Mapping River Floodplain Vegetation Structure

with a consistent mapping scale
by aerial images and LiDAR data
in object-based random forest

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Ernst Kuilder

GIRS-2012-23
centre for geo-information
October 2012

MAPPING RIVER FLOODPLAIN VEGETATION STRUCTURE

with a consistent mapping scale by color-infrared aerial images and LiDAR data in object-based random forest classification

ERNST T. KUILDER

A thesis submitted in partial fulfillment of the degree of Master of Science

Laboratory of Geo-information Science and Remote Sensing
Wageningen University and Research

October 2012 – version 1.1

Ernst T. Kuilder: *Mapping River Floodplain Vegetation Structure,* with a consistent mapping scale by color-infrared aerial images and LiDAR data in object-based random forest classification, A thesis submitted in partial fulfillment of the degree of Master of Science, © October 2012

STUDENT REGISTRATION NUMBER:
880127485120

THESIS CODE:
GRS-80436

THESIS REPORT:
GIRS-2012-23

SUPERVISORS:
Dr. Ir. L. Kooistra
Dr. Ir. C.A. Mücher

LOCATION:
Wageningen, the Netherlands

DATE:
October 2012

The Rider

Every once in a while someone along the road lets us know how far behind we are. A man shouts: 'Faster!' He probably thinks bicycle racing is about going fast.

—Tim Krabbé

ABSTRACT

Vegetation exerts friction on water and obstructs flow, limiting the capacity of river floodplains to discharge water. Predictions of the discharge capacity of river systems require data on the state and distribution of river floodplain vegetation. Very high resolution aerial images and LiDAR (AHN2) datasets with a national coverage provide opportunities to produce vegetation maps automatically. As such the entire area of the river floodplains in the Netherlands may be mapped with high accuracy and regular updates, capturing the dynamic state of the vegetation.

In this study, the novel datasets are used to map the vegetation of 936 ha of the floodplain on the north-side of the river Nederrijn near Wageningen into ten vegetation structure classes. The method follows object-based image analysis principles. Objects are defined in FNEA segmentation and subsequently labeled using the ensemble-tree classifier *random forest*. The mapping scale is controlled by selecting segmentation parameters from quantified discrepancies between reference polygons and segmented objects. Effects on the mapping scale of different reference polygons and different segmentation data is investigated.

The results show that it is important to be able to select the right segmentation parameters to control the mapping scale. A discrepancy measure with reference polygons is a suitable method to do this objectively. The use of the same reference polygons on different data selects different segmentation parameters which result in a similar mapping scale. The application of *random forest* on the objects resulted in an estimated classification accuracy of 43% on the basis of an independent field validation and 86% on the basis of the built-in *out-of-bag* estimate of *random forest*. Variable importance measures of *random forest* showed that the AHN2 LiDAR dataset is a valuable addition to the spectral information contained in the aerial images in the classification.

Once we act, we forfeit the option of waiting until new information comes along. As a result, no-acting has value. The more uncertain the outcome, the greater may be the value of procrastination.

— Peter L. Bernstein (Bernstein, 1996)

FOREWORD AND ACKNOWLEDGMENTS

A couple of weeks ago I was playing scientist in the form of gathering data in the floodplain along the Rhine nearby Wageningen. I created a bunch of *stratified random sampling points* behind my computer to test the quality of my work. I took my hiking boots and my gps device and took off, determined to visit all the points and describe the vegetation present.

The first couple of points went ok. I ordered the points from west to east to cut travel time. At around point twenty six I ran into trouble. A fence. And a sign: *no trespassing, nature area*. I stood there for a while, but in the name of science I decided to climb the fence and continued in the direction of point twenty seven and twenty six. I quickly learned, not deep into the nature area, why this field was fenced. The field contained horses. Three horses. Running around single file like maniacs. At visual contact they grinned and came running towards me. At first I did not felt scared and thought that if I expressed this self-esteem, the horses would back-off soon. Well, they did not back-off. The horses, more looking like large ponies at close sight, kept on coming towards me and I started to feel they might be planning on running me over. Now feeling scared I ran into the safe haven of spiky bush vegetation.

The spiky bush vegetation was not at all a desirable place to be. It was spiky. The ponies chose not to follow the subject of their terror in there and continued running on a different bearing. And I continued my quest towards twenty seven and twenty six. I can tell you now: twenty seven was a sandy river-shore and twenty six was a difficult boundary case between shrubs and grass.

When I returned to the fence to get out of napoleon-complex-pony-territory, I had to face one more encounter. The same horses were standing between me and the fence.

They stood still and looked at me, but let me through.

I would like to thank the ponies from refraining to molest me. Although I had to save myself at the incident with the ponies in the field, this thesis would have been very different without the help of others.

I am thankful to my supervisors Sander Mücher and Lammert Kooistra. When I needed it, they supplied me with guidance and

Brain contamination: contaminating the thoughts in your brain and spoiling original ideas, is a concept I learned from the artistic duo Gilbert & George

confirmed that I was on the right track without infringing any major brain contamination.

The cover page is a photograph provided by my room-mate Tesse Bijlsma[1]. It contains the old brick factory near wageningen, a distinct feature of Dutch floodplains. If you watch closely you will find this brick factory in some of the aerial images in the chapters to follow.

This document is typeset in the open source document preparation system LaTeX. This particular document looks so great because of André Miede's ClassicThesis template[2] which is designed around Robert Brinhurst's *The elements of typographic style* (Bringhurst, 1992).

A great feature of LaTeX and this template are the references. If you are reading this thesis in digital form: the colored text contains links. Blue directs you to figures, sections or the acronym list. Green directs to the bibliography and red to the outside world.

All the graphs are made in the open source statistical environment R (R Core Team, 2012) and the ggplot2 package (Wickham, 2009). ggplot2 is greatly recommended for making nice looking graphs in R. Before printing, the graphs have been slightly touched in the vector graphics editor Inkscape[3], another great opensource application.

Most of the data used in this study has been provided by the *geodesk* at the laboratory of geo-information and remote sensing of Wageningen University. Rijkswaterstaat has supplied me with their *ecotope* map and the *pilot vegetation legger*.

Before you start reading the content, one more thing. Written scientific products are not generally appreciated by a broad audience. This thesis is no exception, it contains specific information which cannot be easily converted to knowledge for the broader audience. Since you have got this far in reading this particular chapter, you are probably not overenthusiastic about the scientific content. To prevent you *the broader audience* from feeling lost I have split my thesis in three bite-sized chunks. These parts are introduced by easy to understand paragraphs containing the principles of the content. If you feel like you are a representative of the broader audience, please read those paragraphs, I have written them for you.

1 http://www.tessebijlsma.nl/

2 available at: http://www.ctan.org/tex-archive/macros/latex/contrib/classicthesis/

3 available at: http://inkscape.org/

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AHN2 Actual Dutch digital elevation model

CART Classification and Regression Tree

CIR Color InfraRed

DCM Digital Canopy Model

DEM Digital Elevation Model

DSM Digital Surface Model

FNEA Fractal Net Evolution Approach

GEOBIA Geo Object Based Image Analysis

GLCM Gray Level Co-occurrence Matrix

GLDV Gray Level Difference Vector

GREN Green Red red-Edge and Near-infrared

IHS Intensity Hue and Saturation

LiDAR Light Detection and Ranging

MAUP Modifiable Aerial Unit Problem

NDVI Normalized Difference Vegetation Index

OBIA Object Based Image Analysis

OOB Out-Of-Bag

RGB Red Green and Blue

UAV Unmanned Aerial Vehicle

PI Photo Interpretation

PBM Previous Boundary Method

RF Random Forest

Part I

INTRODUCTION

Situated between the river and the dikes are the river floodplains. An area home to valuable land. The land contains soils with the perfect mixture of sand and fine clay for agriculture. It features stunning sceneries making it a desirable place to live. A major transport line is always close by and it gets increasing attention from nature conservationists for its ability to connect one nature reserve with the other. Most of all we expect it to handle increasing amounts of water, preventing it from spilling over the dikes to areas with increasing economic activity and an increasing amount of people. This calls for management based on detailed and up to date information.

1 SITUATION

Floodplains of the Dutch rivers; the Rhine, the Meuse and their tributaries, serve several potential conflicting functions. Currently the Rhine tributaries should have the capacity to discharge and store 16 000 m^3/s entering the Netherlands from Germany (Silva et al., 2001). This figure is expected to be increased to 18 000 m^3/s due to climate change in the near future (Makaske et al., 2011). Simultaneously the present nature value of the floodplains are institutionalized by EU legislation (European Union, 1992) and national laws (Department of LNV, 2000). EU birds directive and the more recent habitat directive serve as the basis for the creation of Natura 2000 areas.

Although the current Dutch government decided to suspend investment in the National Ecological Network (Rijksoverheid, 2011), many of the Dutch nature areas and particularly the floodplains are already part of the National Ecological Network or Natura 2000. Legislation states that existing species and habitats in these areas should be maintained or improved and floodplains have to serve as nature corridors connecting national and international nature reserves.

