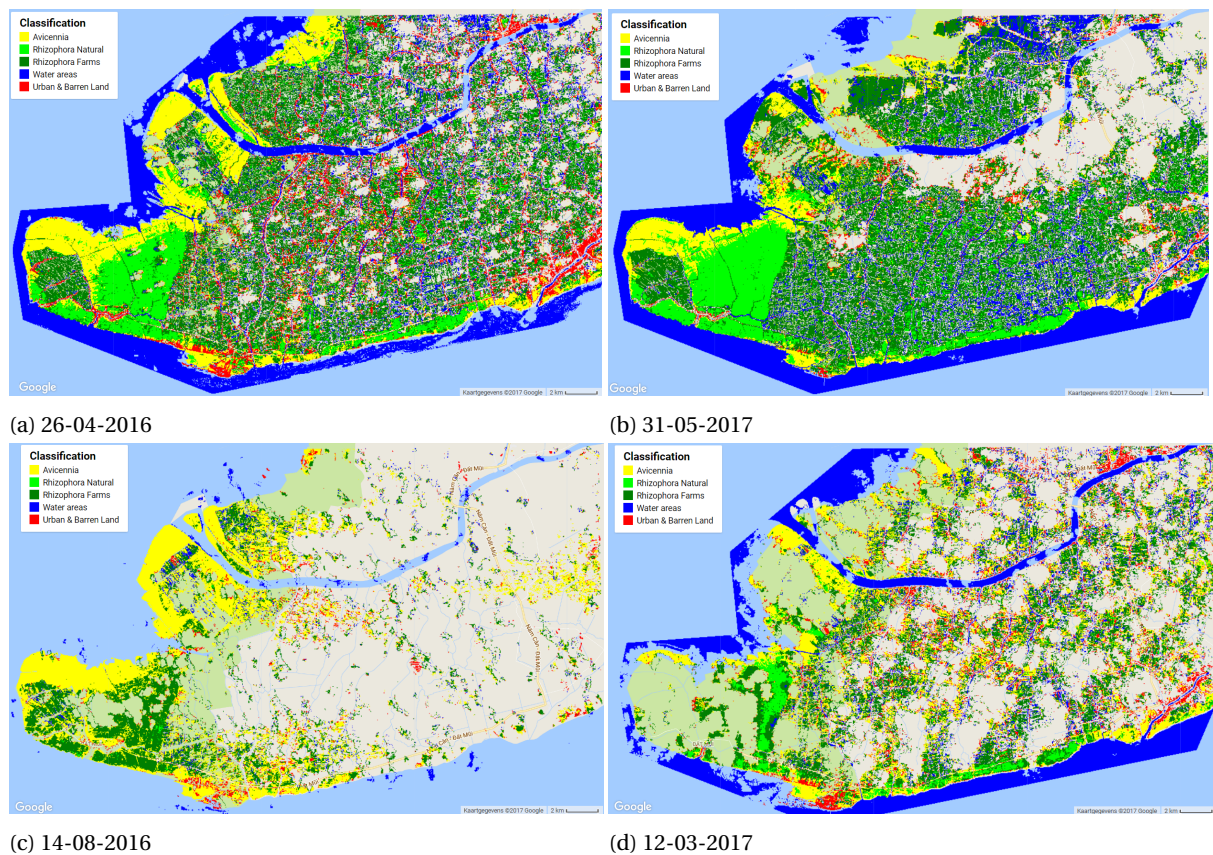


Figure 6.1: Single Sentinel-2 image classification results

It can be seen that the clouds, although they are mostly masked, have a very big effect on the result. Along the edges of the cloud mask often pixels are classified as Urban class. This is caused by the fact that spectral properties of urban land cover are similar to those of clouds. The image of August 2016 has so much masked pixels that the few pixels left are very unlikely to be correct. The class of natural *Rhizophora* mangrove in the area of the National Park is totally absent. This kind of images, just as the image of June 2017 which is very similar to August 2016, have caused the outliers that were visible in the time series, for example in figure 3.23.

Sometimes the results of the single images are very good. For example the southern part of the image

from May 2017. Here, the Rhizophora Natural class is clearly visible along the southern coast and also in the National Park. There are no weird alternations between the different Rhizophora classes inside this National Park. This single image would give a good monthly land cover product in this part of the ROI. Nevertheless, the ROI is not completely covered and those images are very scarce with this date being the only one between April 2016 to September 2017. Although single images are able to give good classification results it is chosen to look into yearly land cover maps. With this method it is assumed that phenology of the different mangrove species does not influence the land cover results. Phenology is proven to give useful information about the different mangrove species from the Temporal Analysis in section 5.3. Still, mangroves do not change in species within a year so therefore yearly land cover maps are still useful to detect mangrove changes in general.

### 6.2.2. Different data input

From the results was found that the accuracy is highest for a classification that used Sentinel-2 optical data fused with the temporal information of Sentinel-1 radar data. To get insight in the differences that occur for the five different data input combinations a comparison is made. All yearly land cover maps from 2017 from the different data input combinations can be found in appendix A.1. Table 6.1 shows the percentages of pixels assigned to the five different classes for each data input. The table shows that the five different data input scenarios result in big differences in classification results, as is also visible in the land cover maps. Radar and optical satellite data input are very different in assigning pixels to different classes.

Class \ Data input	Sentinel-2	Sentinel-1	Sentinel-1 & 2	S1temporal	Sentinel-2 & S1temporal
Rhizophora Farms	47.17%	23.69%	49.40%	28.45%	44.85%
Avicennia	12.41%	1.95%	11.32%	8.62%	10.72%
Water	29.00%	34.89%	28.09%	34.44%	33.02%
Urban & Barren Land	4.51%	21.09%	4.47%	5.81%	4.38%
Rhizophora Natural	6.91%	18.39%	6.72%	22.68%	7.03%

Table 6.1: Overview of percentage of total pixels per class

#### 6.2.2.1. Radar input

Table 6.1 shows that both the Sentinel-1 raw radar input as the Sentinel-1 temporal information overestimate the Natural Rhizophora class. The percentage pixels in this class is much higher compared to the data input scenarios that include Sentinel-2. This is at the expense of the other mangrove classes Rhizophora Farms and Avicennia. Especially the raw Sentinel-1 data input scenario shows only Natural Rhizophora, Water and Urban & Barren Land. Clearly, there's no possibility discriminating the different mangrove species using raw radar data and all mangrove pixels are assigned to Natural Rhizophora. Avicennia even has only 2% of the pixel percentage. The result of Sentinel-1 temporal is already better able to discriminate Avicennia versus Rhizophora mangrove species. For example the difference between pure Rhizophora and pure Avicennia is found in the northern coast of the National Park. This was already explained in the temporal analysis results in section 5.3. Still, the Natural Rhizophora class is much overestimated at the expense of Rhizophora Farms. This is especially visible if we zoom into an area with mostly extensive shrimp mangroves, see figure 6.2. This is the same location as case study 1. showing the land cover result of 2017 for the different data input combinations. The Sentinel-1 input shows the combination of only Natural Rhizophora, Water and Urban & Barren Land. The Sentinel-1 temporal input shows some Rhizophora Farms but still overestimates the Natural Rhizophora compared to the Sentinel-2 input result. Also the effect of the combination of 10m resolution and 50m resolution pixels is visible. This causes a more 'blocky' classification output.

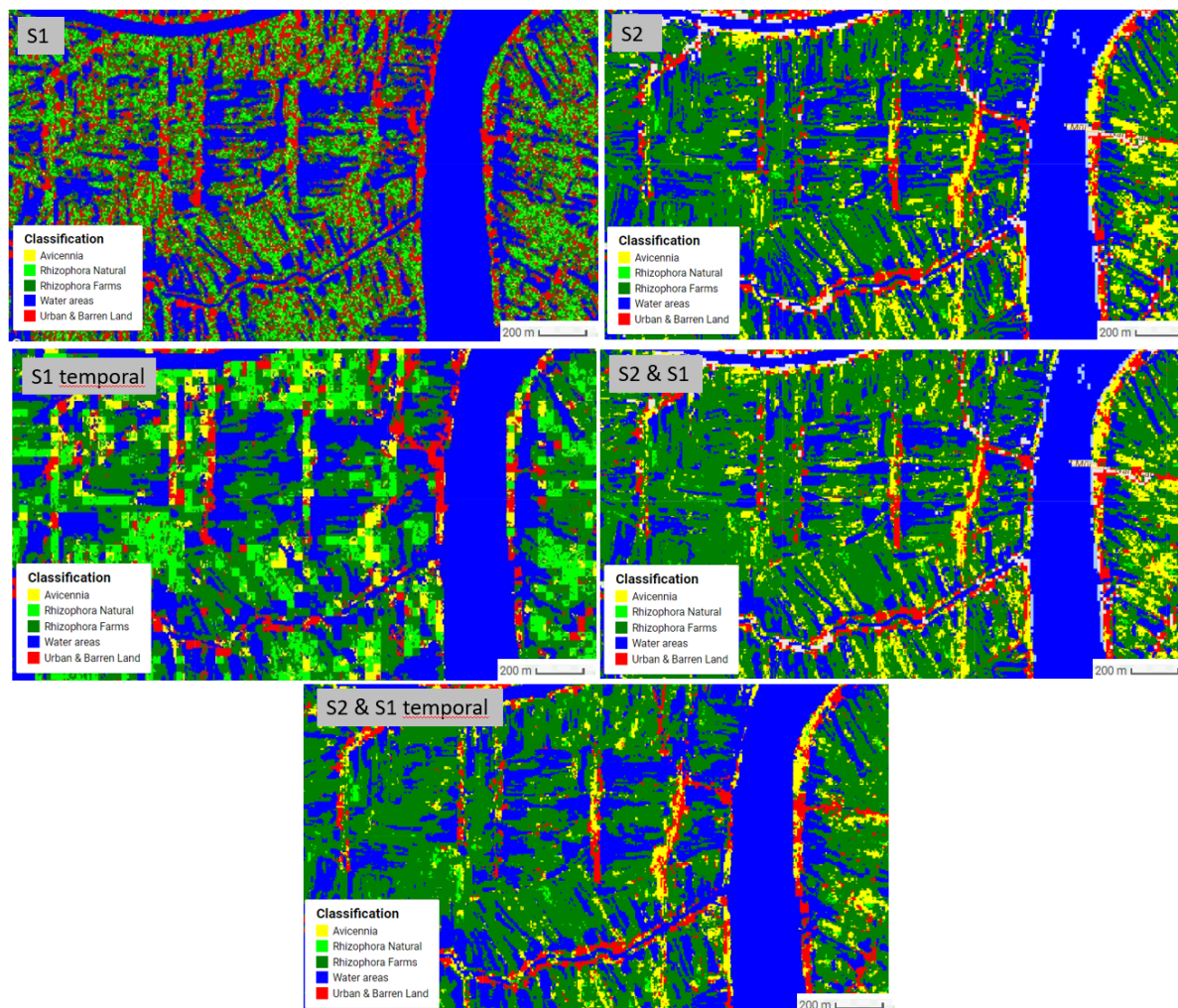


Figure 6.2: Land cover result in 2017 from different data combinations zoomed into location with extensive shrimp farms

#### 6.2.2.2. Optical input

From figure 6.2 it can be seen that all classification results including Sentinel-2 optical data are very similar. Table 6.1 shows the similarities in percentage per class for all data inputs including Sentinel-2. The unsupervised clustering results (section 5.4) already showed that optical data is dominant in differentiating clusters compared to radar data. The only drawback of optical data is the presence of clouds, which is visible by the variation inside the National Park for all land cover maps that include Sentinel-2 optical data input. From the overall classification accuracy's (figure 5.2) it is found that adding raw radar data to optical data does not improve the classification accuracy at all. The best data fusion is with the temporal information of the radar with optical input which improves the overall accuracy to 87%.

