A machine learning assessment of multi-resolution remote sensing data for Natura 2000 dune habitat classification

M.S. Günes

Technische Universiteit Delft

TU Delft Delft University of Technology

Challenge the future

A machine learning assessment of multi-resolution remote sensing data for Natura 2000 dune habitat classification

by

M.S. Günes

to obtain the degree of Master of Science

in Applied Earth Sciences (Geoscience and Remote Sensing)

at the Delft University of Technology,

to be defended publicly on Friday October 2, 2020

Student number: 4305728

Thesis committee:

Dr. S. Lhermitte,	TU Delft, daily supervisor
Dr. I. V. Smal,	TU Delft
Dr. M. A. Eleveld,	Deltares & TU Delft, company supervisor
Dr. ir. S. de Vries,	TU Delft

TU Delft Delft University of Technology

Deltares Enabling Delta Life

ABSTRACT

Coastal dunes habitats are in the transition zone of marine and terrestrial ecosystems and therefore have high biodiversity in terms of animal and vegetation species. These species provide important ecosystem services to the human kind. Due to the development and operation of Maasvlakte 2, an increase of nitrogen emission is expected in the province Zuid-Holland of the Netherlands. When this is taken up by the vegetation and soil in the dunes, encroachment by shrubs will dominate. This will lead to a decrease in species diversity of dune habitats in protected areas of the Natura 2000 network, which is an unfavourable situation conforming to the Nature Protection Law (Wet Natuurbescherming). The main loss is expected to be in grey dunes and humid dune slacks of the Natura 2000 areas nearby Maasvlakte 2. This research focuses on the classification of grey dunes, white dunes, humid dune slacks, woody vegetation, water and their sub-classes in the Natura 2000 areas Solleveld & Kapittelduinen, Voornes Duin and Goeree & Kwade Hoek. Remote sensing in combination with machine learning have been proven to be useful in the monitoring of dune habitats. Open access satellite data provide high temporal and geographic coverage, but are often limited in their spatial resolution. Contrarily, high spatial resolution products are limited in their spectral resolution. In this thesis, the effects of spatial resolution, spectral features and feature combinations on the classification of dune habitats using aerial images and a Landsat-7 image (2012) are analysed. The classifications are performed with a Random Forest classifier, using both a pixel-based and region-based approach. In contrast to the traditional pixel-based method, the region-based method incorporates spectral and statistical information of neighbouring pixels. The best performance is achieved with the region-based approach on a 4m resolution aerial image with an accuracy of 87%. The 10m region-based classification is the second best with an accuracy of 83%. The misclassified pixels are mostly classified as the other sub-class of the same main-class. The region-based method should not be applied to satellite images with low spatial resolution as it causes more loss of detail. More spectral bands contribute to a better classification performance, but they contribute less than the spectral bands that were already used. Also, Landsat-7 images were used to create yearly habitat maps for 2012 and 2019 to subsequently analyse the habitat changes between those years, although the accuracies of the Landsat-7 maps are an issue. The main changes in habitat are found in the woody vegetation (shrubs, trees) and humid dune slacks. The woody vegetation decreased in area, causing an increase in grey dunes area in Voornes Duin and Goeree & Kwade Hoek, whereas the opposite has happened in Solleveld & Kapittelduinen. Loss of area in humid dune slacks is detected in all three regions of interest. The local dune manager confirmed that the detected changes in woody vegetation are mostly correct and caused by dune management activities. The perceived changes in humid dune slacks are less reliable. Finally, a modified training dataset is used to classify Sentinel-2 images in 2019. These results compared reasonably well with the few locations we visited during our fieldwork campaign.

Keywords: airborne, spaceborne, remote sensing, coastal dune habitats, nitrogen emission, machine learning, Random Forest, region-based, multi-resolution, Landsat-7, Sentinel-2.

PREFACE

This thesis is written to obtain the degree of Master of Science in Applied Earth Sciences, track 'Geoscience and Remote Sensing' at the Delft University of Technology. My passion for remote sensing has been around for a long time and is strengthened even further during this research. I am fortunate for being able to combine my fascination for nature with my passion for remote sensing in this work.

During this project, I've met wonderful people whom I want to thank for their contribution to my work. First of all, I would like to thank my company supervisor, Marieke Eleveld, for introducing me to this topic and giving me the opportunity to do my thesis at Deltares. I really enjoyed the fieldwork campaign together. They were definitely the best two days of this project. Also, I have learned a lot from you in every stage of this process. Thanks a lot. Moreover, I would like to thank Bert van der Valk, Frank van der Meulen, Stéphanie IJff and Gerrit Hendriksen from Deltares, Mennobart van Eerden and Kees Borst from Rijkswaterstaat, and Jasper Ohm from Provincie Zuid-Holland for their diverse perspectives and assistance throughout this project. Furthermore, I would like to thank Camilla Marcatelli, Harald van der Werff and Caroline Lievens from the University of Twente for giving me permission to use the field spectrometer and training me to use it correctly in the field. Harrie van der Hagen, Hans Huisman, Niek Koppelaar, Jasper Ohm, Stefan Poot and Mathijs Broere, thank you for granting me access to the dune areas for the fieldwork. Stefan, thanks for taking the time to validate the classification maps with me.

I want to show my gratitude towards my daily supervisor and chair of the graduation committee, Stef Lhermitte, for his guidance throughout my project and keeping me focused on the research. Thank you, Stef. I would like to thank my second supervisor, Ihor Smal, for sharing his knowledge and ideas about machine learning with me. Finally, I would like to thank Sierd de Vries for his comments on this work related to coastal engineering.

Last but not least, I want to thank my parents, sisters, brother-in-law and friends for always supporting and encouraging me in every phase of life. Esila, Benjamin and Kayan, you are the lights of my life. Tante Merve loves you a lot. I want to thank my dearest boyfriend with all my heart for his critical look at my work. I wouldn't make it without his endless trust, support and love throughout this journey.

M.S. Günes
Delft, October 2020

List of Figures

2.1 Distribution of the major dune coasts of the world, including well-developed dune systems (green) and small sand dunes interspersed with sandy shores, rocky headlines, sand marshes, or mangroves (red) (Mclachlan and Defeo, 2018). 6
2.2 Coastal dune in long-shore direction. The coastline is west of the viewing direction. Image taken on 20th of August 2019 near Spanjaards Duin. 6
2.3 Coastal dune in cross-shore direction. The coastline is in the viewing direction. Image taken on 20th of August 2019 near Spanjaards Duin. 6
2.4 Sand dune succession. The visualisation is not to scale. After Jackson (2014). 7
2.5 Vegetation gradients per zone (After Mclachlan and Defeo (2018)). 8
2.6 Grey Dunes high plant growth. Image taken on 20th of August 2019 at Van Dixhoorndriehoek (Solleveld & Kapittelduinen). 9
2.7 Grey dunes with humus layer. Image taken on 21st of August 2019 at Voornes Duin. 9
2.8 Humid dune slack type A: open water. Image taken on 21st of August 2019 at Voornes Duin. . . 10
2.9 Humid dune slacks grassland. Image taken on 21st of August 2019 at the Parnassia Vlak, Oostvoorne in Voornes Duin. 10
2.10 Possible change paths in dune habitats between 2012 and 2019. 11
2.11 The observation cube of remote sensing (Toth and Jóźków, 2016). 12
2.12 The electromagnetic spectrum zoomed in on the visible domain (Radio2Space, 2013). 13
2.13 Typical spectral signatures of different kinds of surfaces (Daneshgar, 2015). 13
2.14 Sentinel-2 and Landsat spectral bands in an atmospheric transmission plot (ESA, 2019d). 15
2.15 The Landsat operating timeline (NASA, 2019a). 15
2.16 An explanatory visualisation of what the Scan Line Corrector should have compensated for and what happens when it does not compensate (USGS, NASA, Landsat 7 Science Team, 2003). . . . 16
2.17 Graphical illustration of bagging, single prediction aggregation and final prediction decision in Random Forest. 18
2.18 Graphical illustration of weight adjustment in boosting. 19

3.1 The official Natura 2000 outlines of the areas. From top to bottom: Solleveld & Kapittelduinen, Voornes Duin, Goeree & Kwade Hoek. 22
3.2 Sketch of measuring method during fieldwork campaign. The black dot is the starting point, the grey dots are the additional measurement points and r is the radius of approximately 2 meters. . 23
3.3 The ASD fieldspectrometer device in the field on field-campaign day 2. 23
3.4 Fieldwork campaign measurement locations day 1. Blue points are *humid dune slacks*, black points are *grey dunes*, yellow points are *white dunes* and green points are *buckthorn formations*. The vegetation in point 12 belongs to *Festuca* species and belongs to both white dunes as grey dune species. The background maps are aerial images of 2019 (WMS). 24
3.5 Fieldwork campaign measurement locations day 2. Blue points are *humid dune slacks*, black points are *grey dunes*, yellow points are *white dunes* and green points are *buckthorn formations*. The background maps are aerial images of 2019 (WMS). Point 14 does not belong to the official Natura 2000 area. 25
3.6 Left: the traditional hierarchy, middle: the hierarchy used in this research with main-habitats (left column) and sub-habitats (right column), right: the legend entries of the corresponding sub-habitat type. 26
3.7 Flightpaths during acquisition of the aerial images in 2012. The colors show the flight dates. . . 28
3.8 Global coverage and revisit time for Sentinel-2 MSI acquisitions. Validity start in October 2019 (ESA, 2019c). 29
3.9 Available images after the filtering processes for Sentinel 2 (2019), Landsat 7 (2011-2013, 2018. 2019) and aerial images (2012). The green bar indicates the acquisition time of the aerial images in 2012. 31

3.10 Median reflectances of the sub-habitat types from the aerial images 2012. *'Dunes'* are abbreviated to *'d.'*, *'formations'* to *'f.'* and *'vegetation'* to *'veg.'*. 32
3.11 Mean spectral reflectances measured in the field with the field spectrometer in Solleveld & Kapittelduinen (a) and Voornes Duin (b). The measurement method is described in section 3.2. 33
3.12 2D visualisation of the high-dimensional dune habitat data (features: Red, Green, Blue, NIR, NDVI, ARVI) random selection for each sub-habitat using t-SNE. 34

4.1 Summarized workflow. Full workflow in Appendix A.2. 38
4.2 Run time and accuracy of Random Forest and XGBoost for a pixel-based flat classification. . . . 41
4.3 Run time and accuracy ratio between Random Forest and XGBoost for a pixel-based classification. Runtime (t) ratio is $\frac{t\_XGBoost}{t\_Random\ Forest}$ and accuracy (Acc.) ratio is $\frac{Acc\_XGBoost}{Acc\_Random\ Forest}$ 41
4.4 Region-based feature extraction scheme. N_f: *number of features.* 42

5.1 Region-based classification using spectral features from center pixel and statistical features (red curve) and using spectral features from all 9 pixels in the window and statistical features (blue curve). 46
5.2 Feature importances for low-spectral pixel-based classification of aerial images for 4m resolution using features 0:Blue, 1:Green, 2:Red, 3:NIR, 4:NDVI, 5:ARVI. 47
5.3 Feature importances for low-spectral pixel-based classification of aerial images for 30m resolution using features 0:Blue, 1:Green, 2:Red, 3:NIR, 4:NDVI, 5:ARVI. 47
5.4 Feature importances for low-spectral pixel-based classification of L7 using features 0:Blue, 1:Green, 2:Red, 3:NIR, 4:NDVI, 5:ARVI. 47
5.5 Feature importances for high-spectral pixel-based classification of L7 using features 0:Blue, 1:Green, 2:Red, 3:NIR, 4:NDVI, 5:ARVI, 6:SWIR1, 7:SWIR2, 8:NDWI. 47
5.6 Classifications on a 4m resolution aerial image (2012) with the (a) region-based and (b) pixel-based method in Solleveld & Kapittelduinen. Background map is the aerial image from 2012. . . 48
5.7 Region-based classifications on (a) 10m, (b) 20m and (c) 30m resolution aerial images (2012) in Solleveld & Kapittelduinen. Background map is the aerial image from 2012. The sub-figures share the same legend shown in sub-figure (c). 49
5.8 Pixel-based classifications in (a) low-spectral and (b) high-spectral category on a Landsat-7 image (2012) in Solleveld & Kapittelduinen. Background map is the aerial image from 2012. The sub-figures share the same legend shown in sub-figure b. 49
5.9 Classifications with the region-based method in (a) low-spectral and (b) high-spectral category on a Landsat-7 image (2012) in Solleveld & Kapittelduinen. Background map is the aerial image from 2012. The sub-figures share the same legend shown in sub-figure b. 50
5.10 Median pixel reflectances of the aerial images (2012) for the different classes in (a) Goeree & Kwade Hoek, (b) Solleveld & Kapittelduinen and (c) Voornes Duin. 52
5.11 Sub-habitat classifications for Landsat-7 images in 2012 and 2019. See Figures A.15, A.16 and A.17 for larger maps. 54
5.12 Main-habitat classifications for Landsat-7 images in 2012 and 2019. See Figures A.12, A.13 and A.14 for larger maps. 55
5.13 Maps for habitats changed between 2012 and 2019 for both (a) sub-habitat and (b) main-habitat. See Figures A.18, A.19 and A.20 for larger maps. 56
5.14 Sentinel-2 final sub-habitat (a) and main-habitat (b) classifications for 2019. From left to right: Solleveld & Kapittelduinen, Voornes Duin, Goeree & Kwade Hoek. See Figures A.21, A.22 and A.23 for larger maps. 59
5.15 Classification stability in percentages for (a) Landsat-7 in 2012 (b) for Sentinel-2 in 2019. The values show the percentages of how many times the final class is chosen from the total available classifications. 61
5.16 Number of available pixels in Solleveld & Kapittelduinen for 2019 Landsat-7 (left) and 2019 Sentinel-2 (right). 62
5.17 Correlation plots between satellite reflectances and field spectrometer reflectances for Landsat-7 and Sentinel-2 satellites (S2: 24-08-2019, L7:22-8-2019 and 14-09-2019). The plots are grouped per band and colored per habitat type. 63
5.18 Sub-habitat map from the province Zuid-Holland 2019 used as external validation. 64

5.19 Spanjaards Duin in 2019. Left to right: Province Zuid-Holland classification, 2019 Sentinel-2 classification, 2019 Sentinel-2 stability map, aerial image (2019) 25cm resolution (PDOK). 65

6.1 Some examples of the effects of most probably sunglint (left) and suspended waters on the classifications in Solleveld & Kapittelduinen. The classification is the region-based 4m resolution classification. The aerial image shown on the first row is the aerial image in 2012. 70

A.1 Modified Mouissie classification of Solleveld & Kapittelduinen (Mouissie et al., 2014). 78
A.2 Aerial image of Solleveld & Kapittelduinen in 2012, Figure A.1. 79
A.3 Modified Mouissie classification of Voornes Duin (Mouissie et al., 2014). 80
A.4 Modified Mouissie classification of Goeree & Kwade Hoek (Mouissie et al., 2014). 81
A.5 Aerial image of Goeree & Kwade Hoek and Voornes Duin in 2012, Figures A.3 and A.4. 82
A.6 Extensive methodology. 83
A.7 Band designations for Sentinel-2 (ESA, 2019a). 84
A.8 Band designations for Landsat-7 (USGS, 2019). 84
A.9 Region-based classification on 4m resolution aerial image in Solleveld & Kapittelduinen. 85
A.10 Region-based classification on 4m resolution aerial image in Voornes Duin. 86
A.11 Region-based classification on 4m resolution aerial image in Goeree & Kwade Hoek. 87
A.12 Main-habitat classifications in 2012 and 2019 of Solleveld & Kapittelduinen using Landsat-7 images. 88
A.13 Main-habitat classifications in 2012 and 2019 of Voornes Duin using Landsat-7 images. 89
A.14 Main-habitat classifications in 2012 and 2019 of Goeree & Kwade Hoek using Landsat-7 images. 90
A.15 Sub-habitat classifications in 2012 and 2019 of Solleveld & Kapittelduinen using Landsat-7 images. 91
A.16 Sub-habitat classifications in 2012 and 2019 of Voornes Duin using Landsat-7 images. 92
A.17 Sub-habitat classifications in 2012 and 2019 of Goeree & Kwade Hoek using Landsat-7 images. . 93
A.18 Habitat change maps (between 2012 and 2019) of Solleveld & Kapittelduinen using Landsat-7 images. The colored pixels show the class in 2019. 94
A.19 Habitat change maps (between 2012 and 2019) of Voornes Duin using Landsat-7 images. The colored pixels show the class in 2019. 95
A.20 Habitat change maps (between 2012 and 2019) of Goeree & Kwade Hoek using Landsat-7 images. The colored pixels show the class in 2019. 96
A.21 Sub-habitat and main-habitat classifications in 2019 of Solleveld & Kapittelduinen using Sentinel-2 images. 97
A.22 Sub-habitat and main-habitat classifications in 2019 of Voornes Duin using Sentinel-2 images. . 98
A.23 Sub-habitat and main-habitat classifications in 2019 of Goeree & Kwade Hoek using Sentinel-2 images. 99

List of Tables

2.1 Overview of the specifications of the open access satellite imagery and aerial imagery used for this project. 16

3.1 Spectral sensitivity of all Bands in the UC-SXp-1-30719036 (Scholz and Gruber, 2008). 27

5.1 Classifications using Random Forest for pixel-based (PB) and region-based (RB) methods. '*Low-spectral*' indicates the bands Red, Green, Blue, NIR and vegetation indices NDVI + ARVI. '*High-spectral*' indicates all 6 bands (Red, Green, Blue, NIR, SWIR1, SWIR2) and indices NDVI, ARVI and NDWI. Performance metrics are *Acc. and* κ indicating respectively the balanced accuracy and Cohen's Kappa. 45

5.2 Confusion matrix of the 4m resolution region-based classification. The columns represent the predicted class and the rows represent the actual class. The overall balanced accuracy is 87%. . 51

5.3 Table with balanced accuracies (%) on an independent test dataset of the model trained on one area and predicted on a different area. The classification method is the region-based method, applied on a 4m resolution aerial image. The columns show the area the model is trained on: all areas, Solleveld & Kapittelduinen (SKD), Voornes Duin (Voorne) and Goeree & Kwade Hoek (Goeree). The rows represent the areas the model is predicted on. 51

5.4 Percentual transition matrix (%) from 2012 (rows) to 2019 (columns) for Goeree & Kwade Hoek. For example: the grey dunes entry shows that, according to Figure 5.13 (b), 84% of the grey dunes in 2012 stay grey dunes, 10% changed to humid dunes, 1% changed to white dunes and 5% changed to woody vegetation. 57

5.5 Percentual transition matrix (%) from 2012 (rows) to 2019 (columns) for Solleveld & Kapittelduinen. For example: the grey dunes entry shows that, according to Figure 5.13 (b), 78% of the grey dunes in 2012 stay grey dunes, 5% changed to humid dunes, 5% changed to white dunes and 12% changed to woody vegetation. 57

5.6 Percentual transition matrix (%) from 2012 (rows) to 2019 (columns) for Voornes duin. For example: the grey dunes entry shows that, according to Figure 5.13 (b), 72% of the grey dunes in 2012 stay grey dunes, 1% changed to water, 16% changed to humid dunes, 2% changed to white dunes and 9% changed to woody vegetation. 57

5.7 Total area change in hectares per region between 2012 and 2019 according to Figure 5.13 (b). An increase is indicated with a plus and a decrease with a minus-sign. 58

5.8 Confusion matrix of the 30m resolution high-spectral pixel-based main-habitat classification. The columns represent the predicted class and the rows represent the actual class. The overall balanced accuracy is 66%. 60

5.9 Validation of the Landsat-7 and Sentinel-2 final classifications on fieldwork points for day 1 in Solleveld & Kapittelduinen shown in Figure 3.4. P stands for point location. 62

5.10 Validation of the Landsat-7 and Sentinel-2 final classifications on fieldwork points for day 2 in Voornes Duin shown in Figure 3.5. P stands for point location. P14 locates outside the official boundaries of the Natura-2000 area and therefore also outside the classification boundaries. . . 62

CONTENTS

List of Figures vii

List of Tables xi

1 Introduction 1
- 1.1 Nitrogen overload in Natura 2000 areas . . . 1
- 1.2 Problem definition and research objective . . . 2
- 1.3 Research questions . . . 3
- 1.4 Thesis outline . . . 3

2 Background information 5
- 2.1 Coastal dunes . . . 5
- 2.2 Remote sensing of dune habitats . . . 12
- 2.3 Machine learning for remote sensing . . . 17

3 Data and area description 21
- 3.1 Regions of interest . . . 21
- 3.2 Fieldwork campaign . . . 23
- 3.3 Dune habitat classes . . . 26
- 3.4 Data selection and pre-processing . . . 27
- 3.5 Data understanding . . . 32
- 3.6 Software . . . 35

4 Methodology 37
- 4.1 Workflow . . . 37
- 4.2 Supervised classification . . . 39
- 4.3 Spatial and spectral resolution analysis . . . 41
- 4.4 Pixel and region-based classification . . . 42
- 4.5 Yearly habitat maps . . . 42
- 4.6 Assessment and validation . . . 43

5 Results 45
- 5.1 Aerial image classifications . . . 45
- 5.2 Satellite image classifications . . . 53
- 5.3 Accuracy assessment . . . 60

6 Discussion 67
- 6.1 Processing and data products . . . 67
- 6.2 Results . . . 69
- 6.3 Comparison with similar research . . . 71

7 Conclusions & Recommendations 73
- 7.1 Conclusions . . . 73
- 7.2 Recommendations . . . 75

A Appendix 77
- A.1 Data labels, Mouissie classifications (2012) . . . 78
- A.2 Methodology . . . 83
- A.3 Band designations . . . 84
- A.4 Classification maps . . . 85

Bibliography 101

1

INTRODUCTION

Nature is not only fascinating, but also a vital part of the living environment for humans, animals and plants. In the previous centuries, nature has declined in biodiversity. Therefore, governmental protection is mandatory, whereby international cooperation is imperative. The Natura 2000 network was set up to facilitate this cooperation. One of the more recent problems related to biodiversity is the increase of nitrogen emission in the atmosphere. A long period of overload of nitrogen (ammonia and nitrogen) oxides is harmful to biodiversity of vegetation. Recent research shows that the overload is mostly sourcing from agricultural fertilizers, but also from traffic and industry (Rijksoverheid, 2019). If this increase is absorbed by the vegetation and soil, encroachment will dominate. This will eventually lead to a decrease in species diversity of vegetation within dune habitats (Ten Harke et al., 1998) in protected areas of the Natura 2000 network. This is an unfavourable situation conforming to the Nature Protection Law (Wet Natuurbescherming) (Overheid, 2019). The provinces are responsible for drawing up management plans for each Natura 2000 area on land or regional waters. The main purpose of these management plans is to monitor the behaviour of the species and habitats to guarantee their long-term survival in the Netherlands. The current methods of monitoring offer insufficient interim visibility on the situation. Remote sensing can play an important role in improving the monitoring quality.

This chapter includes introductory information about the context of the project in terms of nitrogen overload in the Natura 2000 areas, followed by a description of the problem statement and research objective, including the role of remote sensing to the problem. The chapter is concluded with the main research question, the corresponding sub-questions and the outline of the report.

1.1. NITROGEN OVERLOAD IN NATURA 2000 AREAS

There is a strong nature protection legislation in Europe coordinated by the European Union (EU). This has been achieved with the help of the Natura 2000 network: the world's largest network that offers crucial protection for Europe's most threatened species and habitats. The two main directives are the Birds and Habitats Directive. The Habitats Directive protects endangered species of plants, animals and their habitats (1992). The areas protected under these directives belong to the Natura 2000 network. There are around 160 Natura 2000 areas in the Netherlands (European Commission, 2019). The first main objective of the directives is to avoid activities that could seriously disturb the species or damage the habitats for which the site is designated. The second main objective is to take positive measures, if necessary, to maintain and restore these habitats and species to improve conservation. This thesis contributes to the second main objective.

Since years, there has been a nitrogen overload in 118 Natura 2000 areas in the Netherlands (Ministerie van Landbouw, Natuur en Voedselkwaliteit, 2019). An overload of nitrogen will favour species that thrive in nutrient-rich environments, which will subsequently outcompete pioneer species living on nutrient-poor soils. Combinations of pioneer species characterize habitat types. The Integrated Approach for Nitrogen (Programma Aanpak Stikstof, PAS) is developed in 2015 to strengthen the Natura 2000 areas by reducing the nitrogen deposition in these areas. In May 2019, the Council of State concluded that the assumptions on which the PAS depends do not meet the requirements construed in the Birds and Habitat Directive (Rijksin-

stituut voor Volksgezondheid en Milieu, 2019).

1.2. PROBLEM DEFINITION AND RESEARCH OBJECTIVE

An increase of nitrogen emission was expected in Zuid-Holland, due to the development of Maasvlakte 2. As a consequence, it is assumed that there will be more deposition in the nearby Natura 2000 areas. The expected loss in dune habitat area has been estimated to be approximately 16 ha in total for grey dunes and humid dune slacks (Van der Meulen et al., 2014). To compensate for this expected decrease of habitat area, dune nourishment Spanjaards Duin was created in 2009 as part of the Natura 2000 area Solleveld & Kapittelduinen. The goal is that nature builds approximately 10 ha of grey dunes and 6 ha of humid dune slacks by itself before the year 2033 (Van der Meulen et al., 2014). For this thesis work, three Natura 2000 areas in Zuid-Holland are investigated: Solleveld & Kapittelduinen, Voornes Duin and Goeree & Kwade Hoek.

PROBLEM STATEMENT

Monitoring is essential to keep record of what is happening in the intervening time. Traditionally, habitat areas are monitored by translating fieldwork observations into maps once every few years. This process is generally very time consuming and labour intensive (Schmidt and Skidmore, 2003). Methods are developed to speed up and facilitate parts of such processes by automating the mapping where possible. Remote sensing plays an important role in this automation, especially in combination with the use of different classification algorithms. In recent decades, remote sensing has become popular in land cover mapping, change detection applications and overall monitoring of the Earth due to the availability of satellite data with high geographical and temporal coverage. Data are available in a wide range of spatial, spectral and temporal resolutions from different satellite and imagery platforms.

The problem with observing dune habitats using remote sensing data is that the scale on which a habitat can be identified using remote sensing data is very different from the scale on which humans identify (variable) due to limitations in spatial resolution of remote sensing data. In order to achieve a high temporal resolution with geographical coverage, the use of open access satellite data (e.g. Sentinel and Landsat missions) is often the only solution as these provide high spectral and temporal resolution. However, these data are limited in their spatial resolution. On the contrary, high resolution remote sensing data (e.g. commercial satellites and aerial imagery) are limited in their spectral and temporal resolutions.

RESEARCH OBJECTIVE

Due to aforementioned trade-offs in resolution between various remote sensing data sources, it is important to analyse the effects and the significance of the resolution, spectral features and feature combinations on classifying dune habitats using both the pixel-based and region-based methods. The region-based method assimilates information from bordering pixels to the training data of the machine learning algorithm. The pixel-based method only focuses on the information of the target pixel. Moreover, the three Natura 2000 areas of interest can differ in their natural process dynamics and therefore show most probably differences in spectral reflectances of habitat types. This is defined as spatial variability and is meaningful to research in order to get an idea of whether there is potential to apply the developed method nationwide in the near future. For this research, multi-resolution remote sensing data in the range of 25cm to 30m is used, including satellite data (with spatial resolutions of 10m, 20m and 30m) and aerial imagery (with a spatial resolution of 25cm). To analyse intermediate spatial resolutions, the aerial images are downsampled from the original resolution.

Rijkswaterstaat asked, i.a., Deltares for advise on monitoring the target habitats, grey dunes and humid dune slacks. In 2012 (spanning 2011 to 2013), Grontmij conducted an investigation on how the increase in nitrogen relates to a change in dune habitat, both in type and volume. This resulted in a vegetation and habitat classification (Mouissie et al., 2014) of previously mentioned research areas. The habitat labels used in this thesis originate from the investigation done in 2012.

1.3. RESEARCH QUESTIONS

The main research question arises from the problem statement and research objective as following:

> ***"What are the effects of resolution of aerial images and optical satellite images on the classification of dune habitat types in coastal Dutch Natura 2000 areas?"***

The two important components in the main research question are underlined: *resolution* and *classification of dune habitats*. The sub-questions are based on these components:

1. How do the specifications of the different data products in terms of spatial, spectral and temporal resolution relate?
2. Which disturbance factors are introduced when simulating the optical satellite images from aerial images by downsampling?
3. How much better does the classification model perform when more spectral features are added?
4. How is the classification performance influenced when data is combined from the three different Natura 2000 areas instead of treating each area separately?
5. How is the classification performance influenced when statistical relationships between adjacent pixels and the center pixel are incorporated?
6. What is the quality of the training and validation data from 2012 (Mouissie et al., 2014) and 2019?

1.4. THESIS OUTLINE

The thesis contains seven chapters. Chapter 2 gives background information about coastal dunes, remote sensing of dune habitats and machine learning for remote sensing. In Chapter 3 the regions of interest, fieldwork campaign, the dune habitat classes, the selected data and the preliminary analysis of these data are described. Chapter 4 describes the methodology followed for this thesis. Chapter 5 describes the experimental results. The discussion of the results and methodology will be in Chapter 6. Finally, the conclusions and recommendations are discussed in Chapter 7.

2

BACKGROUND INFORMATION

In the previous chapter, the problem statement and research objective are introduced. This chapter explains the necessity to monitor dune habitats more in detail and how remote sensing can significantly contribute to the monitoring. Moreover, background information about coastal dunes and dune habitats, remote sensing platforms, resolution and machine learning is reviewed.

2.1. COASTAL DUNES

Biodiversity is not just the variety of plants and animals on Earth or, in this context, in a dune habitat. It is also the connection between these existences. When one disappears, others are inevitably affected. Biodiversity provides ecosystem services, which are crucial for the existence of human beings, animals and plants. Yet, there is a worldwide biodiversity loss for which humans are primarily responsible (Ontario Parks, 2019).

Due to the construction and use of Maasvlakte 2, there is an expectation for losses of existing high quality dunes in the nearby areas (Van der Meulen et al., 2014). Spanjaards Duin has been constructed as a compensatory man-made dune area. The target habitats for compensation include grey dunes and humid dune slacks. Although this thesis is more focused on the engineering investigation of how the resolution of remote sensing products affect the monitoring of the different habitat types at a distance, it is always beneficial to have a basic understanding of the ecological side of the project.

2.1.1. DUNE SYSTEM

In the Concise Oxford Dictionary of Current English a dune is defined as "a hill of loose sand built by aeolian processes (wind) or the flow of water." (Seidl and McMordie, 1982). Worldwide, dune systems can be found mostly in desertous areas or behind (former) coasts (Tsoar, 1994) (see Figures 2.2 and 2.3). Figure 2.1 gives an overview of the distribution of major dune coasts in the world. There are also sub-aqueous dunes (dunes under water), but these are out of the scope of this project. The focus will be on aeolian coastal dunes.

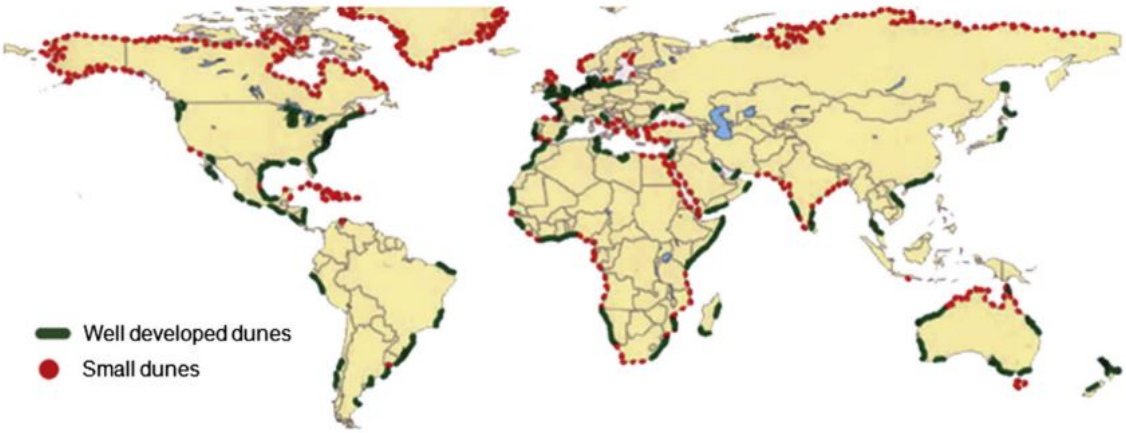


Figure 2.1: Distribution of the major dune coasts of the world, including well-developed dune systems (green) and small sand dunes interspersed with sandy shores, rocky headlines, sand marshes, or mangroves (red) (Mclachlan and Defeo, 2018).

.

Coastal dunes are formed by aeolian transport, which is the movement of sand in air. Initiation of this transport depends on the characteristics of the sediment (material, size, etc.), wind velocities, beach morphology, moisture content, and the degree of roughness elements present (Sloss et al., 2012). Higher up the shore, perennial grasses, such as Marram grass (Ammophila arenaria; beach grass) can act as a sand trap, resulting in the formation of a foredune, see Figure 2.3. Other than clusters of vegetation, litter can help trapping sand. In some cases, this could contribute to the formation of coastal dunes as well.

Figure 2.2: Coastal dune in long-shore direction. The coastline is west of the viewing direction. Image taken on 20th of August 2019 near Spanjaards Duin.

Figure 2.3: Coastal dune in cross-shore direction. The coastline is in the viewing direction. Image taken on 20th of August 2019 near Spanjaards Duin.

A simple cross section of a coastal dune is visualized in Figure 2.4. When dunes reach a certain height, they start to erode and become the source of the landward aeolian transport. This shows that dunes in a coastal system grow in phases: embryo dunes become foredunes, foredunes become white dunes, white dunes become grey dunes etc. Due to spring and storm tides, the seaward sides of the dunes are limited to grow back

seaward. Strong landward winds lead to sand erosion at the seaward side, causing deposition on the leeward slope (landward side) of the growing dunes. This causes the dune to move further landward.

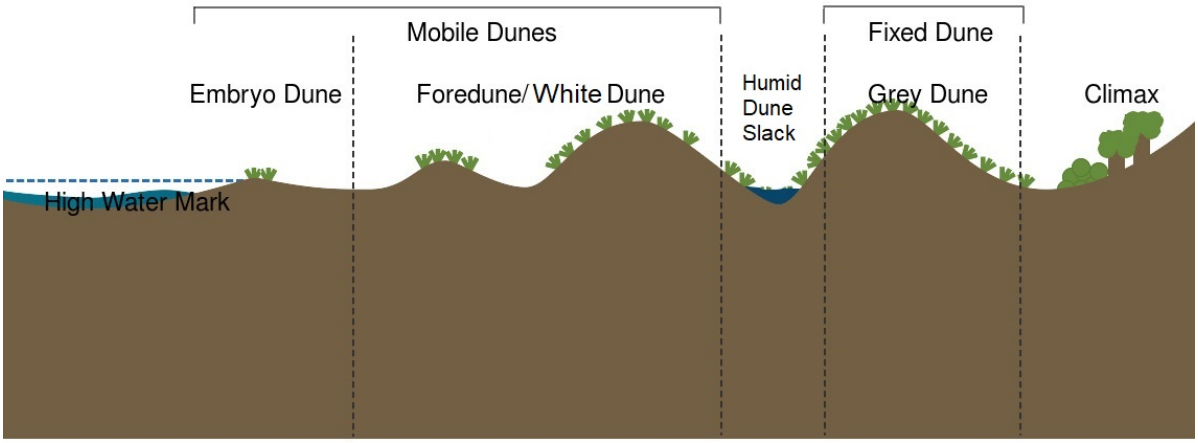


Figure 2.4: Sand dune succession. The visualisation is not to scale. After Jackson (2014).

Other than the sand transport and the vegetation to trap the sand for a good basis dune, the groundwater table is important in dune ecosystems. It determines the lowest level to which wind can erode sand and has a big influence on the development of vegetation. Based on the depth of the water table, the dunes are distinguished in four zones by Ranwell (1972) (Mclachlan and Defeo, 2018):

1. a semi-aquatic habitat where the water table is less than about 0.5m below the surface of the sand,
2. wet slacks, where the water table is about 1m deep,
3. dry slacks, in which the table is 1-2m deep,
4. the open dune habitat, where the water table characteristically lies more than 2m below the surface.

The depth of the water table has been further investigated for Spanjaards Duin and documented in the Spanjaards Duin annual report of 2018 written by Deltares.

A water table more than 3m below the dune surface cannot contribute to the vegetation requirements, as the capillary rise of water is no more than 40 cm. In the case that vegetation cannot reach the water table, they are dependent on rainfall and on condensation. This is often not the case with humid dune slacks (Mclachlan and Defeo, 2018). In Figure 2.5 the important vegetation gradients are shown per zone in the coastal dune cross-section. The behaviour of the dune while growing is summarized in some of the gradients, e.g. canopy height increases and becomes denser when moving more landwards, vegetation closer to the shore have higher salt tolerance, etc.

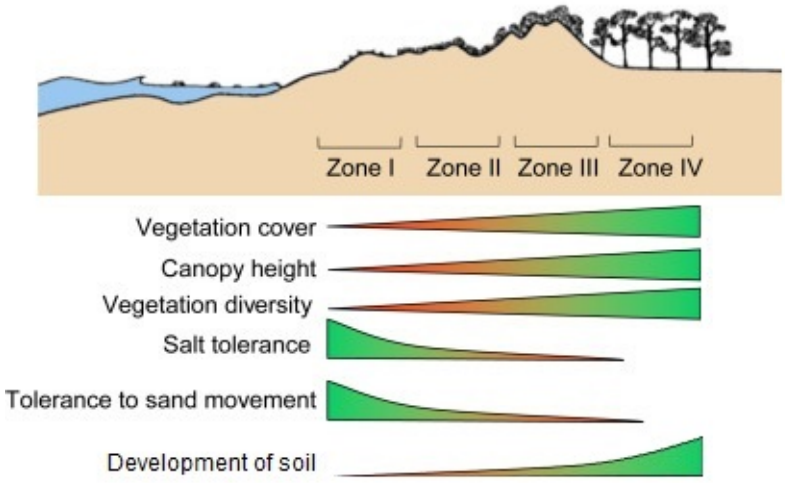


Figure 2.5: Vegetation gradients per zone (After Mclachlan and Defeo (2018)).

In Figure 2.4 the dunes are classified into mobile dunes and fixed dunes in the short term. The focus will be on the white dunes, humid dune slacks and grey dunes.

2.1.2. DUNE HABITAT TYPES

Coastal dune habitats have high biodiversity in terms of animal and vegetation species. They cover 34% of the world's coasts and provide important ecosystem services to humans (Prisco et al., 2013). As previously mentioned, dune vegetation is sensitive for nitrogen deposition. Increased nitrogen deposition will speed up grass encroachment leading to a loss of protected area. It is causing fragmentation and a dramatic recession in the distribution and quality of dune habitats. Coastal dune management and conservation have become demanding issues and deserve much attention (Prisco et al., 2013).

In the following paragraphs, the target coastal dune habitats and the possible coastal dune developments within the timeframe of this research are briefly described. All information is taken from the profile descriptions of the habitat types as defined by the Ministry of Agriculture, Nature and Food Quality (Dutch: Ministerie van Landbouw, Natuur en Voedselkwaliteit) (Ministerie van Lanbouw, Natuur en Voedselkwaliteit, 2008).

GREY DUNES

Grey dunes (H2190) arise behind the foredune (see Figure 2.4) in places where the wind dynamics are sufficiently low as the creation of closed vegetation with herbs and mosses is possible here. Natural dynamics caused by gravity, precipitation and grazing by rabbits ensures the maintenance of the habitat type. Due to the positive influence of shifting sand, bare sand spots (blow-out spots) in closed vegetation also belong to the habitat type. The ecological variation of the habitat type is large, which is related to the lime content (in the top layer of the soil) and the thickness of the upper humus layer. In Figure 2.6, grey dunes are shown with high plant growth. Figure 2.7 shows a grey dune scene with a humus layer. The humus layer is characteristic for grey dunes, with its short and open vegetation landscape.

Figure 2.6: Grey Dunes high plant growth. Image taken on 20th of August 2019 at Van Dixhoorndriehoek (Solleveld & Kapittelduinen).

Figure 2.7: Grey dunes with humus layer. Image taken on 21st of August 2019 at Voornes Duin.

Grey dunes can be divided into three subtypes based on the lime content of the soil: A (rich in calcium), B (poor in calcium), C (transition zone to humid dunes). The type of soil causes different sorts of vegetation to grow. The transitions between the subtypes are gradual. The vegetation of subtype C is alternating with vegetation from subtype A and B, forming complexes or successions of zones. Grey dunes often meet white dunes with Marram grass (Figure 2.3), humid dune slacks (Figures 2.8 and 2.9) and shrubs. Low vegetation (shorter than 50 cm), no or very little shrubs, existence of blow-out spots and rabbit holes are the main characteristics of grey dunes. The optimal functional size is tens of hectares for subtype A and B, and a few hectares for subtype C.

To retain the quality of the habitat, it is essential to keep the vegetation short and landscape open. Without the removal of biomass and with transfer of seeds, coarse grasses will grow high (grassification) at the expense of herbs and other pioneer species that depend on the open structure of the habitat. Moreover, development of shrubs and eventually trees will take place (encroachment). Biomass removal is generally obtained through rabbit grazing. With relatively little rabbit activity and an increase of atmospheric nitrogen deposition, additional management, such as grazing with cattle or manually removal of the biomass (dune management activities), is required. Grey dunes are very sensitive for nitrogen deposition.