1.1 VEGETATION AND DISCHARGE CAPACITY

Vegetation structure and its seasonal and non-seasonal succession dynamics play an important role in the discharge capacity. Hydrological discharge is the amount of water flow, commonly defined as $Q = A \cdot \bar{u}$. With Q the discharge in m^3/s, A the cross-sectional area in m^2 and $\bar{u}$ the average flow velocity in m/s. The cross sectional area A of rivers consists of the river bed and in case of high water, the area between the dikes: the floodplains. The channel bed (Warmink, 2012), the soil, anthropogenic obstructions and vegetation exert friction on water flow. Obstructions lower A and friction lowers $\bar{u}$, limiting the discharge capacity of river floodplains.

Vegetated areas make up most of the floodplain area in the Netherlands and therefore have a great contribution on the total hydraulic roughness. This is a dynamic component. When vegetation grows, its hydraulic roughness increases, lowering the discharge capacity (Makaske et al., 2011; Järvelä, 2002).

1.2 CURRENT MAPPING PRACTICE

To determine the hydraulic roughness for hydrological models and to serve several spatial planning and ecological studies, Rijkswater-

staat (the Dutch river authority) constructs an ecotope map. The map contains spatially explicit ecological units based on biotic, abiotic and anthropogenic aspects (Rijkswaterstaat, 2008). By manually digitizing the vegetation structure from a dedicated aerial imaging campaign a Photo Interpretation (PI) map is created.

This construction follows two steps. First the individual objects such as grass areas, forests and water bodies are digitized on the basis of previous digitizations. The outlines of previous years are only changed when the change is greater than a certain threshold.

Second, the digitized plots are taken into the field were they are assigned labels. This controlled and robust mapping method aims at very stable-in-time maps. Changes in the maps of different years reflect actual physical changes and are not the result of the mapping methodology.

The PI map consists of 18 classes (Appendix A) and is combined with information about management, flood frequencies, inundation depth and morphological dynamics to create the ecotope map. Rijkswaterstaat has made the ecotope map twice, in 1997 and 2005, and is planning on making one every six years to connect to the EU water directive (Rijkswaterstaat, 2008).

Uncertainty

Knotters et al. (2008) estimated the 2005 map accuracy of 46 ecotopes, excluding urban, water or non-inundated areas. They estimated that the overall map accuracy of the ecotope map of 2005 was low at 47.8% with a standard error of 3.5%. The estimate was based on a stratified random sampling of 1041 validation points. The validation dataset was assumed to be perfect, there was no correction for possible acquisition errors. This causes a slight underestimation of the real accuracy.

Uncertainty in vegetation presence and state propagates to incorrect roughness coefficients. Monte Carlo simulations showed that in the case of the Dutch river system the uncertainty in vegetation maps leads to uncertainty in expected flood levels in the order of decimeters (Straatsma and Huthoff, 2011).

Previous boundary method

The sensitivity of the models to the roughness parameter also means that changes in the mapping result have great effects on the outcome. Hydrological models with the vegetation state as input are calibrated against flood levels. To prevent an unrealistically fluctuating flood level estimate, changes in the map should reflect physical vegetation changes. Distortions to such calibrations are *soft* boundaries between for instance a natural bush area and a grass field with shrub encroach-

ment. Shifts in the delineation and labeling of these areas between different mapping years may occur without any physical change.

To accurately calibrate the models and prevent fluctuations in flood levels, it is important that such areas are consistently delineated and labeled. In other words, overall reliability is more important than spatial and thematic accuracy.

This is implemented in the ecotope maps by the Previous Boundary Method (PBM) (in Dutch: "Oude Grenzen Methode"). This method optimizes the trade-off between accuracy and reliability by only changing the outline of a mapping unit if the mapper believes a border has shifted more than 10 m compared to the boundaries of the previous map. (Janssen, 2001).

1.3 VEGETATION MANAGEMENT

To prevent a decrease in discharge capacity due to vegetation growth, vegetation needs to be periodically removed. Periodic removal of vegetation is costly and many actors in the floodplains are unwilling to cover the expenses. Therefore alternative measures such as cyclic rejuvenation have been developed (Baptist et al., 2004). These measures aim at lowering the management costs while sustaining the nature value and discharge function of floodplains by applying activities such as biomass production, high standard meat production and fiber harvesting.

The alternative measures cause additional uncertainty about the floodplains discharge capacity due to dynamically changing vegetation. Further, the effect on vegetation of sudden events such as floods cannot be monitored with the current monitoring approach, limiting the implementation of cost effective management measures (Baptist et al., 2004).

1.4 DEVELOPMENTS IN MAPPING TECHNIQUES AND DATA

Combining structural information from airborne Light Detection and Ranging (LiDAR) with spectral information from either airborne or spaceborne sensors has proven to be a suitable method to monitor floodplain vegetation for hydrological models (Forzieri et al., 2010, 2011; Geerling et al., 2009; Straatsma and Baptist, 2008).

Geerling et al. (2009) and Straatsma and Baptist (2008) aimed at creating a PI map of a floodplain with an automated segmentation method of LiDAR data and spectral information from a Compact Airborne Spectral Imager. Geerling et al. (2009) concluded that such an approach results in a faster, repeatable and more accurate floodplain roughness map which enabled regular updates of hydrological models. They estimated the accuracy of the classification of a relatively small floodplain into eight classes (Appendix A) at 74% based on a

348 random validation points. The validation dataset was acquired by manually interpreting the data used to classify and as such not completely independent.

Geerling et al. (2009) tested their method on a floodplain along the river Waal as a proof of concept. Nation-wide datasets such as the high point-density elevation dataset for structural information: the Actual Dutch digital elevation model (AHN2) and aerial photographs provide opportunities for up scaling to nation-wide products. Mapping products with enhanced thematic detail, more frequent updates and the possibility to incorporate regional data for the monitoring of sudden and regional change.

2 PROBLEM DEFINITION

This research aims at bridging a knowledge gap between reliable and robust manual mapping of floodplain vegetation structure, and automated monitoring with increased accuracy and detail. High resolution aerial images and LiDAR data with a national coverage is available to be used with well developed techniques to correctly classify the data into vegetation structure classes.

The problem to be tackled lies in the lack of controllability. Users are in control in a manual process, both over the delineation of boundaries as over assigning labels to the delineated areas. Although being in control does not necessarily give better results, it does give reliable results. Stable mapping products over time, with the right mapping scale independent of the used data. Current (semi-) automated methods lack this control, resulting in mapping products of one dataset, which differ more than the physical change from mapping products of another dataset. This currently makes (semi-)automated mapping products less robust and thus less suitable for hydrological models.

2.1 DATA FUSION

Nowadays, very-high resolution satellite data with a spatial resolution of less than one meter (IKONOS, Quickbird, Worldview-2) and aerial images with resolutions of half a meter and less are widely available for the whole of the Netherlands. This data may be supplemented with regional images after specific events such as floods or land use changes.

Stereo-aerial images may be substituted by the Dutch AHN2, a very-high resolution nation-wide LiDAR dataset which provides information about the ground elevation and the structure of vegetation. Fusion of these datasets towards time-stable vegetation structure information is difficult. Besides differences in amount and width of spectral bands, and differences in the moment when the data was captured, there are differences in spatial resolution.

For a review of the concepts underlying the problems and opportunities related to the fusion of multi-sensor, multi-temporal, multi-resolution and multi-frequency data for remote sensing applications, please refer to Pohl and Van Genderen (1998) and Zhang (2010). This thesis fuses data at the decision level. Data layers of different sensors, acquisition time and resolution are combined into classification variables. Insights in the importance of these variables provides knowl-

edge about the value and suitability of datasets for mapping of floodplain vegetation structure.

2.2 OBJECT-BASED VERSUS PIXEL-BASED

Instead of analyzing individual pixels, Object Based Image Analysis (OBIA) groups pixels into meaningful objects and analyses the objects for classification. On top of the information contained in the individual pixels, objects contain information about the relevant context of a pixel. Alternative to OBIA is to use a moving window to incorporate contextual information of pixel's direct neighborhoods. Downside of such an approach is that the neighborhood of a pixel is not necessarily meaningful and thus not necessarily relevant for classification (Stuckens et al., 2000), it does not embrace spatial concepts (Blaschke and Strobl, 2001). Besides this, OBIA gives the user control over the mapping scale and can handle the implicit variability that comes with very-high resolution imagery (Jyothi et al., 2008; Liu and Xia, 2010). More important for applications where reliability is more important than accuracy, OBIA separates the identification from the classification which is in line with the manual approach of delineation of boundaries and the assignment of labels in the field. The user has more control over the final mapping result since it has choices in both steps. Simultaneously, this means that there should be objective mechanisms to identify objects of the right scale, and consistently assign the correct labels.

2.3 IDENTIFYING OBJECTS OF THE RIGHT SCALE

The larger an object, the better the contextual information of pixels and thus the more accurate the classification result. As shown by Liu and Xia (2010), this relation holds until the objects include too many pixels and lose their physical meaning. In example, pixels belonging to a grass area are grouped to pixels of a forested area. If the mapping goals is to map the area in vegetated and non-vegetated areas, this would be correct. Though if the goal is to map the grass and forest areas, this would result in a classification error. Therefore, a common problem in OBIA is identifying objects of the right scale in the image segmentation step.