### 6.3. Comparison with external research results

From the overall accuracy results is found that the final classification is able to separate the pre-defined classes with 87% accuracy. This is a high value compared to other research in monitoring mangrove ecosystems using remote sensing data [32], especially when incorporating differences within the mangroves. Separating mangrove from non-mangrove can reach accuracy's above 90% but when including different species such high accuracies have not been reported before [32, 57]. Therefore, it might be questioned if the differences found within the mangrove vegetation are indeed related to the different species. The spectral signatures of the pre-defined classes coincide with the spectral characteristics found from a field spectrometer for the same mangrove genus, namely *Rhizophora* and *Avicennia* (figure 2.9). The radar backscatter coefficient from C-band radar increases with different growth stages (figure 2.11), especially VV-polarization continues



to increase to a higher biomass saturation threshold, in other words, to more mature stage of the mangrove forest. Those properties in the classification ensure that the different classes are indeed caused by the different species of mangroves.

Furthermore, a distinction it is made between mangroves from *Rhizophora* species in different environments: the protected forest called natural environment and the extensive shrimp farms. This separation has been done before in many other researches, where classes are assigned to the percentage of mangrove in relation to water [32, 57]. Examples of some of those results are found in figures 2.12 and 6.3. The classification result below shows the same class of 'pure mangrove' at the locations of our *Rhizophora* Natural and *Avicennia* classification results. Mixed mangroves (31-69% mangrove) and most aquaculture ( $\leq 30\%$  mangrove) locations coincide with the class *Rhizophora* Farms. Also both our classification results and figure 6.3 show the settlement/urban area along the canals.

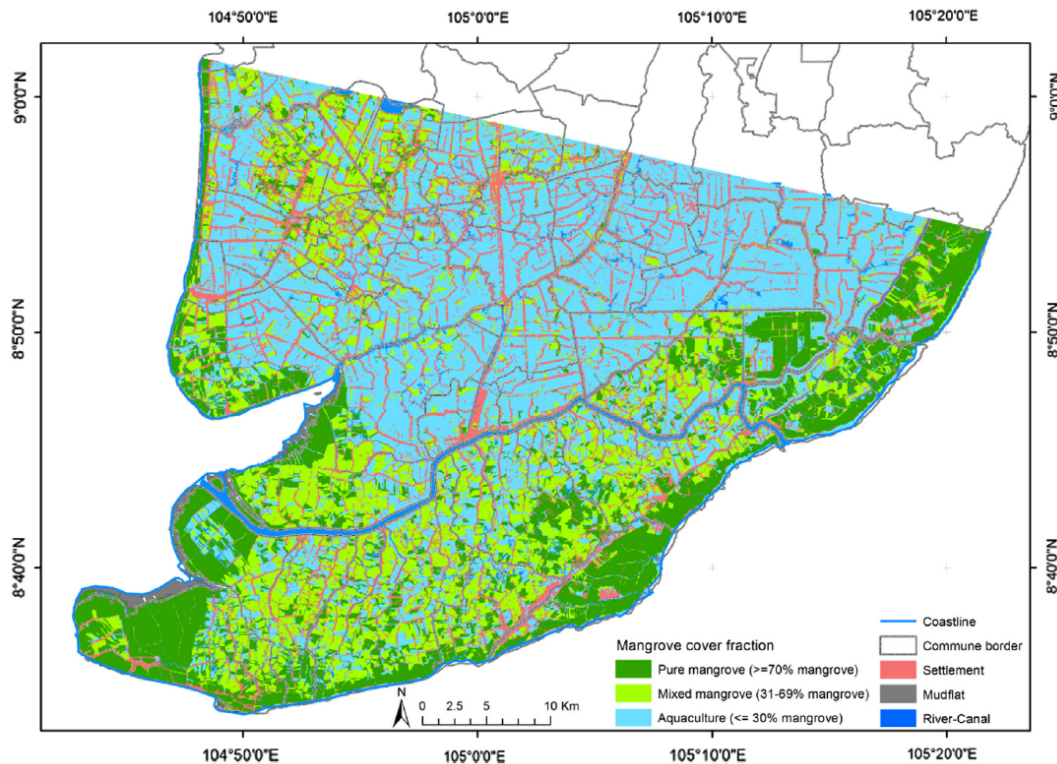


Figure 6.3: Classification results from Vo et al [58]

Another external classification result which includes both different mangrove species classes and mixed mangrove/water and aquaculture classes. The classification of Kuenzer et al [32] is made from a fusion of optical (SPOT) and radar (TerraSAR) data. The land cover 'mixed mangroves' occurs at a location where the classification results in case study 5.8 showed that discriminating the two mangroves species was very bad. This is thus caused by a mixing of the two different species which makes it hard to choose the best result from the available pre-defined classes in our classification algorithm.

Overall, the classification results from Sentinel satellite imagery are found to be reliable in extracting mangrove properties from a combination of optical and radar data. The two most common mangrove species in this district are well separated and also the separation of 'pure' mangroves versus 'mixed' mangroves is successful. Those two separation are reached with high accuracy (87%) which is a new achievement in the field of mangrove remote sensing. Other research results never mentioned the species in the mixed mangrove and aquaculture areas. Our fieldwork campaign showed that the majority of mangroves in those fish and shrimp farms are *Rhizophora Apiculata* species, since they are growing fast and thus easily cultivated. It is assumed that all 'mixed' mangroves from external classification results is thus *Rhizophora* mangrove species.

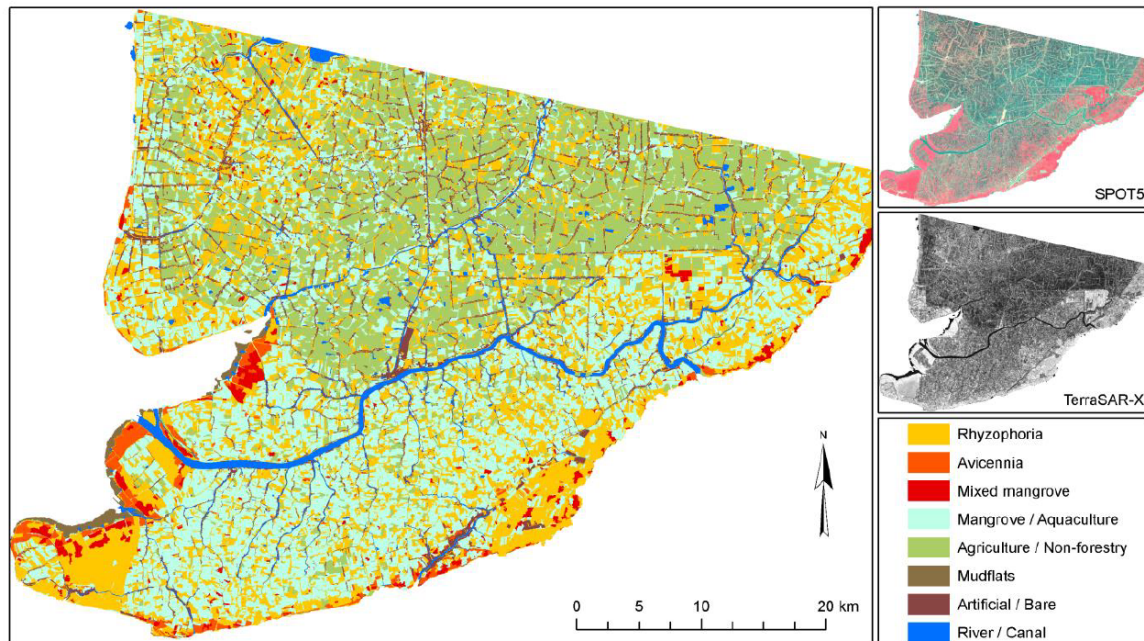


Figure 6.4: Classification results from Kuenzer et al [32]

## 6.4. Future applications

Section 2.1 extensively described the different reasons why we need to study mangroves in the Vietnamese Mekong Delta. The main reason why mangroves need to be monitored was its need for coastal protection. Mangroves are found to be a natural protector of erosion with attenuating waves, collecting and stabilizing sediment supply and reducing storm surges. Figure 5.4 showed the ability of the classification result to detect areas of coastal erosion. Between the reference mangrove cover from 2000 erosion is visible towards our 2017 resulting land cover product. With future yearly land cover products a more timely investigation of areas affected by erosion can be obtained.

Another application of the results is a governmental approach. In 2008 ~70% of Ca Mau's total land was occupied by semi-extensive shrimp farms. In the Introduction was already mentioned that according to provincial regulations farmers must have 60% of their total area to be mangroves and 40% aquaculture [7, 21]. An object-based classification approach by Vo et al already tried to make a quantitative estimation of mangrove fractions within the aquatic shrimp farming systems [57]. A quantitative estimation of mangrove fractions is not feasible with the results from this research but changes within shrimp farms are possible to detect. The classification results show clear land cover changes in areas with shrimp farms, as discussed in the different case studies in section 5.1.5).

A more broad future application of the method from this research is classifying areas outside the ROI. Not only other areas but also nationwide coverage or even worldwide coverage. With the use of Google Earth Engine a global processing tool is available. As soon as a reliable atmospheric correction will be available for Sentinel-2 data, the Google Earth Engine consists of nationwide and even worldwide coverage of optical and radar data from Sentinel. The processing skills of Google Earth Engine are big enough for world wide analysis. Examples are the Global Forest Change map or the Aquamonitor [10, 24]. Still, it is not very likely that the method is directly applicable to obtain reliable world wide land cover maps. First of all, the used ground truth data covers only a tiny amount of global area. Ground truth data of mangroves should be obtained for a much bigger area. Also, the classes in this research are chosen according to the most common land covers in the ROI, especially the two dominant mangrove species: *Rhizophora Apiculata* and *Avicennia Alba*. For nationwide and worldwide coverage much more mangrove species are present. *Rhizophora* and *Avicennia* were relatively easy to separate but for other mangrove species the differences are less significant [29]. Global separation of mangroves species seems very hard. However, monitoring mangroves in general is feasible. In the section (4.5) the Mangrove Forest of the World (2000) was already mentioned. This database is made from Landsat images with supervised classification techniques and shows the global coverage of mangroves in 2000.

Although many improvements are needed it is possible to create such mangrove cover map using Sentinel imagery in the future. Both Sentinel-1 as Sentinel-2 mission consists of satellites that are designed to stay active for at least seven years. The future radar satellites Sentinel-1/2 C& D will extend the operational monitoring component of Copernicus at least until the end of 2030 [13, 14]. With this continuous flow of data much larger time series can be created than used in this research. This makes the extraction of temporal information more reliable. Also, changes in time series are easier to detect. With more available optical images also temporal information from NDVI time series can be used as an input feature in the classification. More simple, the availability of both Sentinel-2A and 2B makes the revisit time much shorter and the chance of cloud-free data much larger. This improvement can be incorporated very soon. The first images of Sentinel-2B are available since September 2017.