HUMID DUNE SLACKS

Humid dune slacks (H2190) is an extensive habitat type. It contains open waters, moist grasslands, low swamp vegetation and reed lands, all of which occur in layers in the dunes. Due to the broad ecological variation, the number of characteristic species is very large. These species are in relatively young succession stages. In Figures 2.8 and 2.9, two scenes of humid dune slacks are shown: open water with transitions to humid dunes of little vegetation growth and moist grassland fully grown with characteristic species, such as Water Mint and Parnassus (Dutch: Krielparnassia).

Figure 2.8: Humid dune slack type A: open water. Image taken on 21st of August 2019 at Voornes Duin.

Figure 2.9: Humid dune slacks grassland. Image taken on 21st of August 2019 at the Parnassia Vlak, Oostvoorne in Voornes Duin.

Humid dune slacks can develop in two ways. Primary humid dune slacks arise by cause of beach cut off from the sea by dunes. This is notably visible on the northeast coast of Goeree & Kwade Hoek. Secondary humid dune slacks originate in the wake of mobile dunes. Nowadays they develop only because the valleys are blown out till the groundwater level. Moreover, secondary humid dune slacks can be created by design measures (Spanjaards Duin).

Within humid dune slacks, there are a wide variety of conditions. For this reason, the humid dune slacks have been split into a number of subtypes. Water depth, vegetation structure and lime content determine the differences between the subtypes. Open water, subtype A, are found in the lowest parts of the dune area, where on average the water is above ground level and will only dry out shortly in the warmer growing seasons. Subtype B, wet grasslands rich in calcium, are wet during winter and dry in the warmer growing seasons. Inconsistent dynamics of the system can also occur for years. In these cases, the dunes will be almost permanently under water or completely dry. This can lead to dramatic shifts in the vegetation composition. However, in a natural dune system with sufficient wet valleys and a lot of variation in the surface level, the population resilience is sufficient to survive these kinds of extremes. Subtype C are wet grasslands poor in calcium. Swamp vegetation, subtype D, is vegetation with high marsh plants such as reed. They are mainly positioned on the edges of dune waters (subtype A). Swamp vegetation lives in shallow waters for most of the year, or permanently. The storage of shrubs, trees and coverage of tall grasses in humid dune slacks is limited. Humid dune slacks are generally sensitive for nitrogen deposition.

DUNE DEVELOPMENT

It is beneficial to conceptualize what the possible dune developments are in the years between 2012 and 2019. This can certainly help analyse the results later in the project. In 7 years time, dramatic changes in the habitat types are not expected. A schematic overview, Figure 2.10, is developed to show the potential changes that could possibly occur in the habitat types in these years.

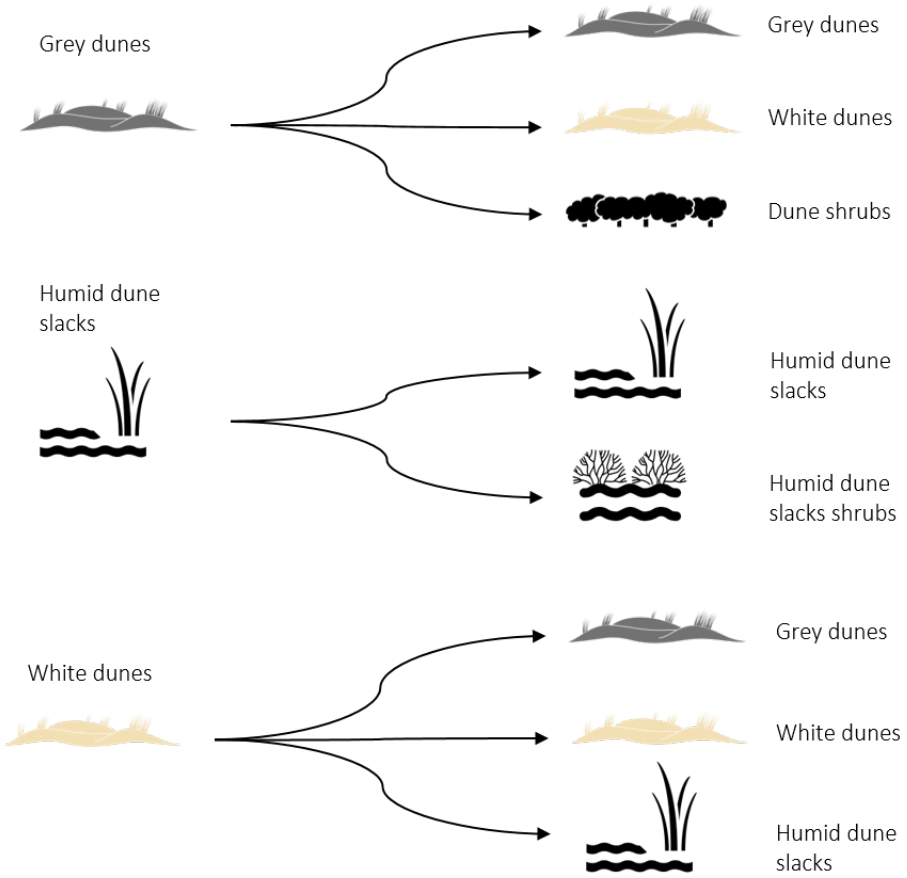


Figure 2.10: Possible change paths in dune habitats between 2012 and 2019.

There are three possible paths: rejuvenation, no change and encroachment. Grey dunes can develop according to all three paths. They can develop respectively to white dunes, stay as grey dunes or become dune shrubs. There is no expectation that humid dune slacks will dry out and develop directly back to grey dunes in this period. They need to first develop back into white dunes. White dunes can develop into grey dunes or humid dune slacks, which is actually the goal at Spanjaards Duin. Encroachment is the unwanted path: the takeover of protected habitat area by dune shrubs. When humid dune slacks are encroached by dune shrubs, they develop into 'humid dune slacks shrubs'. In these 7 years, it is expected that predominantly dune shrubs have been developed due to increase in nitrogen emission. Moreover, in the engineered dunes of Spanjaards Duin, transformations from white dunes to humid dune slacks are occurring, which are expected to be detectable in the results.

2.2. REMOTE SENSING OF DUNE HABITATS

Remote sensing is the acquisition of information on a target without making any physical contact with the target (Baghdadi and Zribi, 2016). It can be subdivided into passive and active remote sensing. Passive sensors only measure the radiation emitted and reflected by the target or its surroundings. In most cases, the source of the radiation is the reflected sunlight. In contrast, active sensors emit radiation themselves and measure the radiation back scattered or reflected back by the target. The focus in this research will be on passive remote sensing.

Remote sensors are installed on many different platforms; spaceborne (Sentinel, Landsat, Cubesat, GPS, etc.), airborne (aircrafts, helicopters, unmanned aerial systems (UAS), free-floating balloons, etc.) and ground-based (hand held cameras, ships, cars, towers, etc.) platforms.

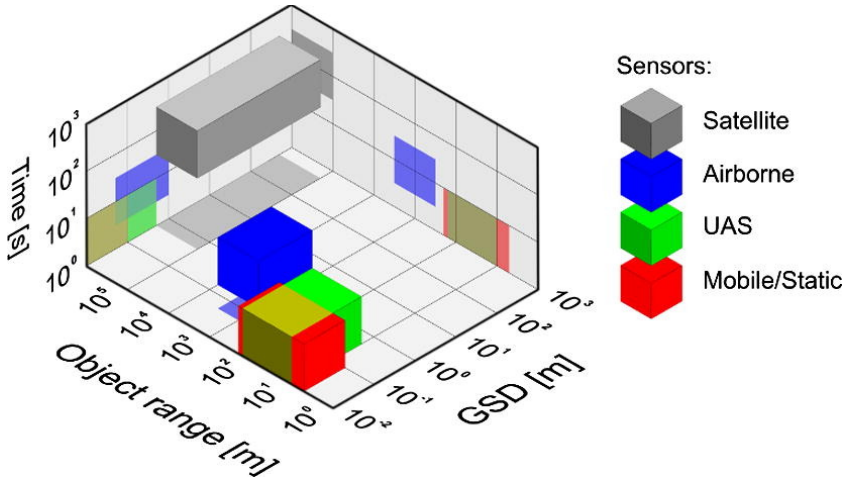


Figure 2.11: The observation cube of remote sensing (Toth and Jóźków, 2016).

The observation range in remote sensing is dependent on mostly four parameters: Ground Sampling Distance (GDS); the spatial resolution, object range; the average distance between the sensor and the target (on Earth), the spectral resolution of the sensors and the revisit time (Toth and Jóźków, 2016). Three out of the four important parameters are visualized as a clear summary on how the different sensors relate to each other based on time, object range and spatial resolution in Figure 2.11.

The focus in this research will be on spaceborne and airborne remote sensing. More specifically, multi-spectral satellite images and aerial images. Satellite images are very popular due to their global coverage and their short revisit time, depending on the orbit in which they are operating. Satellites operate in different types of orbits: polar, non-polar/low-Earth orbit and geostationary (NASA, 2019b). Orbit altitudes can vary with the different types of orbits.

2.2.1. OPTICAL MULTI-SPECTRAL REMOTE SENSING

Electromagnetic energy travels as waves through the atmosphere and space in different wavelengths and frequencies. The various types of waves are summarized in Figure 2.12. The human eye can only detect energy in wavelengths of the visible light (Red, Green and Blue). Energy on other wavelengths also contain a lot of information. To be able to detect these waves, proper instrumentation is needed. The instrument which can detect other waves in the electromagnetic spectrum is called a radiometer. Multiple radiometers together form one multi-spectral instrument. Each multi-spectral instrument is able to detect multiple spectral bands, defining the spectral resolution of the instrument (Stars-project, 2019). Different land cover types have different spectral signatures with respect to the various wavelengths (Zhao et al., 2020). A higher spectral resolution is convenient for differentiation between land cover types. Therefore, multi-spectral satellite images are used in this research.

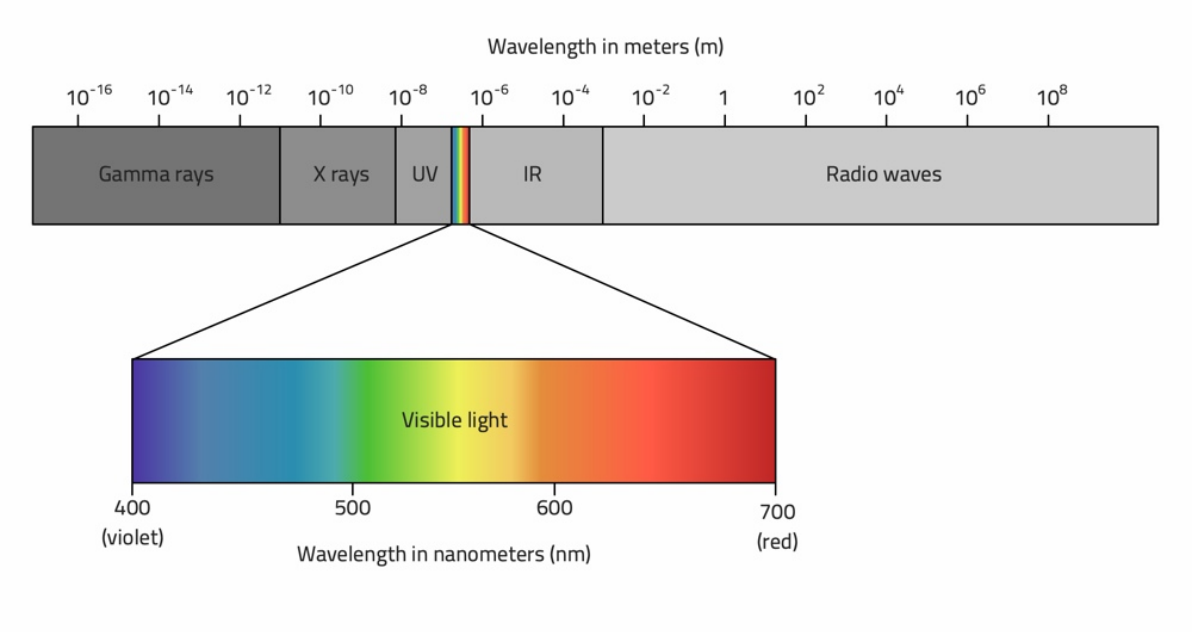


Figure 2.12: The electromagnetic spectrum zoomed in on the visible domain (Radio2Space, 2013).

2.2.2. SPECTRAL DIFFERENTIATION OF DUNE HABITATS

The spectral reflectance is the energy fraction reflected back from the surface of a target and is measured as a function of wavelength. Each type of surface on the Earth has a different spectral signature as each material reflects different in the each spectral domain. Reflectivity of vegetation and soil is, among other things, dependent on the colour, internal and external structure of the vegetation, and texture of the surface. The typical spectral signatures for wet soil, dry soil, vegetation, clear lake water and turbid river water are plotted in Figure 2.13.

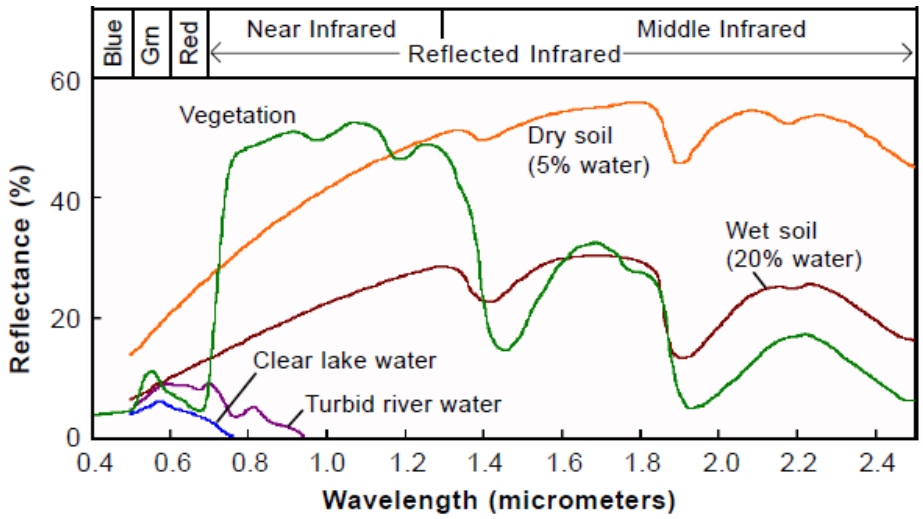


Figure 2.13: Typical spectral signatures of different kinds of surfaces (Daneshgar, 2015).

For the classification of coastal dunes, spectral characteristics in signatures of all three types of land covers in Figure 2.13 are important: vegetation, soil and water. The spectral signature of healthy green vegetation is recognizable by the strong reflectance in the Near Infrared (NIR) domain and relatively weak reflectance in the visible domain of the spectrum. Vegetation reflects NIR light to avoid evaporation as a result of possible heating. Chlorophyll in leaves absorb blue, but more red light to be able to perform photosynthesis, which is needed to generate fuel for the activities of the organism. Healthy vegetation appears green due to higher reflection in the green band of the visible domain. The degree of reflection in the NIR domain gives information

about the internal structure and health of the leaf, and is therefore mostly the domain to distinguish different types of vegetation. The transition zone between the Red and NIR domain, is called the Red Edge. The strong absorption in the Red and reflection in the NIR domain is indicative for vegetation (see Figure 2.13) and can be used to differentiate vegetation from e.g. soil. The Normalized Difference Vegetation Index, NDVI, is one of the most frequently used vegetation indices in remote sensing to detect vegetation and makes use of this Red Edge zone. Water molecules strongly absorb in the Middle Infrared domain (MIR) of the spectrum. The reflection increases with decreasing leaf water content. This range of the spectrum is mostly for vegetation stress detection as a result of drought. The two dips in the MIR domain are due to strong absorption of energy by water molecules (O-H bond in water), cellulose and other organic compounds in leaves (Curran, 1989).

Other than vegetation, there is water, wet sand and dry sand in the targeted dune areas. Soil follows a near-linear trend in the visible and NIR domains. In the MIR domain, depressions are located at the water absorption bands, more specifically at wavelengths 1.4, 1.9, 2.2 and 2.7 μm. The water content in the soil lowers the reflectance of the entire spectrum. The color of the sand on the Dutch coastal areas is mostly white and will therefore reflect more in the visible spectrum. The particle size of the sand, texture, surface roughness and presence of mineral/organic matter, are other factors influencing the reflectance. Clear water has little to no reflectance in the NIR and MIR domains. The highest reflectance is in the Blue and Green portion of the visible domain. Turbid waters, suspended waters with sediment or discolored waters due to e.g. algal bloom, have a higher reflectance in the visible domain of the spectrum and also in the first wavelengths of the NIR domain.

2.2.3. RESOLUTION OF REMOTE SENSING DATA

In geosciences, three types of resolution play an important role: spatial, spectral and temporal resolution. The specifications of the platform and the sensor determine the resolution of the data (Liang et al., 2012). Spatial resolution is the measure of the smallest object that can be detected by a sensor (on a remote sensing platform). Spectral resolution is the number of bands in which the sensor can measure incoming radiance (GEOIMAGE, 2019). For many data products, spectral resolution is often closely linked to their spatial resolution; the higher the spectral resolution, the lower the spatial resolution and vice versa. Temporal resolution is not related to the sensor specifications per se. It corresponds to the frequency of revisits and is dependent on the orbit of the platform around the planet (Ose et al., 2016). Before selecting a specific data product for an application, the spatial, spectral and temporal resolutions are examined. In Table 2.1 a brief overview of the used platforms and their resolutional information is given.

2.2.4. REMOTE SENSING DATA

There are several multi-spectral satellites operating from which the most notorious ones are Sentinel-2 and the Landsat satellites. The multi-spectral sensors on these satellites differ in spatial, spectral and temporal resolution. In Figure 2.14, the spectral bands of Landsat-7, Landsat-8 and Sentinel-2 are compared to each other in an atmospheric transmission graph. Sentinel-2 has more spectral bands than Landsat-7 and 8, and therefore has a higher spectral resolution. Moreover, the spatial resolution of the Sentinel-2 satellite differs per spectral band. For this study, only bands in the visible (Blue, Green and Red), Near-Infrared (NIR) and Short Wave-Infrared (SWIR) spectrum are used. The data products and product specifications are discussed in Section 3.4.

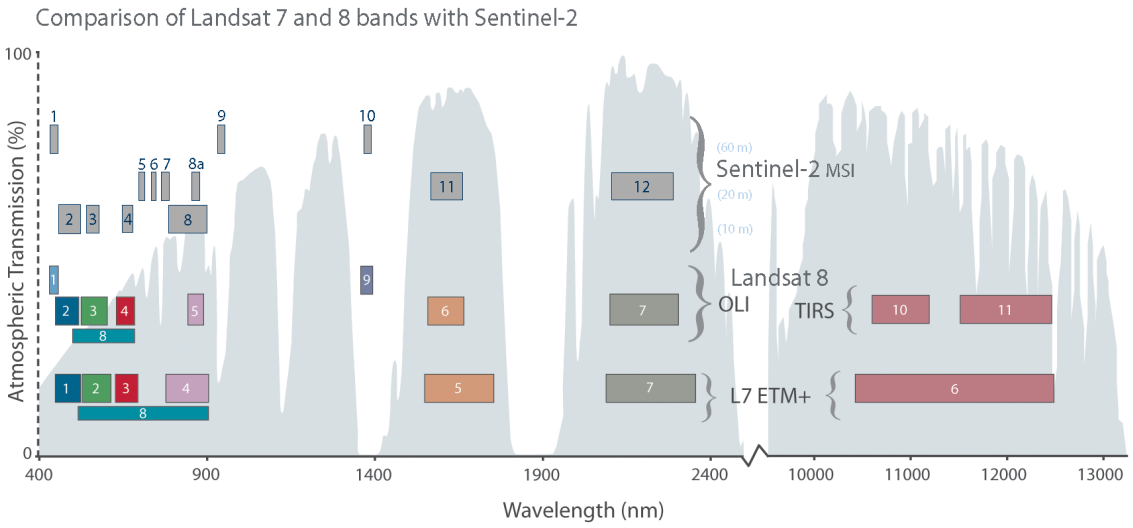


Figure 2.14: Sentinel-2 and Landsat spectral bands in an atmospheric transmission plot (ESA, 2019d).

LANDSAT

The Landsat program, owned by National Aeronautics and Space Administration (NASA), is the longest operating mission for acquisition of satellite imagery of the Earth since 1972 till present. Landsat counts nine satellites and two of them are still operational: Landsat-7 and Landsat-8. Figure 2.15 shows an overview of the operating times of the different Landsat satellites. The time of interest in this thesis is from 2012 till present. The only operational satellite in both 2012 and 2019 is Landsat-7. This is the main reason why, despite the Scan Line Corrector failure, it is decided to continue with Landsat-7 data.

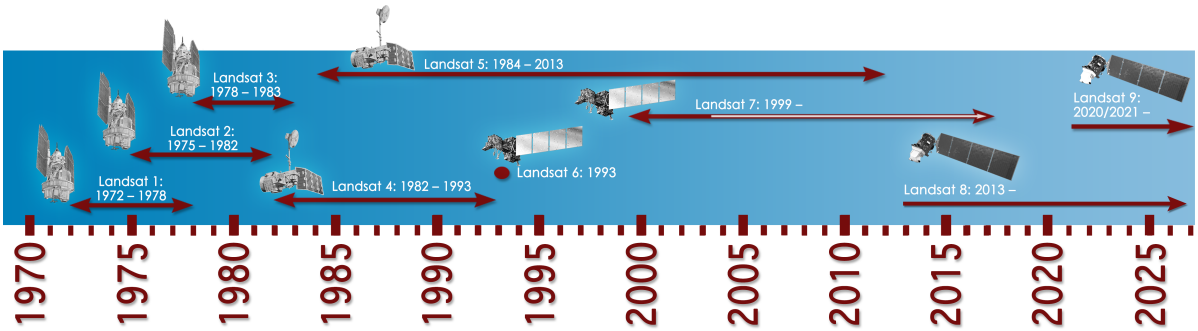


Figure 2.15: The Landsat operating timeline (NASA, 2019a).

From May 31, 2003 suddenly lines of data disappeared on the Landsat-7 data products. This was due to a malfunctioning Scan Line Corrector (SLC), which normally should compensate for the forward motion of the satellite (U.S. Geological Survey, 2017). With an incorrect functioning SLC, the image scanning results in zig-zag patterned products, see Figure 2.16. The usability of the data is examined due to this significant data loss. Several approaches are developed to fill the data gaps in the products as effort to compensate for the loss. These approaches are mostly based on interpolation algorithms (USGS, NASA, Landsat 7 Science Team, 2003). Interpolated data are not real physical measurements and generally only useful for visualization purposes. For scientific research, it is recommended to use the original data as there appear no issues with the geometric and radiometric accuracy of the products. Hence, it has been decided to continue with the data products of Landsat-7. The white space on the arrow of Landsat-7 in Figure 2.15 shows the period of the SLC failure.

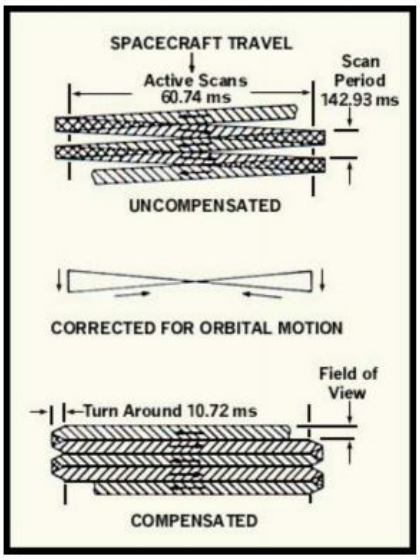


Figure 2.16: An explanatory visualisation of what the Scan Line Corrector should have compensated for and what happens when it does not compensate (USGS, NASA, Landsat 7 Science Team, 2003).

SENTINEL 2

Copernicus is European Union's Earth Observation Programme, which is coordinated and managed together by the European Union and the European Space Agency (ESA). The space component of the Copernicus programme is dominated by the Sentinel space missions (ESA, 2019b). Each Sentinel mission consists of a constellation of two satellites for a shorter revisit time and better global coverage.

Sentinel-1, Sentinel-2, Sentinel-3 and Sentinel-5p are operational now. Sentinel-2 is a sun-synchronous polar-orbiting multi-spectral high-resolution optical imaging satellite for land services. It measures in 13 spectral bands from visible to the SWIR domain of the spectrum (see Figure 2.14).

AERIAL IMAGERY

Airborne remote sensing is relying on images of the Earth taken from the air. This could be from aircrafts, helicopters, unmanned aerial vehicles (UAVs), balloons, etc. Orthorectified aerial images of the Netherlands are accessible for projects of Rijkswaterstaat. True color and false color imagery are available in 25cm resolution.

SUMMARIZED OVERVIEW OF DATA PRODUCTS

In Table 2.1 an overview of the important specifications of the used data products are given. This table describes both the freely available Landsat and Sentinel missions, and the aerial images which are used for the resolution analysis. Pre-processing steps and more details are described in Chapter 3.5.

Table 2.1: Overview of the specifications of the open access satellite imagery and aerial imagery used for this project.

Platform + Sensor	Landsat 7 [ETM+]	Sentinel 2 [MSI]	Aerial imagery
Revisit time	16-days	5-days	on request
Resolution	RGB: 30 m NIR: 30 m SWIR: 30 m Panchromatic: 15 m TIR: 60 m	RGB: 10 m NIR: 10 m SWIR: 20 m Red edge: 20 m -	RGB: 25 cm NIR: 25 cm - - -
Data availability (time)	Apr 1999 till present	Jun 2015 till present	on request

2.3. MACHINE LEARNING FOR REMOTE SENSING

A general definition of machine learning is given by Arthur Samuel in 1959 as "*the field of study that gives computers the ability to learn without being explicitly programmed.*" (Samuel, 1959).

Machine learning has been around for decades now. It is fused into daily life applications e.g. the spam filter in our e-mails. It is great for problems for which a lot of rules or tuning is required, for which there is no good solution at all using a traditional approach, for problems with constant changing environments and for simply getting insight about complex data. Machine learning algorithms are used to reveal patterns and relationships within data, even if they are multivariate and nonlinear. The algorithms do not need any prior knowledge about the nature of these relationships.

Different types of machine learning algorithms exist, each more applicable to certain types of problems. The algorithms can be classified according to the amount of data and type of supervision they get during training (Géron, 2017). The four major categories are: supervised, unsupervised, semi-supervised and reinforcement learning. Only supervised techniques are used in this research and, therefore, only these techniques are further elaborated.

2.3.1. SUPERVISED MACHINE LEARNING

The general goal of supervised machine learning is to describe the input to output relationship of a dataset with a mathematical model. Supervised learning assumes that training examples are labeled (assigned to a class or a value) beforehand. The expectation is that this model will 'generalize' to unseen data (Fürnkranz et al., 2012), such that it can be used to make predictions or decisions. To test how well a model generalizes, a random subset of the available data is left out during training, the so-called test set. Overfitting is a well-known problem in supervised learning and occurs when a complex model fits too closely to the training dataset. It can be recognized when the model achieves a much lower accuracy on the test dataset than on the training dataset. Typical supervised learning tasks are classification (predicting classes) and regression (predicting values). The most important and most popular supervised learning algorithms are: k-Nearest Neighbors, Linear Regression, Logistic Regression, Support Vector Machines (SVMs), Decision Trees, Random Forests and Neural Networks.

2.3.2. CLASSIFICATION

Machine learning algorithms are very useful and successful techniques for a large number of applications in many disciplines of the Earth's system and other planetary disciplines (Barchi et al., 2020, Cabral et al., 2018, Rashno et al., 2017). This research will be focused on a classification application. The most popular and promising supervised classification algorithms in the field of remote sensing are Random Forests and Neural Networks. In recent years, XGBoost became also popular for outperforming Random Forests in accuracy and training time for both regression and classification (Chen and Guestrin, 2016). In this thesis, Random Forest and XGBoost are compared. More information about both algorithms is given in the next paragraphs.

RANDOM FOREST

Ensemble learning receives a lot of interest in recent years. It is based on generating many classifiers and combining their results; classifiers can be of the same type or fundamentally different (Liaw and Wiener, 2001). Boosting and bagging of decision trees are two popular methods (Breiman, 1996, Schapire et al., 1998). Random Forest is an ensemble learning model based on the bagging of multiple decision trees. In bagging, each tree is constructed independently using a random batch of the dataset and is not influenced by the predictions of the other trees in the forest. Afterwards, the single predictions are collected and the majority is chosen to be the final prediction. On top of the randomness that results from sample bagging, random selection of features is also implemented in the Random Forest algorithm. This extra randomness has to minimize the correlation between the trees much as maintaining the strength, which improves the overall accuracy (Breiman, 2001).

The formal definition of Random Forest is given in Definition 2.3.1 (Breiman, 2001).

Definition 2.3.1. *A Random Forest is a classifier consisting of a collection of tree-structured classifiers* $\{h(\boldsymbol{x}, \Theta_k), k = 1, ...\}$ *where the* $\{\Theta_k\}$ *are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input* $\boldsymbol{x}$.

A decision tree (tree-structured classifier) is a very intuitive model and is similar to how humans categorize objects: a series of yes or no questions about the data leads to the predicted class. Figure 2.17 shows a simplified Random Forest including the bagging process, aggregation of the single predictions and the final prediction. Only a subset of the data is seen by each tree due to the use of sample bagging and feature bagging (feature randomness). Therefore, the number of trees in a forest should not be too small, or this will result in a decrease in the predictive power of the Random Forest.

The law of large numbers describes that when performing an experiment a large number of times, the average of the results obtained from the experiments should be close to the expected value (Hsu and Robbins, 1947). In Random Forest, this means that the larger the number of trees, the more it converges towards the expected value. From this, one can conclude that more trees in a Random Forest will never lead to more overfitting. However, there is a potential of overfitting by wrongly tuning other hyperparameters (Segal, 2003).

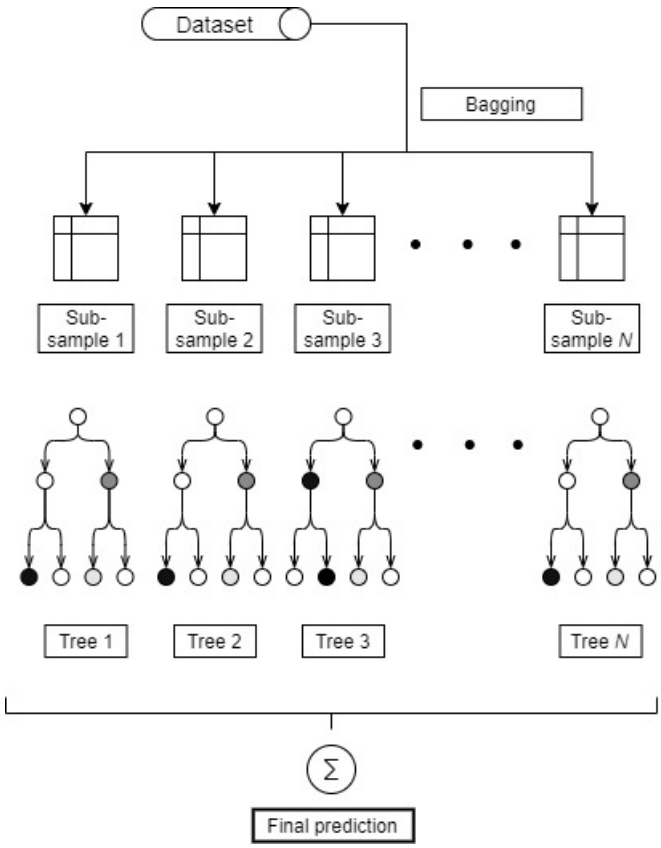


Figure 2.17: Graphical illustration of bagging, single prediction aggregation and final prediction decision in Random Forest.

XGBOOST

XGBoost (eXtreme Gradient Boosting) is a very popular ensemble learning method based on boosting, and is introduced by Chen and Gaustrin (Chen and Guestrin, 2016). In contrast to bagging (Random Forest), the individual trees in boosting are not constructed independent from each other. The weights are influenced by the predictions and errors of the previous tree (see Figure 2.18). Wrong prediction and high error means higher weight and vice versa. Cases difficult to predict get more focus in boosting algorithms. Gradient Boosting Machines are similar to XGBoost, but XGBoost is a more optimized algorithm. XGBoost is popular for its significantly lower training time, and better performance in most problems, compared to other ensemble learning methods.

XGBoost has three types of parameters: general parameters, booster parameters and task parameters. There are two ways to control overfitting in XGBoost: control model complexity or make training robust to noise. Key hyperparameters to tune are: *learning rate, number of estimators, maximum tree depth, colsample_bytree: subsample ratio of columns when constructing each tree, subsample ratio* of the training instances and *gamma: minimum loss reduction required to make a split.*

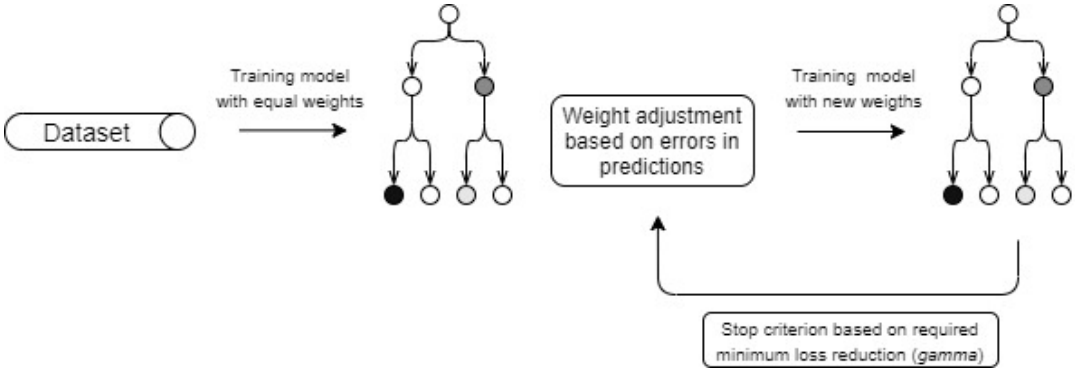


Figure 2.18: Graphical illustration of weight adjustment in boosting.

3

DATA AND AREA DESCRIPTION

In this chapter the regions of interest, description of the used data, preliminary data analysis and corresponding pre-processing steps are discussed. Moreover, the fieldwork campaign and target classes are described.

3.1. REGIONS OF INTEREST

With the arrival and use of Maasvlakte 2, a change in the dune landscapes of Voornes Duin and Solleveld & Kapittelduinen is expected due to an increase in nitrogen emission. This increase ensures that vegetation growing on nutritious soils will outcompete characteristic vegetation species in the dunes. The Natura 2000 area Goeree & Kwade Hoek is a reference area due to the presence of high quality grey dunes and humid dune slacks. Therefore, this area is designated as region of interest (ROI) as well. Figure 3.1 shows the official outlines of the Natura 2000 areas according to the Habitats Directive (European Commission).

3.1.1. SOLLEVELD & KAPITTELDUINEN

This Natura 2000 area owes its name to the two nature monuments in the north (Solleveld) and in the south (Kapittelduinen) of the area. Officially, the area is designated as Natura 2000 site in September 2011. The majority of the dunes in Solleveld are decalcified. Therefore, this area differs from most other dune areas in the Netherlands. Kapittelduinen consist for the most part of very young man-made dunes, such as Spanjaards Duin.The area consists mainly of white dunes, dune forests, grasslands (grey dunes), shrubs and small lakes (humid dune slacks).

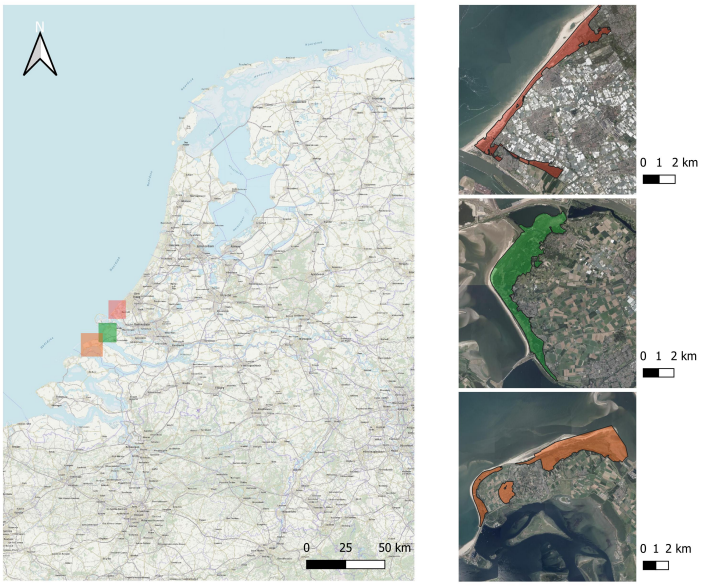


Figure 3.1: The official Natura 2000 outlines of the areas. From top to bottom: Solleveld & Kapittelduinen, Voornes Duin, Goeree & Kwade Hoek.

3.1.2. VOORNES DUIN

Voornes Duin is a Natura 2000 area since February 2008. Voornes Duin belongs to the youngest, calcium rich dunes that were developed in the Late Middle Ages. The landscape changed dramatically during the twentieth century. Overgrazing caused too much sand exposing of the 19th century. The government intervened in 1910, which resulted in a ban on grazing, logging and in planting of forest.

In the middle and east of the Netherlands, groundwater is often used as a source of drinking water. Near the coast, groundwater is too salty to be used as drinking water, with the exception of groundwater in dune areas. Voornes duin could not be used for water extraction due to a relatively small size of sand packages in the geological layers. Therefore, many animal and plant species got the chance to accommodate in Voornes Duin. The poorly permeable clay and peat layers also ensured that there are still many wet and humid environments (Natuurmonumenten, 2020). Besides many humid dune slacks, several small ponds and marshes, the area is also home to two large lakes: Breede Water and Quakjeswater. Moreover, there is a lot of forest and dune grasslands (grey dunes) with shrubs (e.g. Buckthorn formations). Many transitions can be recognized in the area; from wet to dry, nutrient poor to rich, calcium poor to rich and fresh to salt water.

3.1.3. GOEREE & KWADE HOEK

Goeree & Kwade Hoek is a Natura 2000 area since February 2008. It comprises a number of dune areas on the northwestern side of Goeree and Kwade Hoek on the sea side. De Kwade Hoek owes its name to the fact that ships got stuck on the sandy shoals during storms. Kwade Hoek is the northernmost part of the intertidal area of the Voordelta and forms the transition from salt marsh to beach.

A large area of the sandbanks run dry at low tide and continue to grow every year. Geological processes that played a role in the development of the Dutch coast can still be observed daily in the area. The area consists of the sea side of the beach, where spontaneous dunes and mudflats have developed. Because these muddy flats of shoals are flooded with seawater every day, they are hardly grown with vegetation. More inland lie salt marshes that are intersected by winding creeks. Wet primary dune valleys have developed behind the dunes. It is therefore a omnifarious and dynamic landscape with primary dune formation, mud flats, salt marshes, humid dune slacks, grasslands (grey dunes), dune forests, white dunes and shrubs.

3.2. FIELDWORK CAMPAIGN

In August 2019, a two-day fieldwork campaign has been conducted in the ROIs Solleveld & Kapittelduinen and Voornes Duin, during which measurements were taken with a field spectrometer to get more insights in the different habitat types. The field spectrometer has a very high spectral resolution. The purpose of this fieldwork is to make use of this resolution and get detailed insight of the spectral reflectances of the different land cover types (habitat types). An additional purpose is that these measurements could be used as possible ground-truthing spectral reflectance data for the classification maps in 2019. The locations for the measurements are selected beforehand using the classification from 2012 (Mouissie et al., 2014) and recent imagery as guidelines. GPS locations are taken at each measurement point. Figures 3.4 and 3.5 show the fieldwork measurement locations and the footprint of the measurements. Figure 3.3 shows the fieldspectrometer on my back during the second day of the fieldvisit.

After assigning a location as a measurement point (black dot in Figure 3.2), the coordinates and the spectral reflectance are measured. The measurements are repeated three or four times around the centre point with a radius of approximately 2 meters (2 big steps). This is done to create a larger measuring area while remaining close to the original point. At each measurement point, the reflectance is captured three times to reduce the chance of an incorrect reading. This data is potentially suitable to be compared to a Sentinel-2 image taken on the closest date to the fieldwork days.

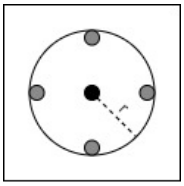


Figure 3.2: Sketch of measuring method during fieldwork campaign. The black dot is the starting point, the grey dots are the additional measurement points and r is the radius of approximately 2 meters.

An important factor for the measurements is the illumination. Like an optical satellite, a spectrometer measures natural light reflected from the surface. It is not favorable that the light intensity, illumination, changes due to, for example, a moving cloud blocking (a fraction of) the sunlight during a measurement. Extra attention is paid to prevent measuring during those conditions. An important operation while doing the measurements is therefore white referencing of the field spectrometer. This is done using Spectralon, which is a white panel with nearly 100% reflection within the wavelength range of 350 nm to 2500 nm. A Spectralon white reference scatters light uniformly in all directions within that wavelength range.

Figure 3.3: The ASD fieldspectrometer device in the field on field-campaign day 2.