Research has primarily been focused on finding an optimum scale for the data at hand in respect to accuracy: maximizing intra-segment homogeneity and inter-segment heterogeneity (Espindola et al., 2006; Gao et al., 2011; Johnson, 2011). Möller et al. (2007) developed a promising tool for comparing manually digitized areas with generated objects to asses segmentation accuracy.

Although all these measures give objective results which are optimized to the data, they are not designed to result in consistent bound-

aries comparable to a method such as PBM . Thus, whereas previous research defined the *right* scale as the scale where the map accuracy is maximized, this research focuses on objectively finding a segmentation scale which is relevant for the mapper and could be kept consistent in following years, independent of the data.

Section 1.2 contains an introduction to PBM: Rijkswaterstaat's solution to consistently delineate soft boundaries

2.4 ASSIGNING LABELS

The outlines of the PI map are manually assigned one of the 18 classes (Appendix A). Together the manual delineation from stereo-aerial images and the classification in the field produces the PI map which relies heavily on human interpretation. This dependence makes the process labor intensive and prone to human error, resulting in a map with low accuracy and high uncertainty (Knotters et al., 2008).

An automated classification method lowers the dependence on expert knowledge about the mapped objects, potentially lowers the subjectivity and may include information about the uncertainty of the assigned label.

Research has shown that fusing data layers and deducting variables from it, provides additional value to the differentiability of the objects (Yu et al., 2006). These variables include textural features (Haralick et al., 1973), and color-space transformations (Koutsias et al., 2000). Conventional parametric classification techniques are not appropriate for classification of multi-source and multi-type data (Gislason et al., 2006). More important for this purpose is that the classification should be reproducible and consistent and that the classification accuracy increases. To keep the mapping procedure robust to changes in data, there is a need to understand the importance of the data layers and variables for classification. As such it is possible to monitor the effects of the used data layers on the classification result.

3 RESEARCH OBJECTIVES AND QUESTIONS

SEGMENTATION

To develop an objective (semi-)automated approach for the segmentation of very high resolution spectral and LiDAR derived elevation and surface data to map river floodplain vegetation structure

1. Could manually delineated reference plots be used to objectively determine segmentation parameters?
2. Could previous mapping results be used as reference objects to objectively determine segmentation parameters?
3. How do manually delineated reference plots compare to previous mapping results in determining segmentation parameters?
4. How do manually delineated reference plots compare to previous mapping results in determining segmentation parameters?
5. Could a comparison with reference objects be used to select segmentation results of different datasets with similar mapping scales.

CLASSIFICATION

To investigate the robustness of object-based random forest classification with added secondary variables for floodplain vegetation structure mapping

1. What is the effect of optimal and sub-optimal segmentation parameters on the classification result?
2. Which datasets are important in the classification of the objects?
3. Which derived variables are important in the classification of the objects?
4. How accurate is the classification of the objects?

4 IMAGE SEGMENTATION

OBIA, or Geo Object Based Image Analysis (GEOBIA) as some argue (Ardila et al., 2012; Chen and Hay, 2011), is an emerging paradigm in remote sensing. A paradigm driven by fine resolution imagery. Often such imagery results in *high resolution situations*, situations where pixels are significantly smaller than the objects of interest. This is the case in floodplain mapping, where objects of interest are several meters wide and long, and the spatial resolution of modern aerial data is half a meter or less.

The goal of OBIA is to intelligently describe the imaged reality using spectral, textural, spatial and topological characteristics. Intelligent in the delineation of spatial units, including implicit information, and intelligent in creating an output that is an understandable summary of the complex content of the data. Segmenting an image to create objects is the prerequisite step of OBIA.

This chapter continues with a literature review of image segmentation aimed to serve as an introduction on the performed analysis. For an extensive literature review on the emergence of OBIA in the geo domain since the 2000's, the readers is referred to Blaschke (2010).

4.1 SEGMENTATION ALGORITHM

Image segmentation has its roots in machine vision of the 1980's (Fu and Mui, 1981; Haralick and Shapiro, 1985). These algorithms are less suitable for earth observation data due to the inherent nature of the data and goals of the algorithms. The algorithms are used to recognize patterns and ease for instance product quality control with artificial surfaces and discrete objects. This differs from the aim of earth observation to identify and discrete spectrally homogeneous segments to map the inherent features of remote sensed images (Blaschke et al., 2004).

The segmentation used for this thesis is Fractal Net Evolution Approach (FNEA) (Baatz and Schäpe, 2000). FNEA can be applied in the commercially available software package eCognition, which generally gives good segmentation results (Meinel and Neuber, 2004), and is widely used in scientific studies (Geerling et al., 2009; Zhang and Huang, 2010; Myint et al., 2011; Benz et al., 2004).

Types of algorithms

Segmentation algorithms commonly used for earth observation are divided in point-based, region-based or edge detection (Schiewe, 2002).

POINT BASED or threshold-based segmentation involves two steps. The first step groups pixels according to their position in the the feature space. The second step spatially defines the segments by combining the members of a group based on their spatial connectivity. For natural scenes, this approach generally gives similar *salt and pepper* effects as a pixel-based approach (Schiewe, 2002).

COMMON EDGE-BASED APPROACHES employ a filter to map boundaries. The edges are connected by a contour generating algorithm to create objects (Schiewe, 2002). Edge detection is especially useful for the mapping of discrete features which contain clear boundaries such as urban areas (Stuckens et al., 2000).

REGION BASED APPROACHES are required for the classification of floodplain vegetation with its continuous nature. Region based algorithms create objects by merging or splitting regions based on the homogeneity of the pixels and shape of the region. These types of segmentation algorithms use region growing, merging and splitting or combined techniques to create the segments.

Region growing starts with seed points. The neighboring pixels are pairwise joined to the seed points. New seed points are added and the process is repeated until a threshold is exceeded or the whole image is segmented. Common thresholds are a maximum allowed heterogeneity of the region, sometimes combined with a shape criterion (Blaschke et al., 2004). These algorithms depend on a set of seed points, which determine the start initial regions. As such, results are only reproducible when the same same points are used (Blaschke et al., 2004).

Region merging and splitting starts with respectively individual pixels or the whole scene and either merges the pixels or splits the objects until a certain maximum homogeneity is met. The challenge for these algorithms is to let the merges or splits happen simultaneous or *pseudo-simultaneous* for the whole scene to ensure an even growth of objects.

Fractal net evolution approach

The FNEA algorithm used for this study is of the region merging type (Baatz and Schäpe, 2000). It may either start with existing objects,

functioning as 'seeds' or it starts with individual pixels, objects consisting of 1 pixel.

Merging does not happen simultaneously, instead object pairs are considered local. It starts with an arbitrary object A, to find the neighboring object B which is most comparable to A. Do the same for B to find C. If $C \neq A$, B is better merged wit another object than A, continue by repeating the loop with B for A and C for B. In this *local mutual best-fitting*, the most homogeneous merge in the vicinity of A is chosen in every step.

Since objects are not considered simultaneous, a distributed treatment order is necessary to prevent growth of objects unequally over the scene. This distributed treatment order chooses the A objects which are used as seeds for the search of the best local merge candidate.

The treatment order in FNEA is not completely random, instead it ensures that every object is handled ones every cycle and distributes subsequent merges as far as possible from each other (Baatz and Schäpe, 2000). The results of subsequent merges are influencing the heuristics decisions of subsequent merges. Initial A pixels are selected random and search for a mutual best-fit in their surrounding. When the segmentation is run again from scratch, other pixels become the initial A pixels.This historicity makes the segmentation not fully reproducible (Blaschke et al., 2004). According to Baatz and Schäpe (2000) this can be seen at objects of low contrast where the precise edges are arbitrary and might fluctuate between segmentations.

4.2 THE ISSUE OF SCALE

Figure 1 shows an aerial image containing objects of different scale. Segmentation of this image may create objects consisting of individual trees and shrubs, hedges, forested areas or the image may be seen as only a couple of objects: one industrial, two water and one big vegetated object.

Thus, the same data may give different results depending on the scale and aggregation of the analysis. This is known as *the issue of scale* (Lam and Quattrochi, 1992) and was first recognized by Gehlke and Biehl (1934). They investigated the effect of geographic and random grouping of census data and subsequently raised the question whether a geographic area is an entity with certain characteristics or merely a property of the characteristic itself.

In remote sensing this translates to the Modifiable Aerial Unit Problem (MAUP) (Marceau and Hay, 1999). Traditionally, the *modifiable areal units* are pixels and the problem is associated to data with varying resolutions. In OBIA, the problem is shifted from data to the segmentation. The higher the resolution of the image the smaller the objects of

Figure 1: True-color aerial image of 2008 containing objects of different scale

interest may be. Simultaneously, the higher the resolution, the more heterogeneous objects of similar scale will be.