## 6.5. Conclusions

This chapter discusses the different results that were described in the previous chapter. Not only the classification results but also the data processing, a comparison with external results and future applications were discussed.

- During the **data processing** some problems are addressed. Improving cloud detection techniques and a better distribution of the ground truth training data from the fieldwork campaign are simple ways to improve final classification results.

Google Earth Engine has great potential but some drawbacks were found. The atmospheric correction for Sentinel-2 data must be further implemented, just as new classification algorithms such as convolutional neural networks. The codes of the Google Earth Engine algorithms are hidden which makes it hard to qualify the outcomes, to find error sources and make adaptations.

- The yearly land cover maps of the final **classification results** are made by single image classifications. Those single images results are not covering the whole ROI due to clouds but at the locations without clouds the transition between the two mangroves species along the edges of the National Park is much clearer. Unfortunately cloud free images are very scarce.

It is not possible to separate different vegetation types using only raw radar backscatter. The radar temporal data is performing better in differentiation the two mangrove genus, *Rhizophora* and *Avicennia*, but not at detecting the mixed farms with mangroves. The temporal data is a valuable addition to raw optical data by increasing the classification accuracy from 83% to 87%.

- Compared to **external mangrove classification results** the reached accuracy of 87% is high since it includes both a mixed mangrove/aquaculture class and is able to separate two different mangrove species. The class 'Rhizophora in extensive shrimp farms' coincides with mixed mangrove (31-69% mangrove) and mangrove/aquaculture classes of external results where the mangrove species is not mentioned. External research results included a class where *Rhizophora* and *Avicennia* genus are mixed, solving the problem of confusions that occur in this research.
- **Future applications** include monitoring longer yearly land cover maps to investigate the areas affected by coastal erosion. Changes inside shrimp farms can be detected with the availability of yearly land cover maps. A wider application is making this method feasible for classifying mangroves on nationwide or global scale. For this, a better investigation of global mangrove species together with global ground truth information is needed. The processing capabilities of Google Earth Engine are present just as future denser time series of Sentinel-1 & 2 when all satellites are fully operational.



# 7

## Conclusions and recommendations

From the results and discussion of the previous chapters, conclusions and recommendations are made. Conclusions are given in section 7.1, answering the different research questions. Recommendations for further research are found in section 7.2.

### 7.1. Conclusions

**The most important conclusion from this research is that combining radar and optical satellite data from Sentinel-1 and 2 satellite missions gives classification results for discriminating mangrove types with 87% accuracy compared to ground-truth data.**

Recalling the main research question:

*What is the best method for discriminating mangrove types in Vietnam using radar and optical satellite remote sensing?*

Sentinel-1 radar and Sentinel-2 optical satellite missions are chosen as remote sensing data since they are free to use and they have both good spatial and temporal resolution. Unsupervised clustering showed the separability of different mangrove densities and also between the mangrove species *Rhizophora Apiculata* and *Avicennia Alba* with different reflectance properties. Temporal analysis of radar backscatter showed that yearly variation is different according to the spatial succession of mangroves. A fusion of optical data with the temporal information of the radar data is found to be the best input for discriminating mangrove and its properties in the area of Ca Mau cape in Vietnam. Classification results of 87% overall accuracy are obtained with this method compared to validation ground-truth data. Case studies show that the classification results can discriminate mangroves in areas with extensive shrimp farms. The classification is less reliable when mangrove species are mixed and when there is less ground truth data available.

#### 7.1.1. Research questions

Many different steps are taken towards this best method that is set-up for discriminating mangrove and its properties. Therefore subquestions of the research question were made and answers can now be provided.

##### 1. *Why do we need to study mangroves in Vietnam?*

The main reason for studying mangroves in the Mekong Delta is the need for coastal protection in this highly vulnerable area where mangroves contribute by reducing wave height and storm surges and holding sediments to protect erosion. Conservation of flora and fauna is also an important driver to monitor mangrove. Finally, ~70% of Ca Mau province's total land was occupied with extensive shrimp farms in 2008, where 60% of the area need to consist of mangroves. Monitoring these mangroves is important to check the regulations and check changes towards intensive shrimp farming.

2. *What are the unique properties to discriminate mangroves?*

Mangroves are found in saline coastal environments and can be recognized by its unique roots and the canopy structure of the different species. Close to the shoreline are shrubs with small leaves that are pioneer species such as *Avicennia Alba*. More inland grows the more mature *Rhizophora* genus that grows much taller with a denser canopy of thick and dark leaves, having the distinctive stilt roots (figure 2.1).

3. *Can we extract those unique properties from satellite imagery? How?*

4. *What differences in mangrove types can be extracted from satellite imagery?*

In optical satellite imagery mangrove vegetation is easily discriminated from other land covers by its high reflectance in the near infrared (NIR) and its low reflectance in the red and green wavelengths (figure 2.3). Dense mangrove forest give higher NIR response than sparse forests. Different mangrove species *Rhizophora Apiculata* and *Avicennia Alba* are easiest to discriminate among different mangrove species (section 2.4.2). In C-band radar imagery the signal is dominated by volume scattering. Dense canopies reflect more backscatter than sparser canopies. Due to multiple scattering in the canopy the signal can easily get saturated. VV-polarization continues to increase to a higher biomass saturation threshold and contains good correlations with canopy parameters (section 2.4.3).

5. *How can space-borne optical and radar data be used for the classification of mangroves?*

Supervised classification is the machine learning task of assigning classes to (satellite) data input where ground-truth data is used as training input for assigning the different classes. From the fieldwork campaign executed in the ROI five different classes are chosen. To investigate the separability of those classes an unsupervised clustering is executed. Unsupervised clustering showed the separability of different mangrove density and also between the mangrove species *Rhizophora Apiculata* and *Avicennia Alba* with different reflectance properties. Temporal analysis of radar backscatter showed that yearly variation is different according to the spatial succession of mangroves. With this data input a workflow is set-up towards classifying mangroves that is visualized in figure 5.9.

6. *How to validate the accuracy of the classification results?*

Validation is done by calculating the confusion matrix for both training as well as validation data. From these confusion matrices overall accuracy's are compared for the different classification input: Sentinel-2 (S2) optical data, Sentinel-1 (S1) radar data, S1 and S2 combined, Sentinel-1 temporal information and finally S2 and S1 temporal information. A fusion of optical data with the temporal information of the radar data is found to be the best data input for classifying mangrove. The classification outcome is also validated by a confidence map showing the stability of the classification for different images within a year. Finally, the resulting land cover map is compared with existing mangrove cover dataset to detect discrepancies and land cover changes (section 4.5).

7. *How can the quality of the classification result be improved by combining the optical and radar data?*

Raw Sentinel-1 radar was not able to separate different mangrove vegetation. This resulted in an overall accuracy of 50% compared with validation data. Calculating extra features as input for the classification is a way of improving the quality of the classification results. For Sentinel-1 radar data a temporal analysis is done at a two year time series to extract the mean and the amplitude of the temporal variations. The differences in amplitude of those variations coincide with the spatial succession of mangrove species from the coastline towards more inland areas. Making a classification with these temporal information from the radar signal improved the classification already to 78%. Single Sentinel-2 raw optical bands, with extra NDVI feature, already showed 83% overall accuracy. The quality of the results does not improve by adding raw radar bands, since this does not add any information. Combining the optical bands with the temporal information of radar does improve the classification accuracy towards 87% (figure 5.2).

8. *Can this method be used to reach nationwide or even worldwide coverage? How?*

The method is implemented in Google Earth Engine which is proven to be an online platform able to perform nation and worldwide analysis. Also the future denser time series of Sentinel-1 & 2, when all satellites are fully operational, are promising for reaching wider coverage. However, the defined method is not expected to directly perform well on global scale since the supervised classification demands ground truth data. In this study a small Region of Interest in Ca Mau province in Vietnam showed good the results because the variability in different mangrove types was relatively small. Obtaining reliable results for bigger coverage demands much more ground truth data and a better investigation on the different mangrove species world wide (section 6.4).

## 7.2. Recommendations

During this research possible topics are addressed that could improve the remote sensing of mangrove properties. They were beyond the scope of this research but are recommended for further research. Section 7.2.1 discusses improvements on the use of the data and the different methods from this research. Section 7.2.2 states topics to continue further research in the field of mangrove remote sensing.

### 7.2.1. Improvements

Some recommendations on the data processing have already been discussed in section 6.1 of the discussion. Those were better cloud detection techniques for Sentinel-2 optical data, better distribution of ground truth data to obtain more reliable training input and future improvements that could be implemented in the Google Earth Engine. This sections recommends more general improvements for both radar and optical data, using also data outside the Sentinel satellite missions and for improvements regarding the methods that are used.

#### 7.2.1.1. Radar data

The radar data from Sentinel-1 in the Google Earth Engine is available in Level-1 GRD scenes. Those scenes are pre-processed with the Sentinel-1 Toolbox including averaging of multiple scenes to create approximately square resolution pixels with reduced speckle. This is at the cost of reduced geometric resolution. A higher resolution can be useful for a better differentiation between the vegetation and water in the shrimp farms. Therefore the investigation of using Level-1C Single Look Complex data instead of GRD scenes is recommended.

The temporal information that is found from radar time series in section 5.3 showed very interesting results regarding mangrove types but also along the coastlines. However, it is not fully understand how those results are explained. Improvements of this understanding can be obtained by using longer wavelengths radar, that are described in section 2.4.3, giving different information on the canopy, trunk and ground conditions of the mangrove types.

#### 7.2.1.2. Optical data

This research has chosen to only use Sentinel-1 and 2 data as input source. Reason for this was the good temporal and spatial resolution but most of all that it was free available. Of course there are more free available such as Landsat and MODIS. Their spatial resolution is worse than Sentinel, especially MODIS data with only 250m pixel size. However, when adding those satellite data to the Sentinel-2 optical data the time series can be very much expanded. This way also temporal information of optical time series can be implemented in the classification workflow.

The seasonal differences of mangroves are mentioned only short in section 3.4. In this research it is chosen to make yearly land cover map that are assumed independent of seasonal differences since the availability of data was too scarce to make a monthly land cover map. Better optical time series together with a more detailed investigation on the phenology of mangroves can help to get a better differentiation between the mangrove species.

#### 7.2.1.3. Methods

From the workflow that is set-up in this study, visualized in figure 5.9, two methods were implemented. The time series analysis is done using HANTS algorithm (section 4.4) executed in Matlab (section 3.5.3). In the discussion section 6.1.2 is already mentioned that implementing the Sen2Cor algorithm for atmospheric correction into the Google Earth Engine (GEE) would be good addition to GEE. The same holds for a time series analysis. This can be the HANTS algorithm that fits the function according to Fourier series but also smooths the data, removes outliers and fills the gaps of missing data. However, a time series analysis that only does the Fourier fit might be already enough to obtain the same information on the seasonal variations. It is recommended that those two additions are investigated to improve the temporal analysis method.

Improvements on the supervised classification are also mentioned in the discussion. Those are the use of new state-of-the-art classifications algorithms such as neural networks and the extension of available ground truth training data input to better train the classifier.