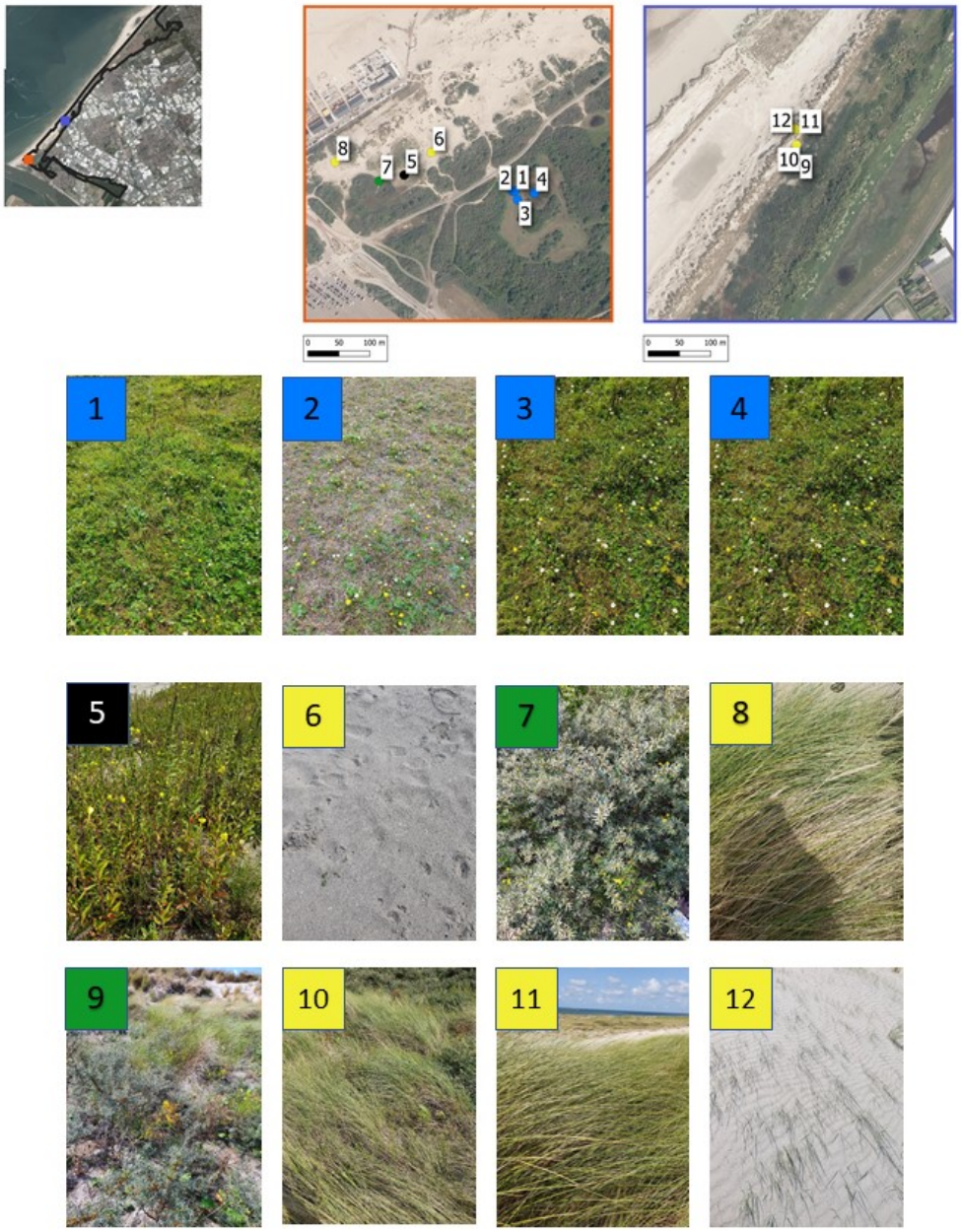


Figure 3.4: Fieldwork campaign measurement locations day 1. Blue points are *humid dune slacks*, black points are *grey dunes*, yellow points are *white dunes* and green points are *buckthorn formations*. The vegetation in point 12 belongs to *Festuca* species and belongs to both white dunes as grey dune species. The background maps are aerial images of 2019 (WMS).

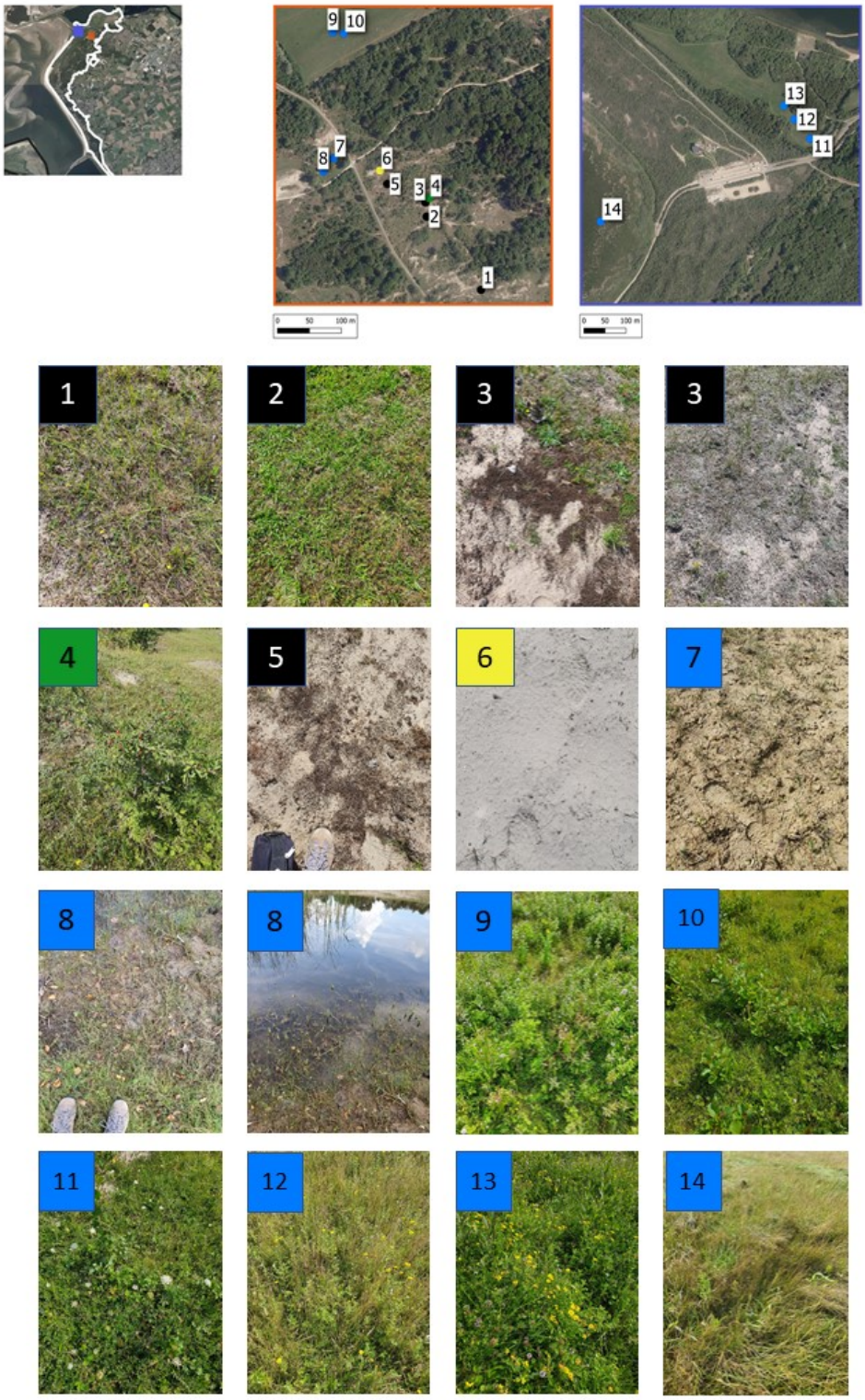


Figure 3.5: Fieldwork campaign measurement locations day 2. Blue points are *humid dune slacks*, black points are *grey dunes*, yellow points are *white dunes* and green points are *buckthorn formations*. The background maps are aerial images of 2019 (WMS). Point 14 does not belong to the official Natura 2000 area.

3.3. DUNE HABITAT CLASSES

In subsection 2.1.2, dune habitats and their subtypes are described based on the traditional definitions (Synbiosys Alterra). The most distinct differences between the subtypes is the moisture content and the calcium content of the soil. In general, the dunes south of Bergen aan Zee are rich in calcium and north of Bergen aan Zee are poor in calcium. Local decalcification in calcium rich dunes can start due to production of organic acids by plants and organic material. Traditional mapping uses additional experiments to determine the local calcium content of the soil. Mostly soils closer to the shoreline are richer in calcium, but there are regions where this is not the reality (e.g. Solleveld & Kapittelduinen; in Solleveld the vast majority of the dunes are decalcified). The decalcification process lowers the pH value of the soil. This ensures adaptations in vegetation type (Bij12, 2019).

From a first visual preliminary investigation on the aerial images of 2012, it is concluded that there is not a clear visual differentiation in soils rich or poor in calcium. Therefore it is assumed that, based on spectral signals in the visible domain, it is hard to differentiate between sub-habitats A and B for both grey dunes as humid dune slacks. The hierarchy described in subsection 2.1.2 are retroactively preserved in the habitat classifications. Pixels are first classified as sub-habitat and afterwards derived to main-habitat. More on hierarchical classification is discussed in subsection 4.5.1.

The classification done in 2012 was mostly focused on the mapping of humid dune slacks and grey dunes. Although these habitats are the main focus in this research as well, other habitats must be added as target classes in order to perform a realistic area covering classification. The additional classes are buckthorn formations (H2160), white dunes (H2120) and wooded dunes (H2180). These classes were covered in the classification of 2012, but not as extensively as grey dunes and humid dune slacks.

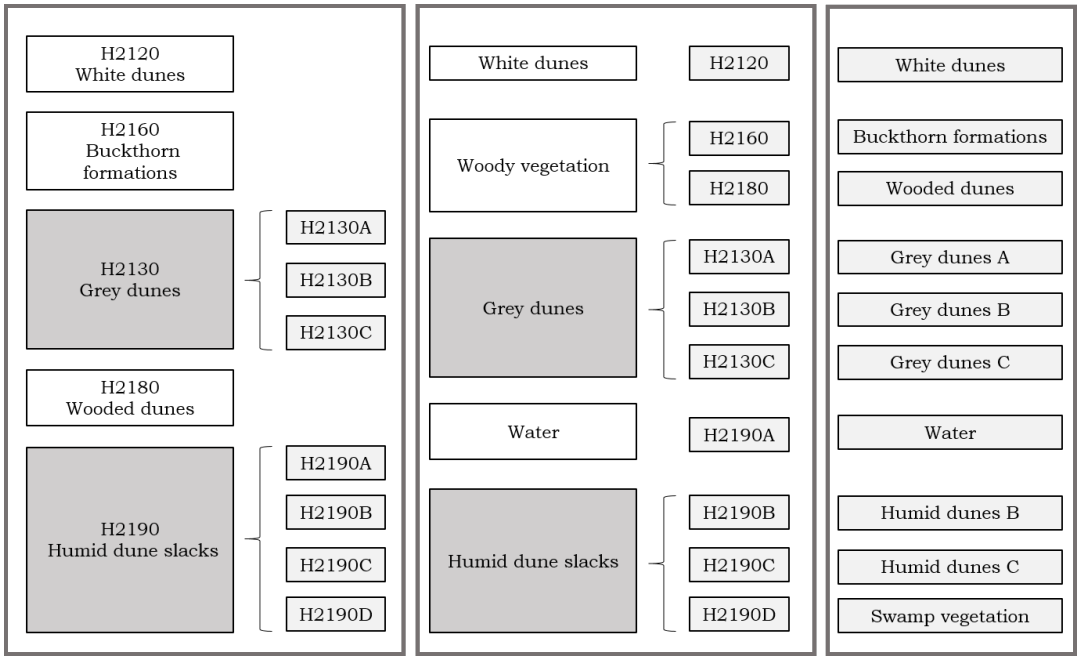


Figure 3.6: Left: the traditional hierarchy, middle: the hierarchy used in this research with main-habitats (left column) and sub-habitats (right column), right: the legend entries of the corresponding sub-habitat type.

The left scheme in Figure 3.6 illustrates the original thematic classification hierarchy, the middle scheme the modified hierarchy applied in this research and the right scheme are the legend entries of the corresponding sub-habitats. In the legends, 'dunes' are abbreviated as 'd.', 'formations' as 'f.' and 'vegetation' as 'veg.'. Important changes are the fusion of wooded dunes and Buckthorn formations in one main class, named 'woody vegetation'. Moreover, humid dune slacks A is designated as a separate main class 'water'. In Figures 3.4 and 3.5 the fieldwork point locations are shown and their corresponding footprint. In these images the differences in the classes can be seen. Important to notice is that vegetation coverage is not homogeneous everywhere, especially not in grey dunes. Also in some areas Marram grass in white dunes are planted in straight lines.

3.4. DATA SELECTION AND PRE-PROCESSING

Data from various sources is used to conduct this research. This includes aerial images, satellite images and data from the classification performed in 2012. This section describes the data selections and the pre-processing steps per data type. Moreover, the feature engineering is described.

3.4.1. DATA LABELS

In 2012 and 2013, Grontmij classified the areas of interest as part of the dune monitoring and evaluation program (MEP Duinen) (Mouissie et al., 2014). MEP Duinen has been set up to monitor the nitrogen deposition, dune vegetation (size and quality of humid dune slacks and grey dunes) and other factors influencing the habitat types, such as sand dynamics, dune management, groundwater and soil chemistry. The classification is partly prepared using the eCognition software to recognize groups of pixels as objects and polygonize them. Afterwards, the polygons are assigned to a habitat in the field according to the guidelines of the syntaxonomy of Vegetation of the Netherlands (Schaminée et al., 1995). After pre-processing, this classification is used to assign labels to the training data.

PRE-PROCESSING

The produced polygons contained many geometric errors. These had to be resolved, because simple operations, such as the merging of polygons, could not be executed. Furthermore, the polygons often overlapped multiple habitat types. While this is accurate enough for ecological approaches, it will cause confusion when used as training data in a machine learning algorithm. For classification problems, the training data should be labeled as correctly as possible. To speed up the data cleaning process, a negative buffer is applied to each polygon and each polygon is afterwards adjusted by discriminating parts that could introduce uncertainty and noise. This part of the pre-processing is done in QGIS.

The additional classes described in section 3.3 were under-represented. More polygons of these classes are created and validated by external professionals, Frank van der Meulen and Bert van der Valk. The training data is point data and consists of 10,000 random points per class generated by stratified random sampling inside the pre-processed polygons. The pre-processed polygons from which the points are sampled are visualized in Appendix A.1.

3.4.2. AERIAL IMAGERY

The aerial images are received from Rijkswaterstaat in 25 cm resolution for both true color and false color images. This includes the visual Red, Green, Blue bands and one NIR band. In 2009, an aerial photography steering committee is set up with representatives from all parties involved. The starting purpose is that one collection and multiple use of this collection contributes to the objective of achieving a government-wide basic provision. Since then aerial imagery have been collected nationwide by one party: beeldmateriaal.nl. The very high resolution products of 10 cm resolution do not have the NIR band, therefore the 25 cm resolution products are used.

The flights are usually between the 1st of April and 15th of July each year. One of the most important requirements is that the position of the sun is at least 30 degrees above the horizon and that there is sufficient light available for taking aerial shots. The photos are acquired during low tide. All photos are free from cloud and cloud shadows. The geometric positioning precision in the Rijksdriehoek-system (RD-system) must be better than 37.5 cm (1.5 pixels) for all pixels at ground level. The measurement error is normally distributed with a required standard deviation of at most 25 cm. In 2012, the photographs are taken using the *UC-SXp-1-30719036*. The UltraCamXp is a camera from Vexel Imaging (a Microsoft Company). In the paper of Gruber (Scholz and Gruber, 2008), the spectral sensitivities of the Blue, Green, Red and NIR band can be found.

Table 3.1: Spectral sensitivity of all Bands in the UC-SXp-1-30719036 (Scholz and Gruber, 2008).

Band	From (nm)	To (nm)
Blue	410	570
Green	480	630
Red	580	700
NIR	690	1000

PRE-PROCESSING

The aerial images are already orthorectified and georeferenced. Unfortunately, the areas are not photographed on the same day and even some parts of one area is covered on different days (and time of the days). Looking at Solleveld & Kapittelduinen, the southern part of this area is captured in May and the northern part in July, which most probably will introduce differences in color reflectances for the same type of land cover. Another bias that will be introduced in this way is the difference in phenology. It is assumed that these differences are not very large and therefore negligible for this study. The aerial images are not radiometrically corrected, which means that the Digital Numbers from the aerial image are not equal to the spectral reflectance. Traditionally, radiometric correction of aerial images are done by placing known reflectance targets in the field or calibrating with field spectral measurements when the photograph is captured. There are techniques in which aerial imagery are fused with satellite imagery to correct correct, but this results in loss of resolution. This technique is therefore not suited for this research. It is assumed that the pixel reflectances are equal to the spectral reflectances.

To be able to study the effects of spatial resolutions between 25cm (aerial image) and 30m (Landsat 7), the original aerial images are resampled to coarser resolutions. These resolutions are chosen as 0.5m, 4m, 10m, 20m and 30m. The choice to resample to 10m, 20m and 30m is to simulate satellite images and 4m is chosen as the intermediate resolution. Resampling is done by averaging the pixels to the target pixel resolution. After resampling, the aerial images are cut into the ROI's to decrease the amount of data to work with. The processing for both resampling and data reduction is done in GDAL command line using bash scripts.

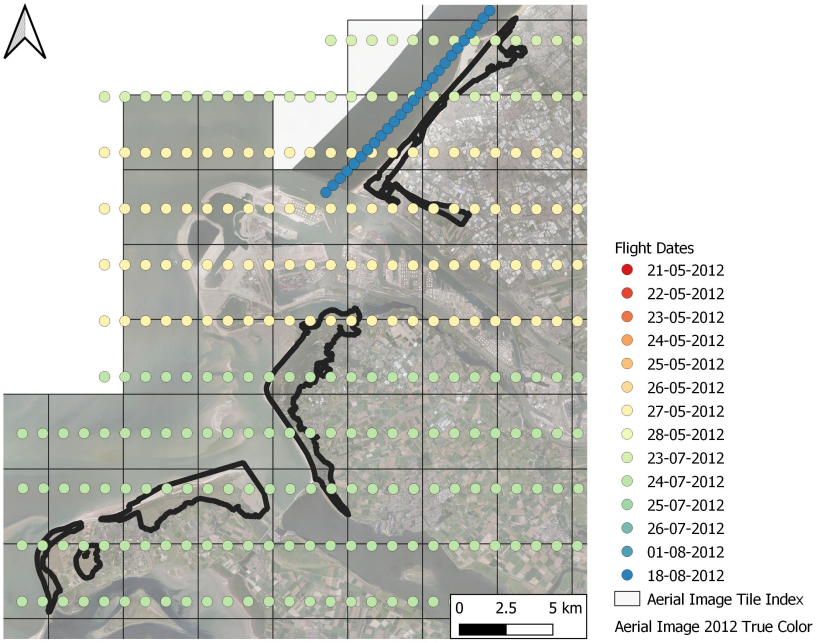


Figure 3.7: Flightpaths during acquisition of the aerial images in 2012. The colors show the flight dates.

3.4.3. LANDSAT 7

The Landsat series of satellites is known for their longest temporal record of Earth observation from space (Roy et al., 2014), see Figure 2.15. Landsat-7 is launched on the 15th of April, 1999. It carries the instrument Enhanced Thematic Mapper Plus (ETM+) and has a global revisit time of 16 days. The ETM+ has 6 spectral bands in the visible, NIR and MIR spectral domains with 30m spatial resolution. Moreover, it carries a thermal infrared channel with 60m resolution and a panchromatic band with a spatial resolution of 15m. See Figures 2.14 and A.8 for more detail. For this research, the thermal infrared and panchromatic bands are not used. The data acquisition and pre-processing is done in Google Earth Engine.

PRE-PROCESSING

Since 2016, Landsat products are reorganized by the USGS into tiered collections. This has been done to have a consistent archive of known data quality and also to support the consistency in time-series analysis. Level-1 scenes with the highest available data belong to Tier 1. Scenes that do not meet the criteria to belong to Tier 1, are assigned to Tier 2. In this research only products from Tier 1 are used. Landsat provides users also with Level-2 products which are atmospherically corrected Level-1 products. The atmospheric correction is done using LEDAPS. Landsat-7 Level-2 products are available in Google Earth Engine. Each data tile has dimensions of 150 $\times$ 150km^2. The image property 'CLOUD_COVER_LAND' gives an indication of the percentage cloud cover over land for the whole scene. Images having a cloud cover larger than 30% are discarded. Moreover, the scenes are cropped to the ROIs (RD New, EPSG:28992). In section 2.2.4 the SLC failure is described. During pre-processing, nothing is done with the missing parts. They are treated as no-data.

3.4.4. SENTINEL 2

Sentinel-2 mission is a constellation of two identical satellites, Sentinel-2A and Sentinel-2B. The satellites carry a Multi-Spectral Instrument (MSI). The MSI has 13 bands in the visible, NIR and Short-wave Infrared (SWIR) spectral domain with 10m, 20m and 60m spatial resolution, see Figures 2.14 and A.7. One satellite covers the Earth in 10 days, but the collaboration of two twin satellites makes it possible to cover in 5 days, see Figure 3.8. Sentinel-2A is launched on the 23rd of June 2015 and Sentinel-2B is launched on the 7th of March 2017. The data acquisition and pre-processing is done in Google Earth Engine.

PRE-PROCESSING

Sentinel-2 products come in different levels. The products available for users are Level-1C Top-Of-Atmosphere reflectance in cartographic geometry (UTM/WGS84 projection) and Level-2A Bottom-Of-Atmosphere reflectance in cartographic geometry (UTM/WGS84 projection). Each data tile has dimensions of 100 $\times$ 100km^2. One can either choose to use the Level-1C products and apply (own) atmospheric correction to the products or use the Level-2A already atmospherically corrected products, which are available in Google Earth Engine. In this research, the available Level-2A products from Google Earth Engine will be used. The bands with spatial resolutions of 10m and 20m are selected. The bands of 20m resolution are resampled to 10m resolution. The ROIs are relatively small compared to the area of the data tiles. Selecting the images on 'CLOUDY_PIXEL_PERCENTAGE' lower than 30% and afterwards checking manually if there are clouds above the ROIs gives the best set of images in this project. Bit masking the areas based on e.g. Cloud Probability Map or Scene Classification Map is also an option, but the areas are relatively small and bit masking them would cause images to have a lot of no-data pixels. The scenes are cropped to the ROIs (RD New, EPSG:28992).

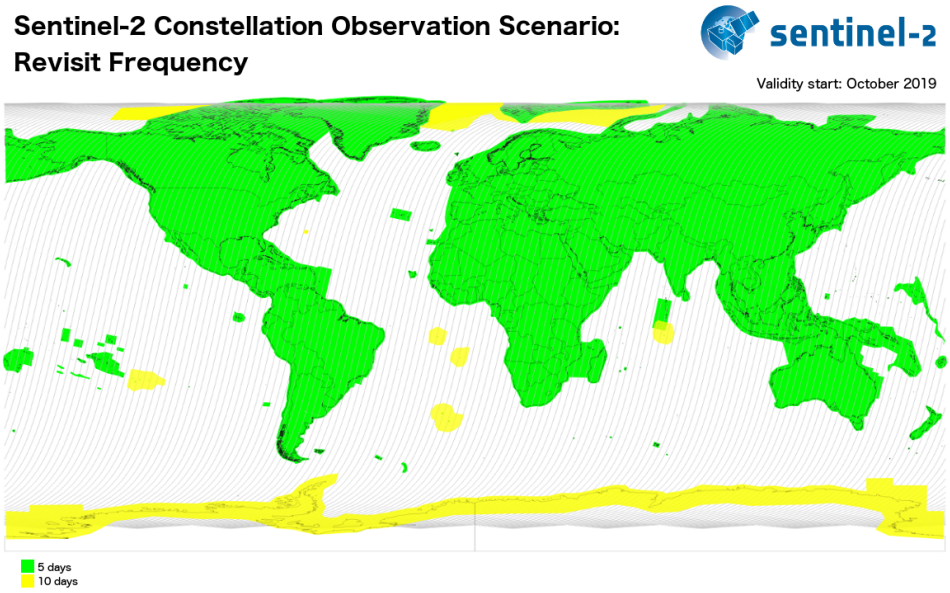


Figure 3.8: Global coverage and revisit time for Sentinel-2 MSI acquisitions. Validity start in October 2019 (ESA, 2019c).

3.4.5. FEATURE ENGINEERING

Developing other features from raw features is called feature engineering and can improve the machine learning algorithm's success rate by including some domain knowledge. In remote sensing, this often refers to vegetation indices. For the aerial images, NDVI and ARVI indices are added. The satellite images have additional SWIR bands which can be used to derive the NDWI index.

NDVI

Normalized Difference Vegetation Index (NDVI) is a well-known vegetation index used to monitor vegetation (Nouri et al., 2017). As introduced in subsection 2.2.2, the NDVI makes use of the chlorophyll absorption in the Red band and the high reflectance of the cell structure in the NIR band. This is translated into the following equation:

$$NDVI = \frac{NIR - Red}{NIR + Red}. \tag{3.1}$$

Figure 2.13 shows the reflectance of healthy vegetation. Healthy vegetation absorbs a lot of the incoming radiation in the Red domain, while the radiation in the NIR domain is more reflected. NDVI varies between -1 and +1. Negative values are most likely to indicate water. Values close to +1 indicate dense healthy green vegetation. NDVI values equal to zero show areas with no vegetation, e.g. urbanized areas or bare soils. When vegetation is dry or unhealthy, the chlorophyll content in the leaves is low, causing a higher reflectance in the visible domain. The bad condition of the leaves also influences the internal cell structure and leads to more absorption in the NIR domain, bringing the NDVI closer to zero.

ARVI

The Atmospherically Resistance Vegetation Index (ARVI) is on average four times less sensitive to atmospheric effects than the NDVI according to Kaufman and Tanre (1992). This index corrects for the atmospheric scattering in the Red reflectance domain by also using the measurements in the Blue domain in addition to the Red and NIR wavelengths. The equation reads as:

$$ARVI = \frac{(NIR – (2 \cdot Red) + Blue)}{(NIR + (2 \cdot Red) + Blue)}. \tag{3.2}$$

NDWI

The Normalized Difference Water Index reflects moisture content in plants and soils, and is determined as in Equation 3.3. Another version of the NDWI uses the Green band instead of the SWIR band, but is used for detection of waterbodies and waterbody volume changes (Du et al., 2016). Negative values of the NDVI (Equation 3.1) already indicate water. Therefore, there was no urge to add an index to the feature collection to detect water bodies. It is more efficient to add an index to detect leaf and soil water content. NDWI as in Equation 3.3 is expected to be useful in detecting humid dune slacks B, C and swamp vegetation.

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR}. \tag{3.3}$$

3.4.6. INPUT DATA FOR TRAINING

This subsection gives the specifications of the different kinds of remote sensing input data used in this research and the corresponding available training imagery.

For the resolution analysis, aerial images and Landsat-7 images are used. Landsat-7 data is divided into 'low-spectral' and 'high-spectral' resolution. The aerial images belong to category 'low-spectral'. Both Landsat-7 and Sentinel-2 are used for the classifications of 2019:

- Aerial images: Blue, Green, Red and NIR in original resolution (0.25m), 0.5m, 4m, 10m, 20m and 30m. The additional features are NDVI and ARVI.
- Landsat-7, low-spectral: B1 (Blue), B2 (Green), B3 (Red) and B4 (NIR) in 30m resolution. Additional features are NDVI and ARVI.
- Landsat-7, high-spectral: B1 (Blue), B2 (Green), B3 (Red), B4 (NIR), B5 (SWIR-1), and B7 (SWIR-2) in 30m resolution. Additional features are NDVI, ARVI and NDWI.
- Sentinel-2: B2 (Blue), B3 (Green), B4 (Red), B5 (Red Edge 1), B6 (Red Edge 2), B7 (Red Edge 3), B8 (NIR), B8A (Red Edge 4), B11 (SWIR-1) and B12 (SWIR-2). Spatial resolution in B2, B3, B4 and B8 is 10m. The other bands have resolutions of 20m. Additional features are NDVI, ARVI and NDWI.

More information about the band designations can be found in Appendix A.3. After filtering the satellite images as described in sections 3.4.3 and 3.4.4, the images are manually checked on clouds missed by the cloud detecting algorithm. It turned out that there are not many usable Landsat-7 images available in the years 2012 or 2019. Therefore, images from 2011 and 2013 are added to the collection of 2012 images, whereas images from 2018 are added to those of 2019. The assumption is that negligible differences in habitats are expected in adjacent years. The aerial images used in this research originate from one single image. The green bar in Figure 3.9 shows the acquisition time range of the aerial images. Sentinel-2 has a relatively larger number of available images due to, among other things, higher temporal resolution, but also the fact that Solleveld & Kapittelduinen is measured over two neighbouring orbits. Moreover, it is clear that almost no images are selected in the winter months January, November and December due to dominating cloud coverage. This is one of the bigger drawbacks of optical remote sensing. All the images indicated in Figure 3.9 are used as input for the supervised classifications of the yearly habitat maps, which are described in Section 4.5 and implemented in Section 5.2.

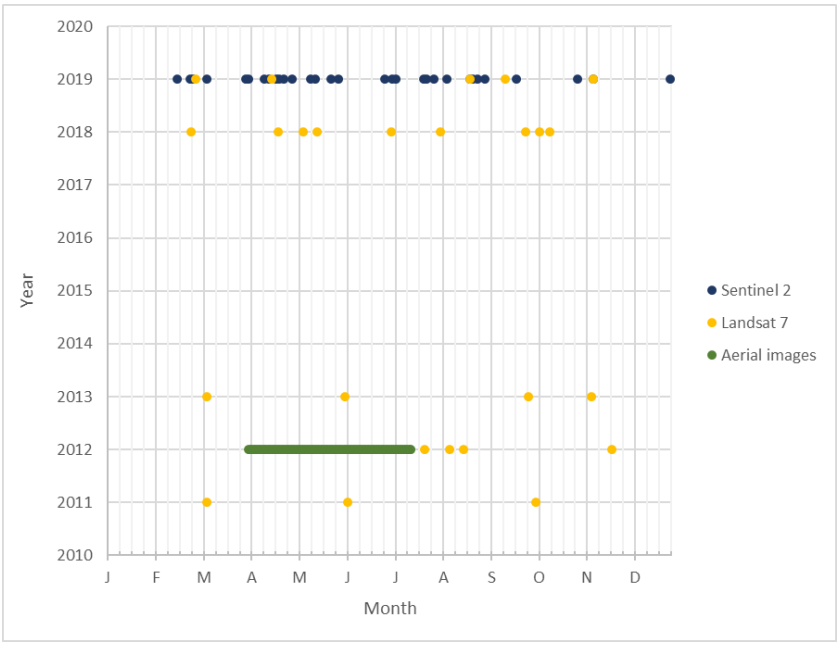


Figure 3.9: Available images after the filtering processes for Sentinel 2 (2019), Landsat 7 (2011-2013, 2018. 2019) and aerial images (2012). The green bar indicates the acquisition time of the aerial images in 2012.

3.5. DATA UNDERSTANDING

This section shows the preliminary data analysis using mostly data from aerial images and the fieldwork campaign. Visual exploration is used to understand the data. The data understanding is only done with a subset of the total data (Solleveld & Kapittelduinen), assuming that the other ROIs will give similar insights.

3.5.1. SPECTRAL SIGNATURES

In subsection 2.2.2, the fundamentals of discriminating the different habitats based on their spectral reflectance are described. Different surface types reflect radiation differently in various channels. In Figure 2.13 the typical spectral signatures for vegetation, soil and water are shown. A similar plot is made using the median spectral reflectances for each sub-habitat from the aerial images in 2012 for the bands in the visible (Blue, Green, Red) and NIR domain of the spectrum. In subsection 3.4.2, it is mentioned that the values of the pixels in the aerial images are not the physical spectral reflectances, but the Digital Numbers are nevertheless assumed as reflectances. The reflectance curves in Figure 3.10 are therefore higher compared to the typical spectral signatures (Figure 2.13).

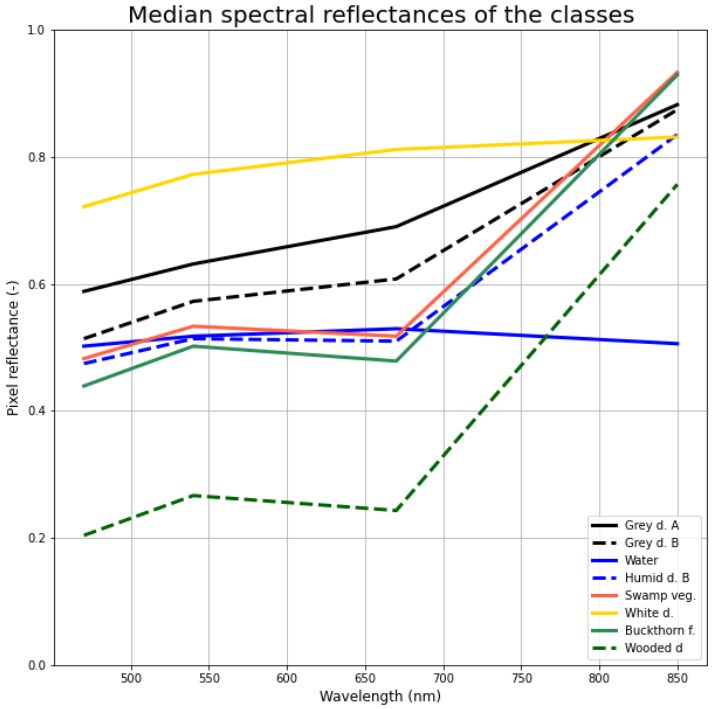
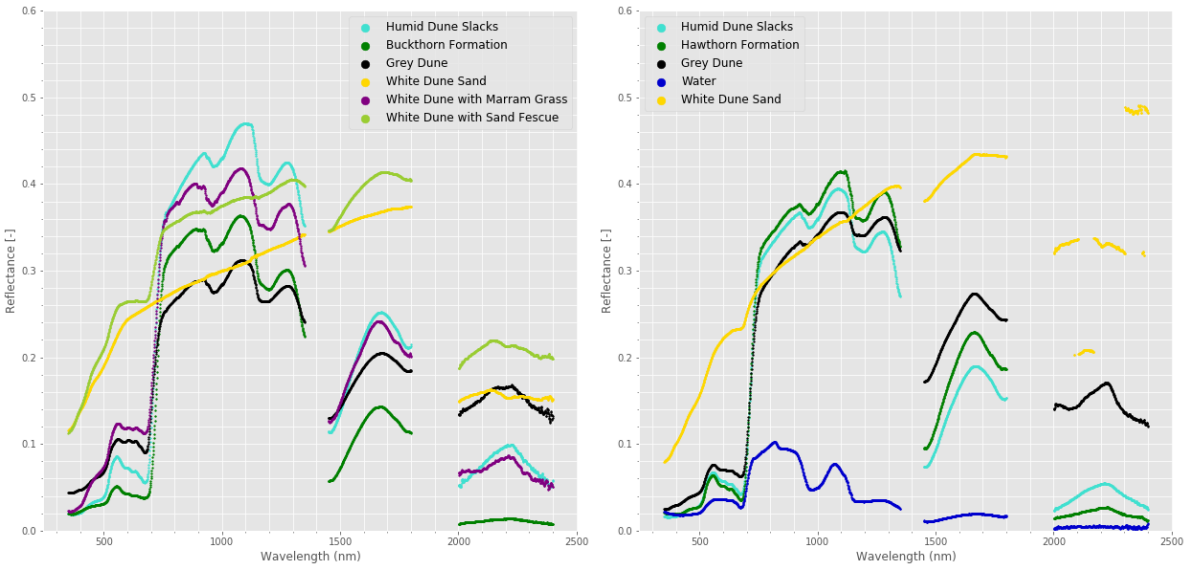


Figure 3.10: Median reflectances of the sub-habitat types from the aerial images 2012. *'Dunes'* are abbreviated to *'d.'*, *'formations'* to *'f.'* and *'vegetation'* to *'veg.'*.

By comparing the spectral reflectances in Figure 3.10 with the typical spectral reflectances in Figure 2.13, some characteristics of the data and the quality can be collected. Important to notice are the reflectances of white dunes and water. White dunes have the highest reflectance relative to the other classes, caused primarily by the bright color of the sand. Water absorbs most of the electromagnetic energy, see subsection 2.2.2, and the reflectance is therefore expected to be one of the lowest compared to the other classes. This is not the case in Figure 3.10. The reason could be caused by the visibility of the bottoms due to the shallowness of the waters, suspension of the waters, discolored waters due to algal bloom or glint on the water surfaces making the surfaces appear brighter. Moreover, swamp vegetation follow closely the reflectances of buckthorn formations and humid dunes B.

In subsection 3.2 the two-day fieldwork campaign is described and the measurement locations with their footprint are shown in Figures 3.4 and 3.5. Similar to Figure 3.10, spectral reflectances of the fieldwork locations are plotted by grouping them together per land cover type and region in Figure 3.11. The water absorption bands are previously mentioned to be at wavelengths 1.4 μm, 1.9 μm, 2.2 μm en 2.7 μm. The bands at 1.4 μm, 1.9 μm and reflections at wavelengths larger than 2.4 μm are removed to denoise.



(a) Solleveld & Kapittelduinen.

(b) Voornes Duin

Figure 3.11: Mean spectral reflectances measured in the field with the field spectrometer in Solleveld & Kapittelduinen (a) and Voornes Duin (b). The measurement method is described in section 3.2.

Based on Figure 3.11 and the figures in which the footprints of the measurement locations are shown (Figure 3.4 and 3.5), the different land covers reflections in all wavelengths are analysed. Obviously, these high spatial and spectral resolution measurements done in the field are expected to show the differences between land covers much more clearly than a satellite image ever could. Yet, it is important to do an initial analysis to see how additional spectral information could contribute to the differentiation of the habitats.

The humid dune slacks measured, in both Solleveld & Kapittelduinen and Voornes Duin, are generally humid grasslands. They have a strong reflectance in the NIR domain and a relatively low reflectance in the MIR domain, meaning that these vegetation are healthy. Moreover, white dunes with only sand has a typical dry soil reflectance. The white dunes with sand fescue (P12 in Figure 3.4) has an overall increased reflection due to the sand on the substrate. It follows the reflection of white dunes in the visible domain, but tends to follow the vegetation reflectance in the NIR and MIR domains (although raised). The sand fescue does have an effect on the reflectance of the footprint. Moreover, the white dunes with Marram grass follows the vegetation curve. In Voornes Duin, grey dunes have a high reflection in the MIR domain, indicating that the leaf water content is relatively low. During the field visits, many patches of dried moss in grey dune areas are detected (P3 and P5 in Figure 3.5). Hawthorn formations (similar to Buckthorn formations) and humid dune slacks have close spectral reflectances in the visible and NIR domain between approximately 700 and 1200 nm. The MIR domain indicates that shrubs have lower leaf water content than humid dune slacks. This is not the case in Solleveld & Kapittelduinen. There are only measurements of shallow water, whereby the measured reflectance of the water is contaminated by the bottom of the water. This can be clearly seen from the reflection of the NIR domain. Clear water should have no reflection in the NIR domain. White dunes without dominating vegetation, water and vegetation are effortlessly distinguished based on the visible domain reflectances. The habitats with vegetation (grey dunes, humid dune slacks) and mixed land covers (white dunes with Marram grass) are somewhat more difficult to distinguish without the additional spectral information from the NIR and MIR domains.

3.5.2. DIMENSIONALITY REDUCTION: T-SNE

In projects that involve large number or data variables, dimensionality reduction is often used to analyse or visualise data in lower dimensions. Dimensionality reduction techniques can be used to visualize high-dimensional data by projecting it into a low-dimensional space. The data from the aerial images contain information from the Red, Green, Blue and NIR band. In total, the data contains six features together with the additional NDVI and ARVI indices. A well-known linear dimensionality reduction technique is the Principal Components Analysis (PCA), which is often used as a pre-processing step in machine learning pipelines.

Laurens van der Maaten and Geoffrey Hinton developed a non-linear dimensionality reduction technique, named T-distributed Stochastic Neighbour Embedding (t-SNE). This technique is mostly used for unsupervised data exploration and visualisation of high-dimensional data. According to Van der Maaten, this technique is better than existing techniques at creating a single map that reveals patterns at many different scales which are important for high-dimensional data that lie on several different, but related, low-dimensional manifolds (Van der Maaten and Hinton, 2008).

t-SNE preserves only small pairwise distances or local similarities, whereas PCA is concerned with preserving large pairwise distances to maximize variance. Figure 3.12 shows a 2D visualisation of the high-dimensional dune data using t-SNE for the Solleveld & Kapittelduinen dataset in default settings. It is very important to mention that the default settings are used, because the visual clusters are strongly influenced by the chosen parameterization. Some combinations of parameter values can lead to false findings by making clusters of non-clustered data.