OBIA has two scale related boundaries, objects must consist of more than one pixel and less than the whole image. Between that, space is relative, there is generally no reference or sense of scale incorporated in the segmentation step of OBIA. Scale becomes a choice fed by intuition about the features under observation (Marceau, 1999).

Hall et al. (2004) showed that landscapes are complex and scale dependent. They performed segmentations and classification of high resolution satellite images to demonstrate that presence and distributions of classes differs between scales. A practical result for floodplain mapping is that the resulting mapping units have an arbitrary scale and the hydraulic roughness deducted from it become arbitrary and thus has little validity. Different degrees of floodplain roughness can be found at different scales.

4.3 OVER AND UNDER SEGMENTATION

Two definitions of scale are used in this study: segmentation scale, defined as the heterogeneity of pixel values in the objects (Section 4.4) and mapping scale, the degree of clutter and aggregation of the mapped features (Section 4.2)

The amount of information contained in an object is subject to the amount of pixels and thus a result of the segmentation scale and the resolution of the data. Over segmentation creates too many and too small objects, under segmentation creates too few and too big objects.

Independent of data resolution, the chance a pixel to be correctly classified, increases with scale until objects contain too many pixels of multiple classes. Liu and Xia (2010) illustrate this from a data perspective. They show that the highest classification accuracy is found at the scale where the errors due to the inclusion of the wrong pixels in an object, equalizes to the increase in information from inclusion of the right pixels.

According to Gao et al. (2011) the loss of accuracy is higher with under-segmentation than with over-segmentation. In other words, it is more harmful to the classification accuracy to make objects too big, than to make objects too small.

4.4 PARAMETERS

Region based segmentation requires user-specified parameters to control the heterogeneity and shape of the objects. These settings implicitly define the map-scale. The basis for the settings are the users desire to map objects of a certain scale, the resolution and dimensionality of the data and the resulting accuracy of the classification.

Scale

FNEA looks for the local mutual best fit, as such it complies to *fuzzy set theory*. An object belongs to a set of objects but its relation to the set is fuzzy instead of binary (Maiers and Sherif, 1985).

The degree of fit between two objects is defined according to the increase in heterogeneity as a result of a merge of two objects.

$$\Delta f = (1 - w) \cdot \Delta h_{color} + w \cdot \Delta h_{shape}$$

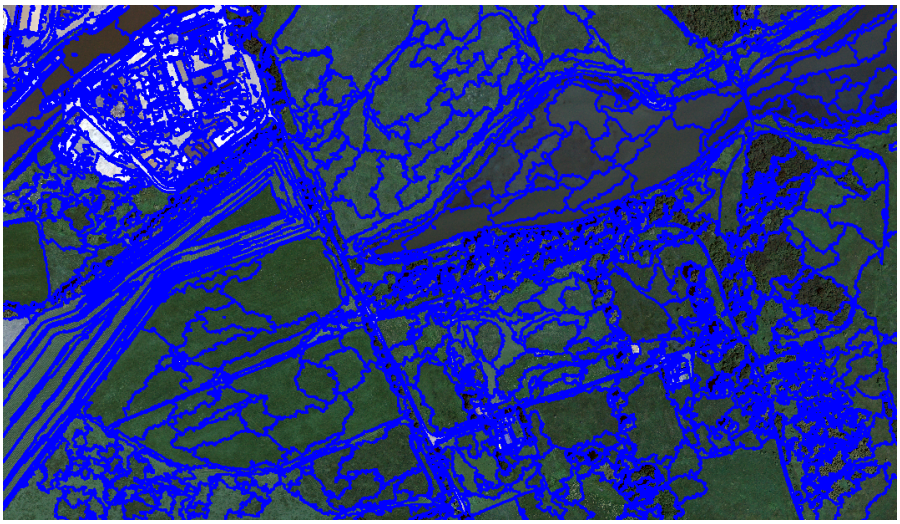
where

Δf	Increase in heterogeneity or *scale*
w	Weight to shape
Δh_{color}	Increase in color heterogeneity
Δh_{shape}	Increase in shape heterogeneity

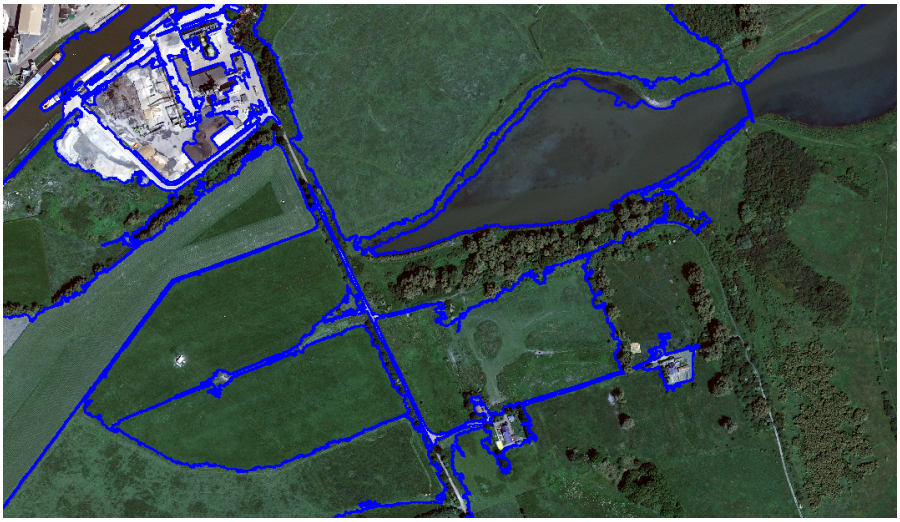
To control the growth of objects, there is a least degree of fit as a homogeneity constrain, f. Since the least degree of fit indirectly controls the scale of the objects, this is called the scale parameter. Figure 2 gives two examples of the possible scale settings. Figure 2a shows many more segmented objects than a human would perceive, a case of over segmentation. In Figure 2b the contrary happens, under segmentation. Part of the water body on the right of the image is grouped to the bush and grass area below of it.

Color

Color heterogeneity constitutes to the heterogeneity of the data layers which may consist of many more data layers than just Red Green and Blue (RGB). Research has successfully included LiDAR data (Hermosilla et al., 2011), combined with hyperspectral data (Geerling et al., 2009), or data layers from data reduction methods such as principal components (Zhang and Huang, 2010).



(a) Scale 100



(b) Scale 1000

Figure 2: Segmentation results of true-color aerial image of 2008 with scale 100 and 1000

The data layers have a user-defined weighted effect on the change in color heterogeneity.

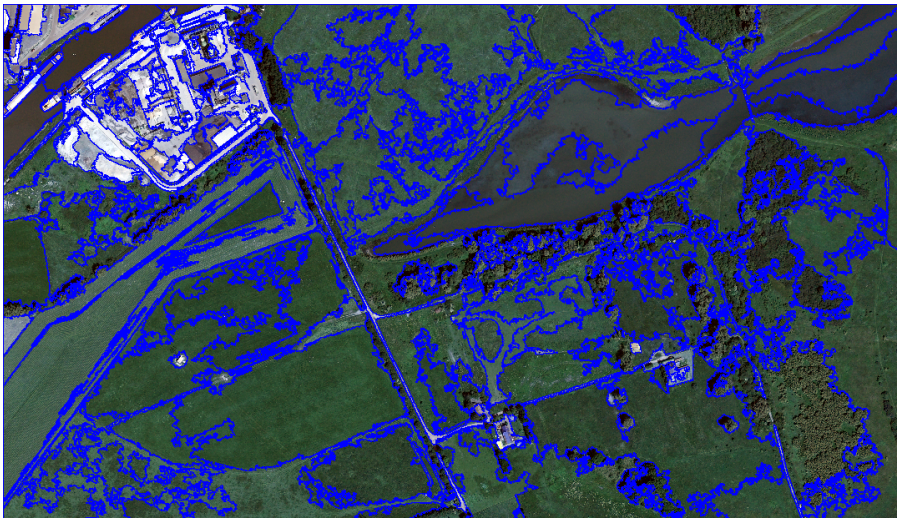
$$\Delta h_{color} = \sum_{l} w_l(n_{merge} \cdot \sigma_{lmerge} - (n_1 \cdot \sigma_{l1} + n_2 \cdot \sigma_{l2}))$$

Increase in color heterogeneity (Δh_{color}) of an object is defined as the weighted (w) sum over all the data layers (l) of the standard deviation (σ) weighted by the number of pixels (n) of the merged (n_{merge}) object minus that of the two merge candidate objects ($n_1 \cdot \sigma_{l1}$ and $n_2 \cdot \sigma_{l2}$).