### 7.2.2. Further research for future applications

In section 2.1.1 the mangrove's potential for coastal protection was described. Also the mangrove rehabilitation programmes from GIZ were mentioned, where the different steps of the programme include planting

the mangroves, a site assessment and finally monitoring [49]. It is recommended to further develop mangrove classification remote sensing methods to fully be implemented in such programmes. The resulting method from this research can serve as a starting point with more and more data from Sentinel satellites being available in the future. With this data hopefully regular, yearly, seasonally or even monthly, land cover classifications can be made that are able to monitor the effect of the mangrove rehabilitation programmes.

As mentioned in the discussion section 6.4 a future application is implementing the method for nationwide or even global scale. More research is needed for this on mangrove species worldwide and available ground truth information for those different species. It might not be possible to separate all global mangrove species but maybe mangroves can be sorted by its maturity, pioneer mangroves versus mature mangroves. An investigation of those possibilities is recommended for any further research.



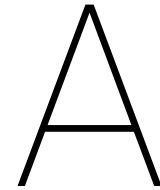
# Bibliography

- [1] Mohammad Abouali. Matlab implementation of harmonic analysis of time series (hants). <https://nl.mathworks.com/matlabcentral/fileexchange/38841-matlab-implementation-of-harmonic-analysis-of-time-series--hants->.
- [2] D.M. Alongi. Present state and future of the world's mangrove forests. *Environmental conservation*, 29(03):331–349, 2002.
- [3] D.M. Alongi. Mangrove forests: resilience, protection from tsunamis, and responses to global climate change. *Estuarine, coastal and shelf science*, 76:1–13, 2008.
- [4] D. Arthur and S. Vassilvitskii. K-means++: the advantages of carefull seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035, 2007.
- [5] M. Belgiu and L. Dragut. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, 2016.
- [6] M. Béland, et al. Assessment of land-cover changes related to shrimp aquaculture using remote sensing data: a case study in the giao thuy district, vietnam. *International Journal of Remote Sensing*, 27(8): 1491–1510, 2006.
- [7] P. Bridson. Black tiger shrimp. ca mau province of southern vietnam and other areas of southeast asia. extensive mixed shrimp and mangrove forestry (silvofishery) independently verified to be compliant with the selva shrimp criteria developed by blueyou consulting ltd. Technical report, Seafood Watch, 2013.
- [8] E.B. Brooks, et al. Fitting the multitemporal curve: A fourier series approach to the missing data problem in remote sensing analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 50(9):3340–3353, 2012.
- [9] CaMau. Map. <http://www.camau.gov.vn/wps/portal/camauportal>.
- [10] G. Donchyts, et al. Surface water changes. <http://aqua-monitor.appspot.com/>.
- [11] N. Duke. Mangroves of the kien giang biosphere reserve vietnam. Technical report, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, 2012.
- [12] ESA. Overview copernicus. [http://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Copernicus/Overview3](http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview3), .
- [13] ESA. Deal sealed for new sentinel-1 satellites. [http://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Copernicus/Sentinel-1/Deal\\_sealed\\_for\\_new\\_Sentinel-1\\_satellites](http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-1/Deal_sealed_for_new_Sentinel-1_satellites), .
- [14] ESA. Copernicus: Sentinel-2 — the optical imaging mission for land services. <https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/copernicus-sentinel-2>, .
- [15] ESA. Sen2cor. <http://step.esa.int/main/third-party-plugins-2/sen2cor/>, .
- [16] ESA. Sentinel-1. [http://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Copernicus/Sentinel-1](http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-1), .
- [17] ESA. Sentinel-2. [http://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Copernicus/Sentinel-2](http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2), .
- [18] ESA. Snap. <http://step.esa.int/main/toolboxes/snap>, .
- [19] Y. Fernandez-Ordonez, J. Soria-Ruiz, and B. Leblon. Forest inventory using optical and radar remote sensing. In *Advances in Geoscience and Remote Sensing*, chapter 26. InTech, 2009.

- [20] Food and Agriculture Organization of the United Nations. *The world's mangroves 1980-2005*. Viale delle Terme di Caracalla, Rome, Italy, 2007.
- [21] S. Gebhardt, L.D. Nguyen, and C. Kuenzer. Mangrove ecosystems in the mekong delta - overcoming uncertainties in inventory mapping using satellite remote sensing data. In *The Mekong Delta System. Interdisciplinary Analyses of a River Delta*, chapter 12, pages 315–330. Springer Environmental Science and Engineering, Dordrecht, Netherlands, 2012.
- [22] C. Giri, et al. Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography*, 20:154–159, 2011.
- [23] N. Gorelick, et al. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27, 2017.
- [24] Hansen, et al. Global forest change. <http://earthenginepartners.appspot.com/science-2013-global-forest>.
- [25] L.T. Hauser, et al. Spectra discrimination of common mangrove species in ca mau peninsula, vietnam: Using a large sample set of hyperspectral in-situ canopy data. In *Proceedings GIS 2015 Conference*, 2015.
- [26] A. Held, et al. High resolution mapping of tropical mangrove ecosystems using hyperspectral and radar remote sensing. *International Journal of Remote Sensing*, 24(13):2739–2759, 2003.
- [27] B.W. Heumann. Satellite remote sensing of mangrove forests: Recent advances and future opportunities. *Progress in Physical Geography*, 35(1):87–108, 2011.
- [28] World Weather & Climate Information. Climate: Average monthly weather in ho chi minh city, vietnam. <https://weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine,Ho-Chi-Minh-city,Vietnam>.
- [29] J. Kamaruzaman and I. Kasawani. Imaging spectrometry on mangrove species identification and mapping in malaysia. *WSEAS Transactions on Biology and Biomedicine*, 8(4), 2007.
- [30] E.S. Kasischke, J.M. Melack, and M. Craig Dobson. The use of imaging radars for ecological applications - a review. *Remote Sensing Environment*, 59:141–156, 1997.
- [31] B. Khairuddin, et al. Degradation mangrove by using landsat 5 tm and landsat 8 oli image in mempawah regence, west kalimantan province year 1989-2014. *Procedia Environmental Sciences*, 33:460–464, 2016.
- [32] C. Kuenzer, et al. Remote sensing of mangrove ecosystems: A review. *Remote Sensing*, 3:878–928, 2011.
- [33] C. Kuenzer, et al. Flood mapping and flood dynamics of the mekong delta: Envisat-asar-wsm based time series analyses. *Remote Sensing*, 5(2):687–715, 2013.
- [34] G. Machado, M.R. Mendoza, and L.G. Corbellini. What variables are important in predicting bovine viral diarrhoea virus? a random forest approach. *Veterinary Research*, pages 46–85, 2015.
- [35] A. Mcivor, T. Spencer, and I. Möller. Storm surge reduction by mangroves. Technical report, UK and Cambridge Coastal Research Unit, Department of Geography, University of Cambridge, UK, .
- [36] A. Mcivor, et al. Reduction of wind and swell waves by mangroves. Technical report, UK and Cambridge Coastal Research Unit, Department of Geography, University of Cambridge, UK, .
- [37] E. Mougin, et al. Multifrequency and multipolarization radar backscattering from mangrove forests. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1):94–102, 1999.
- [38] Sam Murphy. Atmospherically corrected time series using google earth engine. <https://github.com/sammurphy/ee-atmcorr-timeseries>.
- [39] L.D. Nguyen, et al. Change detection of land use and riverbank in mekong delta, vietnam using time series remotely sensed data. *Journal of resources and ecology*, 2(4):370–374, 2011.
- [40] P. Nguyen Hong and H. Thi San. *Mangroves of Vietnam*. IUCN, Bangkok, Thailand, 1993.

- [41] van der Philip, Lugt. An assessment of cloud detection methods concerning high altitude snow and glacial environments with sentinel-2. Master's thesis, Delft University of Technology, September 2017.
- [42] E. Pujiono, et al. Rgb-ndvi color composites for monitoring the change in mangrove area at the maubesi nature reserve, indonesia. *Forest Science and Technology*, 9(4):171–179, 2013.
- [43] E.W. Ramsey and J.R. Jensen. Remote sensing of mangrove wetlands: Relating canopy spectra to site-specific data. *Photogrammetric Engineering & Remote Sensing*, 62(8):939–948, 1996.
- [44] S. Reddy, M. Agrawal, and R.C. Prasar. Automatic extraction of mangrove vegetation from optical satellite data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B8, 2016.
- [45] W.G. Rees. *Physical Principles of Remote Sensing*. Cambridge University Press, Cambridge, United Kingdom, 2013.
- [46] J. Reiche, et al. Combining satellite data for better tropical forest monitoring. *Nature Climate Change*, 6: 120–122, February 2016.
- [47] F.G. Renaud and C. Kuenzer. *The Mekong Delta System. Interdisciplinary Analyses of a River Delta*. Springer Environmental Science and Engineering, Dordrecht, Netherlands, 2012.
- [48] A. Shelestov, et al. Exploring google earth engine platform for big data processing. *Frontiers in Earth Science*, 5, 2017.
- [49] L. Steurer and D. Meinardi. Mangrove management: A manual to appropriate mangrove conservation and planting in the mekong delta. Technical report, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.
- [50] Silke Tas. Coastal protection in the mekong delta. Master's thesis, Delft University of Technology.
- [51] P.T. Thinh, et al. Tool box for mangrove rehabilitation and management. Technical report, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.
- [52] P.M. Thu and J. Populus. Status and changes of mangrove forest in mekong delta: Case study in tra vinh, vietnam. *Estuarine, Coastal and Shelf Science*, 71:98–109, 2007.
- [53] P.H.S. Tong, et al. Assessment from space of mangroves evolution in the mekong delta, in relation with extensive shrimp-farming. *International journal of remote sensing*, 25(21):4795–4812, 2004.
- [54] N.T. Tue, et al. Carbon storage of a tropical mangrove forest in mui ca mau national park, vietnam. *Catena*, 121:119–126, 2014.
- [55] T.T. Van, et al. Changes in mangrove vegetation area and character in a war and land use change affected region of vietnam (mui ca mau) over six decades. *Acta Oecologica*, 63:71–81, 2015.
- [56] VietnamEnergy.vn. Prime minister has agreed on policy to invest a 100mw wind power project in ca mau. <http://nangluongvietnam.vn/news/en/nuclear-renewable/prime-minister-has-agreed-on-policy-to-invest-a-100mw-wind-power-project-in-ca-mau.html>.
- [57] T.Q. Vo, et al. Remote sensing in mapping mangrove ecosystems - an object-based approach. *Remote Sensing*, 5:183–201, 2013.
- [58] T.Q. Vo, C. Kuenzer, and N. Oppelt. How remote sensing supports mangrove ecosystem service valuation: A case study in ca mau province, vietnam. *Ecosystem Services*, 14:67–75, 2015.
- [59] J. Zhou, L. Jia, and M. Menenti. Reconstruction of global modis ndvi time series: Performance of harmonic analysis of time series (hants). *Remote Sensing of Environment*, 163:217–228, 2015.