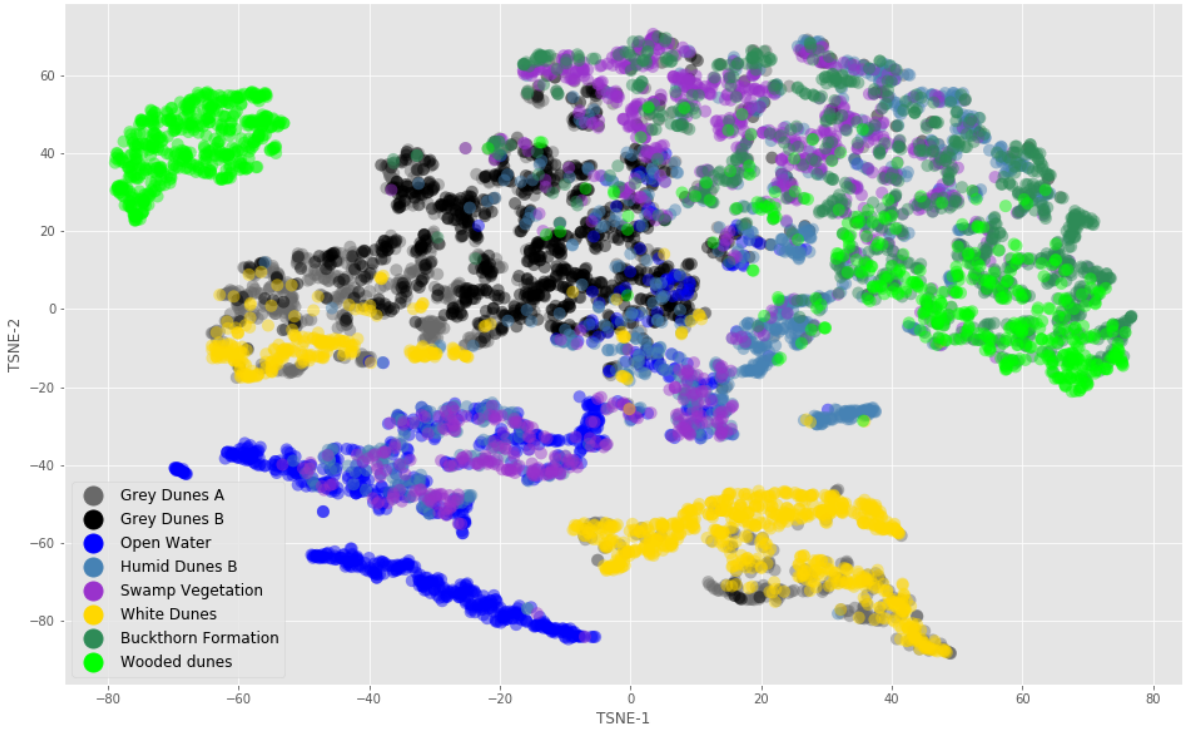


Figure 3.12: 2D visualisation of the high-dimensional dune habitat data (features: Red, Green, Blue, NIR, NDVI, ARVI) random selection for each sub-habitat using t-SNE.

There is certainly some clustering in the cloud of points. Classes with similar spectral signatures tend to have overlapping clusters. In accordance with what emerged from the analysis of the fieldwork spectra, concerning the separability of the sub-habitats based on the visible domain of the spectrum, there is a separate cluster for grey dunes. Grey dunes A and B are largely overlapping. Water has some overlapping with grey dunes and some white dune pixels. The reason for this is most probably shadow of the Marram grass causing a dark reflectance similar to that of water in the visible domain. Swamp vegetation (H2190D) overlaps with other H2190 sub-classes, which includes water (H2190A). Moreover, Buckthorn formations and wooded dunes are overlapping slightly, which is expected because the only difference between these two classes is the canopy height. In the left upper corner, there is a separate cluster of wooded dunes. This is most probably due to the difference in phenology between the northern and southern part of Solleveld & Kapittelduinen. White dunes have a cluster separate from the grey dunes, which is the sand part of the white dunes.

3.6. SOFTWARE

The processing, analysis and visualisation of the data is done in different software programs. In this section, the used software and packages are shortly described.

3.6.1. PYTHON

Python is a well-known and widely used free programming languages created by the Dutch programmer Guido van Rossum. It is nowadays used for many purposes, e.g. geospatial analysis, machine learning, data science etc. The most frequently used packages in this project are Pandas, GeoPandas, Numpy, Scikit-learn, Fiona, Shapely and Matplotlib.

3.6.2. GOOGLE EARTH ENGINE

Google Earth Engine (GEE) is a cloud-based geospatial processing platform for global scale. The main components GEE earns its popularity from are the petabyte-scale archives of publicly available remotely sensed imagery, huge computational power running on Google's servers optimized for parallel processing of geospatial data, APIs for JavaScript and Python for making requests to the Earth Engine servers and its online code editor. A lot can be done in the code editor without downloading the data on the desktop, but in this project aerial images are part of the data and uploading them on GEE is not very time practical. For that reason, data products are filtered in GEE on date, cloud coverage etc. and clipped afterwards to the ROI to limit data size. These clipped data products are downloaded from GEE in the projected coordinate system RD New (EPSG:28992) and further processed in Python.

3.6.3. QGIS

QGIS is a GIS application mostly used for visualization, managing, editing and analysing geo-spatial data. It is a project of the OSGeo (Open Source Geospatial Foundation) and is a 100% volunteer driven project. All corrections to geometries of shapefiles, map visualizations, single raster calculations and similar tasks are done using QGIS 3.4.13. in this project.

3.6.4. POSTGRESQL/POSTGRES

PostgreSQL is a powerful, open source database system that is build on the original SQL (Structured Query Language) language and extended with many features. PostgreSQL is used for generating random points in the shapefiles of the ROIs (see Appendix A.1). This is less time consuming than doing the same operation in Python.

3.6.5. GPS LOGGER

The GPS Logger app from BasicAirData is an app for recording positions, but it is also capable of recording paths. This app is used during the fieldwork to log the different measurement points.

4

METHODOLOGY

An initial data analysis was performed in Chapter 3, by looking at the different spectra measured during the fieldwork and by dimensionality reduction using t-SNE. This gave insight on how the different habitats relate to each other on spectral level. In this chapter, the methodology for the resolution analysis and the final classifications are described. The assessment and validation is a crucial part of the methodology as it helps making a qualitative comparison of the two chosen methods, pixel-based and region-based, and algorithms, Random Forest and XGBoost.

4.1. WORKFLOW

The methodology is summarized in Figure 4.1. A more extensive version can be found in Appendix A.2. In the summarized workflow, the pre-processing steps are excluded considering that a large part is discussed in Section 3.4. The specifications of the various data products in terms of resolution are also discussed in that Section. However, detailed information about the image resampling process and removing duplicates caused by resampling of the images are discussed in this Chapter, including the appointment and explanation of the disturbance factors due to the resampling. Furthermore, Random Forest and XGBoost are compared using a subset of the data to decide for the best performing supervised ensemble learning classification algorithm. That algorithm is used for further analysis during the research. The spatial and spectral resolution analysis is done with both pixel-based and region-based classification methods. The conclusion from the analysis determines which resolutions, both spatial and spectral, and method can best be used to classify dune habitats. The examination of the spatial variability is important to test the applicability of the method in other areas. Furthermore, the development steps of the yearly habitat maps and change maps for both 2012 and 2019 are described, including the hierarchy mentioned in section 3.3. Finally, the different performance metrics to assess and validate the results are described.

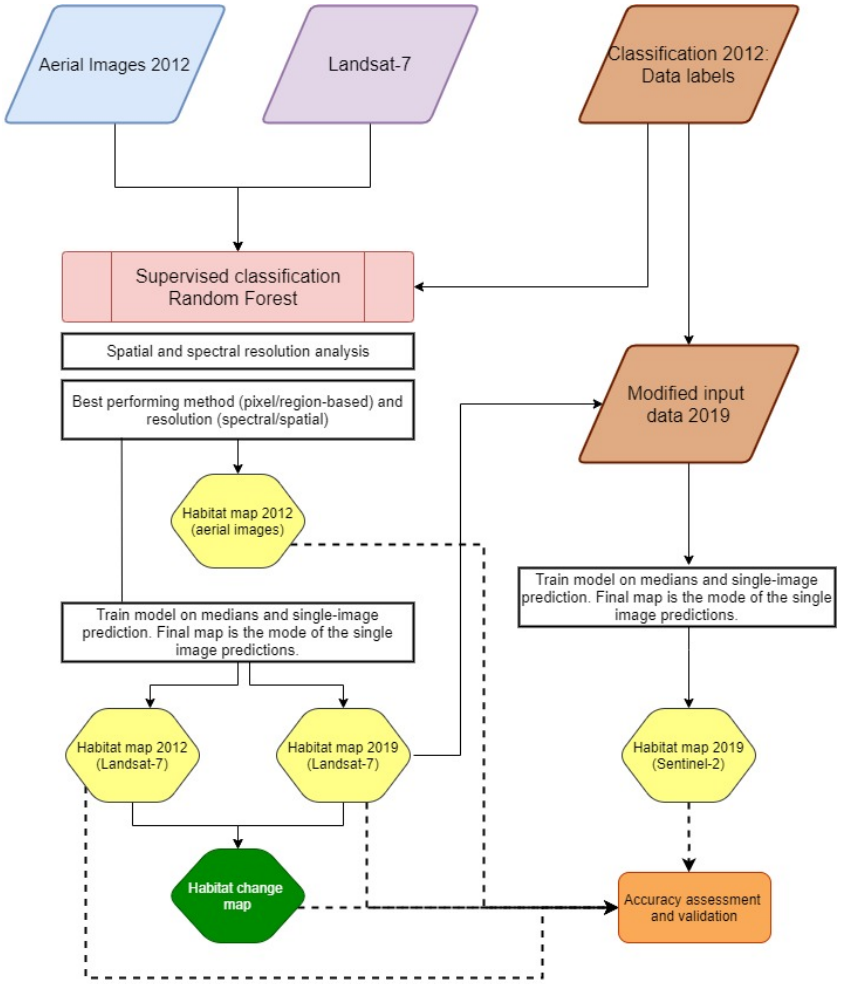


Figure 4.1: Summarized workflow. Full workflow in Appendix A.2.

4.2. SUPERVISED CLASSIFICATION

In Section 2.3 the fundamentals of the algorithms used in this research are described. Random Forest and XGBoost are both used for supervised learning. During a supervised classification, the algorithm is supervised by the user by assigning representative labels to pixels. In this research the training labels are based on the fieldwork classifications done in 2012 in the ROIs described in section 3.1. The results of the supervised classifications are dependent on the algorithm used, the tuning of the hyperparameters, the quality of the images, the quality of the input data and training labels, etc. This section will focus on the process of resampling the aerial images, the possible disturbance factors introduced and the comparison between Random Forest and XGBoost by testing them on a subset of the complete data.

4.2.1. RESAMPLING THE AERIAL IMAGES

In subsection 3.4.2 the pre-processing steps for the aerial images are described. It is mentioned that the aerial images are deteriorated by resampling. The aerial images are resampled after clipping them to the extent of the ROI for reduction of data load. Afterwards, they are clipped to the official Natura 2000 area boundaries. Image resampling means *'changing the number of pixels in an image'*. In other words, the spatial resolution is deteriorated. The images are resampled from 25cm to 50cm, 4m, 10m, 20m and 30m. The 10m, 20m and 30m resolution images are requisite for comparison to how the algorithm is performing with satellite images from Landsat-7 and Sentinel-2. The 50cm resolution is chosen as one of the resolution to see whether a tiny change in the original high resolution has any significant effects on the classification performance. Moreover, the 4m resolution is chosen to be the intermediate resolution between 50cm and 10m.

The resampling and clipping to the ROIs are done using GDAL, which is a translator library for raster and vector geo-spatial data formats (GDAL, 2019a). GDAL offers several (interpolation) algorithms for resampling purposes: nearest neighbour, bilinear, cubic, average, mode, etc. Satellite sensors average the sensed data in the area of 1 unit spatial resolution. To be able to simulate the aerial images as satellite data, the 'average' resampling method is used. Other resampling methods would belong to category 'interpolation' (GDAL, 2019b). Although, interpolation works great for visualisation purposes, it will not be used for further analysis in this research.

DUPLICATES

As the resolution decreases, more training points will position in one pixel. This will result in repetitions, duplicates, in the training dataset. Duplicates in a machine learning model will lead, in most cases, to a biased model. In other words, having a duplicate means that the class belonging to the duplicate is seen more often by the algorithm and receives therefore more weight. The model will tend towards classifying that class more than the other classes. Removing duplicates is not a problem if the resulting dataset contains enough data points. The problem is usually caused by the imbalance in classes introduced when removing duplicates. Likewise, this will lead to a biased model. In some cases, it can even cause the minority class to be ignored. In this research, it is assumed that the training dataset is a reflection of reality and an imbalanced dataset will therefore lead to a worse result than having duplicates in the training dataset. Duplicates in the test dataset, and between training and test dataset are not allowed. Duplicates in the test set will lead to a biased interpretation of the model's performance. This is solved by removing duplicates from the entire dataset, splitting it into a training and test dataset. Afterwards, the training dataset is supplemented with duplicates by random sampling from the unique values in each class to fight the imbalance. Each class has the same number of training samples. The imbalance in the test dataset does not have much effect if the performance is measured by the balanced accuracy (see section 4.6).

4.2.2. RANDOM FOREST ALGORITHM

The Random Forest classification is done in Python's Scikit-learn package. Scikit-learn has ready to use machine learning algorithms, including the Random Forest classifier. In subsection 2.3.2 the Random Forest algorithm is theoretically introduced. In this subsection, the hyperparameter settings and the overfitting phenomena are discussed. As stated in Probst et al. (2019), *"It is well known that in most cases Random Forest works reasonably well with the default values of the hyperparameters specified in software packages."*. Notwithstanding, the performance can be improved by tuning the hyperparameters, which have to be set by the user before training the model.

The most important concept related to parameter tuning, is overfitting. When a model overfits, the parameters are tuned in a way that the prediction is optimized for the training dataset. This means that the model

will perform extremely well predicting the training dataset, but it will probably not be able to generalize to the test dataset or any other unseen data related to the application. That is also how overfitting is recognized; a gap in performance predicting the training- and test dataset. Subsidiary, there is the phenomenon of underfitting, which occurs when the model is too simple to explain the variance of the data. This can be recognized by poor performance on the training dataset. Underfitting is often solved before the model is even tested on the test dataset.

Hyperparameter tuning is generally more of a trial and error process than a theoretical one. The ideal method to find the best hyperparameter settings for a specific case would be to test every possible combination and assess the performance of the model based on these settings. Scikit-learn provides a couple of hyperparameter optimizers. RandomizedSearchCV and GridSearchCV are two popular methods to search for the optimal combination of hyperparameters. CV stands for cross validation. Using either the randomized search or the grid search, the user defines the ranges of the grid of hyperparameters. When using the randomized search, random samples from the grid are used to run the algorithm. Afterwards, the performance is assessed using K-fold cross validation. With grid search, all combinations in the grid are used and the performance is assessed again using K-fold cross validation. Scikit-learn has compared these two methods in terms of performance and run time. Randomized search has a lower run time, while having a slightly worse performance than the grid search. Both methods were too computational expensive for this research. Instead, it is decided to do heuristic tuning.

The Random Forest classifier fits a (user defined) number of decision trees on subsets of the dataset. There will be replacement in the subsets when building the trees if bootstrapping is applied, which means that some samples can occur in the trees multiple times or, on the contrary, not occur at all. In subsection 2.3.2 is mentioned that the individual trees in the Random Forest are independent from each other and have a certain depth. If the maximum depth (*max_depth*) is not set by the user, the nodes are expanded completely until all leaves are pure. A pure leave is when all the samples that belong to that node are labeled by the same class. The number of features used during splitting of the tree is the square root of the total number of features. The number of trees necessary to reach the optimal performance depends on the dataset and its properties. Literature shows that the biggest performance gain occurs in the first 100 trees, which is the default in Scikit-Learn (Probst et al., 2019). Adding more trees does not influence performance, but increases computational complexity.

In this research, the goal is not necessarily to find the best tuned model, but to find one that performs sufficiently well for comparison analysis. First, the algorithm has been run in the default setting to see how it performs. This caused overfitting due to deep trees. Literature proposes having at least 100 trees for training, so the number of trees is not changed. On the other hand, the maximum depth of the trees is lowered in stages. A maximum depth of 12 removes the overfitting for a resolution of 25cm. To be able to reproduce the results later, a random state is defined (*random_state = 42*) to control the randomness. All other hyperparameters are kept to their default setting. The models final prediction class is the majority vote of the predictions.

4.2.3. XGBOOST ALGORITHM

The eXtreme Gradient Boosting classification is done using the Scikit-Learn API of the XGBoost Python package. The biggest difference between Random Forest and XGBoost is that XGBoost uses boosting in contrast to Random Forest using bagging, explained in subsection 2.3.2. The XGBoost classifier has different and a larger number of hyperparameters to tune than Random Forest. XGBoost is also more sensitive for overfitting than Random Forest. The number of trees is set to 75 and the maximum tree depth for base learners is set to 6. The objective is set to the softmax objective, which is used for multiclass classification, and the number of classes are defined. The learning rate is set to 0.35 heuristically. All other hyperparameters are in their default setting. XGBoost includes boosting specific parameters, regularization on leaf weights, a minimum loss parameter (gamma) and more. These parameters are not adjusted and will not be elaborated further.

To see whether XGBoost is advantageous in this research, it is compared to Random Forest in both performance and run time. A pixel-based classification is done for each resolution. The results are shown in Figure 4.2 and 4.3.

Figure 4.2 shows a plot of the run time and the balanced overall accuracy for pixel-based classification using Random Forest and XGBoost. These tests have been running on a Microsoft Windows 10 Enterprise x64-based Desktop PC with Intel(R) Core(TM) i5-3360M CPU 2.80GHz processor. The code is written in the Scikit-learn

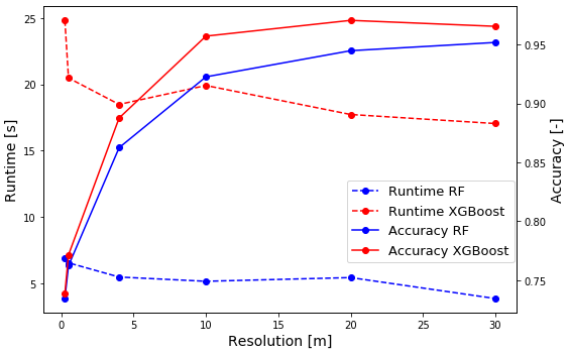


Figure 4.2: Run time and accuracy of Random Forest and XGBoost for a pixel-based flat classification.

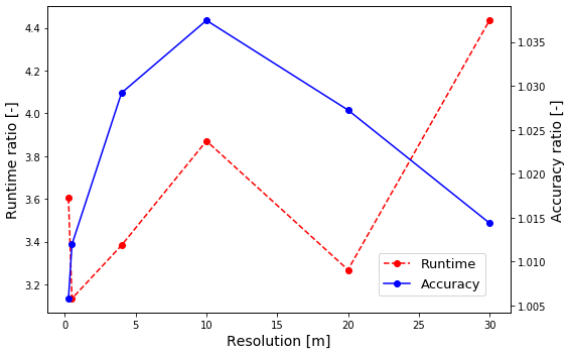


Figure 4.3: Run time and accuracy ratio between Random Forest and XGBoost for a pixel-based classification. Runtime (t) ratio is $\frac{t\_XGBoost}{t\_Random\ Forest}$ and accuracy (Acc.) ratio is $\frac{Acc\_XGBoost}{Acc\_Random\ Forest}$

API for XGBoost. It is preferred over the Core XGBoost library as the code is more compatible to that of Random Forest.

A very outstanding feature in Figure 4.2 is that the runtime of Random Forest is lower compared to XGBoost. On the contrary, XGBoost has a higher accuracy, which is most noticeable at 4, 10 and 20m. As the resolution decreases (the pixel size increases) the accuracy curve of Random Forest starts to meet the accuracy curve of XGBoost. It is great to know that XGBoost outperforms Random Forest, but Figure 4.3 shows that the runtime of XGBoost is approximately 3 to 4.5 times higher than Random Forest in comparison to an average of 2% higher accuracy. An increase of 2% does not outweigh the higher runtime. For this reason, it is decided to continue with Random Forest and to omit XGBoost.

4.3. SPATIAL AND SPECTRAL RESOLUTION ANALYSIS

The classification performances are examined in a spatial and spectral resolution analysis using the balanced accuracy and Cohen's kappa (see section 4.6). This is only possible with labeled data and therefore only with data from 2012. There is not enough ground-truth data for 2019 to assess the performances with. The spatial resolutions are split into different scales, see subsection 4.2.1, and a distinction in spectral resolution is made between low-spectral and high-spectral resolution, introduced in subsection 3.4.6. Aerial images and Landsat-7 images from 2012 are used to do this analysis.

Low-spectral contains the bands Red, Green, Blue and NIR, and vegetation indices NDVI and ARVI. The high-spectral category contains the same bands and the two SWIR bands of Landsat-7. The indices NDVI, ARVI and NDWI are derived from these bands. This differentiation in spectral resolution is done to explore how the classification performs better when adding more spectral features. The spatial and spectral resolution analysis is done on the pixel-based and region-based methods which are explained in more detail in Section 4.4.

4.3.1. SPATIAL VARIABILITY

It is beneficial to check how the performance of the model changes per region. In Figure 3.11, a difference in reflection is detected between grey dunes in Solleveld & Kapittelduinen and Voornes Duin. During the fieldwork, it is also noticed that Voornes Duin is relatively more humid than Solleveld & Kapittelduinen. It is expected that this will have a significant effect on the performance of the classifications depending on data from which area the model is trained. The spatial variability looks at how the classification performance is affected in three scenarios of model training: when data from the three different Natura 2000 areas are combined, when data from each Natura 2000 area is treated separately and when the model is trained on a different area than it is tested on.

4.4. PIXEL AND REGION-BASED CLASSIFICATION

Traditional pixel-based classification is based only on the spectral information of one pixel. Another method is object-based classification. This is the classification of objects or polygons with homogeneous spatial or spectral characteristics. When dealing with high resolution data in particular, object-based classification is preferred to eliminate salt-and-pepper effects.

The advantage of object-based classification is that it uses additional contextual information (Jobin et al., 2008, Whiteside et al., 2011). In this research, the 'objects' are the habitat types. Dune habitats are not very distinctive types of land cover. There are transition areas, different habitats that look similar, etc. Hence, segmentation of the data into polygons would be a very complex task. A combination between pixel-based and object-based is a better option: the region-based classification. This region-based classification is derived from the sliding window method and draws similarities to the convolutional layers in the popular Convolutional Neural Network architecture. Convolutional layers extract features based on pixel values, but also use other regional statistics of pixels.

4.4.1. REGION-BASED

The region-based classification is performed with a 3×3 pixel-window. All values of each band from the 8 pixels around the center pixel are appended as features to the center pixel. Auxiliary, the statistics median, minimum, maximum and standard deviation of all 9 pixels are chosen to be the regional features. These statistics are derived using built-in filters from Scipy and appended to the third dimension b as additional bands. Figure 4.4 shows this feature transfer process to a feature vector with length N_f. Only the feature extraction of a single band is shown for simplicity. The process is iterated over each band b, both spectral and statistical. With the region-based method, the importance of the statistical relationships between adjacent pixels and the center pixel are analysed. In other words, using local statistics, the texture of the landscape is analysed and examined on how important it is while classifying dune habitats.

This method makes the outer boundary pixels of the image unusable due to no-data values from outside the image appended to the feature vector. These outer boundary pixels are therefore removed.

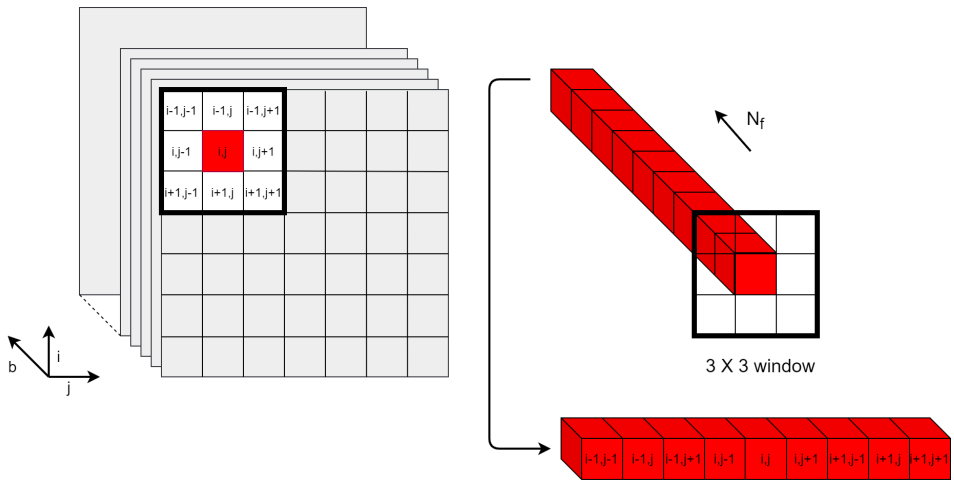


Figure 4.4: Region-based feature extraction scheme. N_f: *number of features.*

4.5. YEARLY HABITAT MAPS

The spectral and spatial resolution analysis results in a combination of resolutions and methods that perform best for classifying dune habitats. For the satellite images, the spatial resolution is fixed, so only the spectral resolution and the method of classification play a role. Besides the dune habitat map in 2012 created with aerial images, yearly habitat maps are created with Landsat-7 images for both 2012 and 2019. Moreover, Sentinel-2 images are used to create yearly habitat maps in 2019 on 10m resolution. This section discusses the creation of the yearly habitat maps, the influence of the hierarchy and how the habitat maps in 2019 are created using Sentinel-2, while there are no data labels available for 2019.

4.5.1. SUB-CLASS VS. MAIN-CLASS

Dune habitats that are the target classes depend on a pre-defined national classification scheme (Synbiosis) to have a common communication language between researchers, management and policy makers (Gavish et al., 2017). They are described in 2.1.2 and illustrated in 3.6. For some management applications, it is not relevant to know which area belongs to what sub-habitat. It is in those cases more important to know what the main-habitat is. Most land cover classification schemes have a hierarchical structure and dune habitats are no exception to this. A traditional hierarchical classification follows a top-down approach, in which first the main-classes are separated by prediction. Afterwards, moving down the hierarchy, the sub-classes are classified. In the case of dune habitats, some sub-habitats can be spectrally similar to sub-habitats of other main-habitats. This may cause the main-habitat to be already classified incorrectly, resulting in the sub-habitat being incorrectly classified anyway. In this thesis, the hierarchical classification will follow a bottom-up approach. Based on the reasoning above, the hypothesis is that this will result in a better main-habitat classification.

4.5.2. FINAL MAPS USING SATELLITE IMAGES

The final maps using the satellite images are produced by training the model on the median of the available images, predicting on every single image available and finally taking the mode of the predicted classes per pixel. The mode is the most frequent prediction class among the single predictions. The training is done on the median, because the median is less sensitive for outliers comnpared to the mean. This method also allows to create confidence maps by looking at the number of times the final class is chosen amongst the available classifications. Moreover, the number of available classifications per pixel plays also an important role, especially with Landsat-7 due to the SLC failure.

There is no training dataset of good quality available for 2019, therefore the Landsat-7 map of 2019 is predicted using the model trained on 2012. This is not applicable to Sentinel-2 images, because Sentinel-2 has a different sensor with a completely different band architecture and resolution specifications. Therefore, a new Random Forest model needs to be trained. This is only possible if a training dataset is available for that year. The training dataset from 2012 is not suitable to use for this as that would assume that nothing has changed in the meantime.

The training dataset is derived from the final Landsat-7 classification in 2019. The training set in 2012 is inspected to see whether it is changed into a different main-habitat class. Training points are only used if the main-habitat class in 2019 is equal to that of 2012. In this way, additional uncertainties are not introduced into the classification. With the final Landsat-7 habitat maps in 2012 and 2019, difference maps are created.

4.6. ASSESSMENT AND VALIDATION

Assessing and validating the results is an essential part of a machine learning project. The training dataset is used to calibrate the classification model, but an additional test dataset is needed to verify the performance of the classification outcome and model. This test dataset should never be used in training and is therefore an unbiased evaluation of the final classification model (Muchoney and Strahler, 2002). In this research, a total of 80,000 points (10,000 per class) are sampled from the Mouissie classification in 2012 (Appendix A.1). From these samples, 80% is used for training and 20% is used for testing purposes. These numbers are used for the tests during the spectral analysis. As the resolution decreases, the number of duplicates will increase. Subsection 4.2.1 explains how this problem is solved. Moreover, while tuning the algorithms, a 3-fold cross-validation is used to asses in between settings. The most universally used form of accuracy assessment is the confusion matrix. This matrix shows the predicted (columns) and actual (rows) classes for a dataset used in the classification (Shanmugam et al., 2003). Using the confusion matrix, the overall balanced accuracy (see Equation 4.1), the user's accuracy (precision) and the producer's accuracy (sensitivity) can be determined.

$$Balanced\ accuracy = \frac{1}{n_{class}} \sum \frac{TP}{TP + TN} \tag{4.1}$$

in which n_{class} is the number of classes in the multi-class classification, TP is the number of correctly classified instances of the actual class and TN is the not-correctly classified instances of the actual class.

The user's accuracy is the proportion of correctly classified pixels for a class. It quantifies the confidence a user can have on a classification map. On the contrary, the producer's accuracy is defined as the fraction of

correctly classified ground-truth samples of this class. It quantifies how well the validation pixels are classified (Kempenaars et al., 2009).

Another measure for the accuracy is the Cohen's Kappa coefficient κ (Cohen, 1960), which indicates how well the model performs compared to random performance (Landis and Koch, 1977). This is called an inter-rater agreement: a measurement of how often raters agree when doing a task. The Cohen's Kappa performance measure is calculated as following:

$$\kappa = \frac{p_0 - p_c}{1 - p_c} = 1 - \frac{1 - p_0}{1 - p_c} \tag{4.2}$$

in which p_0 is the proportion of the observed agreement, and p_c is the proportion of agreements expected by chance (Sim and Wright, 2005). It tells how much better a classifier is performing over a classifier based on random guessing conforming to the frequency of each class (Landis and Koch, 1977). The range is usually between 0 and 1. Perfect agreement is represented by 1, indicating that the raters agree in the classification of every case. No agreement is represented by zero, indicating agreement no better than expected by chance or random guessing (Kempenaars et al., 2009).

ADDITIONAL VALIDATION

Moreover, the fieldwork locations visualised in Figures 3.4 and 3.5 are strong, but sparse validation data. There is a recent dune habitat map created by Provincie Zuid-Holland (PZH). Although, this map is not validated yet, it will be used briefly for validation in 2019. Furthermore, the created yearly maps using satellite images are discussed with the local manager of these dune areas (Stefan Poot).

CLASSIFICATION STABILITY

In addition to the validation methods for maps made with either Landsat-7 or Sentinel-2 satellite images, it is important to appraise the stability of the final classification outcomes. This is done by assessing the percentage of how many times the final classification outcome is chosen from the total number of available classifications. This is especially necessary for the Landsat-7 maps, because not every area has the same number of classifications due to the SLC-failure. Moreover, satellites move in orbits and therefore the coverage is not completely equal everywhere in the ROIs.

5

RESULTS

In Chapter 4, the methodology is discussed based on the summarized workflow. The results of the different steps are presented in the coming chapter. First the spatial and spectral resolution analysis is performed on both the pixel-based and region-based approaches. Then, the spatial variability is analysed. Next, the yearly habitat maps and change maps as a result of the Landsat-7 classifications are shown and analysed based on transition matrices. Thereafter, the final classification maps based on the Sentinel-2 images are presented and analysed. Finally, the accuracy assessment is discussed and the correlation plots of the fieldwork measurements with the satellite observations are analysed.

5.1. AERIAL IMAGE CLASSIFICATIONS

As described in section 4.4, the spatial and spectral resolution analysis is done using the traditional pixel-based method and the region-based method with the Random Forest Classifier. The analysis is done on Solleveld & Kapittelduinen. The best combination of resolutions and method is also used to predict aerial images of all three ROIs, see Appendix A.4.

5.1.1. RESOLUTION ANALYSIS

The aerial images play an important role in the spatial resolution analysis, while the Landsat-7 image plays an important role in the spectral resolution analysis. The Landsat-7 image captured on 24-07-2012 is used for the spatial and spectral resolution analysis. In Table 5.1 the accuracies and Cohen's Kappas are shown for the different classification tests using the aerial images and Landsat-7 image. With the current setup, it is not possible to classify the high resolution aerial images on the region-based method due to memory limitations.

Table 5.1: Classifications using Random Forest for pixel-based (PB) and region-based (RB) methods. '*Low-spectral*' indicates the bands Red, Green, Blue, NIR and vegetation indices NDVI + ARVI. '*High-spectral*' indicates all 6 bands (Red, Green, Blue, NIR, SWIR1, SWIR2) and indices NDVI, ARVI and NDWI. Performance metrics are *Acc. and* κ indicating respectively the balanced accuracy and Cohen's Kappa.

	Aerial Imagery [Low-spectral]				Landsat 7 [Low-spectral]				Landsat 7 [High-spectral]			
	PB		RB		PB		RB		PB		RB	
Resolution [m]	Acc.	κ	Acc.	κ	Acc.	κ	Acc.	κ	Acc.	κ	Acc.	κ
0.25	0.73	0.70	-	-	-	-	-	-	-	-	-	-
0.50	0.75	0.71	-	-	-	-	-	-	-	-	-	-
4	0.78	0.77	0.87	0.86	-	-	-	-	-	-	-	-
10	0.75	0.78	0.83	0.86	-	-	-	-	-	-	-	-
20	0.67	0.72	0.77	0.80	-	-	-	-	-	-	-	-
30	0.60	0.66	0.68	0.70	0.60	0.64	0.69	0.72	0.65	0.68	0.67	0.68

The purpose of these tests is to analyse whether decreasing the spatial resolution and increasing the spectral resolution affects the accuracy. Accuracy in these cases is based on a test dataset which comprises 20% of the total dataset (see section 4.6). Table 5.1 is used to analyse the influence of the spatial and spectral resolutions.

The performance of the pixel-based method peaks at 4m resolution. Subsequently, it performs worse as the resolution decreases. There is no result for the region-based classifications in the higher resolutions, but there is a relationship between the pixel-based and region-based performance. The region-based method consistently performs 2-10 % better than the pixel-based method. This can also be assumed for the 0.25m and 0.5m resolutions. With this information, it is concluded that, for the aerial images, 4m is the spatial resolution with the best performance for both pixel-based and region-based methods to classify dune habitats.

The region-based classification includes both spectral and statistical information from the 9 pixels (1 center pixel and 8 adjacent pixels), see subsection 4.4.1. Region-based performs significantly better than pixel-based. This increase in performance is caused either by the spectral features of the adjacent pixels, the statistical features or both. Figure 5.1 shows how the region-based classification performs in two cases: when isolating the statistical features with only spectral features from the center pixel (red curve) and when incorporating all spectral features from the 9 pixels with the statistical features (blue curve). It is clear that the statistical features are the most important features causing the region-based to perform better than pixel-based. The performance improves when the spectral features of the adjacent pixels are added to the classification, but the effect is not as big as achieved with the statistical features.

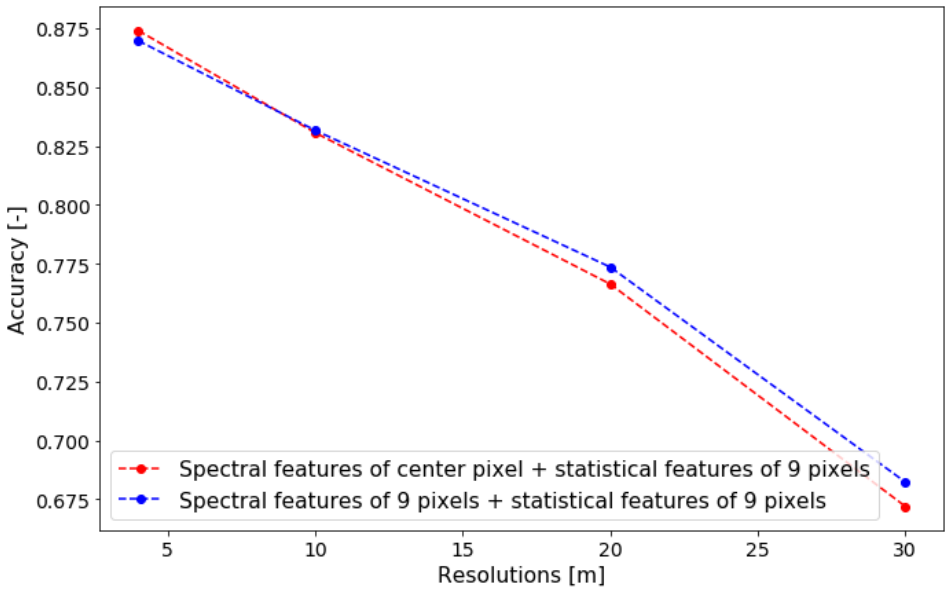


Figure 5.1: Region-based classification using spectral features from center pixel and statistical features (red curve) and using spectral features from all 9 pixels in the window and statistical features (blue curve).

Both pixel-based and region-based classifications are performed for a low-spectral and high-spectral image. For pixel-based, adding the two SWIR bands causes the spectral classification to perform 5% better. The region-based classification adds 2% of performance improvement. Likewise, the region-based classification of the low-spectral image performs 9% better than the pixel-based approach. From this, the same conclusion as before arises. The region-based approach leads to a larger improvement than using more spectral features. However, the SWIR bands are important for the pixel-based approach.

The most important features for the region-based classification are the statistical features (see Figure 5.1). It is also important to see which features are the most decisive for the pixel-based classifications. These feature importances are plotted in Figures 5.2, 5.3, 5.4 and 5.5.

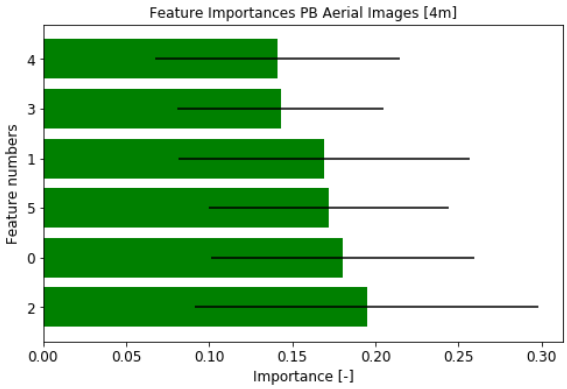


Figure 5.2: Feature importances for low-spectral pixel-based classification of aerial images for 4m resolution using features 0:Blue, 1:Green, 2:Red, 3:NIR, 4:NDVI, 5:ARVI.

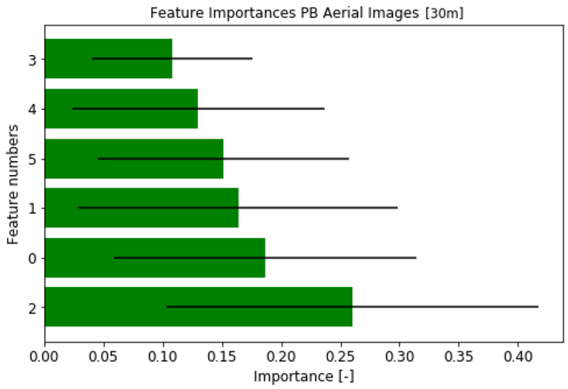


Figure 5.3: Feature importances for low-spectral pixel-based classification of aerial images for 30m resolution using features 0:Blue, 1:Green, 2:Red, 3:NIR, 4:NDVI, 5:ARVI.

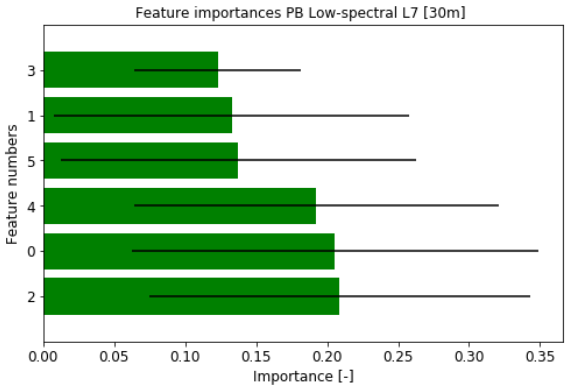


Figure 5.4: Feature importances for low-spectral pixel-based classification of L7 using features 0:Blue, 1:Green, 2:Red, 3:NIR, 4:NDVI, 5:ARVI.

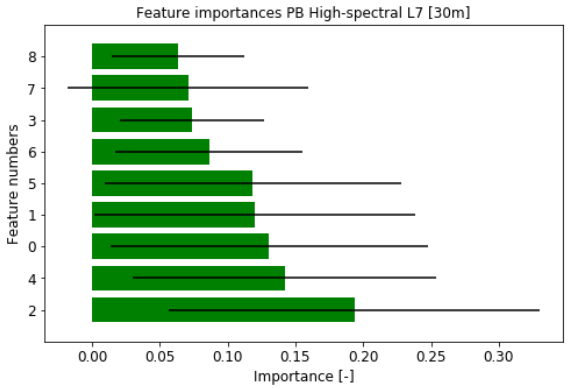


Figure 5.5: Feature importances for high-spectral pixel-based classification of L7 using features 0:Blue, 1:Green, 2:Red, 3:NIR, 4:NDVI, 5:ARVI, 6:SWIR1, 7:SWIR2, 8:NDWI.

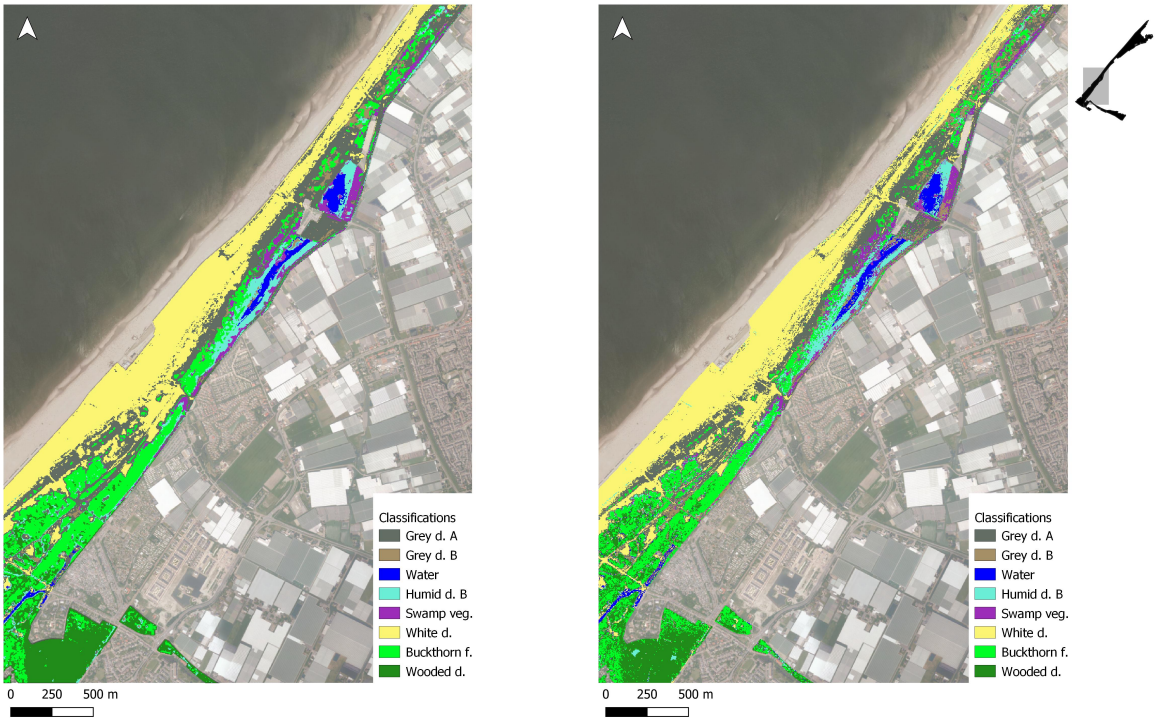
The feature importances are shown for the aerial images at 4m and 30m resolution to see whether a difference in resolution changes the importance of features. The Blue, Red and ARVI features are in the top 3 most important features in both 4m and 30m resolutions, but their positions differ. The importance factor of the Red band is much higher in the 30m resolution than in the 4m resolution. The Blue band is the most important feature in the 4m resolution model. The importance factor of the Blue and Red band are not very different in the 4m resolution model. In the 30m resolution the second and third most important features are respectively the ARVI and Blue features.