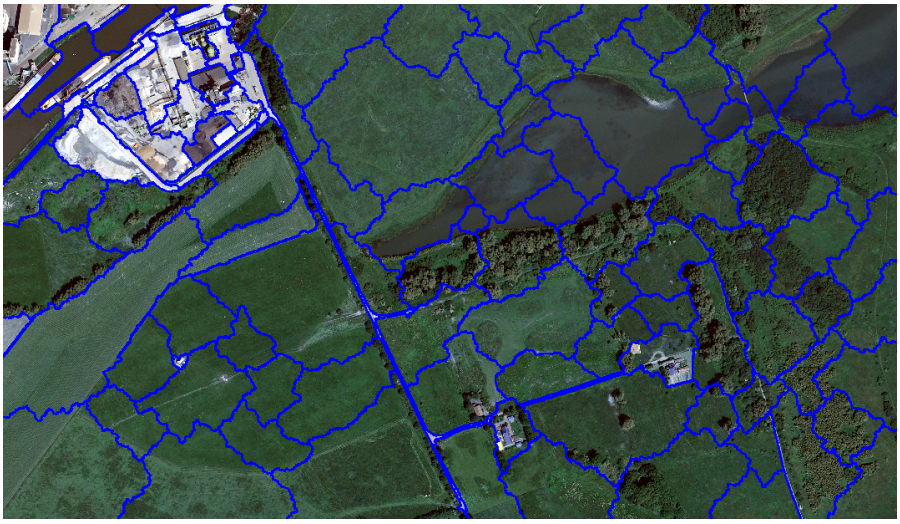
Shape

Color is the only source of information for the segmentation and is thus also the most important, but there is a need to keep the shape of objects resembling those of the real world to keep the result appealing.

Figure 3 shows the effects of no recognition of the shape of natural objects by a segmentation with a weight of zero assigned to weight in Figure 3a. Figure 3b shows the contrary with objects which are almost circular.



(a) Shape 0%



(b) Shape 80%

Figure 3: Segmentation results of true-color aerial image of 2008 with shape 10% and 80%

$$\Delta h_{shape} = (1 - w_{comp}) \cdot \Delta h_{smooth} + w_{comp} \cdot \Delta h_{comp}$$

This is implemented in FNEA by the user-defined shape heterogeneity criterion (Δh_{shape}) which is built-up of smoothness (Δh_{smooth}) and compactness (Δh_{comp}).

$$\Delta h_{smooth} = n_{merge} \cdot \frac{l_{merge}}{b_{merge}} - \left(n_1 \cdot \frac{l_1}{b_1} + n_2 \frac{l_2}{n_2} \right)$$

$$\Delta h_{comp} = n_{merge} \cdot \frac{l_{merge}}{sqrtn_{merge}} - \left(n_1 \cdot \frac{l_1}{\sqrt{n_1}} + \cdot \frac{l_2}{\sqrt{n_2}} \right)$$

Δh_{smooth} reflects the deviation of the perimeter of the object(l) and the perimeter of the smallest bounding box around the object. Δh_{comp} comprises the compactness of an object, the ratio between the size and the length of the perimeter.

The effects of a zero weight assigned to compactness and 100% weight to compactness is illustrated by Figure 4. A low compactness setting gives smooth, more rectangular objects (Figure 4a). With the same scale and shape, a high compactness setting gives compact, square or circular objects (Figure 4b).

4.5 MEASURES TO SELECT OPTIMAL PARAMETER VALUES

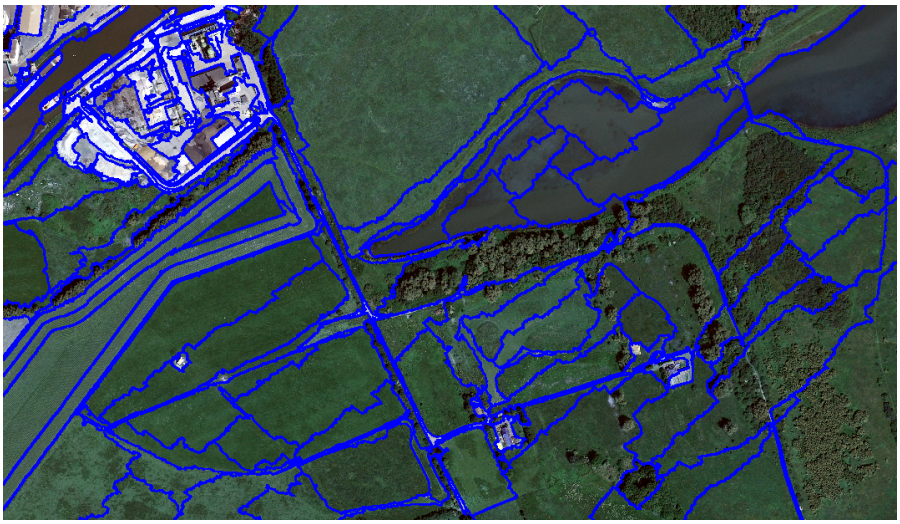
Objectively selecting the optimal parameter values of segmentation algorithms has become its own sub-field within OBIA (Blaschke, 2010). Zhang (1996) distinguishes three types: analytical, empirical goodness and empirical discrepancy measures.

ANALYTICAL METHODS involve the inner workings of the algorithm and are unavailable for common region growing algorithms including FNEA due to their complex nature and heuristic decisions.

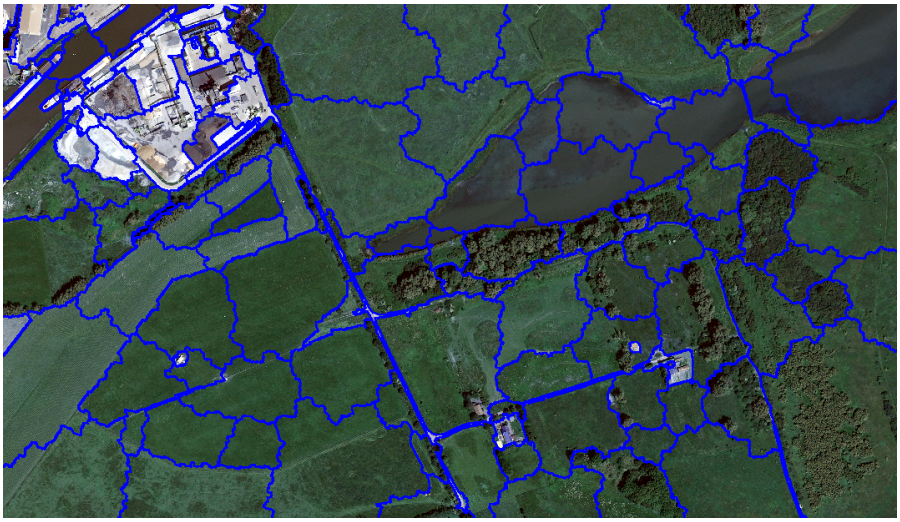
EMPIRICAL GOODNESS MEASURES are based on objective measures of "goodness" such as intra-segment heterogeneity and autocorrelation (Espindola et al., 2006; Gao et al., 2011). They are set apart from empirical discrepancy measures because they do not require expert knowledge such as reference segments.

EMPIRICAL DISCREPANCY aims at quantifying the differences between reference objects and segmented objects. This is described as the *problem of matching objects* (Zhan et al., 2005). Many of such measures have been developed recently. For a non-exhaustive comparison, please see Liu et al. (2012). Not included in Liu et al. (2012), but used in this study in adapted form, is the empirical discrepancy method of Möller et al. (2007).

All the measures use reference polygons, these can be individual objects or digitized regions. The measures then quantify the degree of



(a) Compactness 0%



(b) Compactness 100%

Figure 4: Segmentation results of true-color aerial image of 2008 with compactness 0% and 100%

over and under segmentation of the segmentation result. In the case of over segmentation there are more objects than the reference object, the reference object may overlap and they may not lay in the same position.

The reference polygons enable the user to explicitly define the desired mapping scale and is as such independent of the resolution of the data and considered a partial solution to the MAUP (Hall et al., 2004).

Möller et al. (2007) does not compare individual segments to reference objects since a segment may touch and overlap multiple reference polygons and vice versa. Instead the segments and the reference are intersected to get combined polygons, smaller or equal to those

of the reference and the segments alone. In other words the segments and the reference have a one-to-many relationship to the intersected polygons.

The intersected polygons are compared on size and position of its center of gravity.

$$RA_{SO} = \frac{A_i}{A_{SO}}$$

where

RA_{SO}	Relative area to super object, with either segment or reference object as super object
A_i	Area of intersected polygon
A_{SO}	Area of super object

$$RP_{SO} = \frac{d}{d_{max}}$$

where

RP_{SO}	Relative position to super object
d	Distance of intersected polygon's center of gravity to the center of gravity of the super object
d_{max}	Maximum d of the intersected polygons belonging to the super object

Combined, these metrics give a comparability called the comparison index. The comparison index gives the alikeness of the intersections to the segments and to the reference, corresponding to the under and over segmentation.

The comparison between intersected polygons and reference gives over segmentation, while the comparison between intersected polygons and segments gives under segmentation. When the whole image is one segment, the intersection will fully resemble the reference and the under segmentation is none and the over segmentation is full. Similarly, when all the pixels are one segment, the intersections will fully resemble the segments and the under segmentation is full and the over segmentation is none. So, when scale increases, the segments will become bigger, under segmentation will become less and over segmentation will become greater. The reference most resembles the segments when over and under segmentation are equal (Möller et al., 2007).