# Results

## A.1. Land cover maps using different data input

Figures A.1 to A.5 show the different yearly land cover maps from 2017 made by the five different data input combinations:

1. Sentinel-2 optical data: Bands 2,3,4,5,6,7,8,8A,11,12 (figure 3.6) and NDVI
2. Sentinel-1 radar data: VV and VH backscatter
3. Sentinel-1& Sentinel-2 fusion: all bands 1. and 2.
4. Sentinel-1 temporal information: Mean  $a_0$  and yearly harmonics term  $A_2$  for VV and VH backscatter at 10m pixel resolution and 50m pixel resolution (8 bands total)
5. Sentinel-2& Sentinel-1 temporal information: all bands 1. and 4., in total 19 features

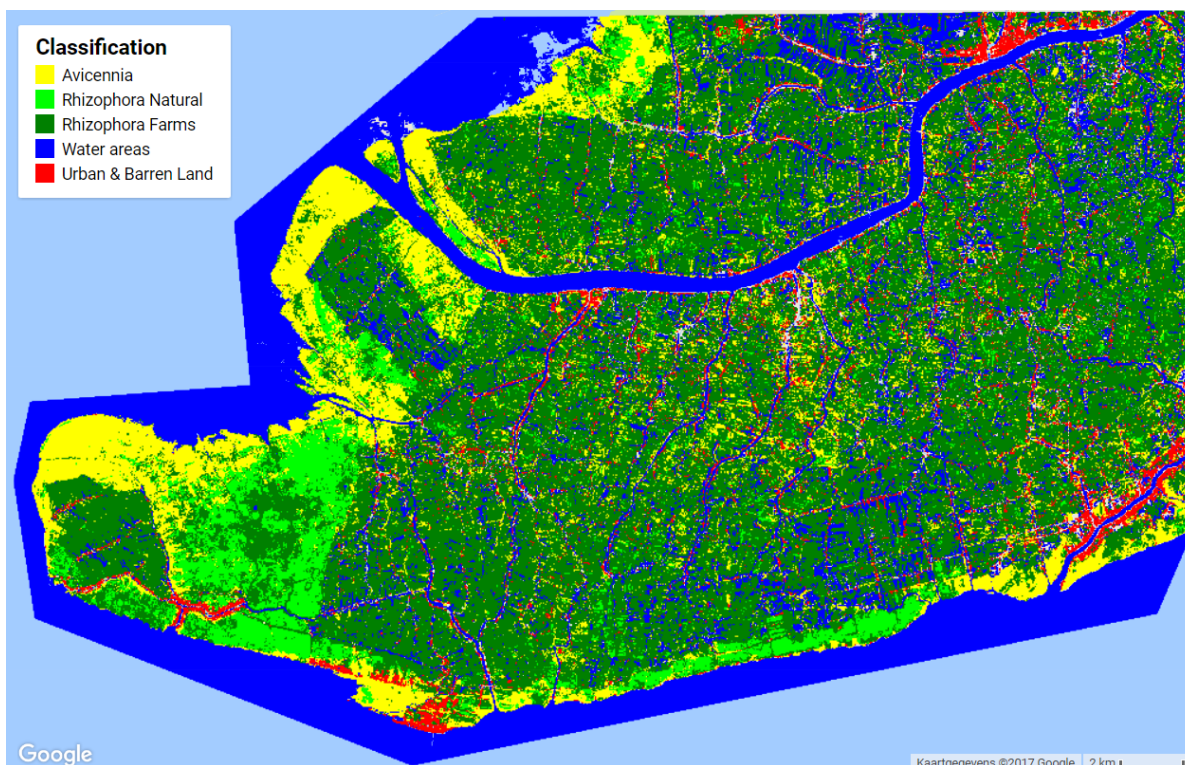


Figure A.1: Land cover map of 2017 using Sentinel-2 data

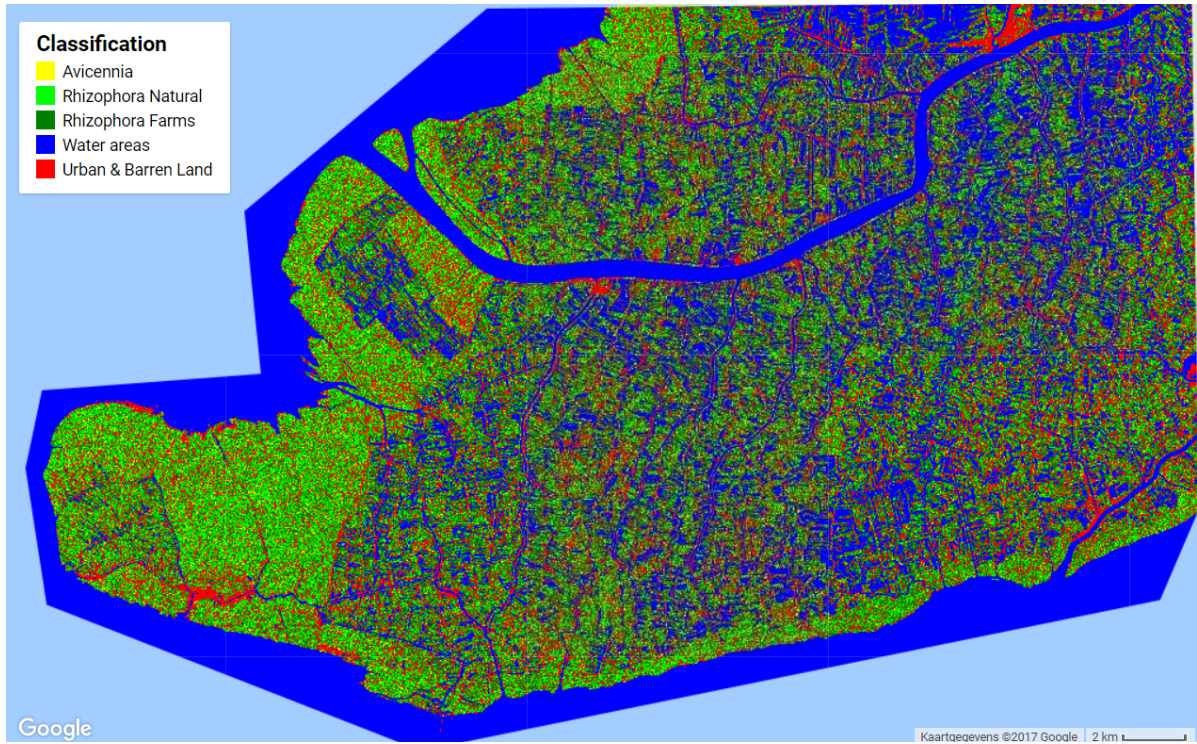


Figure A.2: Land cover map of 2017 using Sentinel-1 data

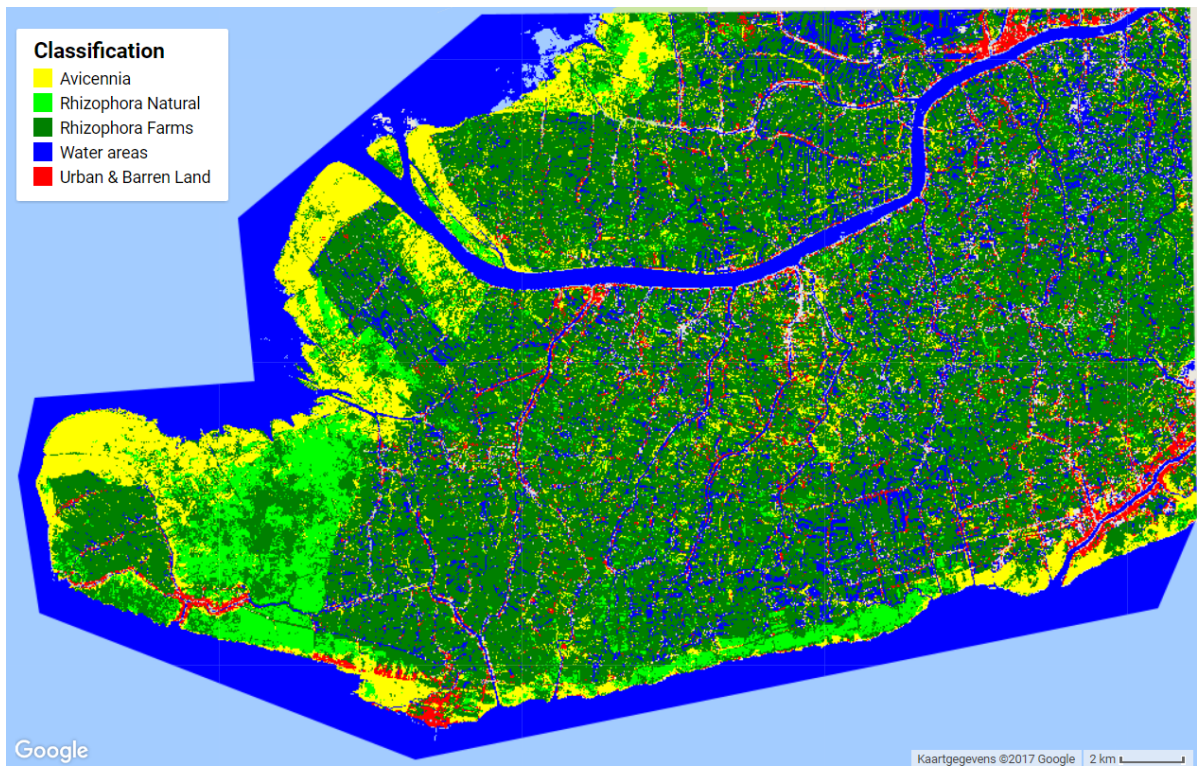


Figure A.3: Land cover map of 2017 using Sentinel-1 & 2 data

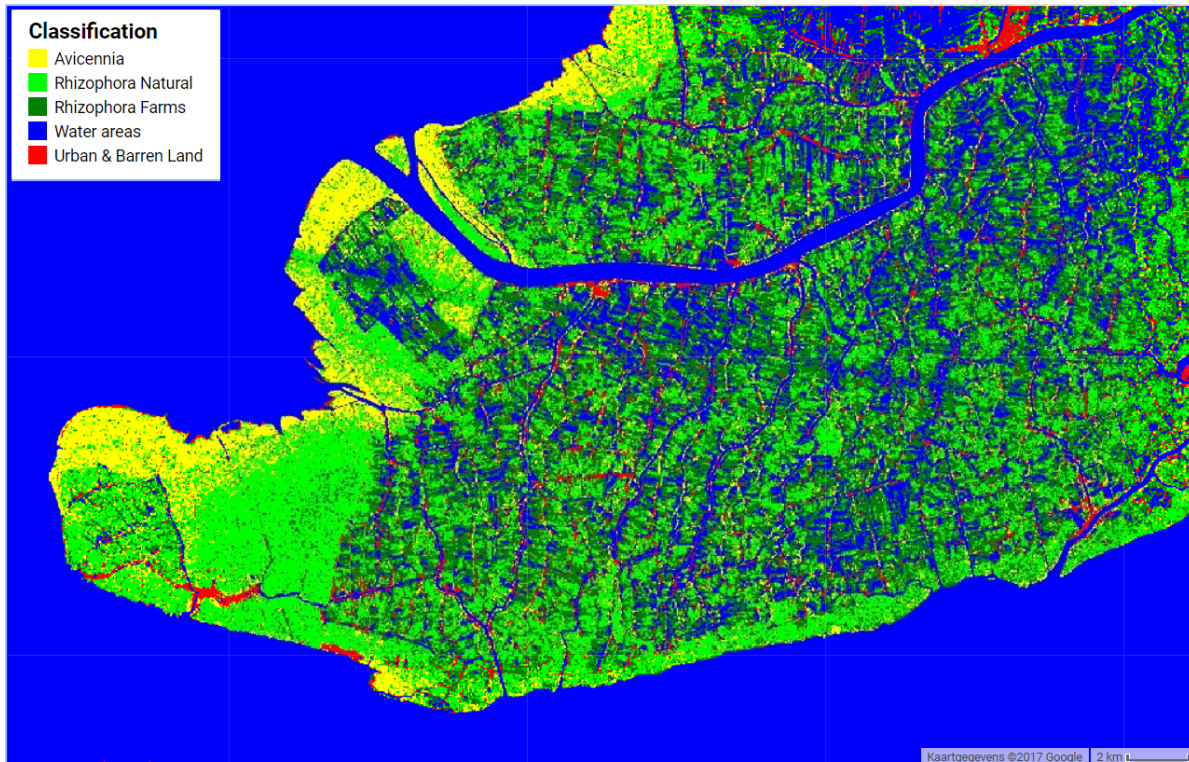