Moreover, Figures 5.4 and 5.5 show that the extra SWIR bands are not amongst the most important features, but they contribute to a better overall performance. The SWIR-1 band is more important than the SWIR-2 band. The NDVI and the ARVI are also amongst the most important features in the high spectral L7 pixel-based classification. This was expected for the NDVI as this is an often used vegetation index in classification of vegetational areas.

In Figure 3.11 (a), the surface spectral reflectance is plotted for Solleveld & Kapittelduinen. This shows that the NIR and the MIR domain play an important role in the distinction between vegetation conserving habitat types. When this is compared to the feature importances above, it is important to notice that features from the visible domain of the spectrum plays a bigger role in the Random Forest model, especially in the aerial images. The NDVI, and therefore also the NIR band, plays a more important role in the satellite images. Downsampling the aerial image to 30m, does not cause drastic differences in order of feature importance, but the most important feature (Red band) has become more important. The importance changed from 0.19 to 0.26.

The most distinct conclusions from Table 5.1 are listed below. Some of these conclusions can be supported using maps. Solleveld & Kapittelduinen is a very elongated region. To be able to spot visual differences, a focus area is plotted instead of the complete area. The conclusions are as following:

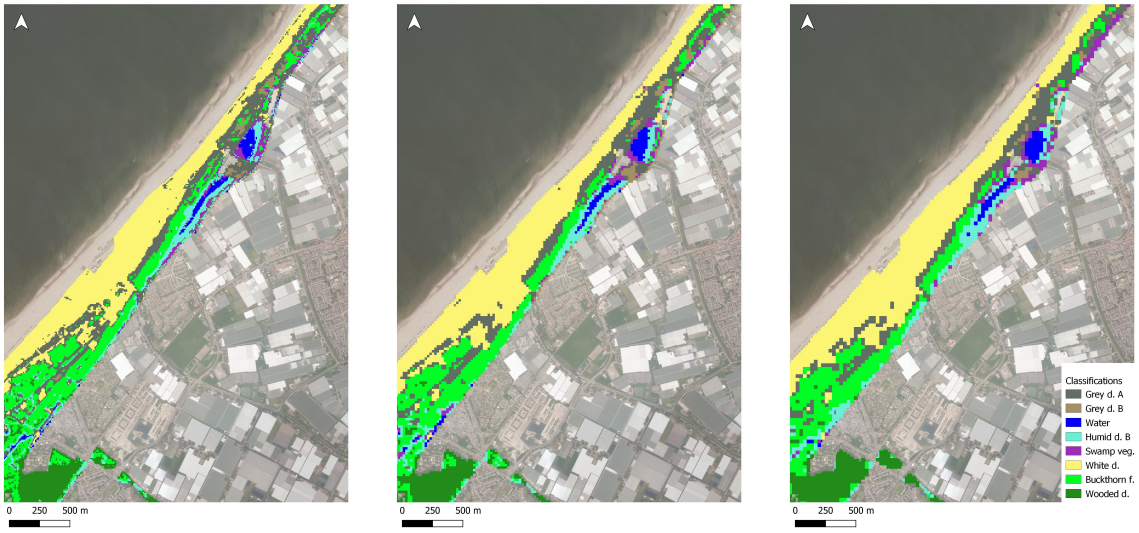
1. The region-based approach has an overall better performance than the pixel-based approach for all resolutions.
2. For both pixel-based and region-based, the performance peaks at 4m resolution.
3. The region-based approach and the addition of the 2 spectral bands (SWIR-1 and SWIR-2), and the NDWI contributed together to an increase of approximately 7% for the classification of the Landsat 7 image with respect to the pixel-based low-spectral classification.
4. The addition of the spectral features of the adjacent pixels in the region-based approach is not as important for a better performance as the addition of the statistical features.



(a) Region-based

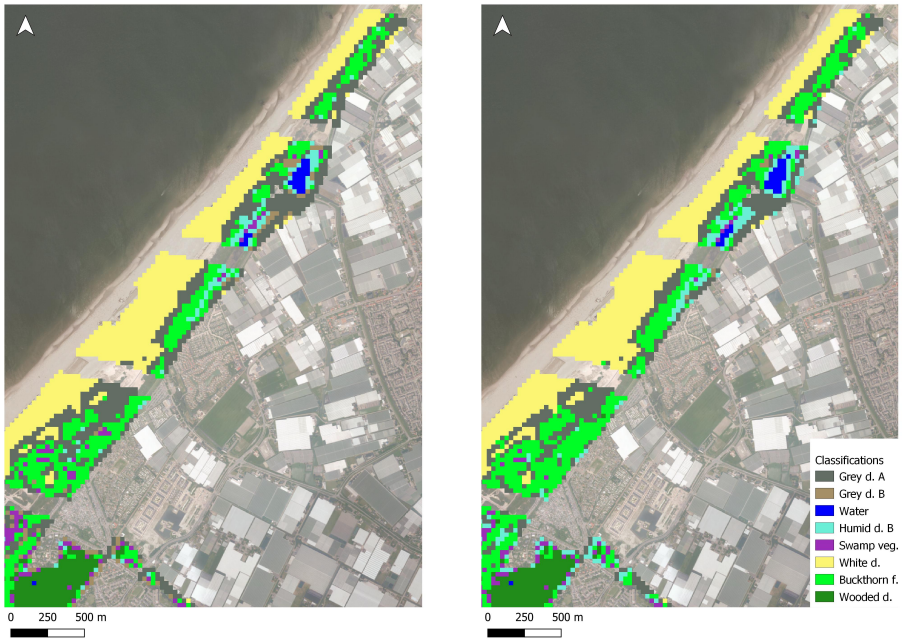
(b) Pixel-based

Figure 5.6: Classifications on a 4m resolution aerial image (2012) with the (a) region-based and (b) pixel-based method in Solleveld & Kapittelduinen. Background map is the aerial image from 2012.



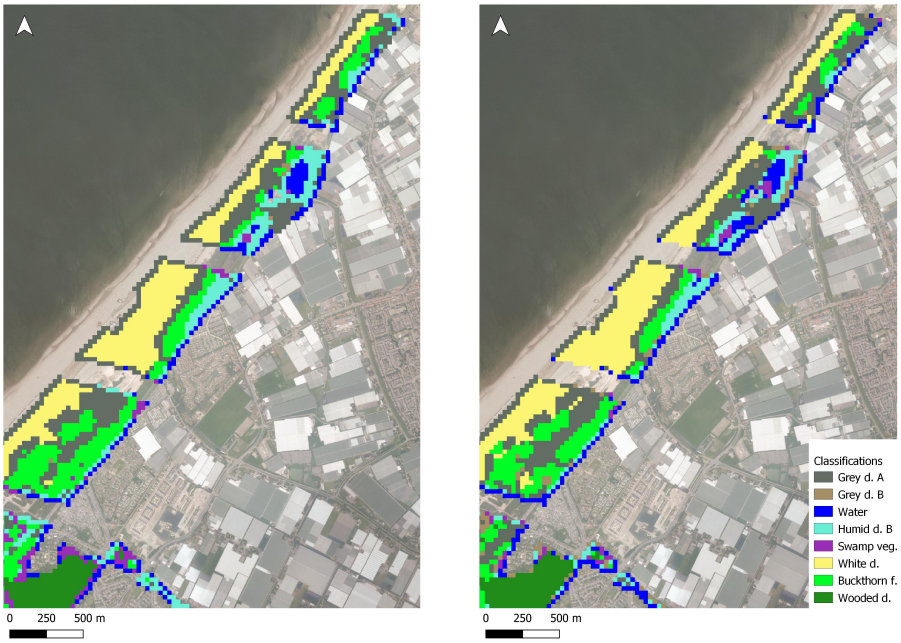
(a) 10m spatial resolution (b) 20m spatial resolution (c) 30m spatial resolution

Figure 5.7: Region-based classifications on (a) 10m, (b) 20m and (c) 30m resolution aerial images (2012) in Solleveld & Kapittelduinen. Background map is the aerial image from 2012. The sub-figures share the same legend shown in sub-figure (c).



(a) Low-spectral (b) High-spectral

Figure 5.8: Pixel-based classifications in (a) low-spectral and (b) high-spectral category on a Landsat-7 image (2012) in Solleveld & Kapittelduinen. Background map is the aerial image from 2012. The sub-figures share the same legend shown in sub-figure b.



(a) Low-spectral (b) High-spectral

Figure 5.9: Classifications with the region-based method in (a) low-spectral and (b) high-spectral category on a Landsat-7 image (2012) in Solleveld & Kapittelduinen. Background map is the aerial image from 2012. The sub-figures share the same legend shown in sub-figure b.

The region-based approach provides a less noisy classification at 4m resolution, causing a better grouping of the habitats, see Figure 5.6. As there is less detail in the true-color images of the lower resolution images, there will be also less detail in their classification maps. Figure 5.7 shows this. The deterioration in performance is directly related to the deterioration in resolution. In Figure 5.8 the pixel-based classification of Landsat 7 low-spectral and high-spectral images are shown. According to the low-spectral resolution classification, there is considerably more swamp vegetation than water or humid dunes in that area. That seems implausible when the classification is compared to the high resolution aerial images.

Region-based classification of Landsat-7 images causes even more detail to vanish, see Figure 5.9. The high-spectral classification does not recognize the swamp vegetation in the southwestern corner. A problem with the region-based approach for the Landsat-7 classifications is that the outer pixels of the image are classified unreliably, due to the SLC-failure.

From Table 5.1 and the analysis above, it can be concluded that for the aerial images, the best classification is done using the 4m resolution region-based approach. For the Landsat 7 classifications (Figures 5.8 and 5.9) the region-based low-spectral classification scored better, but if the loss of detail using the region-based method on this scale is taken into account, the better classification is achieved using the high-spectral pixel-based classification.

In Table 5.2, the confusion matrix for the 4m resolution region-based classification of the aerial image for the area Solleveld & Kapittelduinen is shown. The individual class accuracies are added to the table as producer's and user's accuracy. The imbalance in the test dataset does not influence the accuracies as each class has its own producer's and user's accuracy.

Table 5.2: Confusion matrix of the 4m resolution region-based classification. The columns represent the predicted class and the rows represent the actual class. The overall balanced accuracy is 87%.

	Grey d. A	Grey d. B	Water	Humid d. B	Swamp veg.	White d.	Buckthorn f.	Wooded d.	Producer's accuracy (%)
Grey d. A	1123	159	0	18	11	45	25	1	81.3
Grey d. B	158	1518	5	35	88	3	6	0	83.7
Water	0	9	647	70	12	0	1	0	87.6
Humid d. B	3	16	44	446	35	0	2	2	81.4
Swamp veg.	0	3	15	15	189	0	8	0	82.2
White d.	72	0	0	0	0	1430	0	0	95.2
Buckthorn f.	31	2	0	24	15	0	759	29	88.3
Wooded d.	0	0	0	8	0	0	59	1717	96.2
User's accuracy (%)	81.0	89.0	91.0	72.4	54.0	96.8	88.3	98.2	

The highest user's accuracies are scored by water, white dunes and wooded dunes. The highest producer's accuracy are scored by white dunes, Buckthorn formations and wooded dunes. Water scores also high. These scores were expected, because the appointed classes have different spectral signatures (see Figure 3.10) and they have different surface roughnesses, which is expressed by the standard deviation. White dunes are mostly plane white sand, which raises the entire spectral reflectance. Water has a very dark reflectance overall, but also in the NIR domain. The overall biggest confusion is that sub-habitats are classified as the other sub-habitat of the same main class (Figure 3.6). Grey dunes A is mostly wrongly classified as grey dunes B. This also holds for the humid dunes and the wooded vegetation. This means that even if the model is able to classify the majority right, it is struggling with detecting changes between the sub-habitats. The worst user's accuracy is scored by the swamp vegetation. The biggest confusion is between grey dunes B and the swamp vegetation. Shadows of Buckthorn formations are sometimes classified as swamp vegetation and these zones are in reality mostly grey dunes B. Most probably, the reflectance of the soil underneath the swamp vegetation is blocked by the vegetation. In that case, the reflectance of the swamp vegetation would be not much different than the reflectance of other vegetation in the area. Also the input data of the swamp vegetation is not always covering a large area. Classifying them on 4m resolution could introduce misinterpretations. The overall accuracy of this classification is 87%.

5.1.2. SPATIAL VARIABILITY

Applying the model on different areas gives insight on the robustness of the model. In this subsection, the robustness on spatial variability of the model is tested. Thus far, the spatial and spectral analyses are done on Solleveld & Kapittelduinen as there is more training data available for this area, see Figure A.1.

Table 5.3: Table with balanced accuracies (%) on an independent test dataset of the model trained on one area and predicted on a different area. The classification method is the region-based method, applied on a 4m resolution aerial image. The columns show the area the model is trained on: all areas, Solleveld & Kapittelduinen (SKD), Voornes Duin (Voorne) and Goeree & Kwade Hoek (Goeree). The rows represent the areas the model is predicted on.

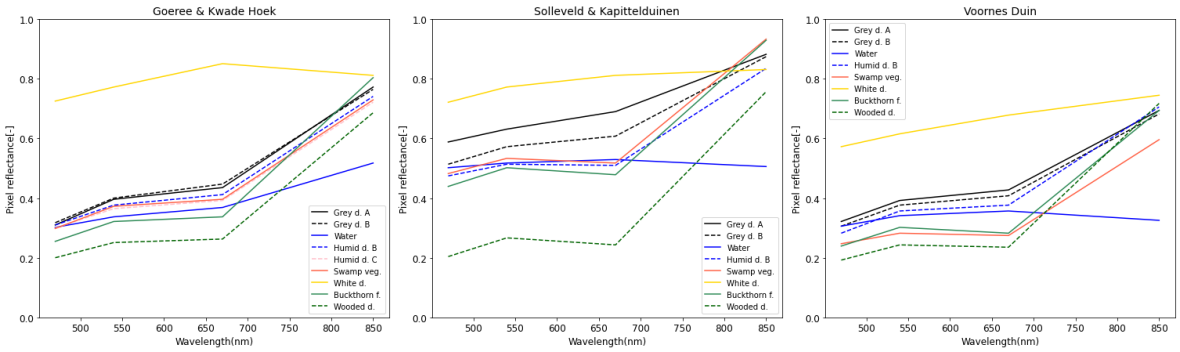
		Trained			
		All	SKD	Voorne	Goeree
Predicted	SKD	80.4	87.0	21.5	24.4
	Voorne	74.4	14.9	77.0	21.0
	Goeree	72.3	16.9	25.8	76.0

In Table 5.3 is shown how the model performs when applied to other areas. It is understandable that training on all regions will give a reasonable score for single areas. It is distinct that Voornes Duin and Goeree & Kwade Hoek perform much worse than Solleveld & Kapittelduinen when trained on all areas. This reveals that the training dataset of Solleveld & Kapittelduinen has a higher quality than the training datasets of Voornes Duin and Goeree & Kwade Hoek. This is verified by the results of training one area and predicting the same area: both Voornes Duin and Goeree & Kwade Hoek perform approximately 10% worse than Solleveld & Kapittelduinen.

In Figure 3.7 is shown that not all areas are captured on the same day. Solleveld & Kapittelduinen are captured on two different dates: the northern part is captured in July and the southern part in May. Also, the northern

part of Voornes Duin is captured in May and the southern part on a different day in July. The training dataset in Solleveld & Kapittelduinen covers the area enough to filter out the possibility on differences in phenology (season) or brightness (sun position). The training datasets for the other regions are less area covering, see Appendix A.1. Goeree & Kwade Hoek is captured on the same day like the larger southern part of Voornes Duin. This may be one of the reasons why the model of Voornes Duin works better for Goeree & Kwade Hoek than Solleveld & Kapittelduinen.

Figure 5.10 visualizes the normalized pixel values, named as 'pixel reflectance' in the plots, for the three ROIs. The pixel reflectances in Figures 5.10 (a) and 5.10 (b) are more similar than they are to Figure 5.10 (c). In Goeree & Kwade Hoek and Voornes Duin, all classes except water and white dunes, have a smaller reflectance range in comparison to Solleveld & Kapittelduinen and also the overall reflectance is lower. From Figure 3.11 and from the fieldwork it can be concluded that the soil in Voornes Duin has an overall higher humidity than the soil in Solleveld & Kapittelduinen. A humid soil has less reflectance than dry soil. Now, it can be assumed that Goeree & Kwade Hoek has a higher soil moisture.



(a) Goeree & Kwade Hoek (b) Solleveld & Kapittelduinen (c) Voornes Duin

Figure 5.10: Median pixel reflectances of the aerial images (2012) for the different classes in (a) Goeree & Kwade Hoek, (b) Solleveld & Kapittelduinen and (c) Voornes Duin.

5.2. SATELLITE IMAGE CLASSIFICATIONS

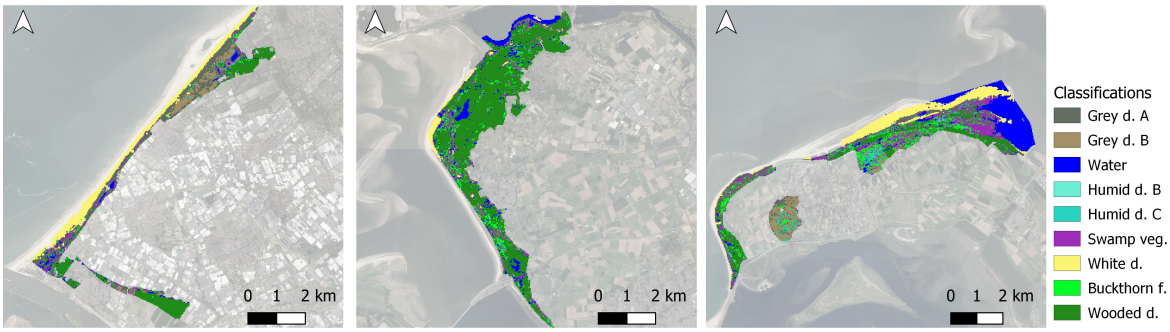
This section focuses on the yearly classifications of the Landsat-7 and Sentinel-2 images as discussed in Section 4.5. The classifications are done on both sub-habitat level as main-habitat level by the retroactive hierarchy approach. The conclusion for satellite images from the spatial and spectral resolution analysis was that the pixel-based method works better for Landsat-7, therefore the maps in this section are done on the pixel-based method. The habitat change maps are also computed on these levels and are only based on Landsat-7 images. The analysis of these maps are supported by habitat transition matrices. Sentinel-2 images are used to get a better resolution classification map in 2019. Finally, the accuracy assessment and validation is done using a confusion matrix, classification stability maps, comparison with spectral reflectances from the field and external validation.

5.2.1. LANDSAT-7 CLASSIFICATIONS

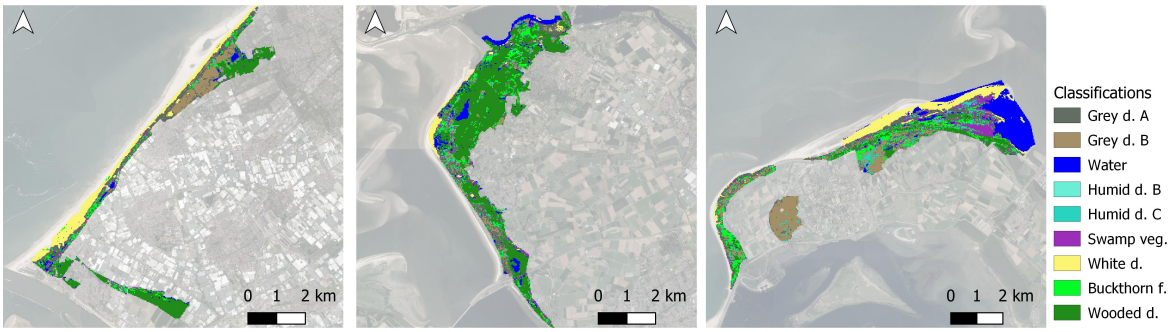
The final habitat classification maps are produced using Landsat-7 and Sentinel-2 images with the Random Forest model on the pixel-based technique, as the region-based causes the details to become less clear on this resolution scale. Area covering maps based on Landsat-7 images are showed in Figure 5.11 and 5.12. Table 5.3 shows that training on all areas of interest provides a relatively better homogeneous classification performance along the coast of Zuid-Holland. The final maps for 2012 and 2019 are therefore produced using all input data from the three Natura 2000 areas. The Random Forest model is trained on the median of all Landsat 7 images available between January 2011 and December 2013. The model is used to predict single images and the final map represents the mode of the single images. The mode is the class that occurs the most often per pixel.

From Table 5.2 it appeared that sub-classes are often wrongly classified on the 4m resolution as another sub-class from the same main-class (Figure 3.6). The same or even stronger effect is expected with lower resolutions. Therefore both sub-class habitat maps and main-class habitat maps are produced based on the retroactive hierarchy approach explained in subsection 4.5.1.

In Figure 5.11 the final Landsat-7 sub-habitat classification maps are shown for respectively 2012 and 2019. In Figure 5.12 the main-habitat classification maps are shown. Larger versions of these maps can be found in Appendix A.4.

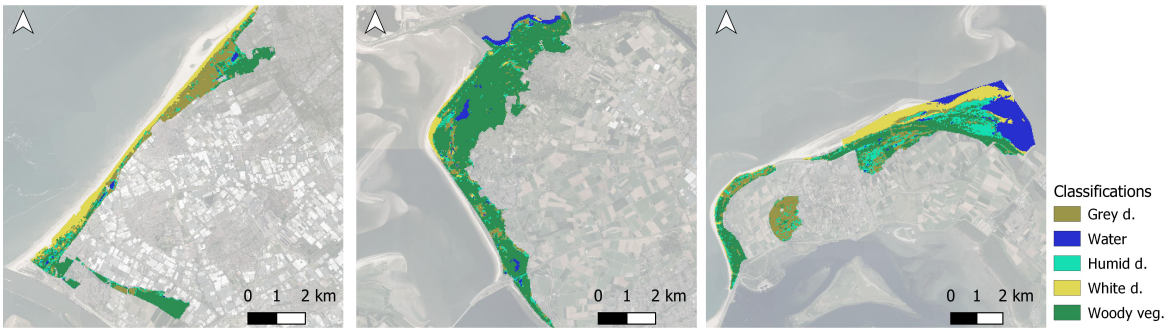


(a) Sub-habitat classification maps in 2012 (Landsat-7).

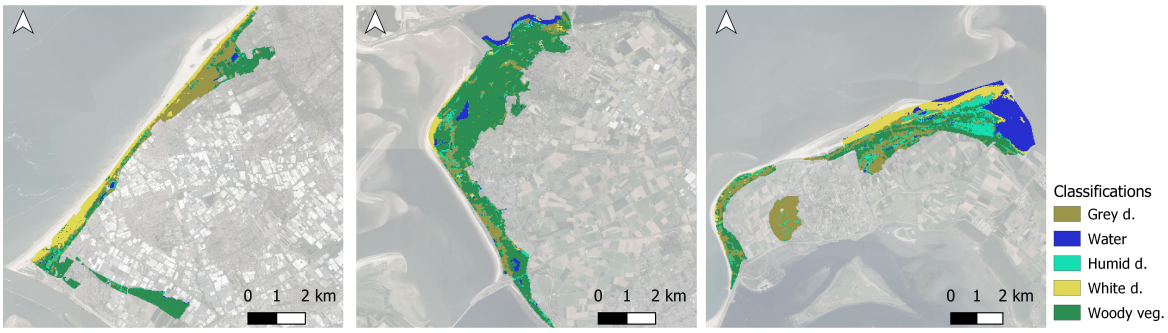


(b) Sub-habitat classification maps in 2019 (Landsat-7).

Figure 5.11: Sub-habitat classifications for Landsat-7 images in 2012 and 2019. See Figures A.15, A.16 and A.17 for larger maps.



(a) Main-habitat classification maps in 2012 (Landsat-7).

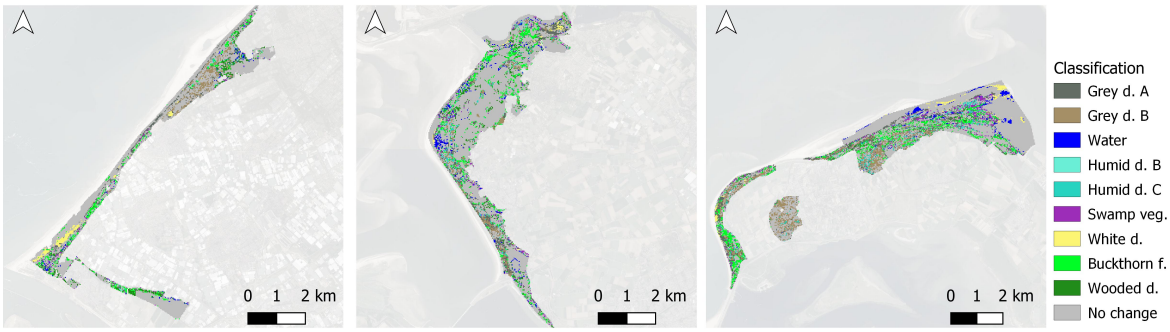


(b) Main-habitat classification maps in 2019 (Landsat-7).

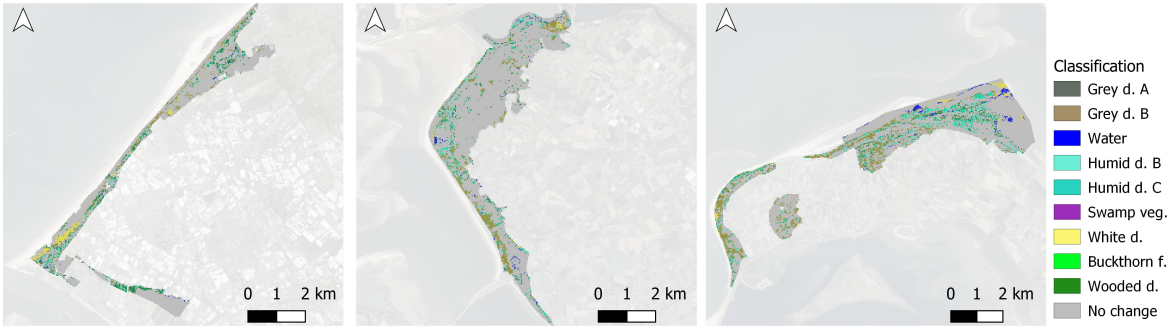
Figure 5.12: Main-habitat classifications for Landsat-7 images in 2012 and 2019. See Figures A.12, A.13 and A.14 for larger maps.

CHANGES IN HABITAT

In Figure 5.13 the changes in habitat between 2012 and 2019 for both sub-habitat level and main-habitat level are shown. The pixels with a color different than grey represent a change in habitat between the years and the color of the pixel corresponds to the class in 2019.



(a) Change in sub-habitat.



(b) Change in main-habitat.

Figure 5.13: Maps for habitats changed between 2012 and 2019 for both (a) sub-habitat and (b) main-habitat. See Figures A.18, A.19 and A.20 for larger maps.

There are clearly changes according to the maps in Figure 5.13 but to quantify these differences, habitat transition matrices are used. The construction of a transition matrix is related to a confusion matrix. The rows and the columns have the same indices (habitats). A confusion matrix informs about the performance of a model, but a habitat transition matrix gives information about the pro rata change of a habitat compared to another year, in this case 2012. Using a habitat transition matrix one assumes that the maps used to calculate the changes are completely correct. This is often not the case. It is important to keep that in mind. The matrix is based on the main-habitat classes (see Figure 5.13 (b)), because the quantification of the change in main-habitats between 2012 and 2019 is much more attractive for environmental management cases, and provides a higher certainty by cause of a better classification performance of the main-habitat classes. In Tables 5.4, 5.5 and 5.6 the percentual transition matrices of the three ROIs are shown. The changes are calculated separately for each ROI, because decisions in management of the environment can vary greatly per region. The numbers in the tables represent the percentage of pixels that changed in 2019 with respect to 2012.

Table 5.4: Percentual transition matrix (%) from 2012 (rows) to 2019 (columns) for Goeree & Kwade Hoek. For example: the grey dunes entry shows that, according to Figure 5.13 (b), 84% of the grey dunes in 2012 stay grey dunes, 10% changed to humid dunes, 1% changed to white dunes and 5% changed to woody vegetation.

	Grey d.	Water	Humid d.	White d.	Woody veg.
Grey d.	84	0	10	1	5
Water	0	90	3	6	1
Humid d.	26	2	49	1	22
White d.	9	8	4	78	1
Woody veg.	19	0	12	1	68

Table 5.5: Percentual transition matrix (%) from 2012 (rows) to 2019 (columns) for Solleveld & Kapittelduinen. For example: the grey dunes entry shows that, according to Figure 5.13 (b), 78% of the grey dunes in 2012 stay grey dunes, 5% changed to humid dunes, 5% changed to white dunes and 12% changed to woody vegetation.

	Grey d.	Water	Humid d.	White d.	Woody veg.
Grey d.	78	0	5	5	12
Water	2	75	14	0	9
Humid d.	20	5	25	6	44
White d.	14	0	1	83	2
Woody veg.	3	1	3	3	90

Table 5.6: Percentual transition matrix (%) from 2012 (rows) to 2019 (columns) for Voornes duin. For example: the grey dunes entry shows that, according to Figure 5.13 (b), 72% of the grey dunes in 2012 stay grey dunes, 1% changed to water, 16% changed to humid dunes, 2% changed to white dunes and 9% changed to woody vegetation.

	Grey d.	Water	Humid d.	White d.	Woody veg.
Grey d.	72	1	16	2	9
Water	2	82	6	1	9
Humid d.	28	5	39	1	27
White d.	14	1	2	81	2
Woody veg.	8	1	3	0	88

In Figure 2.10 the expected and possible habitat changes are illustrated. In the concept of this thesis, the most important change is the change to shrubs (woody vegetation) and from shrubs to grey dunes or humid dunes. It is known that bushes are occasionally cleared to fight the encroachment. The change from woody vegetation to grey dunes represents this clearance path and is, according to the transition matrices, 19% in Goeree (western and mid-northern coast), 3% in Solleveld & Kapittelduinen and 8% in Voorne (southern coast). According to Figure 5.13 (b) there is no indication of a man-induced change from shrubs to grey dunes in Solleveld & Kapittelduinen in contrast to the other areas. Although, there is clearly an unusual increase in white dunes in the southern part of Solleveld & Kapittelduinen (Van Dixhoorndriehoek). This area used to be a mix of grey dunes, humid dunes and shrubs. This might be caused by cutting sod (removal of top soil layer with vegetation) activities in combination with removing the shrubs. The low percentage of change in shrubs is remarkable, because the activities by the local management are started here exactly due to the encroachment. The remaining sand is classified as white dunes.

The increase in woody vegetation (encroachment) mainly comes from humid dunes and grey dunes. This may have been caused by the increase of nitrogen emission in the air due to the operation of the Maasvlakte 2 (see Chapter 1). This seems to be mostly the case in Solleveld & Kapittelduinen. The increase from white dunes to grey dunes is another remarkable change. It is possible that white dunes have been taken over by grey dunes, but it is also possible that Marram grass is growing a lot. The shape of the spectral reflectance curve of 'White Dune with Marram Grass' (see Figure 3.11) starts to resemble that of grey dunes, but is slightly raised by the high reflectance of the sand ('White Dune Sand' in Figure 3.11). When the Marram grass is growing very dense, the reflection from the sand is blocked. This can cause confusion for the classification model and in that case white dunes are classified as grey dunes.

Humid dunes seems to be the most unstable class for all three areas. Only 25% to 50% stays unchanged between 2012 and 2019. This is possible to a large extent between grey dunes and humid dunes, especially

in Goeree & Kwade Hoek where a lot of water channels are positioned in a grey dune area, which also makes the transition easier. Table 5.5 claims that 44% of the humid dunes in 2012 changed to woody vegetation in the southern part of Solleveld & Kapittelduinen. Some will be correct, as it is expected that trees and bushes grow denser around waters, principally if those areas are not used for recreational purposes. However, a large part originates from a bad classification result. Shadow zones are often wrongly classified as humid dunes or water. When these wrongly classified pixels in 2012 are classified correctly in 2019, than it will indicate it as a change while this is completely incorrect. The same holds for the transition from humid dunes to grey dunes in less moist soils. A scenario for this is that the shadow zones in 2012 are classified as humid dunes, in the years up to 2019, the bushes are removed by dune management causing the pixels to be classified as grey dunes in 2019. These are false differences.

Table 5.7: Total area change in hectares per region between 2012 and 2019 according to Figure 5.13 (b). An increase is indicated with a plus and a decrease with a minus-sign.

	Grey d.	Water	Humid d.	White d.	Woody veg.
Goeree	+144.3	+4.9	-39.2	-33.7	-76.2
SKD	-4.9	+2.3	-39.5	-5.8	+47.9
Voorne	+85.1	+0.7	-1	-0.5	-84.4

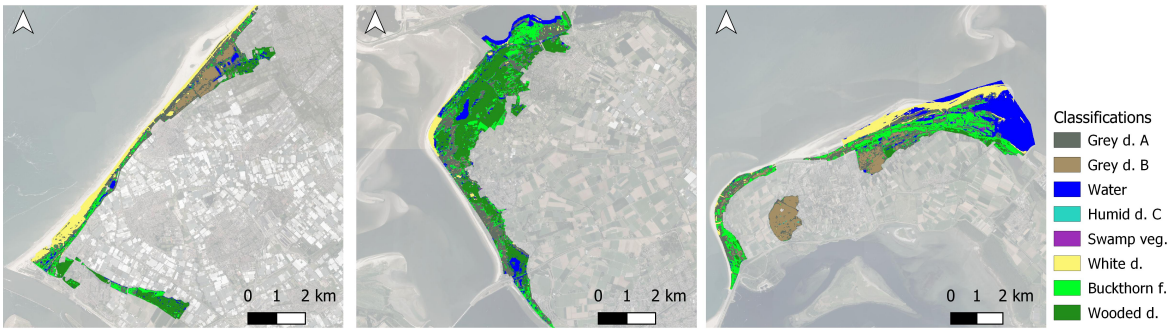
Based on a 30 by 30 pixel size, the total differences in hectares per area are calculated and shown in Table 5.7. The increase in grey dunes in Goeree & Kwade Hoek and Voornes Duin is mainly due to the management described above. The decrease in woody vegetation in Goeree and Voorne corresponds nicely with the increase in grey dunes and the environmental management of removing the shrubs. However, there has been an increase in woody vegetation in Solleveld & Kapittelduinen.

5.2.2. SENTINEL-2 CLASSIFICATIONS

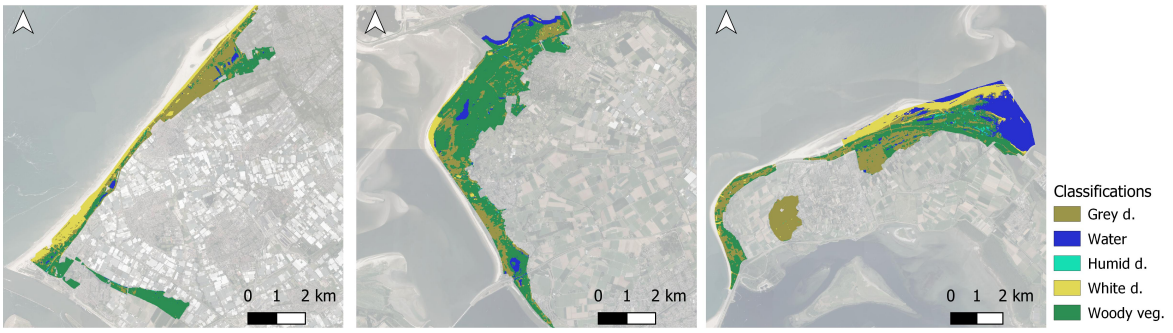
Sentinel-2A and B are since 2017 fully operational. In comparison to Landsat-7, they provide higher temporal, spatial and spectral resolution images. The only problem is that there are no data labels available for 2019 to train the Random Forest model with. This is solved by creating a new training dataset consisting of the unchanged labels from 2012 (Mouissie) based on the Landsat-7 main-habitat classifications in 2019 (Figure 5.11 (b)). This is explained in more detail in subsection 4.5.2. The high-spectral pixel-based classification technique is used. The Random Forest model is trained on the median of the available images in 2019 and the final map is generated by taking the mode of the predicted single images.

HABITAT CLASSIFICATION MAPS

In Figure 5.14 the final Sentinel-2 sub-habitat and main-habitat classification maps are shown for 2012 and 2019. Larger versions of these maps are shown in Figures A.21, A.22 and A.23. In section 5.2.1, Tables 5.4, 5.5 and 5.6 are analysed. One of the conclusions was that humid dune is the most unstable main-class and changed a lot into grey dunes and woody vegetation. This has affected the training data derived from the Landsat-7 2019 classification. There are a few humid dunes training points that have not changed between 2012 and 2019 according to the Landsat-7 final classification. Although the imbalance is compensated, this shortage has consequences for the final classification using the Sentinel-2 images.



(a) Sub-habitat classification.



(b) Main-habitat classification.

Figure 5.14: Sentinel-2 final sub-habitat (a) and main-habitat (b) classifications for 2019. From left to right: Solleveld & Kapittelduinen, Voornes Duin, Goeree & Kwade Hoek. See Figures A.21, A.22 and A.23 for larger maps.

It is not different than expected that the sub-habitat structures are more defined in Figure 5.14 than in Figure 5.11. This is a result of the improvement in spatial resolution. Grey dunes C and humid dunes B are not identified as class. Moreover, a remarkable amount of pixels are classified as water in a dense forest area in the northern and southern part of Solleveld & Kapittelduinen. The classification is possibly more unstable at these locations. More on this in section 5.3.

5.3. ACCURACY ASSESSMENT

Multiple classifications with the use of different input images are done for this research. Both aerial images and Landsat-7 images are used for the maps in 2012. For 2019, Landsat-7 and Sentinel-2 images are used. The accuracy assessment of the aerial image classifications are done in Table 5.1 and 5.2. In this section the accuracy assessments of the classifications done with satellite images are reported.

5.3.1. LANDSAT-7 CLASSIFICATION CONFUSION MATRIX

The final overall accuracies are calculated on the test dataset which is not shown to the model during training, making it an independent dataset suitable for quantifying the model performance. The final confusion matrix refers to the final Landsat-7 main-habitat classification in 2012. The balanced accuracy for the sub-habitat classification is 46% and 66% for the main-habitat classification based on the test dataset derived from the classification done in 2012 (Mouissie). The Cohen's Kappa scores are respectively 41% and 56%.

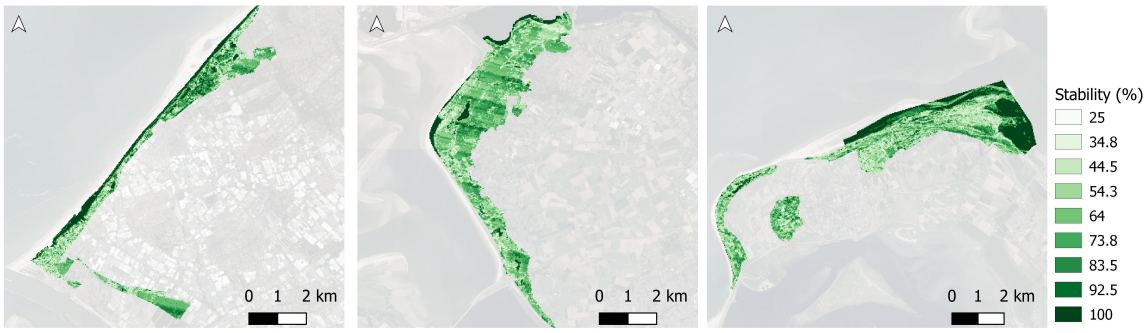
Table 5.8: Confusion matrix of the 30m resolution high-spectral pixel-based main-habitat classification. The columns represent the predicted class and the rows represent the actual class. The overall balanced accuracy is 66%.