4.6 ALTERNATIVE APPROACHES

Region merging and splitting may also be performed on existing segmentation results. As such it is possible to create a work flow that

simulates that of the PBM. Objects of previous floodplain vegetation maps may be used as seed objects in a segmentation algorithm. Existing segments are only altered if the new data shows significant more or less heterogeneity than certain user-specified thresholds. Or a user may decide to only alter segments belonging to certain classes. As such the methodology becomes a *multi-scale* analysis and involves *object-relationship modelling*. With segments of the re-segmented objects of a different scale than the original. For this the reader is refered to Burnett and Blaschke (2003) who performed such an analysis using the FNEA algortihm.

5 OBJECT-BASED CLASSIFICATION

Segmentation of the source data creates objects. Within the framework of OBIA, the objects may be further refined in subsequent segmentation steps, regression may be applied, or the objects are categorized into predefined classes (supervised) or a number of undefined groups (unsupervised).

The machine learning algorithm Random Forest (RF) (Breiman, 2001) is used in this study to perform a supervised classification of the segmented objects into ten vegetation structure classes. Such a classified map is valuable for several spatial planning and ecological studies in the river floodplains.

5.1 THE CURSE OF DIMENSIONALITY

Objects contain more information than the digital numbers of the pixels alone. Up to hundreds of additional *object features* provide additional means to differentiate between classes. The object features may be summaries of the data sources, including texture metrics but also geometric characteristics of the objects and relations of objects to other objects.

The object features, or *variables* (v) once used in classification, violate the assumptions of parametric statistical techniques. Variables are often not normally-distributed and may be categorical. Additionally, the high amount of variables often leads to a *small n, large v* situation, where the amount of variables is smaller than the amount of observations (n) used to train the classifier. This is called the *curse of dimensionality* and a major cause of overfitting with parametric or simple non-parametric classification techniques such as k-nearest neighbors (Verleysen and François, 2005). For these reasons common parametric classifiers such as maximum likelihood are deemed inappropriate (Duro et al., 2012).

5.2 RANDOM FOREST CLASSIFIER

RF makes no assumptions on the distributional characteristics of the independent variables nor the response variables and may handle situations where v greatly exceeds n (Cutler et al., 2007). This makes RF a suitable classifier to perform an object-based classification with limited amount of training data and a high amount of object-features.

RF builds trees by taking a random subset of variables and a random subset of training data. The amount of trees, the number of vari-

ables and the number of training objects of each tree are all parameters of the algorithm, though their settings have little influence on the result (Breiman, 2001; Gislason et al., 2006; Liaw and Wiener, 2002).

The data is recursively split into increasingly homogeneous regions. At each step the most optimal variable and value is selected that results in subgroups of the data with the least impurity. This impurity is defined by the *gini* measure, similar to Classification and Regression Tree (CART). The splitting continues until all the variance is explained (Cutler et al., 2007).

During classification, all the objects are pushed through the trees and the trees cast a vote according to the class of the terminal node. The objects get classes assigned on the basis of a majority vote. For a more thorough review of RF please refer to the 15^{th} chapter of Friedman et al. (2001).

Additional information about the forest

Breiman (2001) states that:

> *A forest of trees is impenetrable as far as simple interpretations of its mechanism go.*

There is no way to fully interpret the trees or the functioning of the variables and observations within the trees. Though RF has build in functions to get a grip on the forest. Observations and variables which were not part of the subset of the tree are considered Out-Of-Bag (OOB). The OOB observations are used to cross-validate the trees, which results in a confusion matrix for the total forest (Cutler et al., 2007).

A similar approach is used to asses the importance of individual variables. The values of the variable is randomly changed for the OOB observations, the modified OOB's are pushed down the tree the resulted prediction. This is compared to the original OOB observations to get a measure of the importance of the variable (Cutler et al., 2007).

Random forest compared to other classifiers

RF is a modern machine learning classifier using an ensemble of decision trees (Breiman, 2001). Ensemble decision trees are among a couple of modern popular alternatives to the traditional maximum likelihood classification.

For a comparison based on several datasets of the practically available alternatives including RF and the popular support vector machines, please refer to Meyer et al. (2003). For a comparison between the ensemble decision tree methods such as RF and a method with an individual decision tree CART, please see Gislason et al. (2006). And for a comparison between RF and other ensemble tree methods using

bagging and boosting, tested on land cover datasets, please refer to Chan and Paelinckx (2008).

All authors agree that RF generally ranks high concerning classification accuracy and that RF is relative insensitive to its parameters and computationally fast.

Breiman (2001) states that a forest is constructed with double randomness of the subset, both random v and random n. In addition, a high amount of trees in the forest, makes the *law of large numbers* applicable. As such, RF is theoretically insusceptible to issues commonly associated with decision trees (Breiman, 2001). Though there is some debate about the tendency of decision trees to overfit (Segal, 2004) and poor performance of RF with high class-imbalance (Lusa et al., 2010).

Random forest and regression

Specifically hydrologists will need a lookup table to attach roughness values to the classes. An alternative would be to use the objects with regression and create a roughness map directly, superseding the use of a lookup table. This has been proven to be a viable approach by Forzieri et al. (2011) and for a Dutch floodplain by Straatsma and Middelkoop (2007).

Although RF is used in supervised classification mode in this study, RF may also be used to perform such calibrated regressions. RF has for example successfully been used to model biomass from high resolution satellite imagery (Mutanga et al., 2012).

5.3 VARIABLES

Instead of using individual pixel values, OBIA uses the object means of the different bands, which lowers the noise and *salt and pepper* effect. Besides the color and elevation means, over a hundred other features are included: variance of spectral bands, a vegetation index, color-space transformations, texture measures, and geometric features.

Similar features have been used by Yu et al. (2006), who also ranked the features by using CART. For their classification of tree species in a mountainous area, they found three texture features in the top 16. Yu et al. (2006) also included ancillary data to calculate topographic features such as distance to nearest water body. These features were also present in the top 16 among elevation and slope derived from LiDAR.

Color-space transformation

Color is often presented as RGB, consisting of the reflection in the red, green and blue parts of the spectrum. There are many other models

which describe light in the visible spectrum. Several of these models have been tested on artificial images by Alata and Quintard (2009).

Specifically an Intensity Hue and Saturation (IHS) transformation has been successfully used in mapping burned areas of landsat images (Koutsias et al., 2000). With IHS, spatial characteristics of composite images are separated to the intensity and the spectral information is kept with the saturation and hue.

Texture

Next to color and shape, texture is one of the main components humans use to recognize objects. It is defined as the spatial frequency of tonal changes within certain area (Ke et al., 2010). These areas are commonly moving windows or in case of OBIA objects and are thus only applicable in high resolution situations.

High resolution situations: cases were the objects of interest are smaller than the resolution of the data (Chapter 4)

Texture in this study is quantified using Gray Level Co-occurrence Matrix (GLCM) and Gray Level Difference Vector (GLDV) (Haralick et al., 1973). GLCM is a matrix of frequency of band values at a specified distance in an object. GLDV is the sum of the diagonals of the matrices (Ke et al., 2010). Different metrics may be deduced from the matrices and vectors, common are mean; variance; homogeneity; contrast; dissimilarity; and entropy.

Texture may be calculated of different bands. Yu et al. (2006) succesfully used these texture metrics of the near-infrared band, while Ke et al. (2010) calculated the metrics on four different spectral bands and on transformations and indices such as principal components and Normalized Difference Vegetation Index (NDVI) (Johansen and Phinn, 2006). Johansen and Phinn (2006) found positive relationships between these texture bands and vegetation parameters such as leaf area index. This suggests that these texture metrics provided valuable information in predicting the class of objects.

Part II

METHODOLOGY

The human mapping capacity sets the bar high when it comes to precision, flexibility and overall map quality. Automating the manual process introduces reproducibility and provides relative cheap opportunities for more up to date information. More important, it may improve the consistency of the maps. Time-consistent mapping methods lead to precise predictions of the maximum amount of water to flow between the dikes. The goal of the automated method is to mimic the human mapping process while adding the benefits of an automated mapping procedure.

6

STUDY AREA

Figure 5: True-color aerial image of 2008 of the study area (north to the right) with the position of the study area in the Netherlands in the insert (north up)

Figure 5 gives the boundaries of the study area. Please note that the north of the image is to the right of the page. The area is confined by a small river port in the west and a railway bridge over the Nederrijn in the east. The acquisition of AHN2 data is not fully completed for this area of the Netherlands, therefore the study area is confined to the floodplains north of the river Nederrijn, including fifty meters of the river. The dikes are not considered part of the floodplains, which is in correspondence to the current ecotope maps.

Classes and landmarks

10 classes are identified within the study area. Representations are depicted in Figure 6. The area is characterized by three major built-up areas (Figure 6f): a small river port west of Wageningen, a paper factory south of Renkum, a major highway bridge (the A50) east of Heteren and the locks and weirs to the north of Driel.

Diversity

The area to the west of the highway near Heteren consists of managed natural vegetation, mainly grazed grass with herbaceous vegetation and patches of bush. This area also contains some old river arms and restored oxbow lakes. To the east of the hightway, the patches are bigger and consist of meadows and agricultural fields.