Figure A.4: Land cover map of 2017 using Sentinel-1 temporal data

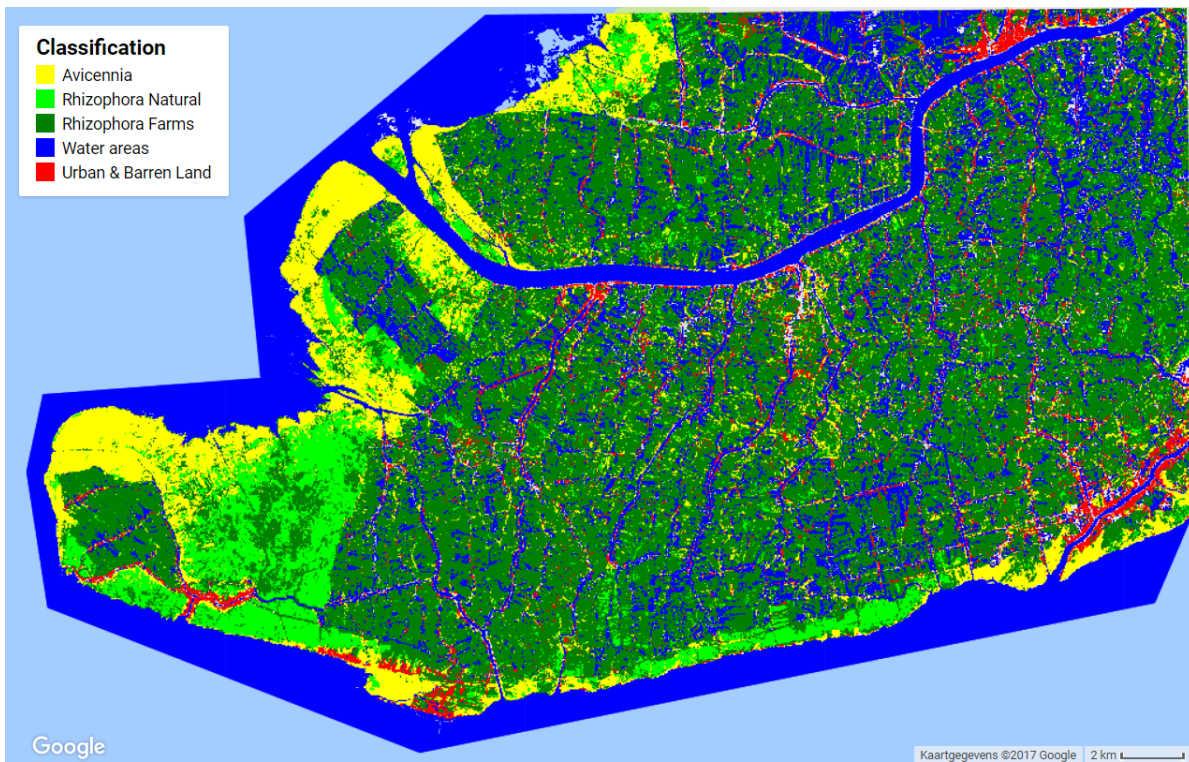


Figure A.5: Land cover map of 2017 using Sentinel-1 temporal and Sentinel-2 data

## A.2. Detailed workflow

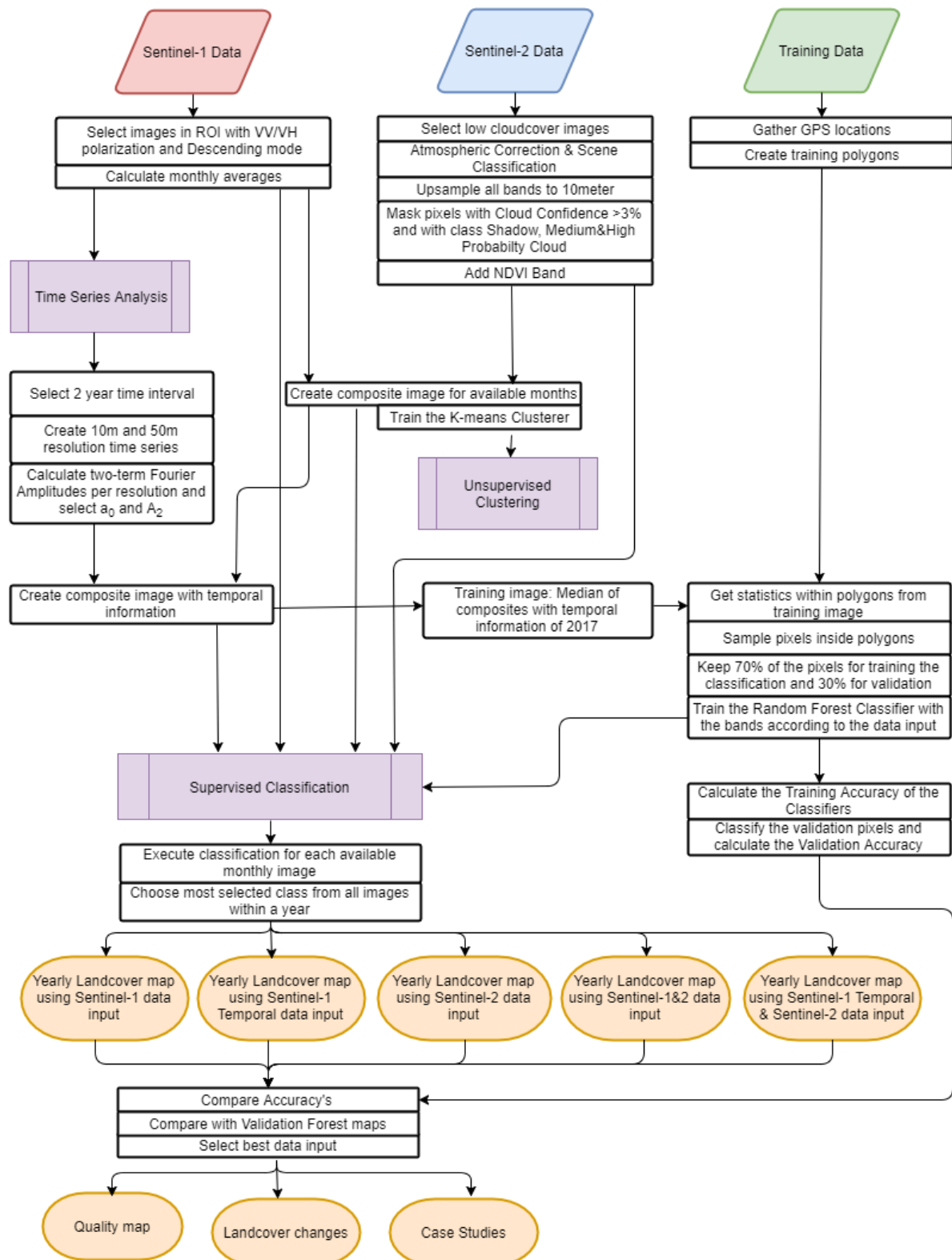


Figure A.6: Detailed workflow for classifying mangroves using Sentinel-1 and Sentinel-2 satellite imagery



### A.3. Temporal analysis

Figures A.7 to A.11 show the spatial results of the terms resulting from the temporal analysis. Only the mean and the second amplitude term were finally implemented in the supervised classification. The first and third amplitude (figures A.10 and A.11) did not show very significant differences between the classes.

Figures A.12 to A.12 show the time series at the locations 3, 4 and 5 indicated in the maps.

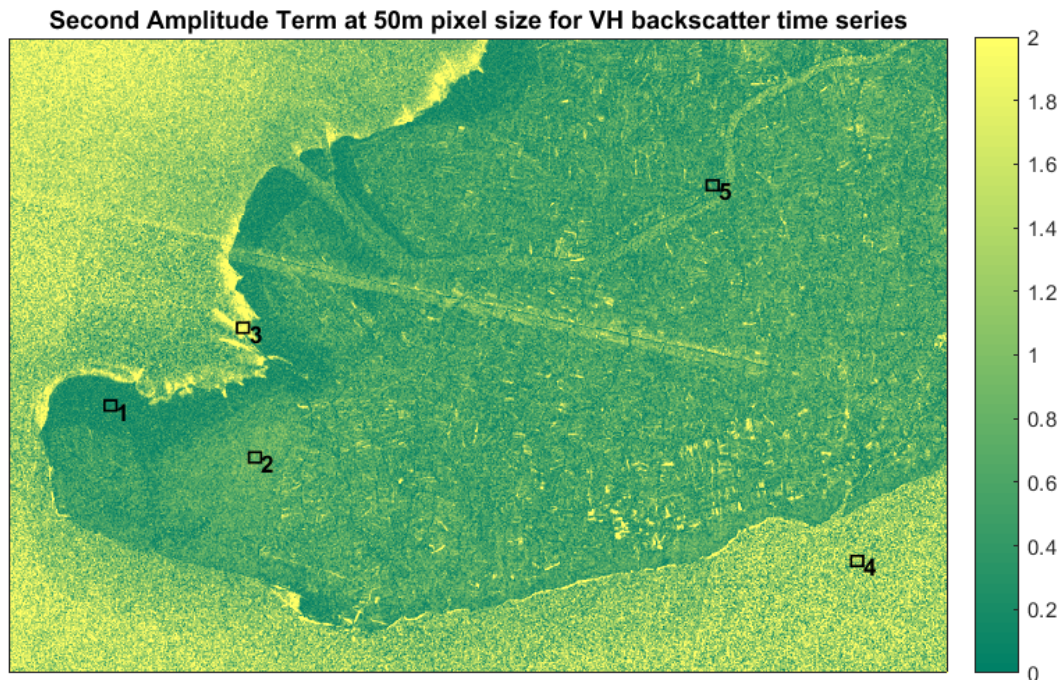


Figure A.7: Second amplitude term for VH cackscatter time series. Black rectangles indicate POI.