	Grey d.	Water	Humid d.	White d.	Woody veg.	Producer's accuracy (%)
Grey d.	480	13	82	9	135	67.0
Water	10	51	19	0	17	52.3
Humid d.	66	44	81	0	58	32.5
White d.	6	0	0	166	0	96.5
Woody veg.	45	16	23	1	390	82.1
User's accuracy (%)	79.0	41.0	40.0	94.3	65.0	

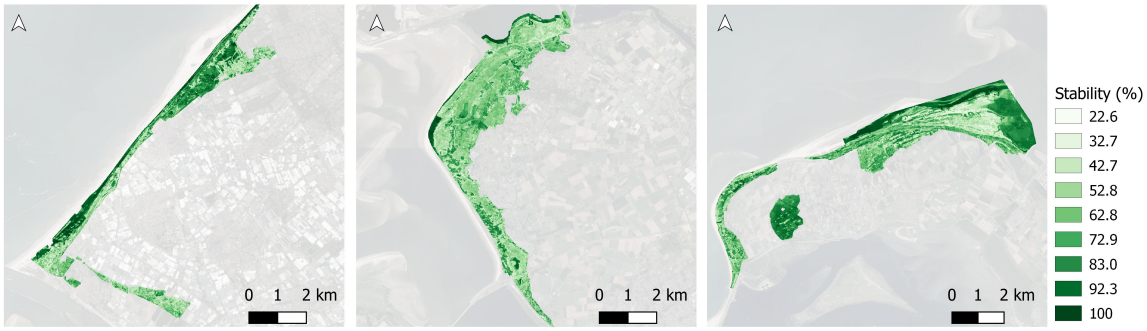
The highest producer's and user's accuracies are scored by white dunes, woody vegetation and grey dunes. The worst is scored by humid dunes and water, which is not expected, because water has a very characteristic spectral reflection. An explanation for this may be that some pixels in the test dataset are located at the edges of the waters, where the spectral reflectance is contaminated by vegetation. According to Table 5.8, water is generally wrongly classified as humid dunes and woody vegetation (shadow zones). White dunes are only confused with grey dunes.

5.3.2. CLASSIFICATION STABILITY

The final maps are produced by selecting the mode out of the single image predictions. Accordingly, the stability percentages are calculated by taking the percentage of number of times the final class is chosen over the number of available classifications. Due to the 'no-data' lines in the Landsat 7 images caused by the SLC-failure not every pixel has the same number of classifications. This effects the classification stability. The classification stability maps are based on the main-habitat classifications.



(a) Landsat-7, 2012.



(b) Sentinel-2, 2019.

Figure 5.15: Classification stability in percentages for (a) Landsat-7 in 2012 (b) for Sentinel-2 in 2019. The values show the percentages of how many times the final class is chosen from the total available classifications.

According to these images, the classifications are the most confident in classifying water, white dunes and some areas where the class is wooded dunes. Less confident areas, such as the southern part of Solleveld & Kapittelduinen and the northern part of Voornes Duin, are generally not caused by the number of available classifications, but the number of single predictions of the final class. In Figure 5.15 (b), the northeastern part of the area Solleveld & Kapittelduinen is relatively much less confident. This is because this part has been captured by a different orbit and there are fewer images of this orbit available than the orbit capturing the western part, see Figure 5.16. This only happens in Solleveld & Kapittelduinen for Sentinel-2. For Landsat-7 the SLC-failure is visible in every area.

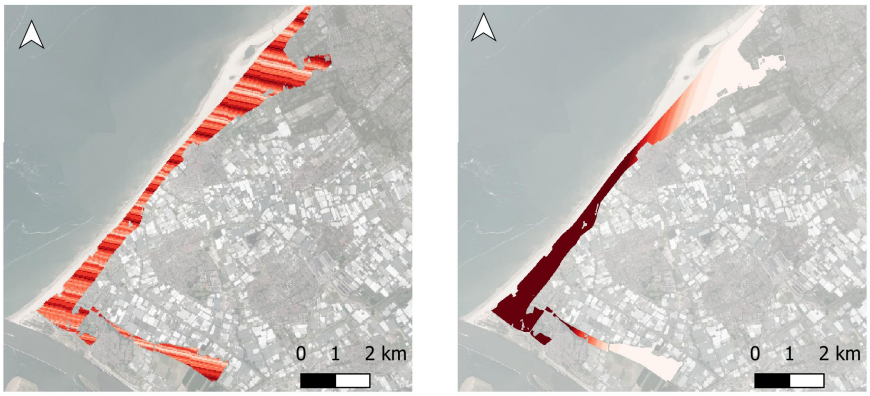


Figure 5.16: Number of available pixels in Solleveld & Kapittelduinen for 2019 Landsat-7 (left) and 2019 Sentinel-2 (right).

5.3.3. COMPARISON WITH FIELD SPECTRA

In Section 3.2 the fieldwork is described, the locations and the appearances of the different footprints are visualized with pictures. These fieldwork locations are also used as validation material for the final classifications of 2019, both Landsat-7 and Sentinel-2. The validations are reported in Table 5.9 and 5.10. Correlation plots are used to better understand and analyse the results in the tables (Figure 5.17).

Table 5.9: Validation of the Landsat-7 and Sentinel-2 final classifications on fieldwork points for day 1 in Solleveld & Kapittelduinen shown in Figure 3.4. P stands for point location.

Solleveld & Kapittelduinen

	Ground-truth	L7	S2		Ground-truth	L7	S2
P1	Humid d.	Humid d.	Grey d.	P7	Buckthorn f.	White d.	White d.
P2	Humid d.	Humid d.	Grey d.	P8	White d.	White d.	White d.
P3	Humid d.	Humid d.	Grey d.	P9	Buckthorn f.	White d.	White d.
P4	Humid d.	Humid d.	Humid d.	P10	White d.	White d.	White d.
P5	Grey d.	White d.	White d.	P11	White d.	White d.	White d.
P6	White d.	White d.	White d.	P12	White d.	White d.	White d.

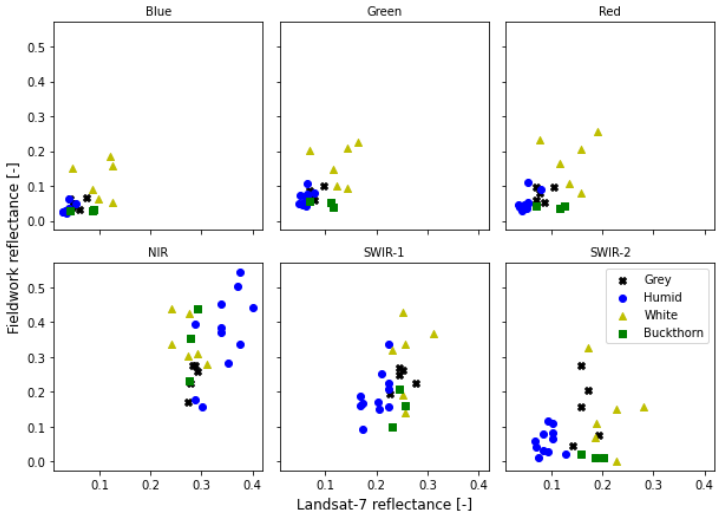
There are difficulties classifying humid dunes in Sentinel-2 images. The pixels are generally classified as grey dunes (which humid dunes resemble the most). This limitation of the model is also deductible from Table 5.9. At P5, the grey dunes consist mainly of grey dune related vegetation, but the subsoil and the surroundings are mostly sand (white dunes). The Buckthorn formation measurements are done on transition areas from white dunes to Buckthorn formations. Unfortunately, it was impossible to do the measurement deeper in the formations. In Figure 3.4 - P9, one can see the Marram grass clumps in the background. It appears that Landsat-7 classification has performed better on these points than the Sentinel-2 classifications.

Table 5.10: Validation of the Landsat-7 and Sentinel-2 final classifications on fieldwork points for day 2 in Voornes Duin shown in Figure 3.5. P stands for point location. P14 locates outside the official boundaries of the Natura-2000 area and therefore also outside the classification boundaries.

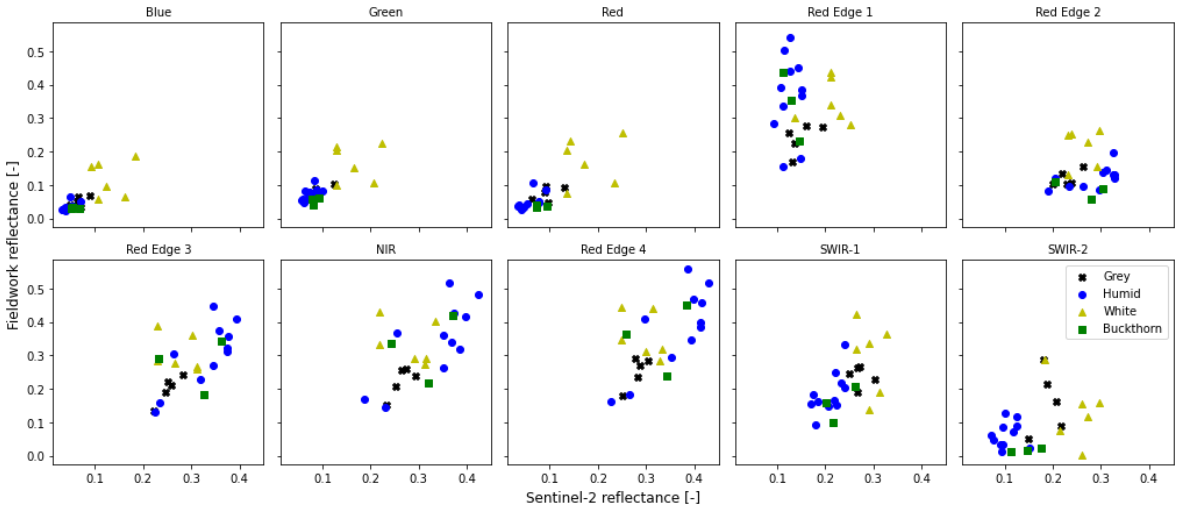
Voornes Duin

	Ground-truth	L7	S2		Ground-truth	L7	S2
P1	Grey d.	Grey d.	Grey d	P8	Humid d.	Humid d.	Humid d.
P2	Grey d.	Grey d.	Grey d.	P9	Humid d.	Woody veg.	Woody veg.
P3	Grey d.	Grey d.	Grey d.	P10	Humid d.	Woody veg.	Woody veg.
P4	Buckthorn f.	Grey d.	Grey d.	P11	Humid d.	Woody veg.	White d.
P5	Grey d.	Grey d.	Grey d.	P12	Humid d.	Woody veg..	Grey d.
P6	White d.	Grey d.	Grey d.	P13	Humid d.	Woody veg.	Woody veg.
P7	Humid d.	Grey d.	Grey d.	P14	Humid d.	nd.	nd.

P4 is measured on a Buckthorn shrub surrounded by a grey dunes field. P6 is a small patch of white dune in a larger grey dune area. Apparently not large enough to be classified as one pixel of white dunes. P7 is a humid area, with low vegetation and the color resembles that of grey dunes. P9 till P13 are located in a humid dune field which has quite high vegetation. This blocks most probably the reflection of the moisture and causes the model to classify these locations as bushes instead of humid dunes.



(a) Correlation between Landsat-7 reflectances and fielwork measurements



(b) Correlation between Sentinel-2 reflectances and fielwork measurements

Figure 5.17: Correlation plots between satellite reflectances and field spectrometer reflectances for Landsat-7 and Sentinel-2 satellites (S2: 24-08-2019, L7:22-8-2019 and 14-09-2019). The plots are grouped per band and colored per habitat type.

The correlation plots are based on the reflectances from images dating close to the fieldwork dates. The images on 24-08-2019 (S2), 22-8-2019 (L7) and 14-09-2019 (L7) are used. Two images are used from Landsat-7 due to the SLC-failure lines. The mean fieldwork reflectances are used to compare with the satellite images.

The correlations vary notably per band and satellite. Blue, Green and Red bands are quite correlated for both satellites and for all habitats, but not for white dunes. This is caused by the fact that different types of white dunes are measured; white dunes with Marram Grass, white dunes without vegetation, white dunes with shells, etc. These different land covers are distinguishable by the spectrometer, which has a higher resolution due to its smaller footprint, but less distinguishable by a satellite sensor. The Landsat-7 satellite is less correlated in the NIR, SWIR-1 and SWIR-2 bands. There is only some correlation for the humid dunes. For the Sentinel-2 satellite, the correlations are mostly weak in the Red Edge 1. White dunes are again not very high correlated and have a large variance, due to the wide variety of the white dune types that are measured. The

poor correlations are mainly caused by the difference in footprint between spectrometer and satellite, but also due to outliers in the spectrometer data (Figure 3.11).

5.3.4. EXTRA VALIDATION

Previously, the main focus was the assessment of the classification and looking at the classification stability by obtaining the balanced accuracies, Cohen's Kappas and classification stabilities. There is not more validation data available for 2019, besides the fieldwork points and comments from the local manager. Although, a new fieldwork map is produces by the province of Zuid-Holland. This map is not yet validated, but is compiled from several reliable sources. It is used as an extra validation for Solleveld & Kapittelduinen in 2019.

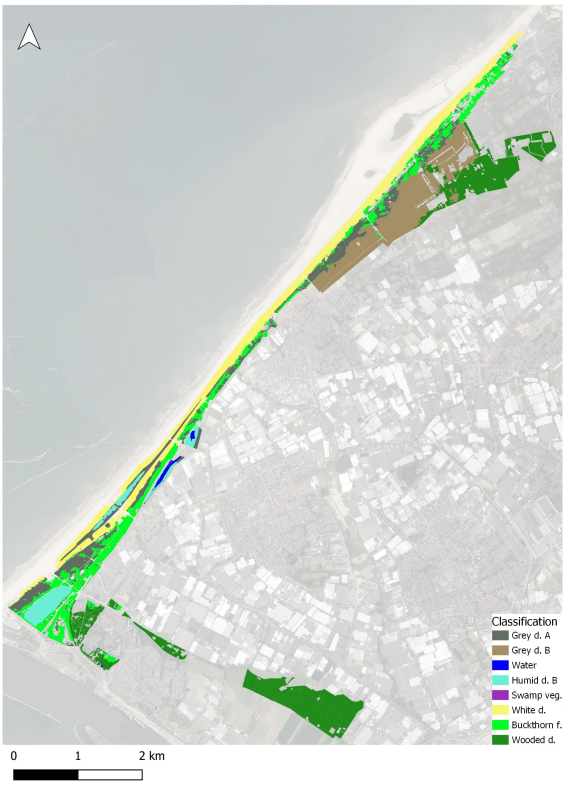


Figure 5.18: Sub-habitat map from the province Zuid-Holland 2019 used as external validation.

First of all, it is important to state that this map is an object-based map, which is different than the final satellite image classifications created in this research (pixel-based). When the classifications of Solleveld & Kapittelduinen from Figures 5.11 (b) and 5.14 (a) are compared to the map above, there is in general much similarity. The considerable differences are mostly in the southern part of the area (Van Dixhoorndriehoek). It was already stated that the dune management took action and removed all vegetation causing the sand to expose. According to the province, this area is classified as grey dunes. In this research, that area is classified as white dunes. Many of the humid dunes in this area are classified as woody vegetation due to the high plant coverage. In the northern part of the area, some blow-out spots (sand) are detected and classified as white dunes. According to the description in subsection 2.1.2, blow-out spots belong to grey dunes. The map of the province has considerable more shrubs than the final classifications of this research.

SPANJAARDS DUIN

As described earlier, Spanjaards Duin is created in 2009 as a new dune area. The goal is to create mainly humid dune slacks, humid dunes and grey dunes. Careful excavation of five depressions have been done to bring the surface closer to the groundwater. The expectation is that these excavations help the area to develop towards a humid dune environment. From the various fieldwork visits, it can be said that there is some dynamics ongoing in the right direction. Some humid dune slacks have already been formed and

are mostly visible in the wetter periods of the year. Unfortunately, these slacks are note visible in the 2019 classifications with Landsat-7 and Sentinel-2. They are classified as white dunes. The reason is that the used images (Figure 3.9) are mainly from periods of the year when the depressions are often dry. However, these valleys are slightly visible in the stability maps, see Figure 5.19. This means that the humid areas are classified correctly in some images, but due to the technique used to produce the final map, these few 'correct' classifications are averaged out.

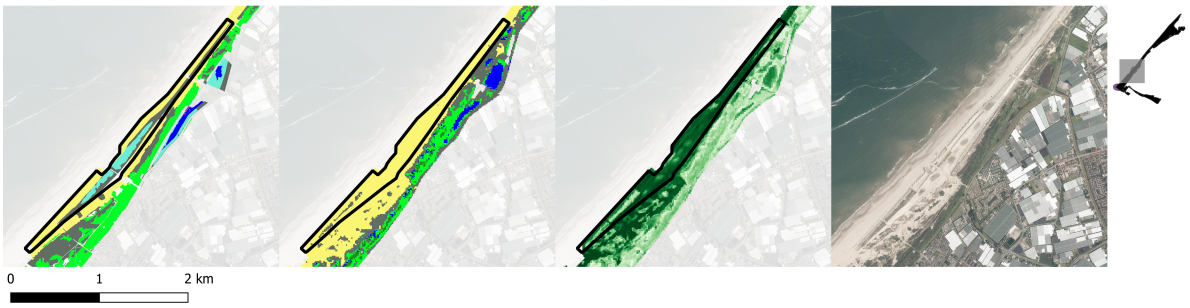


Figure 5.19: Spanjaards Duin in 2019. Left to right: Province Zuid-Holland classification, 2019 Sentinel-2 classification, 2019 Sentinel-2 stability map, aerial image (2019) 25cm resolution (PDOK).

6

DISCUSSION

This chapter discusses the processing of the data, the results more in depth, comparison with other similar research and the contribution of this thesis to similar researches in the future.

6.1. PROCESSING AND DATA PRODUCTS

Data-processing is the largest part of this thesis. Difficulties, choices and assumptions during the research related to data-processing are discussed in this section. The processing is closely linked to the data products.

6.1.1. ATMOSPHERIC CORRECTION

The effects of the atmosphere need to be corrected to do physical analysis with the remote sensing data. For this research, different vegetation and water indices are used: NDVI, ARVI and NDWI. NDVI and NDWI are popular indices in remote sensing related studies. ARVI is chosen due to its low sensitivity for atmospheric effects (Kaufman and Tanre, 1992). The atmospheric interferences in the satellite images used in this research are assumed to be sufficiently corrected for land areas with the used correction models. Due to their positioning, coastal dunes are possibly impacted by both marine and terrestrial aerosols. This can impact the atmospheric correction. No extra attention has been paid to correct more for inland waters, besides the use of ARVI.

In Figures 5.2 to 5.5 the feature importances of the aerial images on 4m, 30m and of the low-spectral and high-spectral resolution satellite image (L7) are shown. ARVI is not one of the most important features, but also not the least important. For the aerial photos of 4m it is in the top 3. A sufficient score is expected for the water classifications. This is true for the aerial image classifications (Table 5.2), but not for the Landsat-7 classifications. The user's and producer's accuracies for the water habitat are one of the lowest in Table 5.8.

6.1.2. CLOUD MASKING

One of the drawbacks of optical remote sensing are clouds. High cloud coverage makes an image unusable or creates a lot more work in the pre-processing steps. The satellite image selection for this research was relatively simple. Google Earth Engine is used for selecting and downloading the images to post-process in Python. The images are filtered on date and clouds. Only images with cloud percentage cover lower than 30% are selected as possible images. This was often enough to filter out the largest part of the cloudy images. However, the cloud percentages are determined by the cloud detector algorithms. There are different types of clouds. The altocumulus cloud and cirrus cloud sometimes tend to escape the cloud detector. To remove these images, a manual check was done after filtering the clouds automatically.

A manual check is obviously not the most favourable way to 'detect clouds', especially not when the stack of images becomes bigger than it is in this research (Figure 3.9). There are several algorithms for singular cloud detection, including bit masking based on the Quality Assessment band (QA) in a.o. Landsat-7 products. This is not applied to the images used in this thesis. The most important reason for this is that the areas of interest are relatively small compared to the spatial resolutions, especially when Landsat-7 images are used. In the classification confidence maps, shown in Figure 5.15, there are no areas with low confidence directly related

to cloud presence.

6.1.3. GOOGLE EARTH ENGINE

Google Earth Engine is used to filter and select available satellite images for the convenience of being able to easily switch between data products. There was no need to switch between the EarthExplorer of the USGS and Copernicus Open Access Hub of ESA. Besides the fact that Google Earth Engine has lots of data and is a cloud-based platform, it also has other functionalities. Several algorithms are implemented into Google Earth Engine. The aerial images are classified using the Random Forest implementation of the Scikit-Learn package in Python. The drawback of Google Earth Engine is that source codes of the algorithms are not accessible. It is possible that there are differences in implementations of the Random Forest algorithm between Google Earth Engine and Scikit-Learn. It is not the proper way to compare analysis of classifications done with two different Random Forest implementations. Therefore, both satellite images and aerial images are predicted using the Scikit-Learn Random Forest implementation. Data acquisition can also be automated for larger projects via for example a Copernicus API in Python.

6.1.4. DATA LABELS

As described in section 3.4.1, the input data consists of random sampled points from a thematic classification map done in 2012 by Mouissie et al. (Mouissie et al., 2014). The polygons were not drawn accurate enough for this research and are therefore improved by adapting the existing polygons or by creating more polygons (Appendix A.1). The training samples are a representation of the classes. It is chosen to sample 10,000 points on purpose to have a sufficient enough representation of the class. From Table 5.3, it is clear that Solleveld & Kapittelduinen performs in training the region and predicting the same region, at least 10% better, than Voornes Duin and Goeree & Kwade Hoek. From the distribution of the training samples over the regions (see Appendix A.1) it is clear that the training samples in Solleveld & Kapittelduinen are better distributed over the area than the other regions. The distribution of the training samples plays an important role in classification performance.

Furthermore, in section 4.2.1, the resampling of the aerial images and the effects on the training data (duplicates) is discussed. The smaller the pixel size (lower resolution), the less unique samples are left. Less unique samples of a class cause a worse representation of that class in the model. This decreases the learning variance of the model leading to more classification errors.

6.1.5. RADIOMETRIC CORRECTION OF AERIAL IMAGES

A large part of the spatial resolution analysis is done using the (resampled) aerial images. The aerial images are not radiometrically corrected and therefore the values of the pixels in the aerial images are not spectral reflectances, but Digital Numbers representing the color with a value from 1 to 256. There was not a fast and accurate way to do a radiometric calibration as these images date from 2012, and therefore it is pretended that the normalized DNs are the spectral reflectances. Deriving the spectral reflectances from satellite images and calibrating the aerial images with these derivations might lead to more distortions in the images due to the difference in spatial resolution and spectral resolution (bandwidths).

If the aerial images would have dated from 2019, they could have been calibrated using the spectral signatures obtained in the field (Section 3.2). In Figure 3.11 the mean spectral reflectances per habitat cover type are shown. In Figures 3.4 and 3.5 the locations of the measurements are shown. Each of these measurements could be used to calibrate aerial images. Unfortunately, no field spectra from 2012 could be obtained.

6.1.6. SCAN LINE CORRECTOR

The 'no-data' lines in Landsat-7 products are caused by the SLC failure described in subsection 2.2.4. The problem is not only the missing lines. These images also have distorted pixels and are mostly positioned on the transitions from data to no-data areas in the image. This is not a dominating problem and this problem is also kept in mind while choosing the images. Still, it is important to be aware of these distorted pixels.

6.1.7. MODIFIED LABELS FOR 2019 SENTINEL-2 CLASSIFICATIONS

One of the conclusions from Table 5.1 was that the 4m and 10m are one of the best performing spatial resolutions for both pixel-based and region-based classifications. Therefore, it is decided to also classify the dune habitats using Sentinel-2 images. There were no data labels available in 2019 to train the Random Forest

model with, so the input data is derived from the 2019 Landsat-7 classification (subsection 4.5.2). According to the 2019 classification done with Landsat-7 images, in all three ROIs there was a decrease in humid dune area. This caused drastic reduction of humid dune slack points in the training labels for Sentinel-2. The classifications contain no humid dunes B and few humid dune slacks C (Figure 5.14). The results for Solleveld & Kapittelduinen is assessed with the unverified map from the province (Figure 5.18) and by checking with the local dune manager. Both in Voornes Duin and Goeree & Kwade Hoek, it is indeed not true that there is almost no humid dune slacks B and C. For the 2012 data labels, the polygons are checked with experts who have a lot of knowledge about these areas. This could also be done for 2019.

6.2. RESULTS

In this section, the results shown in Chapter 5 are discussed in terms of new insights, potential confirmation of expectations, limitations of the method that most probably effect the results and implications of the results.

6.2.1. RANDOM FOREST AND XGBOOST

In Section 4.2.3, a performance test on a subset of the total dataset is done to see whether XGBoost is more advantageous in this research. Chen and Guestrin (2016) claim that this method is used by data scientists to achieve state-of-the-art results on many machine learning challenges. It is also known that XGBoost wins a lot of the Kaggle competitions. Kaggle is an online community of data scientists and machine learning practitioners. However, in Figure 4.2 is seen that although it performs better in accuracy, it performs worse in run time. This was not expected, because XGBoost is popular for its speed besides its performance in accuracy.

XGBoost and Random Forest are both ensemble learning methods, based on boosting and bagging respectively. Boosting methods need more hyperparameter tuning as each tree is dependent on its predecessor and is also more sensitive to overfitting. This means that, in order to make XGBoost outperform Random Forest in similar researches, one should keep in mind that it is a requirement to attentively tune the hyperparameters. In Section 4.2.2 is explained why there is little time spend on hyperparameter tuning. Random Forest has thus the advantage that it requires little tuning and has relatively high performance. It is an easy to understand algorithm (Figure 2.17) and it is often used in remote sensing of coastal dunes researches, e.g. in Gavish et al. (2017) and Timm and McGarigal (2012).

6.2.2. SPATIAL AND SPECTRAL RESOLUTION ANALYSIS

The results of the spatial and spectral resolution analysis are summarized in Table 5.1. It is concluded that the best performing resolution is 4m in combination with the region-based method resulting in an accuracy of 87%. The statistical features play a more important role than the spectral features of the adjacent pixels (Figure 5.1). The region-based method uses a 3×3 window to calculate statistical relationships between the center and the adjacent pixels. Apparently, this window size worked best for the 4m resolution, whereby the model uses information from a $12m \times 12m$ frame. When classifying dune habitats, the detection of groups of vegetation is important. Moreover, out of the four statistical features (minimum, maximum, median and standard deviation) it is expected that the standard deviation plays the most important role. The standard deviation quantifies the surface texture in this case. Higher or lower resolutions than 4m partially fail in detecting these above-mentioned characteristics. A too high spatial resolution (0.25m, 0.5m) will detect the sand in between the vegetation. Moreover, the shrubs, for example, won't be detectable using a too low-spatial resolution (10m, 20m, 30m). A more popular classification method that also uses contextual information is object-based classification (Jobin et al., 2008, Whiteside et al., 2011). The conclusion from these researches is that object-based performs better than pixel-based in most cases and that contextual information is advantageous for the classification. Also in this research, the statistical information is almost always more important than the conceptual spectral information (Figure 5.1).

Furthermore, the region-based method is not applicable for low-spatial resolutions as details blur out even more (for Landsat-7 images the information is derived from a frame of $90m \times 90m$), although Table 5.1 claims that region-based performs better for the used Landsat-7 image. The region-based method performs better on the low-spectral Landsat-7 image than on the high-spectral. This is a very unexpected result. Even tough region-based classification on 10m spatial resolution performs better than pixel-based, it is chosen to do only pixel-based classifications with satellite images to stay consistent. The high-spectral pixel-based classification performs approximately 5% better than the low-spectral. It is also striking to see that although

the high-spectral pixel-based performs better, the SWIR-1 and SWIR-2 bands are one of the least important features. The NDVI is the second most important feature for the classification of high-spectral Landsat-7 image and the NDWI the least important (Figure 5.5). This implies that the information NDVI gives the model, is already enough information about the dune habitats.

6.2.3. MISCLASSIFICATIONS

The most common misclassifications are shallow waters, suspended waters, water with algal bloom or sunglint. These waters are often misclassified as grey dunes or white dunes. See Figure 6.1 for some examples. Shadow zones of vegetation are not a huge problem in lower spatial resolution images, especially when the satellite orbit is sun-synchronous. However, it causes a problem when the classification is based on a high resolution image (0.25m, 0.50m). These are common problems in optical remote sensing and, therefore, different approaches have been developed to deal with them. For example, Tolt et al. (2011) use a Digital Surface Model to help detecting shadows and Overstreet and Legleiter (2016) propose a glint removal method with the use of field measurements of water depth.

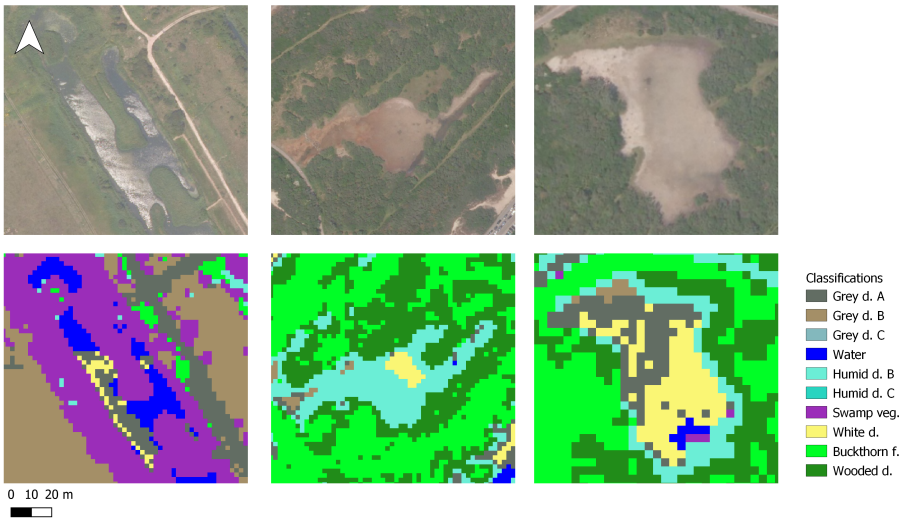


Figure 6.1: Some examples of the effects of most probably sunglint (left) and suspended waters on the classifications in Solleveld & Kapittelduinen. The classification is the region-based 4m resolution classification. The aerial image shown on the first row is the aerial image in 2012.

6.2.4. HIERARCHICAL APPROACH: SUB-CLASS VS MAIN-CLASS

In the paper of Gavish et al. (2017) a hierarchical classification based on Random Forest is introduced. To each main-class, a separate model is fitted. This resulted in an integrated model architecture to a single prediction. The hierarchical approach outperformed the flat approach (classifying directly the sub-class) for complex hierarchies, however for a simple hierarchy, the flat approach outperformed the hierarchical approach. From, among other things, the dimensionality reduction done in subsection 3.5.2 it was concluded that there is clearly some distinction in the sub-habitat types, but there is also overlap between these sub-habitat types of the same main-habitat. Moreover, from Table 5.2, it was concluded that most misclassifications on sub-habitat level are not wrong in main-habitat level. To give an example, grey dunes A is often misclassified as grey dunes B or when wooded dunes are wrongly classified, they are classified as Buckthorn formations. The combination of the conclusion that a flat approach outperformed the hierarchical approach for a simple hierarchy and the results described above, has led to the decision to use a retroactive hierarchy. The retroactive hierarchy implies that first the sub-class is classified and from there the main-habitat is derived. There are two reasons for why it is ultimately decided to create both sub-habitat maps and main-habitat maps. First of all, the classification maps are more accurate (improvement of 20%, see Section 5.3) on main-habitat level and the habitat change maps (Figure 5.13) are more important to analyse on main-habitat level than on sub-habitat level in the context of this research.

6.2.5. CHANGES IN HABITAT BETWEEN 2012 AND 2019

In Figure 5.13 the habitat change maps are shown. These maps are based on the 2012 and 2019 classifications of Landsat-7 images shown in Figure 5.11 and 5.12. Tables 5.4, 5.5 and 5.6 are based on these change maps. Before doing this research, it was expected that the detected changes are mostly due to dune management activities, e.g. removing shrubs to fight encroachment. According to Table 5.7 this is mainly done in Goeree & Kwade Hoek and Voornes Duin (decrease in woody vegetation). The changes in these maps are checked by the local manager and largely confirmed. In Solleveld & Kapittelduinen there is a gain in woody vegetation, which is mainly east of Spanjaards Duin and in the northern part of the area (see Figure 5.13).

6.3. COMPARISON WITH SIMILAR RESEARCH

Many researches in the field of coastal habitat classifications using remote sensing, are done with Random Forest as the classifier model (Gavish et al., 2017, Timm and McGarigal, 2012). The ones using aerial imagery or data from an UAV also use additional information, such as Digital Elevation Models (DEM). Gavish et al. (2017) researched hierarchical classification of coastal dune habitats in Natura 2000 areas by comparing it to flat classification. Flat classification means classifying the sub-classes directly without an initial discrimination between main-classes. They claim that pixel-based classification using 10m spatial resolution data and high-spectral resolution data (8 and 4 spectral bands) show similar results as a 2m and 5m spatial resolution remote sensing data product. Similar to this thesis, the data was initially not at 10m resolution. The original resolution is resampled using cubic convolution interpolation (not using the average as was done in this research). Moreover, a DEM, a canopy height model and more environmental variables are used as extra information for the classifier. One of the most outstanding results is that the environmental variables have considerably higher variable importances. The performance of both flat classification and hierarchical classification in that research exceeds 90%. In the paper of Timm and McGarigal (2012) aerial orthophotography imagery (0.5m spatial resolution) is used together with a DEM and high resolution satellite images (4 spectral bands, 2.4m spatial resolution) to classify coastal dune and salt marsh ecosystems at Cape Cod National Seashore using Random Forest. The data products are resampled to 1m spatial resolution. Moreover, focal statistics (pixel range and mean for 2×2 and 4×4 radii pixels) are derived and added as contextual information to fight, among others, effects of small scale heterogeneity to the classification results. According to the results described in this paper, the focal statistics did not contribute to a better overall performance. The overall accuracy of the final classification of the dune areas is 75.1%.

There are some similarities and differences in the use of data products, methodology and results between the research done in this thesis and the researches of Gavish et al. (2017) and Timm and McGarigal (2012). First of all, it is very important to mention that, in contrast to these researches, no additional environmental features (e.g. DEM, canopy height model, soil cover information, etc.) are used in this thesis besides the statistical information derived by the region-based method (standard deviation, median, minimum and maximum) and the spectral information of the pixels. This research showed that the best classification is achieved on a 4m resolution aerial image with the region-based method resulting in an accuracy of 87% and Cohen's kappa of 86%. The idea of the statistical features created with the region-based method is comparable to the focal statistics used in Timm and McGarigal (2012), where only the pixel range and mean value is used. Whereas in this thesis, the standard deviation, median, minimum and maximum are used. In Figure 5.1 is shown that the statistical features have a greater contribution to the performance. Moreover, from Table 5.1 it can be seen that the 10m resolution is the second best performing resolution, whereas in Gavish et al. (2017) is claimed that higher resolutions of 2m and 4m show similar results.

The use of DEM or other height models in classifying dune habitats is understandable as the dune surface height relative to the sea level plays an important role in the location of dune types (Figure 2.4). This research scores high, without the use of any environmental variables, in separating sub-habitat types (87%). The statistical features play a more important role than only spectral features. In remote sensing of dune habitats, as far as the author is aware of, no studies have so far been conducted in which such accuracies are achieved using only aerial imagery and contextual statistics, e.g. standard deviation, mean, minimum and maximum. High accuracies are achieved in sub-habitats, with the misclassification staying in the same main-habitat. This indicates that main-habitats are even better separable. The methods used are also easily applicable to any land cover classification related study.

7

CONCLUSIONS & RECOMMENDATIONS

In this chapter, conclusions are drawn based on the results and discussion. Moreover, recommendations are given for further improving classification of dune habitats in future researches. Conclusions are presented as answers to the main research question and the sub-questions.

7.1. CONCLUSIONS

In the introduction, the main research question was formulated as following:

> **"What are the effects of resolution of aerial images and optical satellite images on the classification of dune habitat types in coastal Dutch Natura 2000 areas?"**

Aerial images and Landsat-7 images are used in this thesis to do spatial and spectral resolution analysis with the traditional pixel-based method and the region-based method. The region-based method uses additional statistical and spectral information from the adjacent pixels. Among the different resolutions tested, the performance of both pixel-based and region-based methods peaks at 4m resolution, resulting in accuracies of respectively 78% and 87%. The second best resolution is the 10m resolution (75% and 83%). The statistical features contribute more to the classification performance than the spectral features from adjacent pixels. The 4m resolution is high enough to detect necessary spectral and spatial information from the land cover, but low enough to smooth out the spectral reflectance of the substrate. The misclassified pixels are often classified as the other sub-habitat of the same main-habitat. Moreover, a higher spectral resolution contributes to a better classification performance whereby the Red, NDVI and Blue features are the most important. Data from different Natura 2000 areas improve the overall applicability of the model in classifying dune habitats. The yearly habitat maps based on Landsat-7 images (30m resolution) show a low classification accuracy on both sub-habitat (46%) and main-habitat level (66%). The habitat change maps are based on these yearly classification maps, although the accuracies are an issue. However, the change maps show reasonably reliable results according to the local dune manager.

7.1.1. RESEARCH QUESTIONS

The following sub-questions have been set up to answer the main research question. In this subsection, these sub-questions are answered.

1. How do the specifications of the different data products in terms of spatial, spectral and temporal resolution relate?

In this thesis, three types of remote sensing data are used: aerial images, Landsat-7 and Sentinel-2 satellite images. The aerial images have the highest spatial resolution (0.25m), but the lowest spectral (4 bands: Blue, Green, Red, NIR) and temporal resolution. Landsat-7 has the lowest temporal resolution with a revisit time of 16 days, the lowest spatial resolution (30m) and a medium spectral resolution (6 bands: Blue, Green, Red, NIR, SWIR-1 and SWIR-2). Sentinel-2 has the highest temporal resolution with a revisit time of 5-days, the highest spectral resolution (10 bands: Blue, Green, Red, NIR, Red Edge 1, Red Edge 2, Red Edge 3, NIR, Red Edge 4, SWIR-1 and SWIR-2) and an intermediate spatial resolution (10m or 20m, depending on the band).

These data products are mentioned and described in Sections 2.2 and 3.4. In general, lower spatial resolution remote sensing data products have higher spectral and temporal resolution.

2. Which disturbance factors are introduced when simulating the optical satellite images from aerial images by downsampling?

For the spatial and spectral resolution analysis, the spatial resolution of the aerial images are deteriorated. Aerial images result from different smaller images taken from an airplane, which are later combined in a land covering mosaic image. Landsat-7 uses a whiskbroom design. This modus makes zig-zag movements and the Scan Line Corrector corrects for the forward movement of the satellite (which failed for Landsat-7). Averaging of the aerial images takes place after image acquisition. With the satellite images, the averaging goes parallel with the scanning of the Earth's surface and is the final product delivered. This causes differences in the direction of the averaging. Furthermore, the resampling of the images has effect on the training data as these were originally based on the 25cm resolution aerial images. Due to the larger pixel sizes of the resampled aerial images, less unique training samples are left for training (subsection 4.2.1). This influences the model's performance.

3. How much better does the classification model perform when more spectral features are added?

The spectral resolution analysis is done with both the pixel-based and region-based methods using a Landsat-7 image. The addition of two spectral bands (SWIR-1, SWIR-2) and the NDWI contributed to an increase of 5% accuracy in comparison with the low-spectral classification (features: Blue, Green, Red, NIR, NDVI, ARVI) with the traditional pixel-based approach. This resulted in a classification performance of 65%. The three most important features are Red, NDVI and Blue. The SWIR-1, SWIR-2 and NDWI contribute to a better performance, but they belong to one of the least important features (Figure 5.5). The region-based method improved the accuracy slightly more, but this method in combination with the low spatial resolution (30m) causes even more loss of detail in the results. Therefore, the classifications of satellite images are done with the pixel-based method and not region-based method.

4. How is the classification performance influenced when data is combined from three different Natura 2000 areas instead of treating each area separately?

During the spatial variability analysis (Table 5.3), three scenarios of model training are tested: the model is trained with data from all three Natura 2000 areas, trained with data from one area and tested on the same area and finally trained with data from one area and tested on one of the other two Natura 2000 areas. The outcomes based on the last scenario show that a model trained on one area results in a poor classification performance in the other area. Although the best performance is achieved when training and testing on the same area, it is advantageous to combine data from all areas to increase the learning variance of the model. The model needs this as the habitats between the different areas vary in terms of spectral reflectance. The better performance of Solleveld & Kapittelduinen in all three scenarios reveals that the training dataset of this area is spatially better distributed along its own boundaries. Moreover, training on Voornes Duin and testing on Goeree & Kwade Hoek gives a better performance than testing on Solleveld & Kapittelduinen. This shows that the habitats in Voornes Duin resemble those of Goeree & Kwade Hoek more than those of Solleveld & Kapittelduinen. From the fieldwork campaign (Section 3.2), it was concluded that Voornes Duin has an overall higher soil moisture compared to Solleveld & Kapittelduinen. Figure 5.10 confirms that the soils of both Voornes Duin and Goeree & Kwade Hoek have higher moisture content than Solleveld & Kapittelduinen.

5. How is the classification performance influenced when statistical relationships between adjacent pixels and the center pixel are incorporated?

In this thesis, classifications are done both with pixel-based and region-based methods. The region-based method uses statistical and spectral information from a 3×3 pixel window. The statistical measures are the standard deviation, minimum, maximum and median of the pixels in this window. It is found that the statistical information from adjacent pixels are more important in the contribution to a better performance of the region-based method than the spectral features of the same adjacent pixels (Figure 5.1). The statistical features give the necessary contextual information about the area, such as surface texture etc. The classification accuracy improved from 78% to 87% on the 4m resolution aerial images. Although, the region-based method improved the accuracy of the Landsat-7 classification as well (Table 5.1), it is not recommended to apply the region-based method on low spatial resolution images (Landsat-7). Details are already limited in low spatial

resolution images. The region-based approach will cause even more loss of detail (see research question 3 and Figure 5.9).