Size

The study area covers 936 ha which is significantly greater than experimental study areas of previous attempts to map floodplain vegetation structure (Geerling et al., 2009; Straatsma and Baptist, 2008) ($<$ 100 ha). This makes it possible to study the spatial variability of scale at which vegetated objects occur and to investigate the possibility to apply an automated mapping methodology nation-wide.

Class orchard

There are only two fruit orchards present in this area. This makes selecting enough training objects to accurately train the classifier difficult. All the other classes are sufficiently present in the study area to classify correctly and to validate the method.

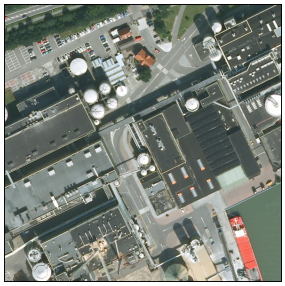
(a) Water

(b) Forest

(c) Orchard

(d) Bush

(e) Field



(f) Built-up

(g) Sand

(h) Herbaceous

(i) Grass

(j) Pioneer

Figure 6: True-color aerial image of 2008 of the 10 classes distinguished in this study, Scale 1:5,000

7 DATA SETS

Two datasets are used as information sources to mimic the manual digitization and classification of river floodplain vegetation according to the flowchart of Figure 7. Spectral data from two high resolution airborne sensors, a Color InfraRed (CIR) of 2008 and a Green Red red-Edge and Near-infrared (GREN) of 2012. Both datasets are used in conjunction with the LiDAR AHN2 data, providing information on elevation and the structure of vegetation.

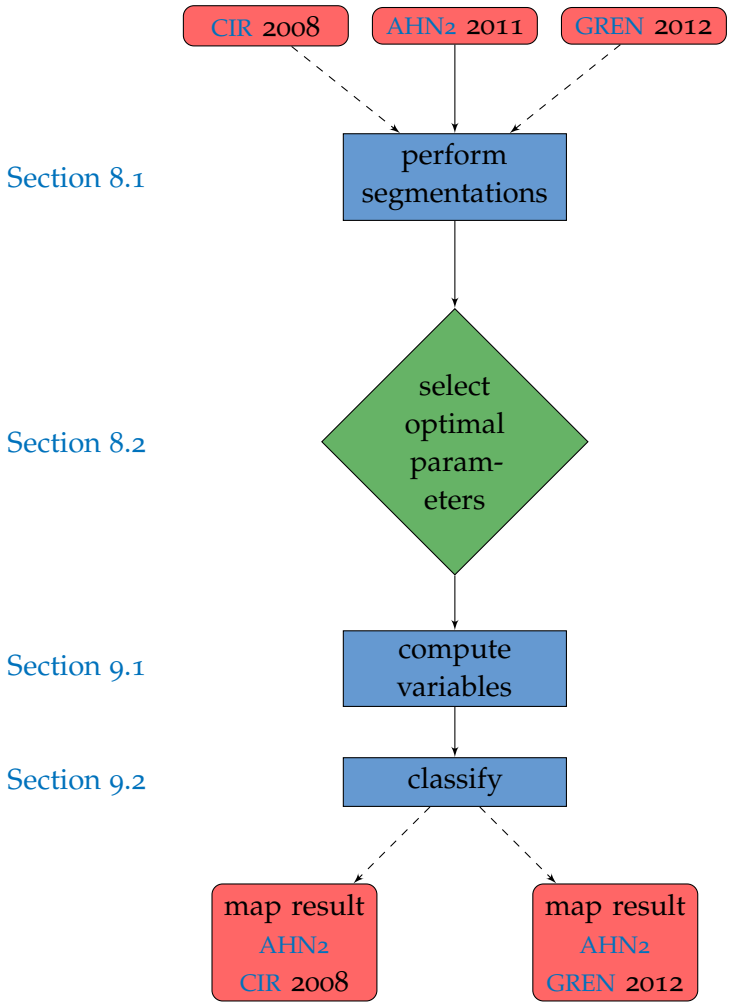


Figure 7: Flowchart from data to classified results using either the color infrared data of 2008 or the green red red-edge and near-infrared data of 2012

The CIR image of 2008, which is nationally available, is used to test the method on the full study area. The GREN dataset only covers a

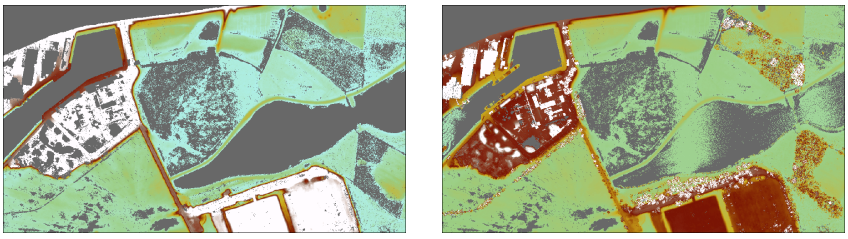
subset of the study area. This dataset is used to test the robustness of the method to different datasets.

Ancillary data was used in the selection of segmentation parameters and to pre-process data. To fill up holes in the AHN2 products, digital topographic information of the Dutch kadaster's TOP10NL dataset was used. The segmented objects were compared to a pilot vegetation map product of Rijkswaterstaat and digitized reference plots.

7.1 AHN2

As the first country in the world, the Netherlands is fully covered by airborne laser scanning altimetry in the Actueel Hoogtebestand Nederland (van der Sande et al., 2010). The sequel of this project is AHN-2 and is currently carried out to be completed in 2012. This dataset has an average point density of 10 points per square meter.

Accuracy assessments of a strip in the South-West of the Netherlands showed vertical offsets of up to 4 cm and horizontal offsets of up to 34 cm (van der Sande et al., 2010). Filter algorithms are applied to filter out all the points which are not part of the ground area. Both the remaining points as the filters points are used to construct a grid with a spatial resolution of 0,5 m. Making a digital elevation model Digital Elevation Model (DEM) and a digital surface model Digital Surface Model (DSM). During classification an additional layer is used to characterize the vegetation height, the Digital Canopy Model (DCM). This is created by subtracting the pixel values of the DEM of the DSM.



(a) DEM (b) DSM

(c) Interpolated DEM (d) Interpolated DSM

Figure 8: Digital elevation and surface models as provided and after interpolation, gray values indicate *no data* pixels

Figure 8a and Figure 8b show a small part of the DEM and DSM as provided. As can be seen, these datasets have missing pixels (gray pixels in Figure 8a and Figure 8b). To ensure a high degree of data quality, many points are filtered out during the creation of the datasets. Especially with scattering surfaces such as water, there are many missing pixels. In order to use these layers in the segmentation process, the holes have to be filled. This has been done in ESRI's ArcGIS.

A simple interpolation gives unrealistic results at boundaries between land and water. Therefore land and water areas were handled different. Polygons of the water bodies part of the TOP10NL dataset were used to assign the lowest value of the pixels within the polygons to the water bodies.

Pixels surrounding the other holes were converted to points and interpolated using natural neighbors. Computing time has been reduced by selecting only two rows of pixels surrounding a hole. The interpolated values were only used where no original data was present. A subset of the results are depicted in Figure 8c and Figure 8d.

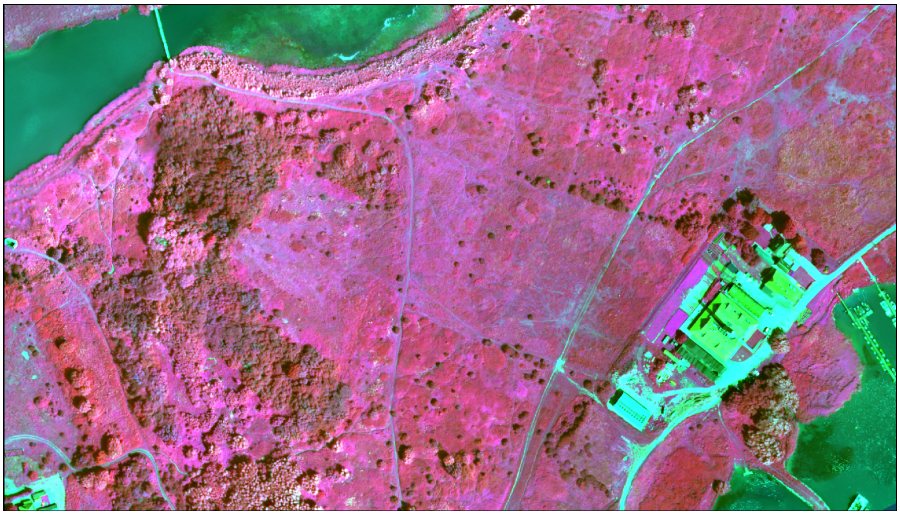
7.2 AERIAL IMAGES

CIR aerial images have been made by Cyclomedia commissioned by the Kadaster in 2008. They are made yearly and available in four bands: Blue, Green, Red and Infrared. Spatial resolution is 0.25 cm with a JPEG 2000 compression. A false color subset is shown in Figure 9a.