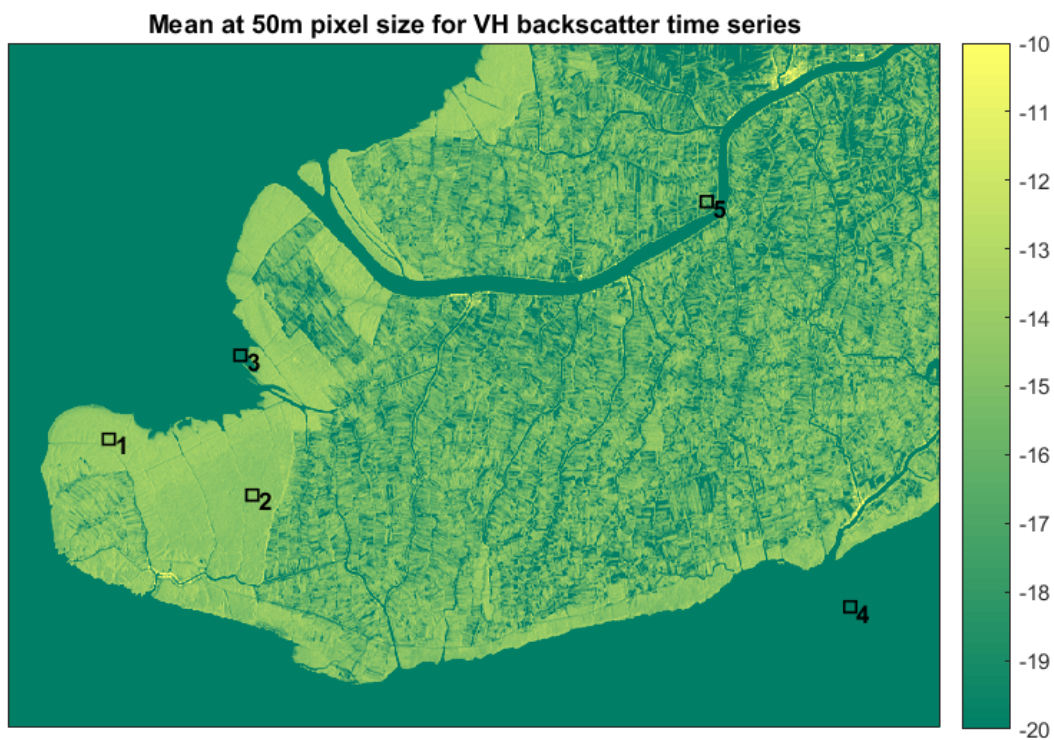
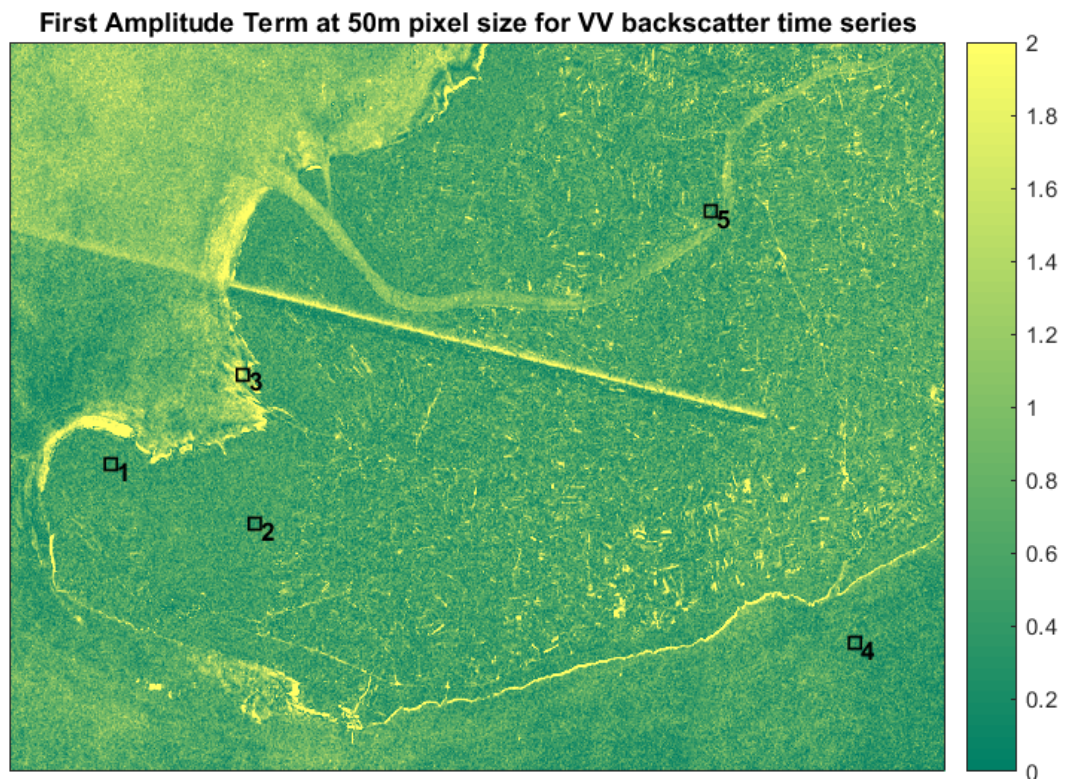
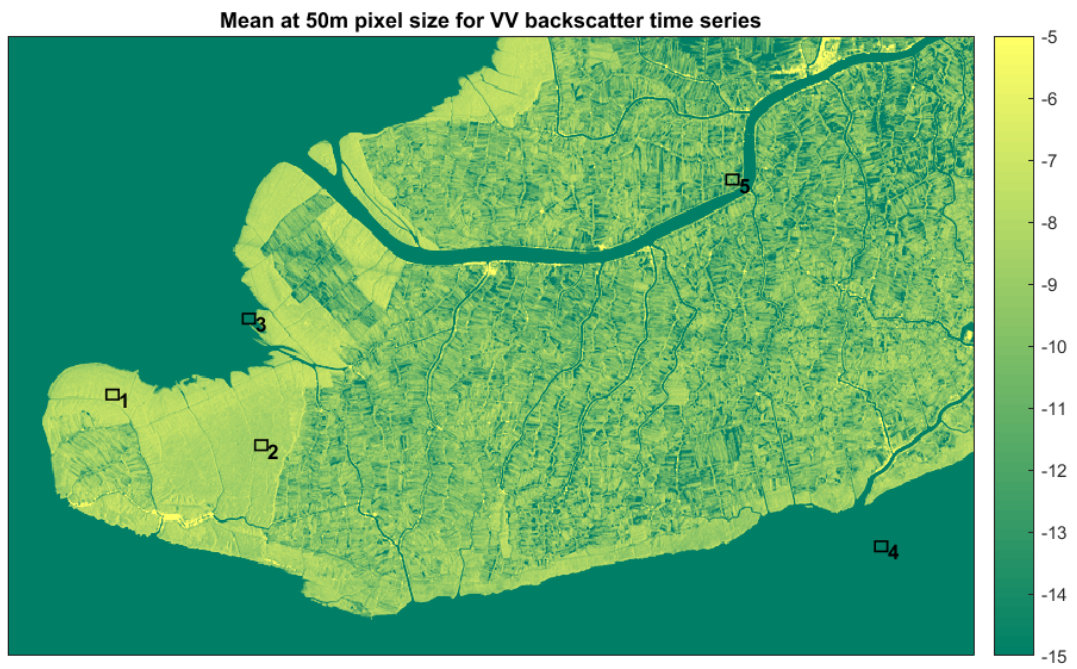


Figure A.8: Mean for VH backscatter time series. Black rectangles indicate POI



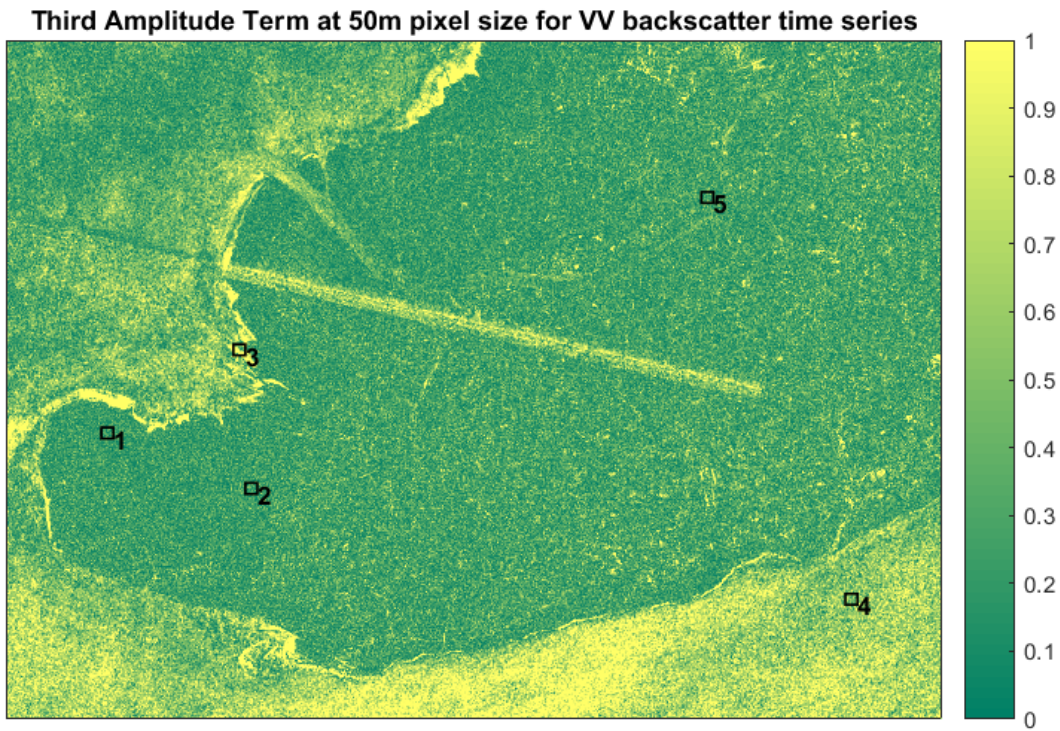


Figure A.11: Third amplitude term for VV backscatter time series. Black rectangles indicate POI