6. What is the quality of the training and validation data from 2012 (Mouissie et al., 2014) and 2019?

The data labels used for training the Random Forest model are from the classification of Mouissie et al. (2014) that was conducted as a baseline for the bigger project of which the research done in this thesis is part of. This classification was more focused on mapping grey dunes and humid dune slacks, whereas this thesis also focuses on other habitat types. Therefore, the distribution of the polygons over the areas was limited. This created even more imbalance when sampling from poorer resolutions. Because of this, the model did not see enough variety in the dune habitats, which made it difficult to classify unseen data. In this research, more polygons are drawn with the supervision of experts of these areas. That improved the quality of the training and validation data. Fieldwork points are, among other things, used to validate the classifications in 2019. The measurement locations are close to trails and transitions between habitat types. This is sufficient for doing fieldwork measurements, but the reflectance of the trail will contaminate the spectral reflectance of the pixel measured by the satellite. This was often the case with shrubs. There is moderate correlation between the satellite images and the fieldwork measurements (Figure 5.17). Moreover, the external validation from the province of Zuid-Holland for Solleveld & Kapittelduinen is not yet validated in the field. Still, it is used to validate the final 2019 classifications. The province has assumed that not much has changed between 2008 and 2019. Furthermore, the final maps are checked with the local dune manager of the areas. It is confirmed that the maps are largely correct.

7.2. RECOMMENDATIONS

During the discussion and conclusion, several possible improvements are mentioned that were beyond the scope or duration of this project. In this section, recommendations on the use of the data or methods proposed in this thesis and interesting topics for further research in remote sensing of coastal dune habitats are discussed.

7.2.1. IMPROVEMENTS

In this subsection some improvements for subsequent studies which (partly) overlap this thesis are briefly discussed.

LANDSAT

The habitat change maps in Figure 5.13 are based on Landsat-7 images from 2011 till 2013, 2018 and 2019. These images have missing data lines caused by the SLC failure. Landsat-8 is operational from 2013 onwards. As images from 2013 are already used with the assumption that there are no drastic changes between 2012 and 2013, Landsat-8 products could be useful. In accordance with the results in Table 5.1 showing that 10m is the second best performing resolution, it could be useful to do pansharpening on lower resolution bands with the higher resolution panchromatic band (15m for Landsat-7).

DATA ACQUISITION AND PROCESSING

The data acquisition is done using Google Earth Engine and the processing is mostly done in Python. Other small processes, random point generation and clipping of the aerial images, are done in SQL and GDAL. To have a good oversight on all processes, it is recommended to do processing on one single platform as far as it is possible. There are APIs available for Google Earth Engine, SQL and GDAL as packages in Python. In the discussion, it is stated that sunglint and shadow zones are two phenomena causing the misclassifications. This is a well-known problem in remote sensing and there are existing technique for removing these bad pixels. The data quality can be improved by removing these pixels during the pre-processing phase.

METHODS

In this thesis, the window to derive statistical features from is fixed to 3×3 pixels. From the results it is clear that this works best for 4m resolution. There is undoubtedly a relationship between the window size and the spatial resolution. It is interesting to see how the performance changes when using different window sizes on different resolutions. Moreover, dune surface height or vegetation height are often used datasets in other coastal dune classification researches, see Section 6.3. It is also interesting to see how the addition of height features to the region-based classification could improve the classification performances of the different habitat types. Furthermore, the statistical features derived using the region-based model are the

standard deviation, mean, minimum and maximum. It could be useful to add other statistical features, such as the entropy. On top of that, the resolutions between 0.5m and 4m are not assessed for both pixel-based and region-based methods. It is interesting to see whether a resolution between 0.5m and 4m performs even better using the region-based method. The step size would preferably be set to 0.5m. In subsection 6.2.1, it is discussed that limited time is spend on tuning the hyperparameters of the XGBoost method whereas, in general, it requires more tuning than Random Forest. It is useful to see if a more optimally tuned XGBoost model would improve the performance in runtime and accuracy.

7.2.2. FURTHER RESEARCH

In the background information, the importance of coastal dunes is discussed. The introduction explains how remote sensing could be used to monitor dune habitats and why it is important for coastal dune management. This thesis focused, among others, on the change of habitats over years (2012-2019) in the context of nitrogen increase. For dune monitoring, it is also interesting to see seasonal changes throughout the year. For example, this could help monitoring the humidity of humid dune slacks as the duration of dry or wet period could give information about the progress of succession. From the spatial and spectral resolution analysis, it turned out that 10m is the second best performing spatial resolution. Therefore, an analysis is done using Sentinel-2 data as these images have high spatial resolutions of 10m/20m, high spectral resolution and temporal coverage. It was outside the duration of this project to collect new ground-truth data for 2019. Sentinel-2 data in combination with more recent ground-truth data could help with the remote sensing of coastal dune habitats. Also, data fusion of Sentinel-1 and Sentinel-2 data could be very interesting. Sentinel-1 data gives insights about the terrain structure and surface roughness. Furthermore, an in depth research on the spatial variability of classifying coastal dune habitats is needed to see how well dune habitat classifications are applicable nationwide.

A

Appendix

A.1. DATA LABELS, MOUISSIE CLASSIFICATIONS (2012)

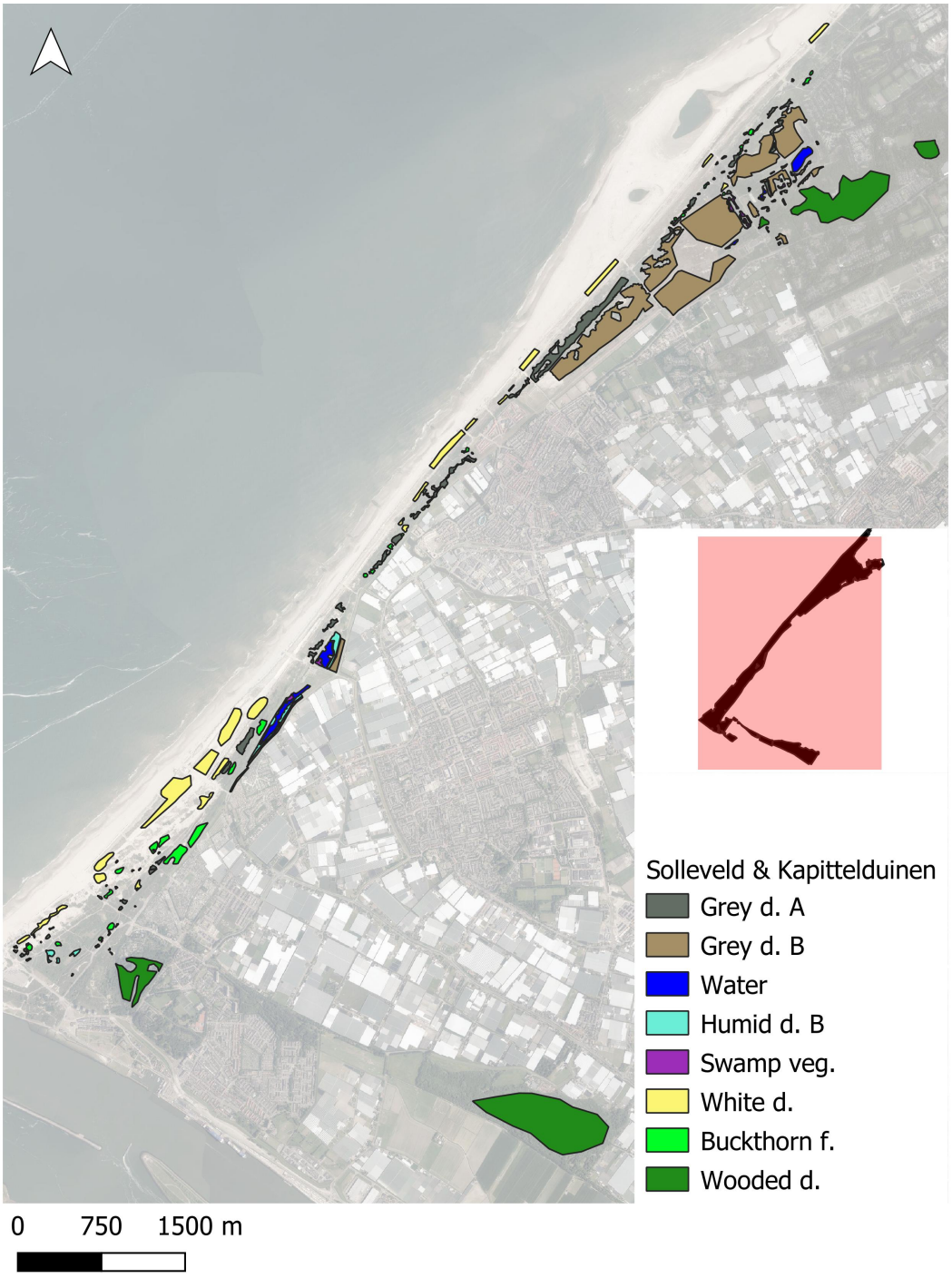


Figure A.1: Modified Mouissie classification of Solleveld & Kapittelduinen (Mouissie et al., 2014).

Figure A.2: Aerial image of Solleveld & Kapittelduinen in 2012, Figure A.1.

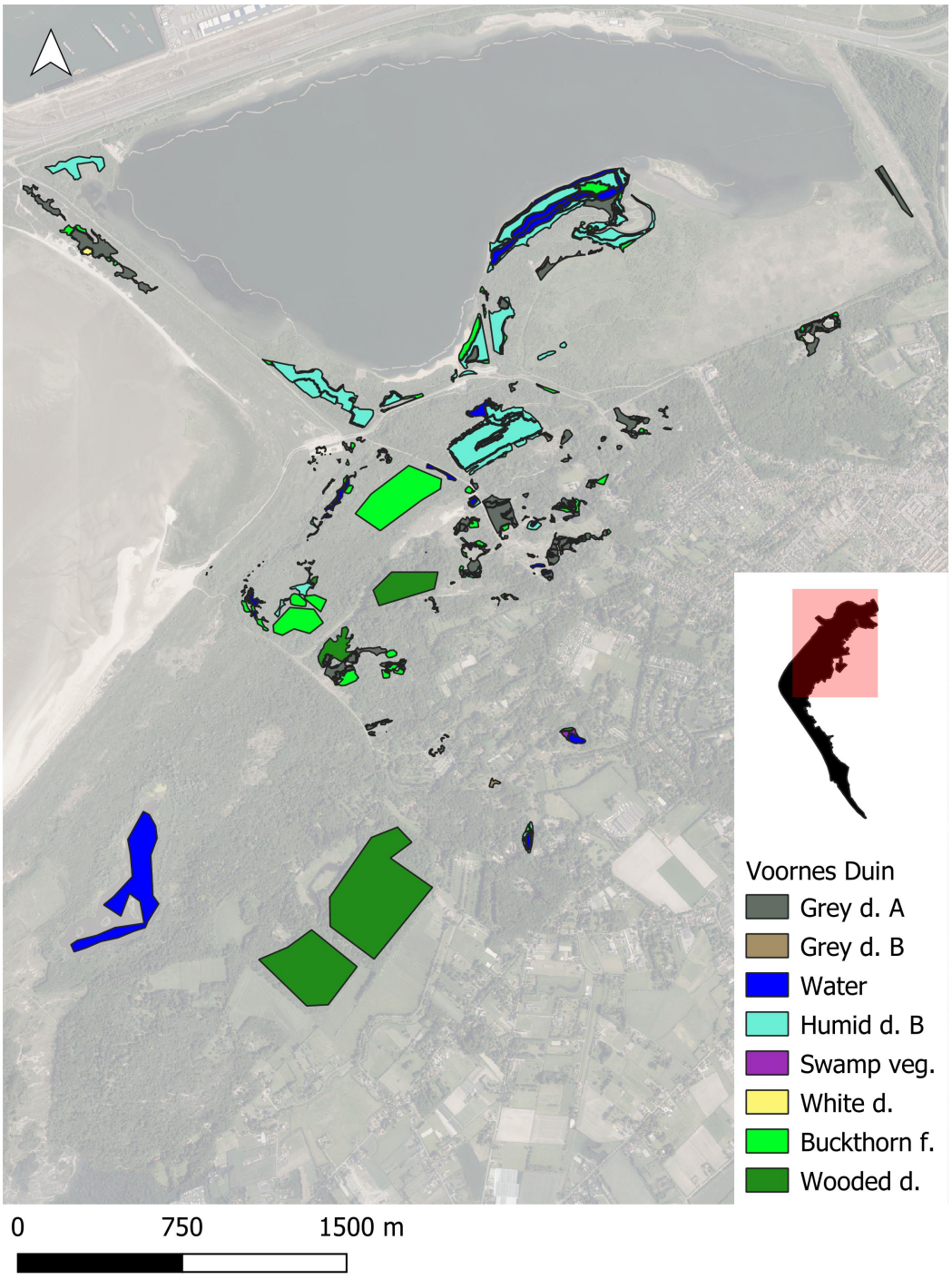


Figure A.3: Modified Mouissie classification of Voornes Duin (Mouissie et al., 2014).

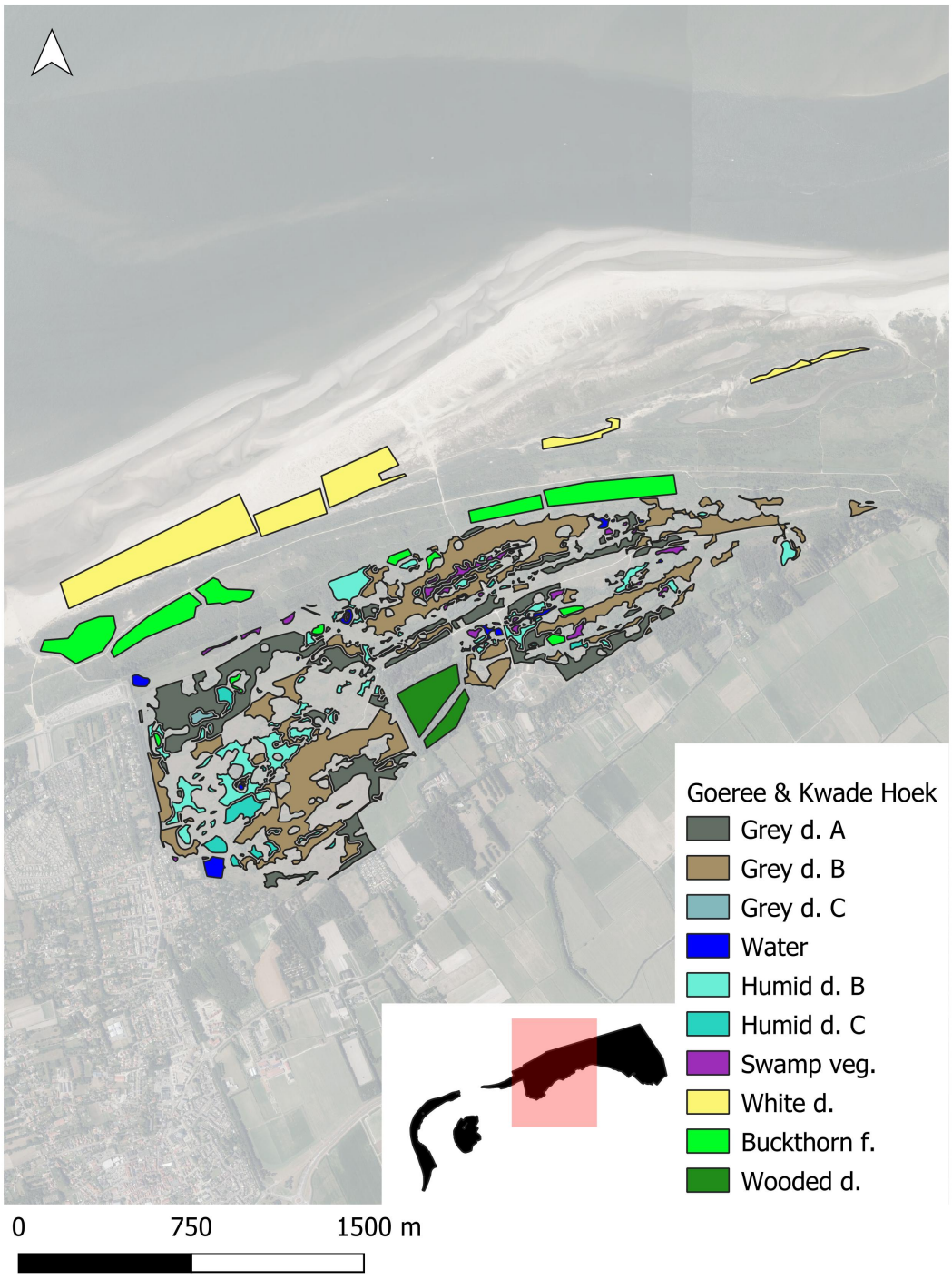


Figure A.4: Modified Mouissie classification of Goeree & Kwade Hoek (Mouissie et al., 2014).

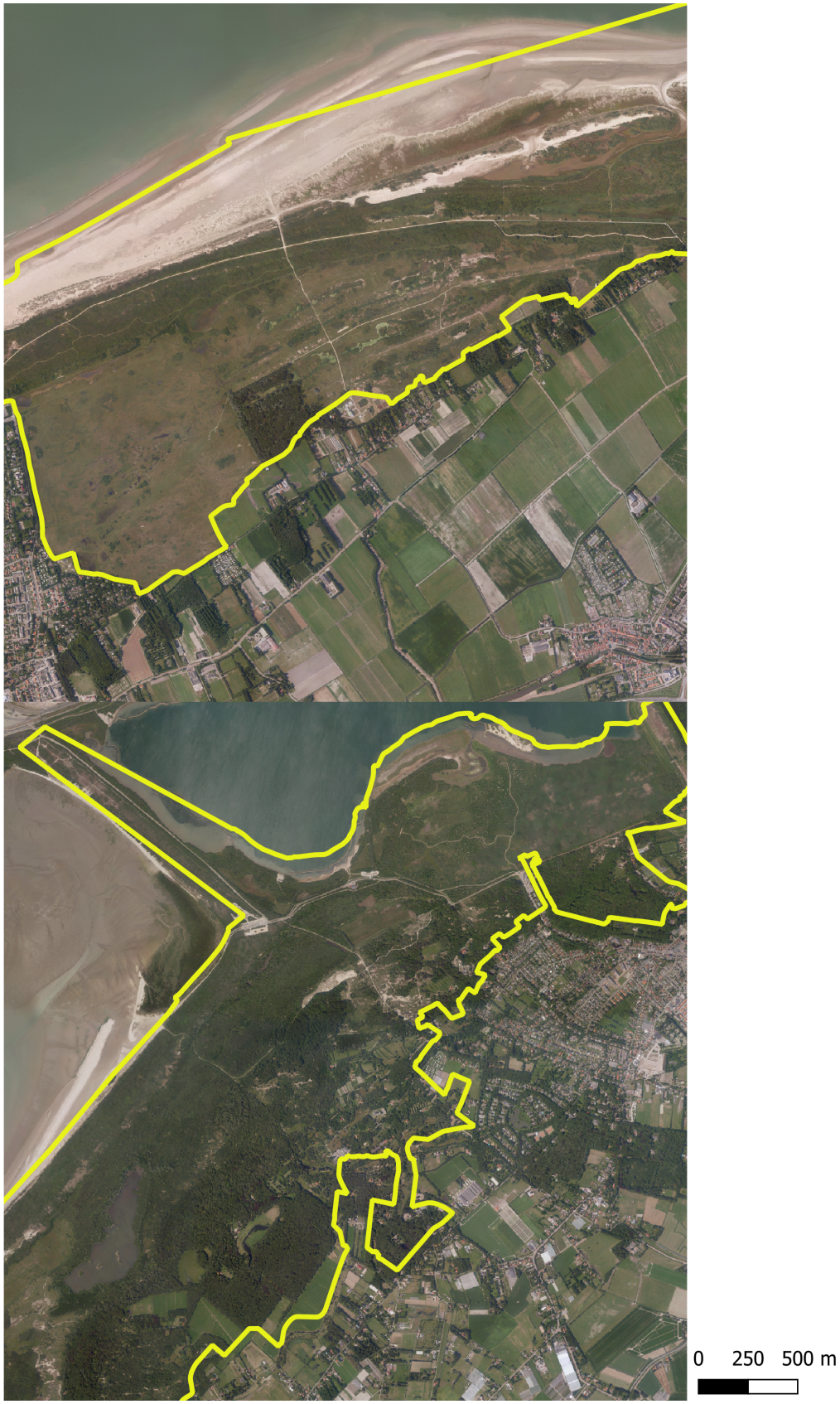


Figure A.5: Aerial image of Goeree & Kwade Hoek and Voornes Duin in 2012, Figures A.3 and A.4.

A.2. METHODOLOGY

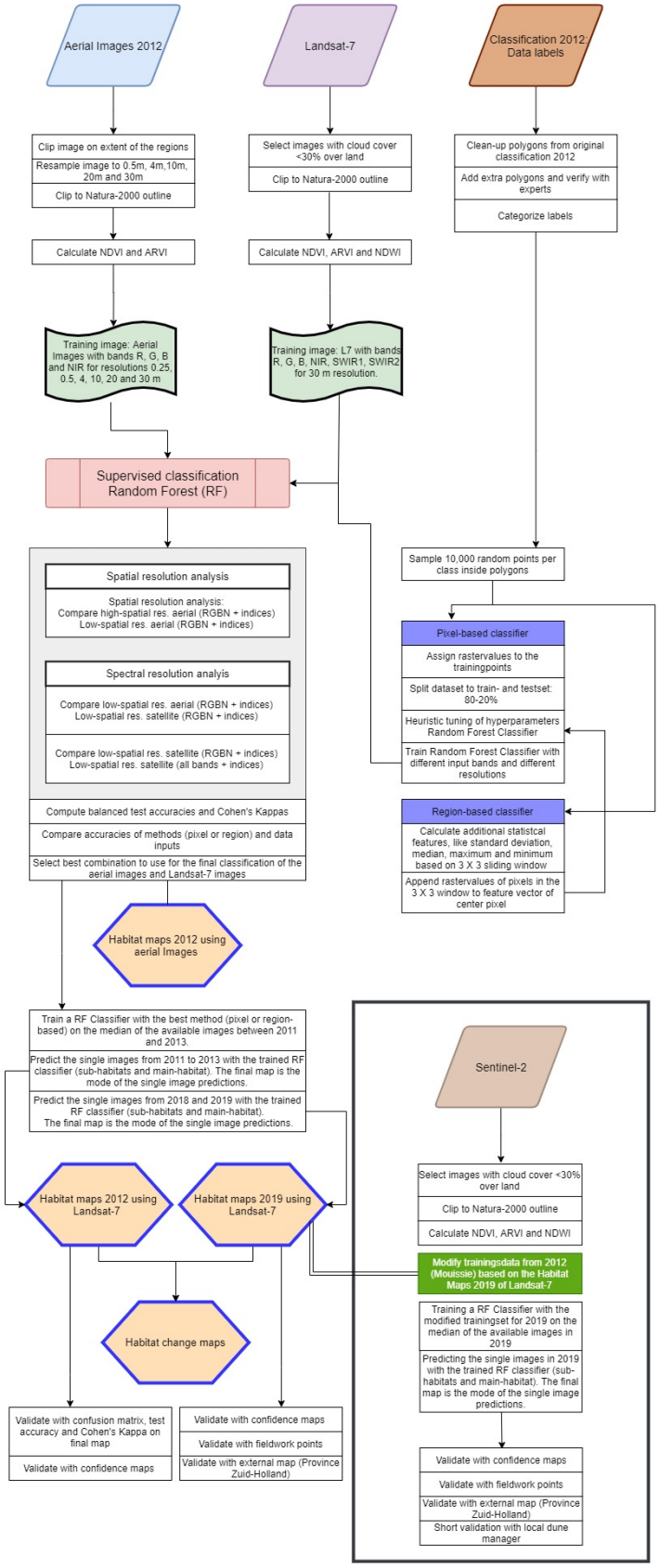


Figure A.6: Extensive methodology.

A.3. BAND DESIGNATIONS

	S2A		S2B		
Band Number	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
1	442.7	21	442.3	21	60
2	492.4	66	492.1	66	10
3	559.8	36	559.0	36	10
4	664.6	31	665.0	31	10
5	704.1	15	703.8	16	20
6	740.5	15	739.1	15	20
7	782.8	20	779.7	20	20
8	832.8	106	833.0	106	10
8a	864.7	21	864.0	22	20
9	945.1	20	943.2	21	60
10	1373.5	31	1376.9	30	60
11	1613.7	91	1610.4	94	20
12	2202.4	175	2185.7	185	20

Figure A.7: Band designations for Sentinel-2 (ESA, 2019a).

Landsat 7 Enhanced Thematic Mapper Plus (ETM+)

Landsat 7	Wavelength (micrometers)	Resolution (meters)
Band 1	0.45-0.52	30
Band 2	0.52-0.60	30
Band 3	0.63-0.69	30
Band 4	0.77-0.90	30
Band 5	1.55-1.75	30
Band 6	10.40-12.50	60 (30)
Band 7	2.09-2.35	30
Band 8	.52-.90	15

Figure A.8: Band designations for Landsat-7 (USGS, 2019).

A.4. CLASSIFICATION MAPS

A.4.1. AERIAL IMAGES

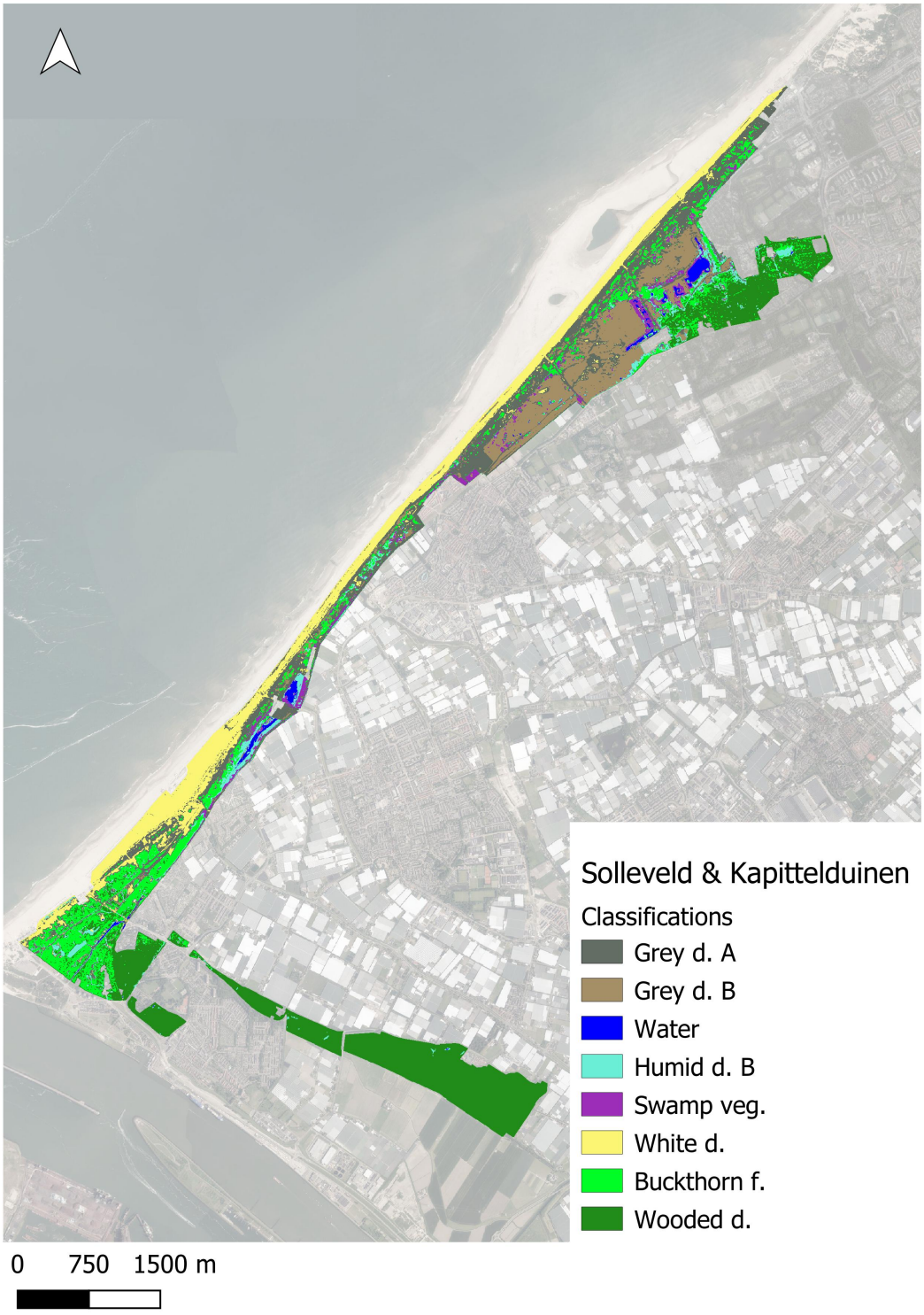


Figure A.9: Region-based classification on 4m resolution aerial image in Solleveld & Kapittelduinen.

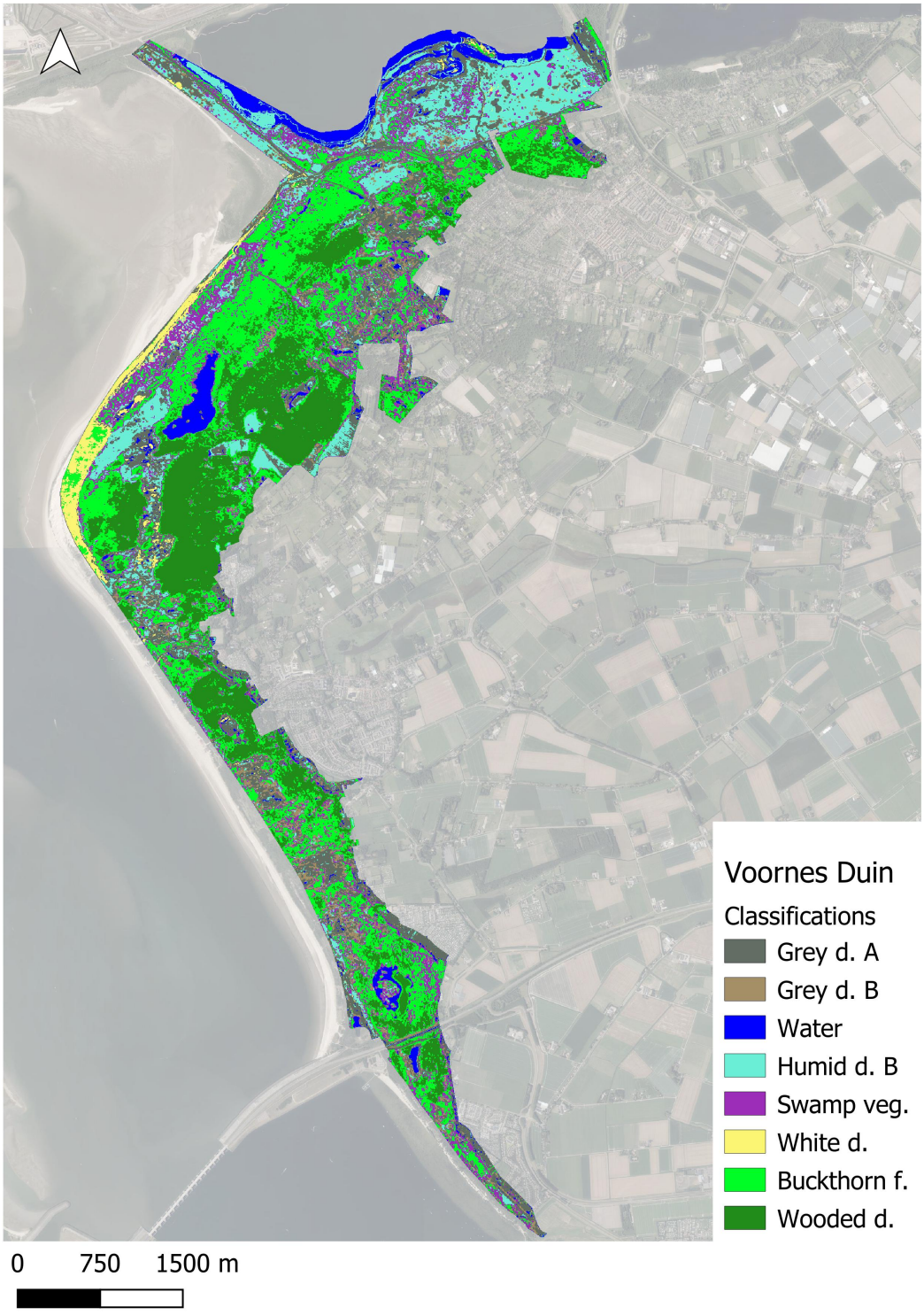


Figure A.10: Region-based classification on 4m resolution aerial image in Voornes Duin.

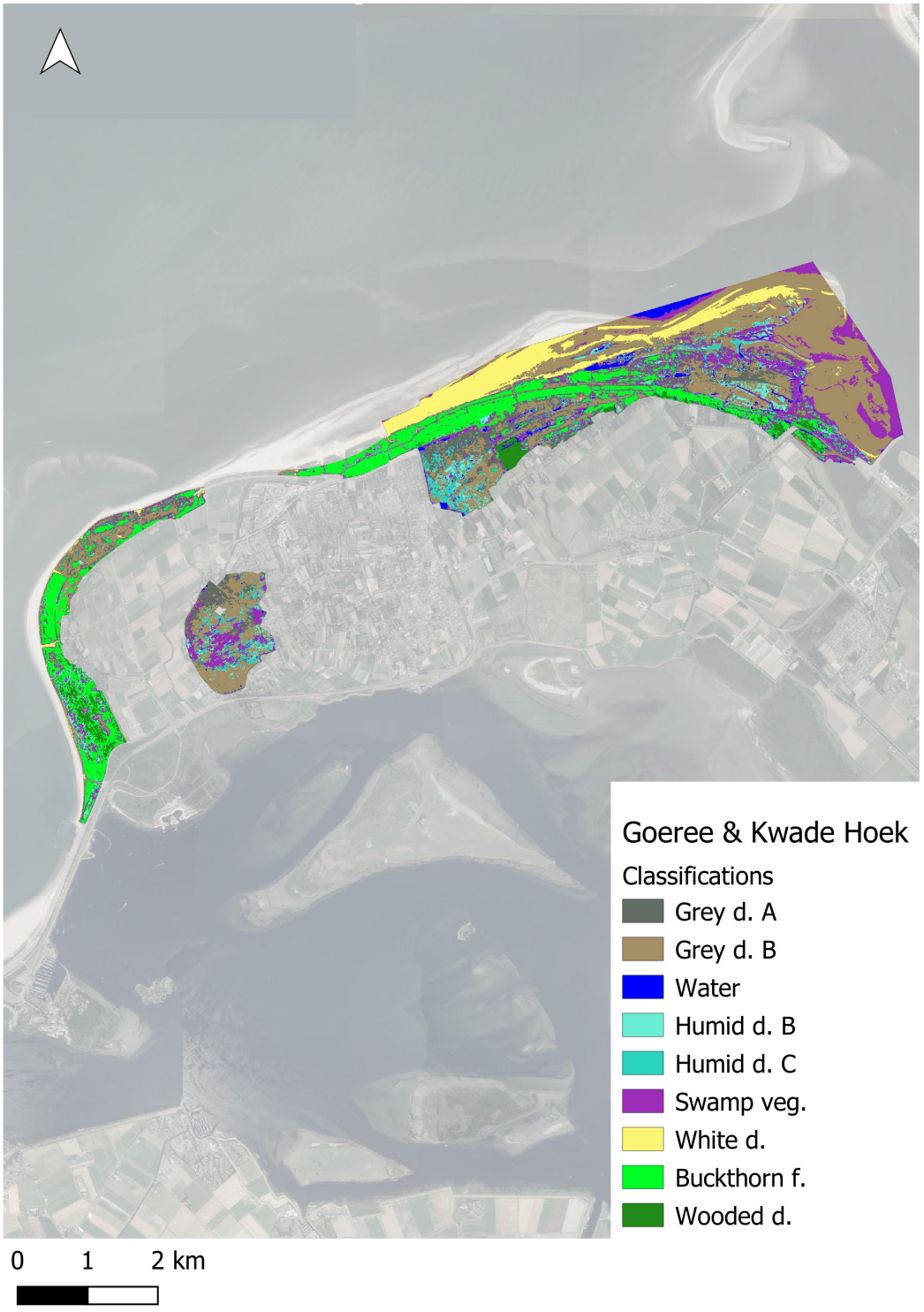


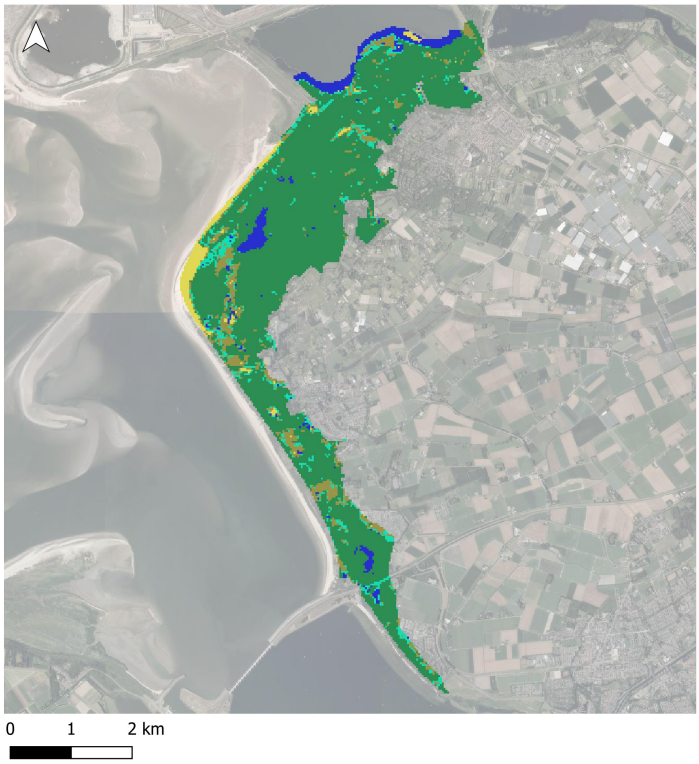
Figure A.11: Region-based classification on 4m resolution aerial image in Goeree & Kwade Hoek.

A.4.2. LANDSAT-7

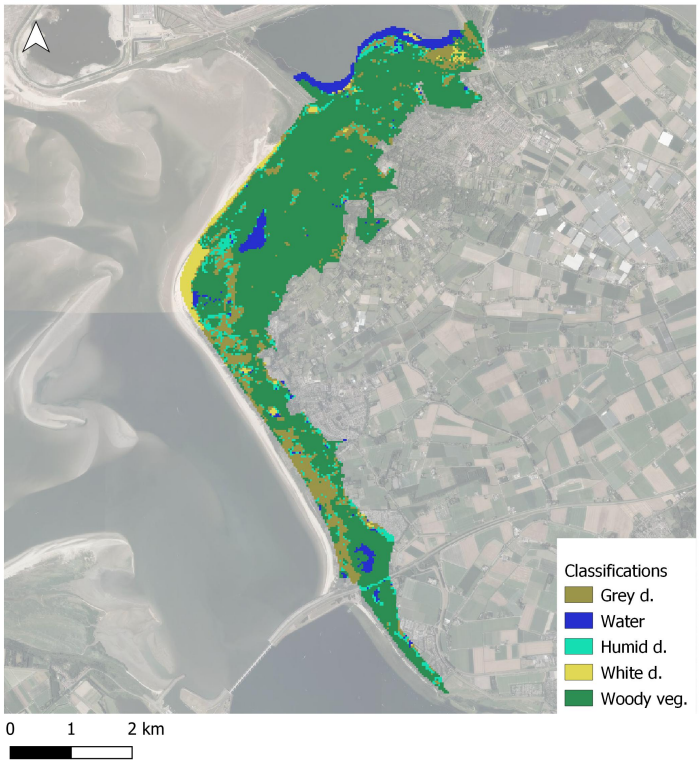
(a) Main-habitat classification map 2012.

(b) Main-habitat classification map 2019.

Figure A.12: Main-habitat classifications in 2012 and 2019 of Solleveld & Kapittelduinen using Landsat-7 images.

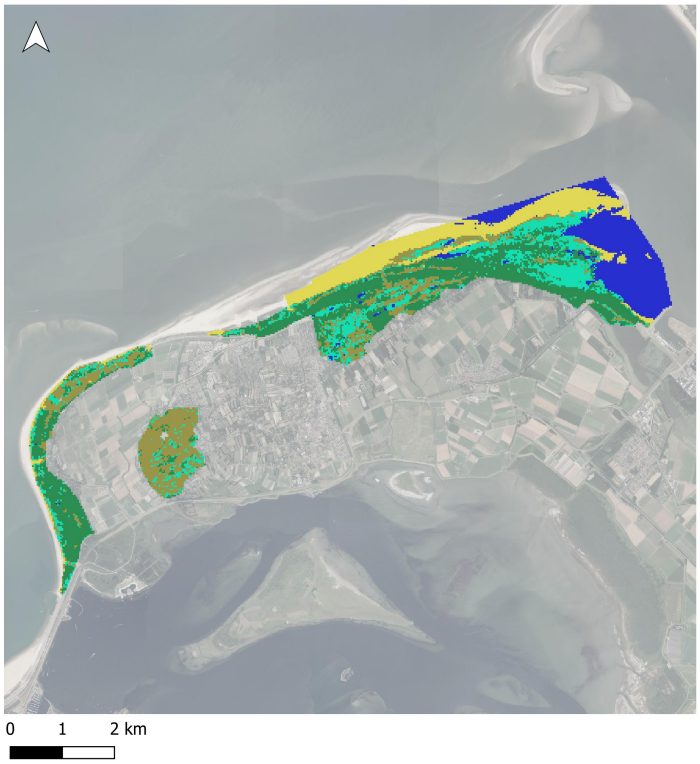


(a) Main-habitat classification map 2012.