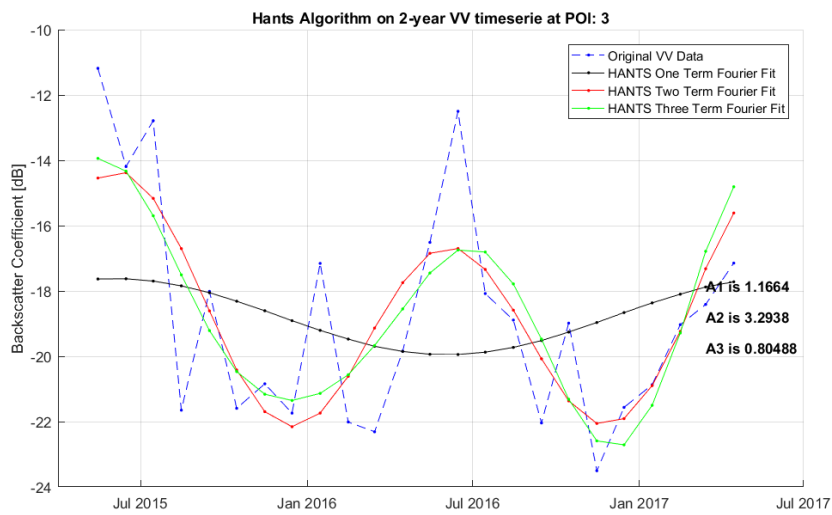


Figure A.12: VV backscatter time series at POI 3

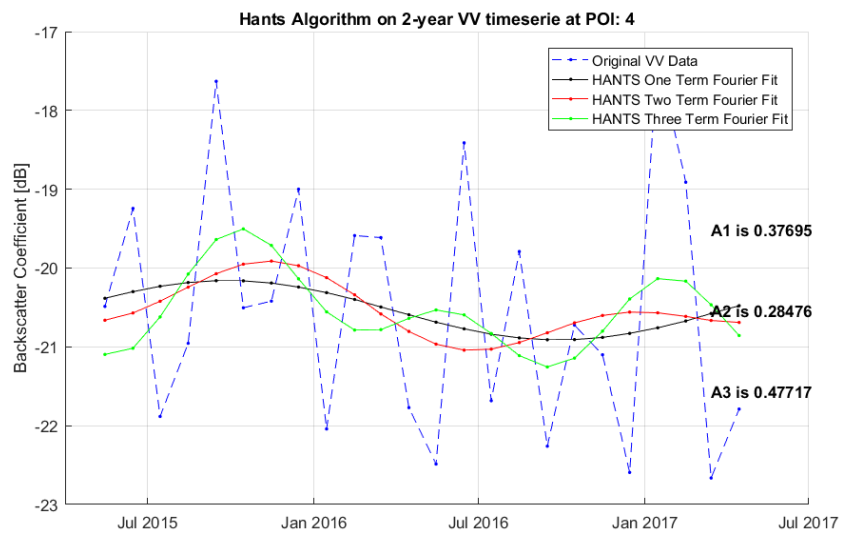


Figure A.13: VV backscatter time series at POI 4

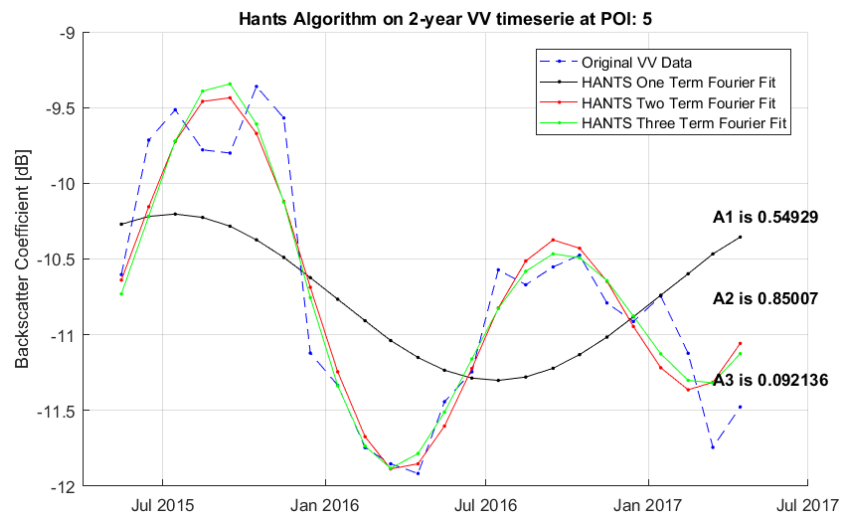


Figure A.14: VV backscatter time series at POI 5

## A.4. Clustering

Figures A.15 to A.17 show the maps resulting from the unsupervised clustering.

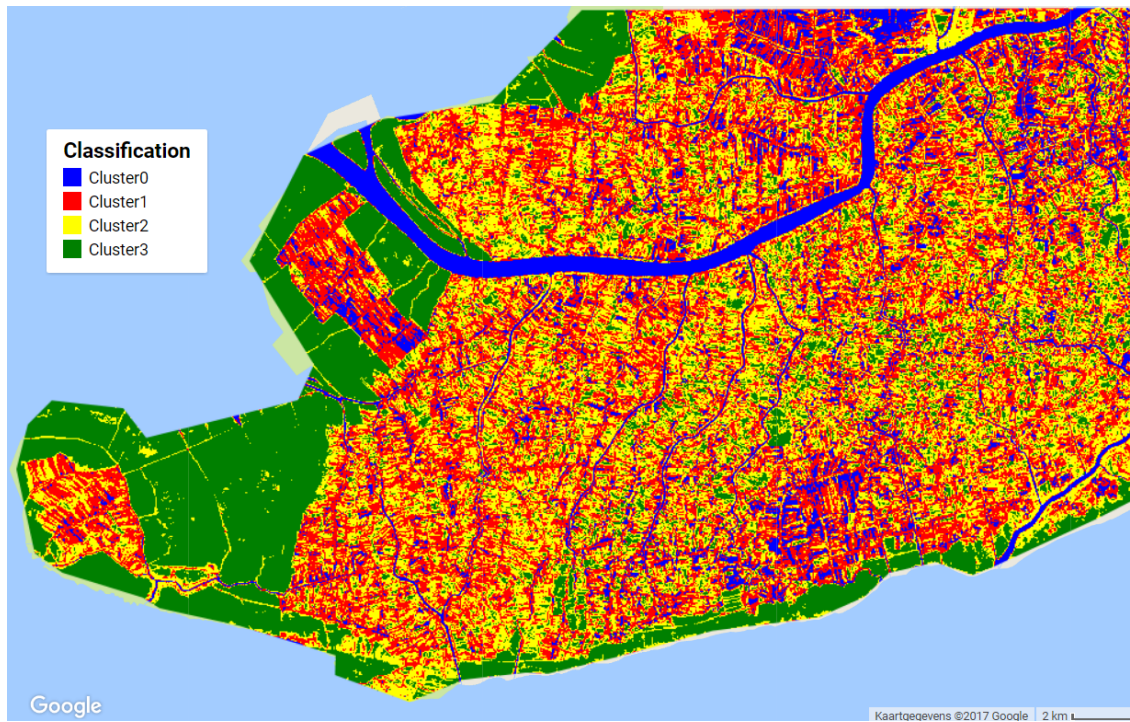


Figure A.15: Cluster analysis for Sentinel-2 median image for  $k=4$

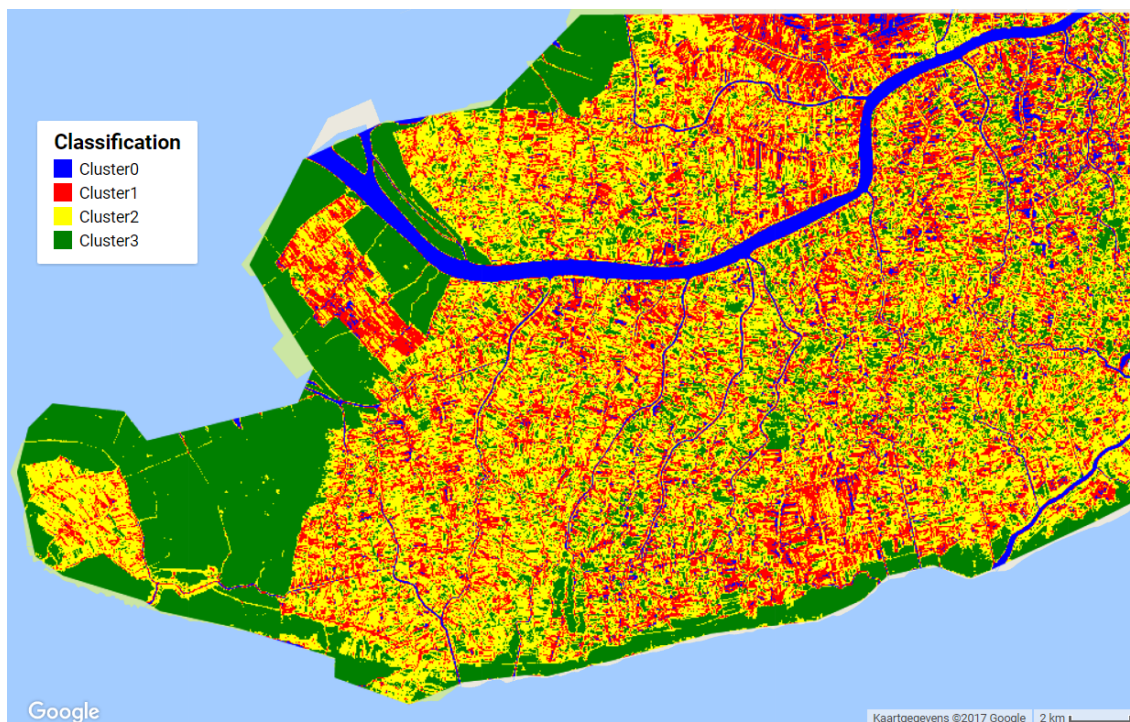


Figure A.16: Cluster analysis for combined Sentinel-1 & 2 median image for  $k=4$

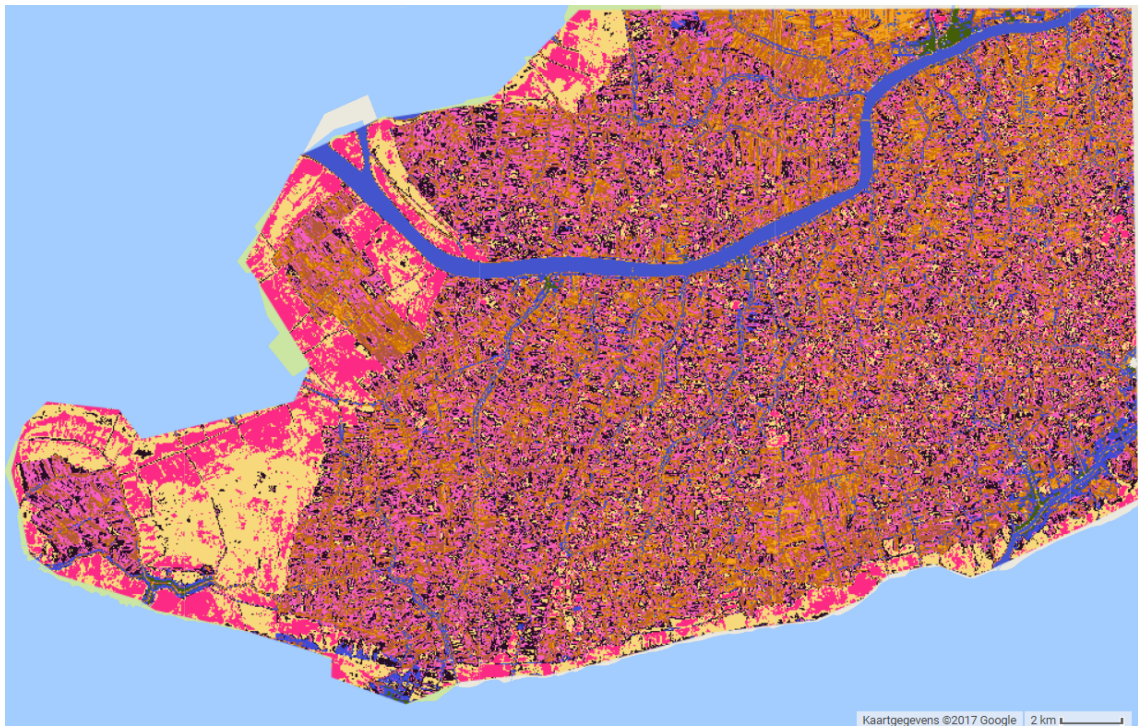


Figure A.17: Cluster analysis for combined Sentinel-1 & 2 median image for  $k=10$  in random colors