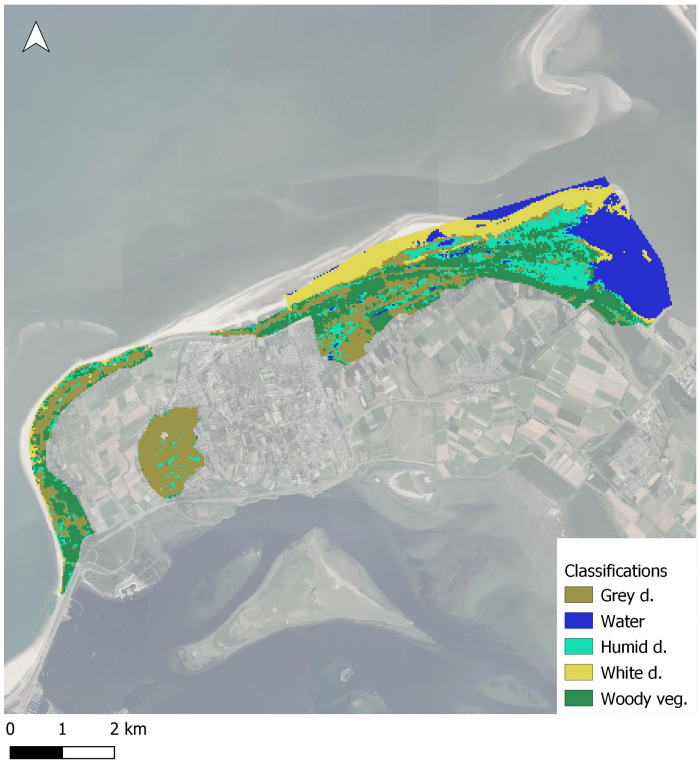


(b) Main-habitat classification map 2019.

Figure A.13: Main-habitat classifications in 2012 and 2019 of Voornes Duin using Landsat-7 images.

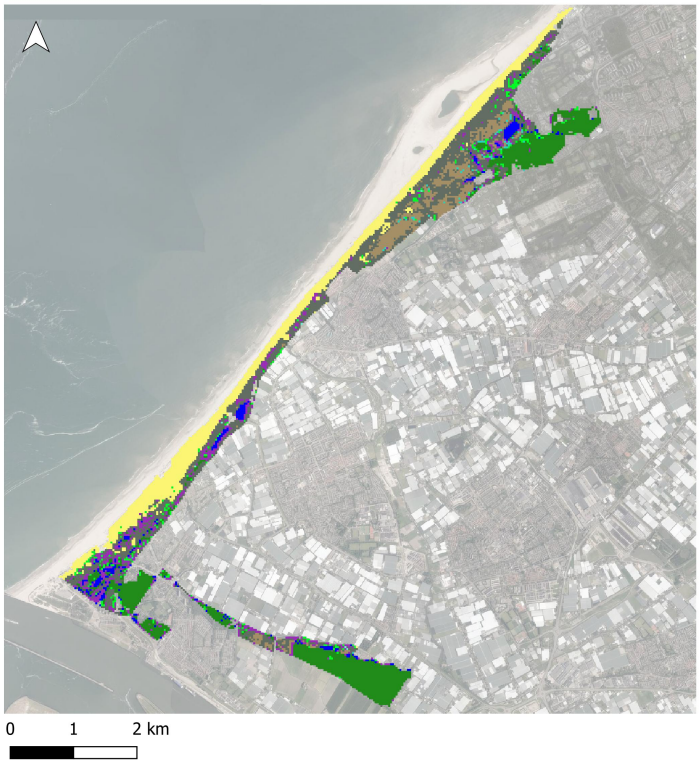


(a) Main-habitat classification map 2012.

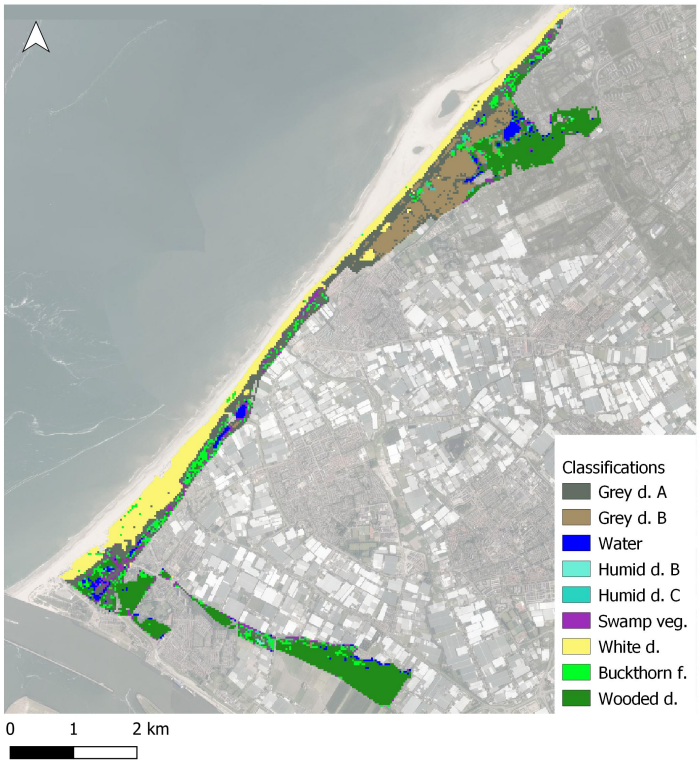


(b) Main-habitat classification map 2019.

Figure A.14: Main-habitat classifications in 2012 and 2019 of Goeree & Kwade Hoek using Landsat-7 images.

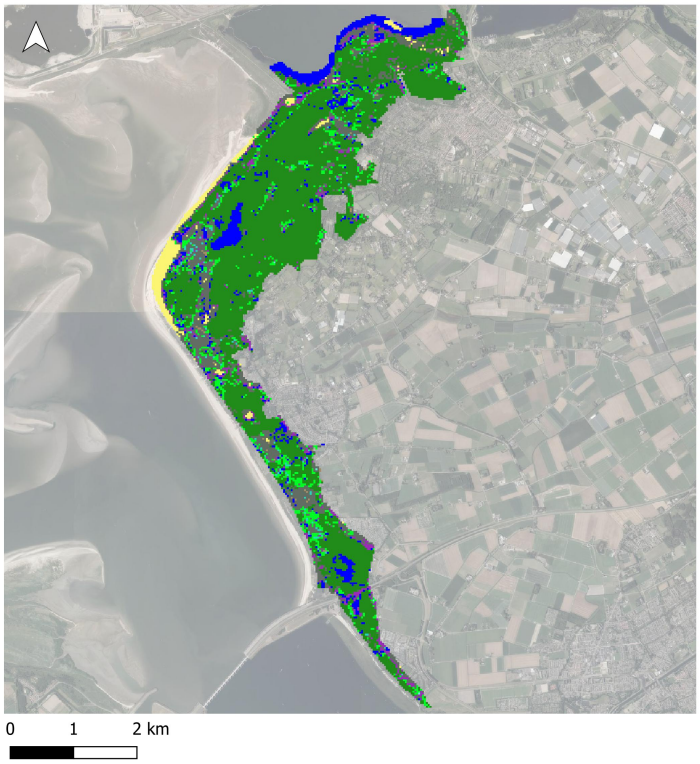


(a) Sub-habitat classification map 2012.

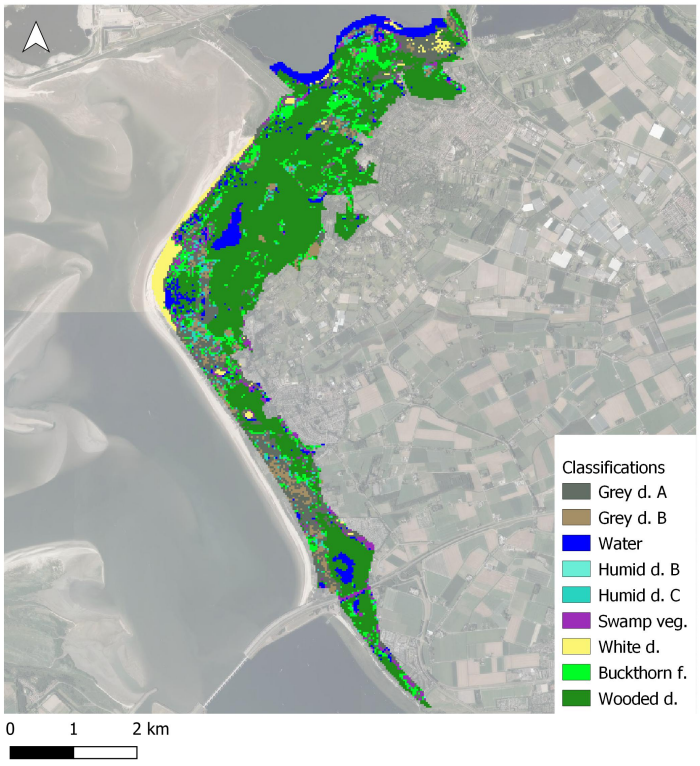


(b) Sub-habitat classification map 2019.

Figure A.15: Sub-habitat classifications in 2012 and 2019 of Solleveld & Kapittelduinen using Landsat-7 images.

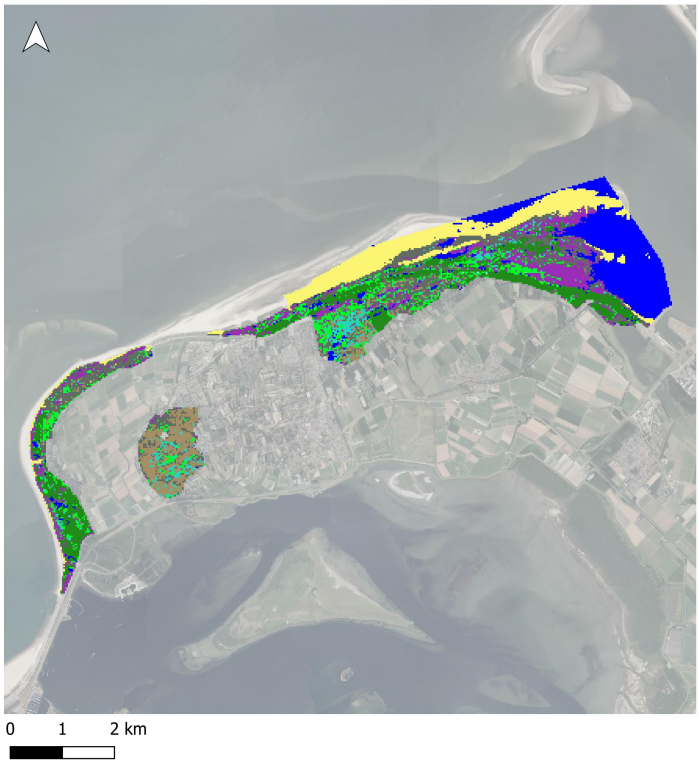


(a) Sub-habitat classification map 2012.

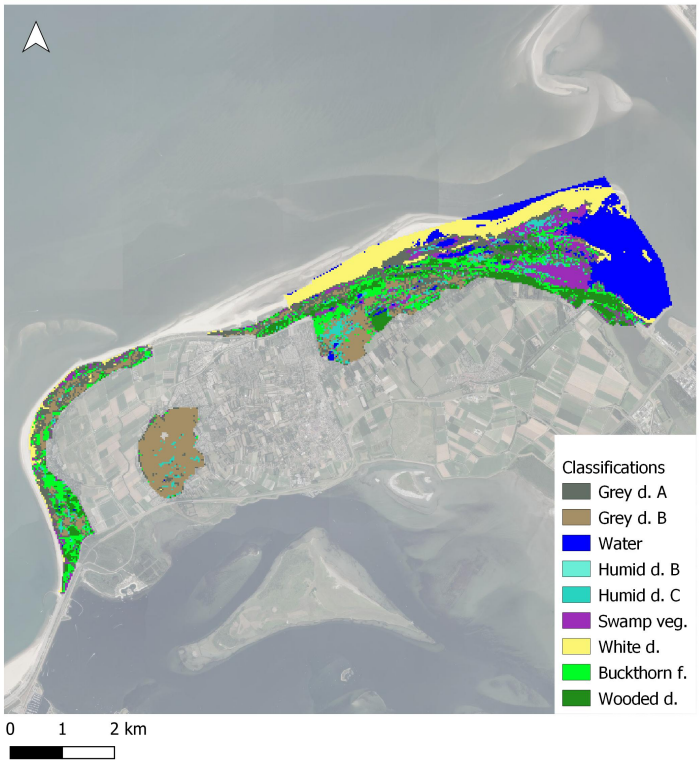


(b) Sub-habitat classification map 2019.

Figure A.16: Sub-habitat classifications in 2012 and 2019 of Voornes Duin using Landsat-7 images.



(a) Sub-habitat classification map 2012.



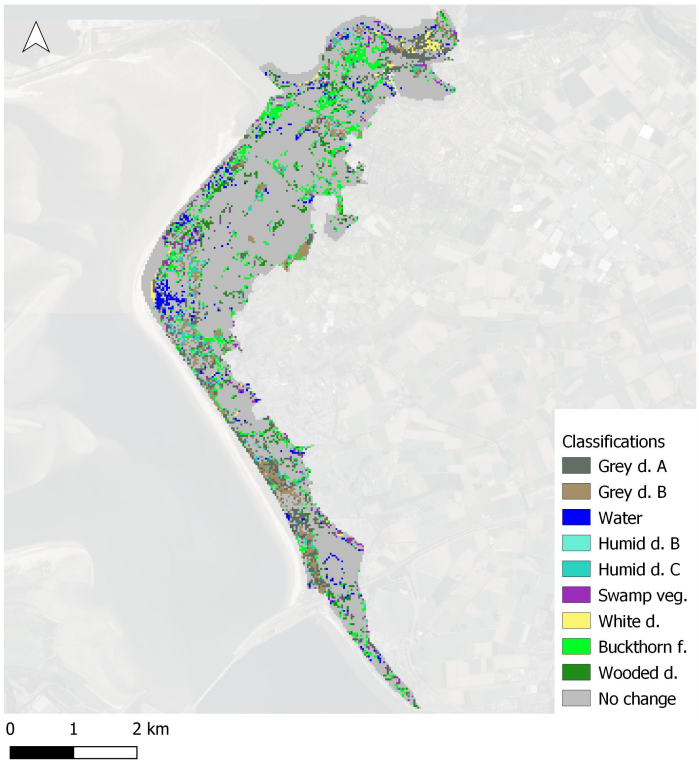
(b) Sub-habitat classification map 2019.

Figure A.17: Sub-habitat classifications in 2012 and 2019 of Goeree & Kwade Hoek using Landsat-7 images.

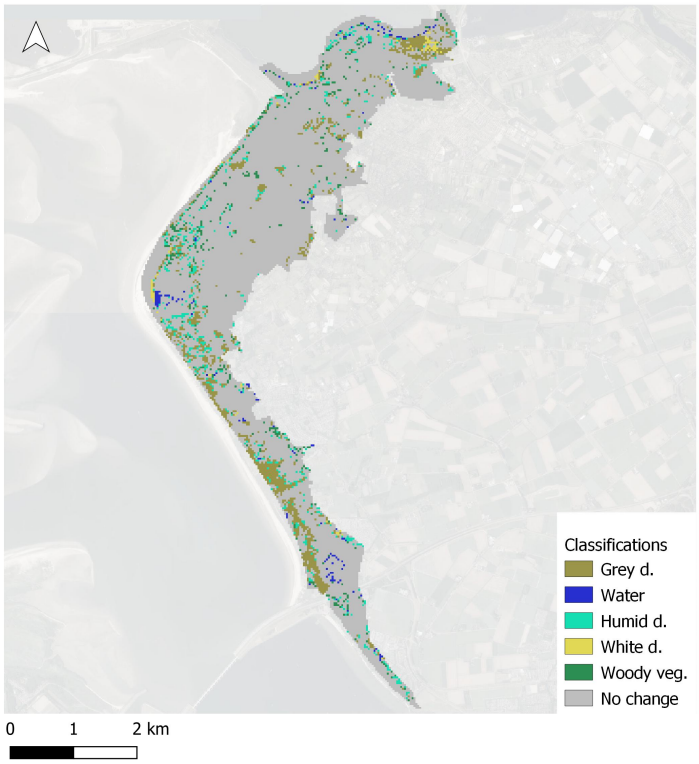
(a) Change in sub-habitat.

(b) Change in main-habitat.

Figure A.18: Habitat change maps (between 2012 and 2019) of Solleveld & Kapittelduinen using Landsat-7 images. The colored pixels show the class in 2019.

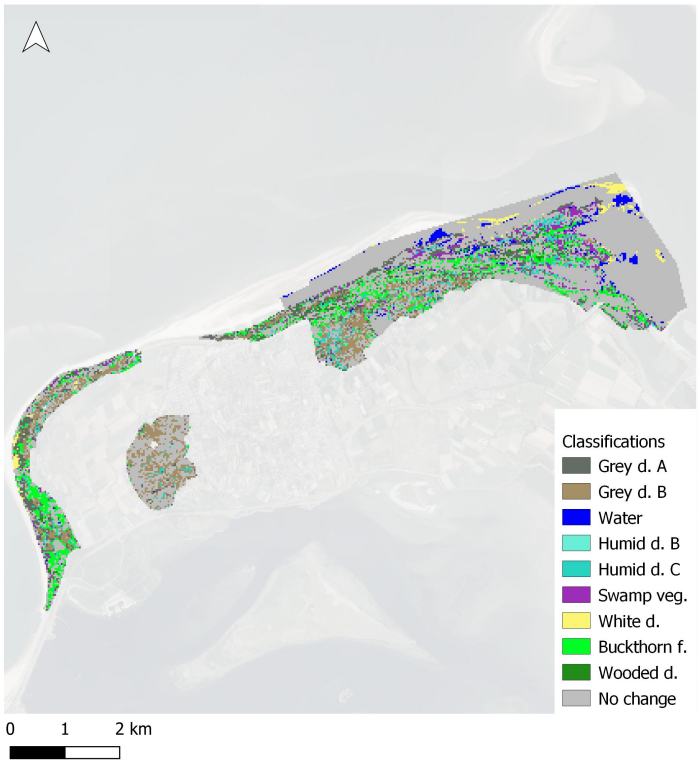


(a) Change in sub-habitat.

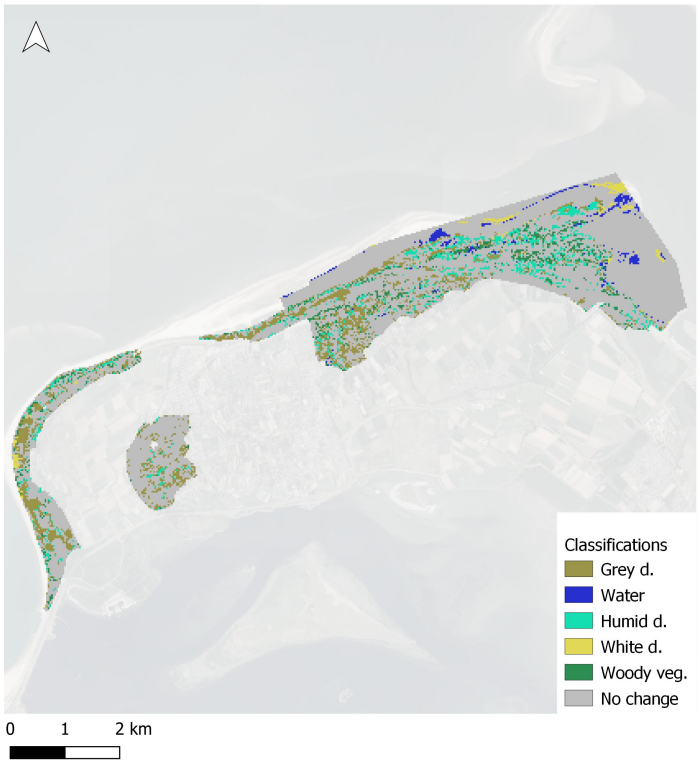


(b) Change in main-habitat.

Figure A.19: Habitat change maps (between 2012 and 2019) of Voornes Duin using Landsat-7 images. The colored pixels show the class in 2019.



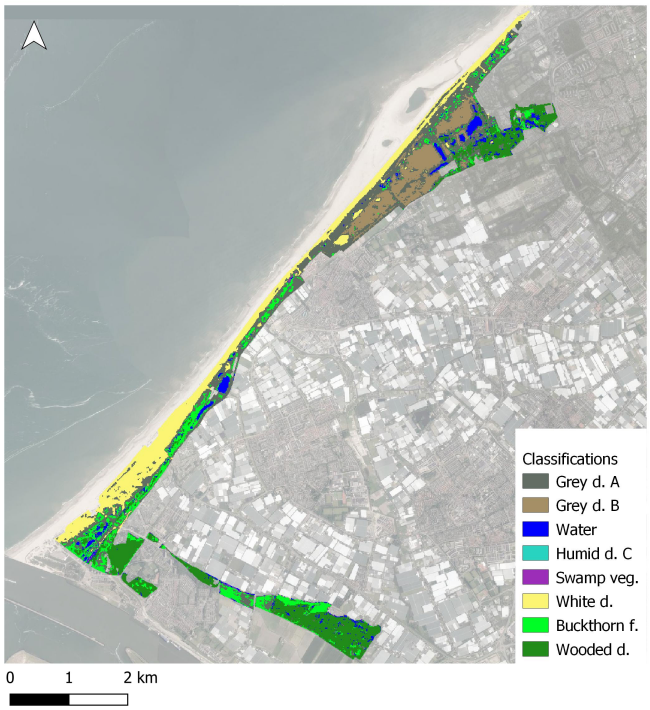
(a) Change in sub-habitat.



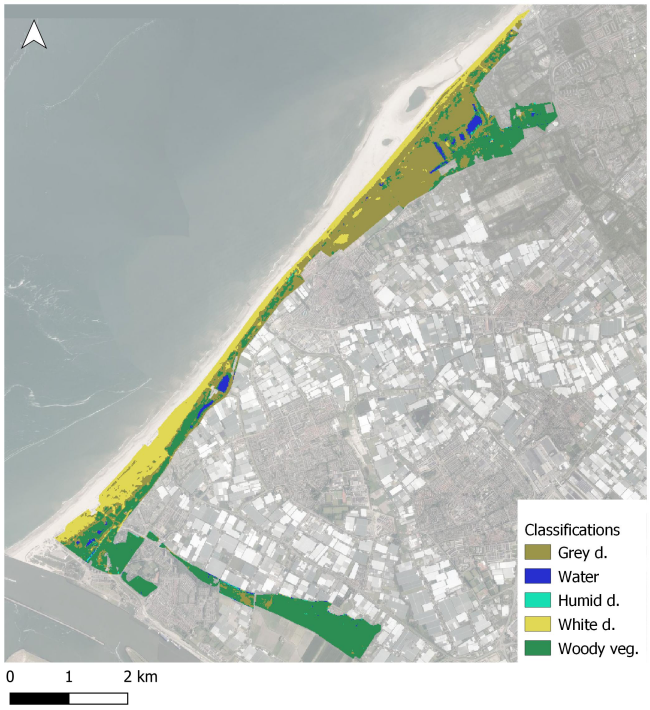
(b) Change in main-habitat.

Figure A.20: Habitat change maps (between 2012 and 2019) of Goeree & Kwade Hoek using Landsat-7 images. The colored pixels show the class in 2019.

A.4.3. SENTINEL-2

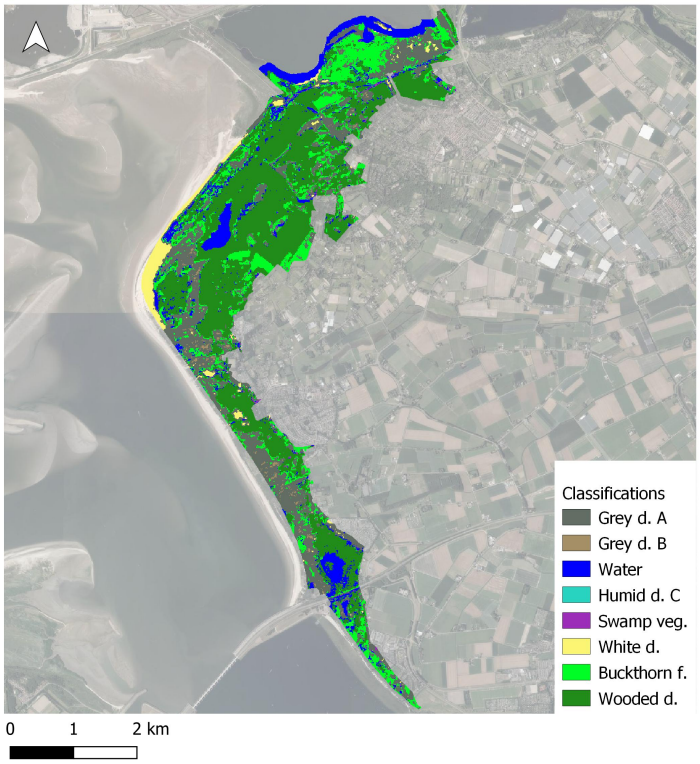


(a) Sub-habitat classification map 2019.

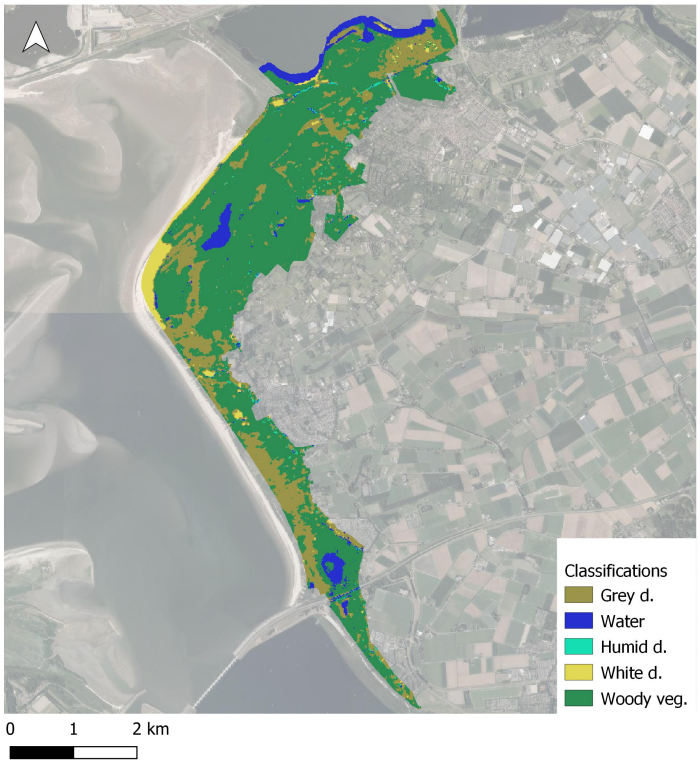


(b) Main-habitat classification map 2019.

Figure A.21: Sub-habitat and main-habitat classifications in 2019 of Solleveld & Kapittelduinen using Sentinel-2 images.

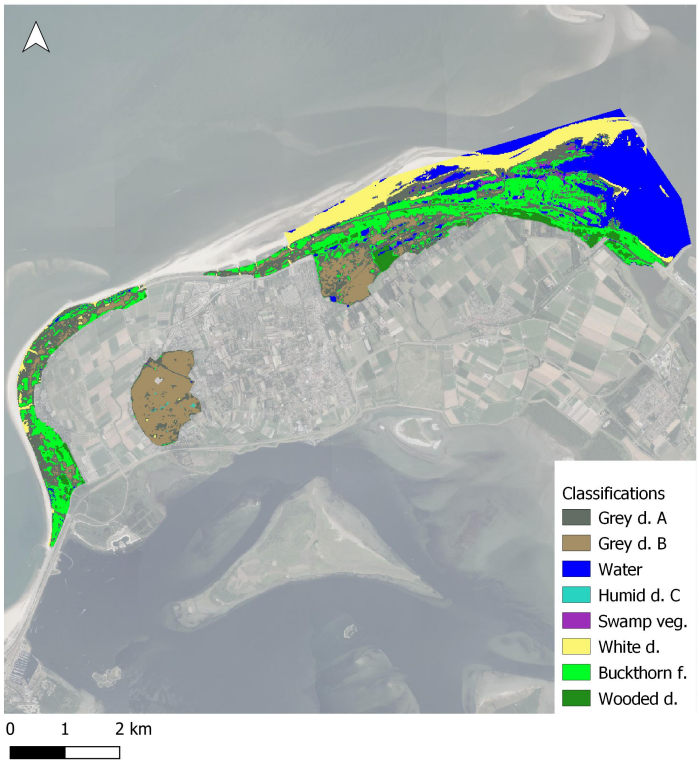


(a) Sub-habitat classification map 2019.

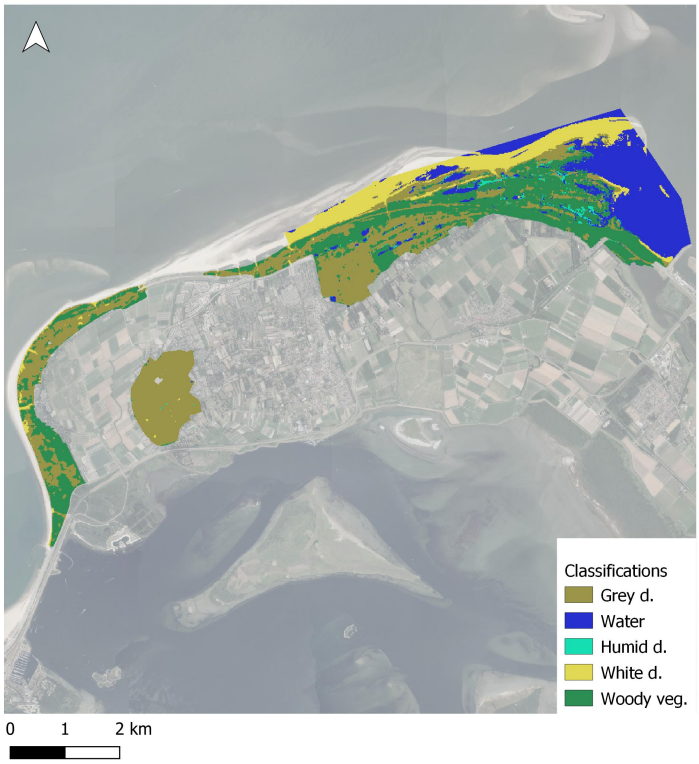


(b) Main-habitat classification map 2019.

Figure A.22: Sub-habitat and main-habitat classifications in 2019 of Voornes Duin using Sentinel-2 images.



(a) Sub-habitat classification map 2019.



(b) Main-habitat classification map 2019.

Figure A.23: Sub-habitat and main-habitat classifications in 2019 of Goeree & Kwade Hoek using Sentinel-2 images.

BIBLIOGRAPHY

Baghdadi, N. and Zribi, M. (2016). Introduction. In Baghdadi, N. and Zribi, M., editors, *Optical Remote Sensing of Land Surface*, pages xxxix – xlii. Elsevier.

Barchi, P., de Carvalho, R., Rosa, R., Sautter, R., Soares-Santos, M., Marques, B., Clua, E., Gonçalves, T., de Sá-Freitas, C., and Moura, T. (2020). Machine and Deep Learning applied to galaxy morphology - A comparative study. *Astronomy and Computing*, 30:100334.

Bij12 (2019). N08.02 Open duin. https://www.bij12.nl/onderwerpen/natuur-en-landschap/index-natuur-en-landschap/natuurtypen/n08-open-duinen/n08-02-open-duin/.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Cabral, J., Sánchez, B., Ramos, F., Gurovich, S., Granitto, P., and Vanderplas, J. (2018). From FATS to feets: Further improvements to an astronomical feature extraction tool based on machine learning. *Astronomy and Computing*, 25:213 – 220.

Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *CoRR*, abs/1603.02754.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30(3):271 – 278.

Daneshgar, S. (2015). *Remote sensing observations for monitoring coastal zones, Volturno River mouth case study*. PhD thesis.

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., and Li, X. (2016). Water bodies' mapping from sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sensing*, 8(4).

ESA (2019a). Band designations sentinel-2. https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath.

ESA (2019b). Copernicus overview. http://www.esa.int/Applications/Observing_the_Earth/Copernicus/Overview4.

ESA (2019c). Revisit and coverage. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/revisit-coverage.

ESA (2019d). Sentinel 2 - MSI. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/overview.

European Commission (2019). Natura 2000. https://ec.europa.eu/environment/nature/natura2000/index_en.htm.

Fürnkranz, J., Gamberger, D., and Lavrac, N. (2012). *Machine Learning and Data Mining*, pages 1–17.

Gavish, Y., O Connell, J., Marsh, C., Tarantino, C., Blonda, P., Tomaselli, V., and Kunin, W. (2017). Comparing the performance of flat and hierarchical Habitat/Land-Cover classification models in a NATURA 2000 site. *ISPRS Journal of Photogrammetry and Remote Sensing*, 136.

GDAL (2019a). Gdal documentation. https://gdal.org/.

GDAL (2019b). Gdal: gdalwarp. https://gdal.org/programs/gdalwarp.html.

GEOIMAGE (2019). SWIR Resolution+ Unlocks Potential. https://www.geoimage.com.au/SWIR%20Series/resolution.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn & Tensorflow*. O'Reilly, first edition.

Hsu, P. L. and Robbins, H. (1947). Complete convergence and the law of large numbers. *Proceedings of the National Academy of Sciences of the United States of America*, 33(2):25–31.

Jackson, A. (2014). Features of Deposition. https://geographyas.info/coasts/features-of-deposition/. Accessed on the 31st of August 2019.

Jobin, B., Labrecque, S., Grenier, M., and Falardeau, G. (2008). Object-based classification as an alternative approach to the traditional pixel-based classification to identify potential habitat of the Grasshopper Sparrow. *Environmental Management*, 41(1):20–31.

Kaufman, Y. J. and Tanre, D. (1992). Atmospherically resistant vegetation index (arvi) for eos-modis. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):261–270.

Kempenaars, P., Deronde, B., Provoost, S., and Houthuys, R. (2009). Synergy of airborne digital camera and lidar data to map coastal dune vegetation. *Journal of Coastal Research*, 10053(2):73 – 82.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Liang, S., Li, X., and Wang, J. (2012). Chapter 1 - A Systematic View of Remote Sensing. In *Advanced Remote Sensing*, pages 1 – 31. Academic Press, Boston.

Liaw, A. and Wiener, M. (2001). Classification and Regression by Random Forest. *Forest*, 23.

Mclachlan, A. and Defeo, O. (2018). Coastal Dune Ecosystems and Dune-Beach Interactions. In *The Ecology of Sandy Shores*, chapter 13, pages 309–329. Academic Press, third edition.

Ministerie van Lanbouw, Natuur en Voedselkwaliteit (2008). Profielen habitattypen en soorten. https://www.natura2000.nl/profielen.

Ministerie van Landbouw, Natuur en Voedselkwaliteit (2019). Programma Aanpak Stikstof (PAS). https://www.synbiosys.alterra.nl/natura2000/gebiedendatabase.aspx?subj=pas&deel=0.

Mouissie, M., Bleeker, A., Hensen, A., Riksen, M. van Dobben, H., and Huiskes, R. (2014). Stikstof, vegetatie en duinbeheer: Eindrapport T0-monitoring (2011-2013) duinen van Goeree tot Solleveld, in het kader van MEP Duinen i.r.t. Maasvlakte 2.

Muchoney, D. M. and Strahler, A. H. (2002). Pixel- and site-based calibration and validation methods for evaluating supervised classification of remotely sensed data. *Remote Sensing of Environment*, 81(2):290 – 299.

NASA (2019a). A Landsat Timeline. https://landsat.gsfc.nasa.gov/a-landsat-timeline/.

NASA (2019b). What is Remote Sensing. https://earthdata.nasa.gov/learn/remote-sensing.

Natuurmonumenten (2020). Natuurgebied Voornes Duin. https://www.natuurmonumenten.nl/natuurgebieden/voornes-duin.

Nouri, H., Anderson, S., Sutton, P., Beecham, S., Nagler, P., Jarchow, C. J., and Roberts, D. A. (2017). Ndvi, scale invariance and the modifiable areal unit problem: An assessment of vegetation in the adelaide parklands. *Science of The Total Environment*, 584-585:11 – 18.

Ontario Parks (2019). Why is biodiversity important? https://www.ontarioparks.com/parksblog/why-is-biodiversity-important/. Accessed on the 10th of June 2019.

Ose, K., Corpetti, T., and Demagistri, L. (2016). 2 - Multispectral Satellite Image Processing. In Baghdadi, N. and Zribi, M., editors, *Optical Remote Sensing of Land Surface*, pages 57 – 124. Elsevier.

Overheid (2019). Wet Natuurbescherming. https://wetten.overheid.nl/BWBR0037552/2020-01-01.

Overstreet, B. and Legleiter, C. (2016). Removing sun glint from optical remote sensing images of shallow rivers: Removing sun glint from river imagery. *Earth Surface Processes and Landforms*.

Prisco, I., Carboni, M., and Acosta, A. T. R. (2013). The fate of threatened coastal dune habitats in italy under climate change scenarios. *PLOS ONE*, 8(7).

Probst, P., Wright, M. N., and Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for Random Forest. *WIREs Data Mining and Knowledge Discovery*, 9(3):e1301.

Radio2Space (2013). Components of electromagnetic spectrum. https://www.radio2space.com/components-of-electromagnetic-spectrum/.

Rashno, A., Nazari, B., Sadri, S., and Saraee, M. (2017). Effective pixel classification of mars images based on ant colony optimization feature selection and extreme learning machine. *Neurocomputing*, 226:66 – 79.

Rijksinstituut voor Volksgezondheid en Milieu (2019). Stikstof. https://www.rivm.nl/stikstof.

Rijksoverheid (2019). Natura 2000. https://www.rijksoverheid.nl/onderwerpen/natuur-en-biodiversiteit/natura-2000. Accessed on the 18th of May.

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

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM JOURNAL OF RESEARCH AND DEVELOPMENT*, pages 71–105.

Schaminée, J., Stortelder, A., and Westhoff, V. (1995). *De Vegetatie van Nederland; deel 1: Inleiding tot de plantensociologie - grondslagen, methoden en toepassingen*. Opulus Press.

Schapire, R. E., Freund, Y., Bartlett, P., and Lee, W. S. (1998). Boosting the margin: a new explanation for the effectiveness of voting methods. *Ann. Statist.*, 26(5):1651–1686.

Schmidt, K. and Skidmore, A. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85(1):92 – 108.

Scholz, S. and Gruber, M. (2008). Radiometric and geometric quality aspects of the large format aerial camera ultracamxp.

Segal, M. (2003). Machine Learning Benchmarks and Random Forest Regression. *Technical Report, Center for Bioinformatics Molecular Biostatistics, University of California, San Francisco*.

Seidl, J. and McMordie, W. (1982). The Concise Oxford Dictionary of Current English. Oxford University Press, 7th edition.

Shanmugam, S., Lucas, N., Phipps, P., Richards, A., and Barnsley, M. (2003). Assessment of remote sensing techniques for habitat mapping in coastal dune ecosystems. *Journal of Coastal Research*, 19(1):64–75.

Sim, J. and Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3):257–268.

Sloss, C., Hesp, P., and Shepher, M. (2012). Coastal Dunes: Aeolian Transport. *Nature Education Knowledge*, 3:10:21.

Stars-project (2019). Multispectral and panchromatic images. https://www.stars-project.org/en/knowledgeportal/magazine/remote-sensing-technology/introduction/multispectral-and-panchromatic-images/.

Ten Harke, M., Van Boxel, J., and Verstraten, J. (1998). Water and solute fluxes in dry coastal dune grasslands: the effects of grazing and increased nitrogen deposition. *Plant and Soil*, 202:1–13.

Timm, B. C. and McGarigal, K. (2012). Fine-scale remotely-sensed cover mapping of coastal dune and salt marsh ecosystems at Cape Cod National Seashore using Random Forests. *Remote Sensing of Environment*, 127:106 – 117.

Tolt, G., Shimoni, M., and Ahlberg, J. (2011). A shadow detection method for remote sensing images using vhr hyperspectral and lidar data. pages 4423–4426.

Toth, C. and Jóźków, G. (2016). Remote sensing platforms and sensors: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:22 – 36. Theme issue 'State-of-the-art in photogrammetry, remote sensing and spatial information science'.

Tsoar, H. (1994). Bagnold, R.A. 1941: The physics of blown sand and desert dunes. London: Methuen. *Progress in Physical Geography - PROG PHYS GEOG*, 18:91–96.

U.S. Geological Survey (2017). Landsat 7. https://www.usgs.gov/media/files/preliminary-assessment-value-landsat-7-etm-slc-data.

USGS (2019). Band designations landsat-7. https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites?qt-news_science_products=0#qt-news_science_products.

USGS, NASA, Landsat 7 Science Team (2003). Preliminary assessment of the value of landsat 7 etm+ data following scan line corrector malfunction. *USGS*, 115:1,5–6. Compiled and summarized by the staff of the U.S. Geological Survey, EROS Data Center, Sioux Falls, SD 57198. 16 July 2003'.

Van der Maaten, L. and Hinton, G. (2008). Viualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605.

Van der Meulen, F., Van der Valk, B., Baars, L., Schoor, E., and Van Woerden, H. (2014). Development of new dunes in the Dutch Delta: nature compensation and 'building with nature'. *Journal of Coastal Conservation*.

Whiteside, T. G., Boggs, G. S., and Maier, S. W. (2011). Comparing object-based and pixel-based classifications for mapping savannas. *International Journal of Applied Earth Observation and Geoinformation*, 13(6):884 – 893.

Zhao, J., Zhong, Y., Hu, X., Wei, L., and Zhang, L. (2020). A robust spectral-spatial approach to identifying heterogeneous crops using remote sensing imagery with high spectral and spatial resolutions. *Remote Sensing of Environment*, 239:111605.