The GREN dataset used to test the method's robustness to data is a completely different dataset as depicted in Figure 9b. It does not contain a blue band but instead measures in the red-edge region of the spectrum. It has a higher resolution and a greater bit depth of 16-bit compared to 8-bit of CIR. During the processing by the provider, the contrast of the vegetation is enhanced which creates saturated pixels in high reflectance areas, as visible in the build-up area in the lower left corner of Figure 9b. This makes for a more textured image with higher contrast in vegetated areas.

Figure 10 also shows the increase in vegetation between 2008 and 2012 in this part of the study area. With greater areas covered by bush and more individual shrubs in the grass areas. This shows the relevancy of an increased mapping frequency. But it also makes comparing classification results between the different datasets difficult. Changes in the distribution between different classes are not necessarily due to the use of different data and a different mapping scale but may also be due to actual change of the vegetation.

(a) 2008



(b) 2012

Figure 9: False color aerial images of the 2008 and 2012 datasets

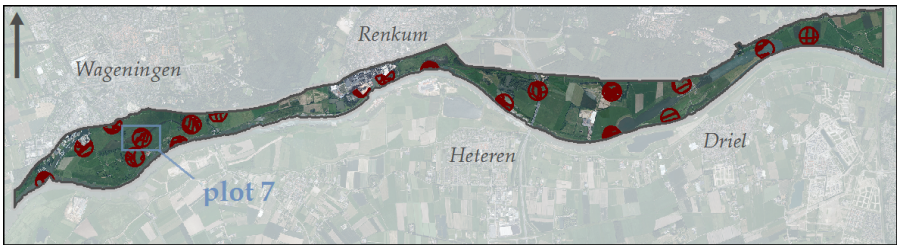
7.3 REFERENCE PLOTS

Two types of reference datasets were used in the selection of optimal segmentation parameters. One reference dataset, Figure 10a is acquired according to the methodology of Möller et al. (2007). The other dataset, Figure 10b is a polygon map made by Rijkswaterstaat as a pilot to map vegetation in the Dutch river floodplains.

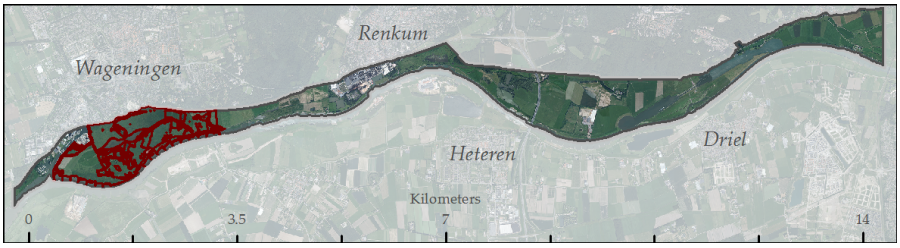
Manually digitized circular reference plots

According to the method of Möller et al. (2007), 19 randomly selected points were used to create circular reference plots with a radius of 150 m.

Figure 11 shows an example, the digitization of plot number seven.



(a) Manually digitized circular reference plots according to Möller et al. (2007), indicating the location of plot 7



(b) Rijkswaterstaat's pilot vegetation legger (Knotters, 2011)

Figure 10: Reference polygons used to select segmentation parameters, as positioned in the study area against the true-color 2008 background

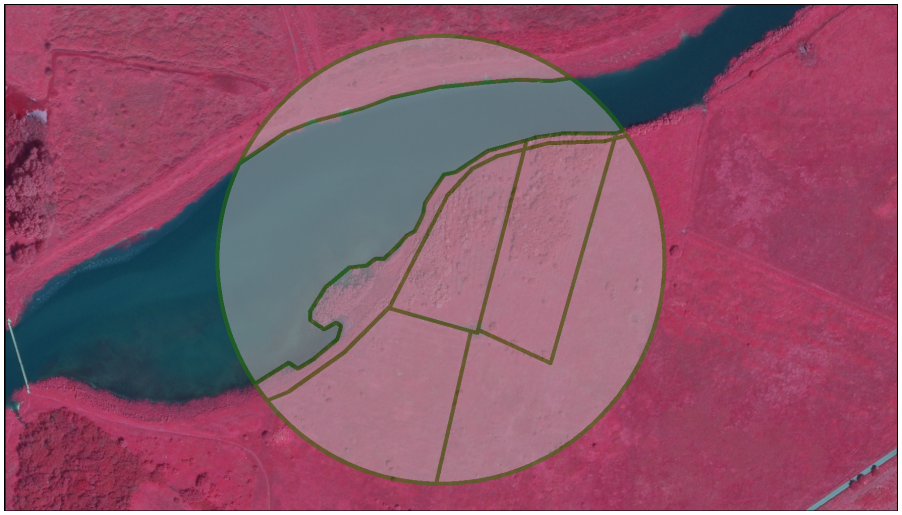


Figure 11: Plot 7 of the 19 circular reference plots used to select segmentation parameters, with false-color aerial image of 2008 as background

The rectangular objects are the result of elevation differences in de DEM. Not all plots were completely situated inside the study area and some covered extensive built-up areas which were deleted from the reference dataset. Many off the data layers have a high variability in pixel values in built-up areas. This heterogeneity results in a high amount of separate objects after segmentation which affect the result of the selection. Since the selection is aimed at mapping the

vegetation, these areas are considered of lower importance and have therefore been removed.

The other areas have been manually digitized with the CIR of 2008 and the AHN2 of 2011 as a background. Totaling 206 objects covering 103 ha, around 11% of the study area.

Plot 7 of the 19 circular digitized circular reference plots, is consistently used as an example in the remaining of this research. It is positioned in the middle of the western part of the study area, part of the legger and contains both naturally shaped as rectangular objects

Rijkswaterstaat's pilot vegetation legger

Figure 12: Rijkswaterstaat's pilot vegetation legger (Knotters, 2011) used to select segmentation parameters, with true-color aerial image of 2008 as background

Rijkswaterstaat is in the process of altering the monitoring of river floodplain vegetation. To provide insights in the effects of a different monitoring on their hydrological models they manually mapped a pilot area between the Nederrijn and the town of Wageningen on the basis of stereo-aerial images of the summer 2008 (Knotters, 2011).

Figure 12 shows the outlines of the pilot study which lie within the study area of this research. The 294 objects are being used as reference polygons.

8

SEGMENTATION BY FNEA

The optimal segmentation result has been selected of 25 segmentation results. The 25 segmentations are unique combinations of five shape settings and five scale settings. All other parameters influencing the identification of objects are kept constant, the effects of these parameters has not been analyzed in this study.

8.1 SEGMENTATION

Segmentation is performed in Definions eCognition developer 8 (Baatz and Schäpe, 2000). It involves two steps. A multi-resolution segmentation and subsequently a spectral difference merge.

Data layers

Table 1 shows the data layers as used with the corresponding weights assigned to it. The GREN image, used in this study does not contain a blue band and has been substituted by the red-edge band.

Table 1: Data layers and associating weights used in segmentation

TYPE	LAYER	WEIGHT
Aerial images	Blue or red-edge	1
	Green	1
	Red	1
	Infrared	1
AHN2	DEM	2
	DSM	2

Only primary data and no secondary layers such as NDVI or DCM have been included. Adding more layers will make the already heavy computations of the segmentation heavier. Second, these layers essentially contain the same information as the layers which are included and are believed to have no beneficiary effects on the segmentation.

Equal weights have been assigned to the CIR and AHN2 data. The spectral information of the CIR has the same influence to the increased heterogeneity of a merge of two objects as the structural information of the AHN2. Altering these weights might have a positive effect on the segmentation but this has not been pursued in this study.

Multi-resolution segmentation

The FNEA algorithm is implemented in eCognition's tool called *multi-resolution segmentation*. This algorithm creates the initial objects.

Table 2 shows all the parameter settings used in this study. Compactness is left constant. All the other combinations of shape and scale have been run.

Table 2: Parameters used in the segmentation

DATASET	SCALE	SHAPE	COMPACTNESS
CIR + AHN2	50-450 incremented with 100	10%-30% incremented with 5%	50%
GREN + AHN2	50-950 incremented with 100	10%-30% incremented with 5%	50%

Spectral difference merge

For more information about the algorithm of FNEA, please refer to Section 4.1

Multi-resolution segmentation creates the objects. FNEA creates objects around seed points. As a result even homogeneous areas such as grass fields or the river, become multiple objects. To make the objects more meaningful, these objects are merged on the basis of color heterogeneity only. This is done via a tool called *spectral difference segmentation* included in eCognition. Color here stands for the digital numbers in all of the data layers and therefore the weights of Table 1 are used here as well. After visual inspection, the allowed spectral heterogeneity has been set at 2.

8.2 SELECTION

Möller et al. (2007) methodology is explained in-depth in Section 4.5

The selection of the optimum segmentation is done with a method similar to that of Möller et al. (2007). The difference between the comparability measure of this study and the comparison index used by (Möller et al., 2007) is only computational. They reduced the dimensionality due to the combination of the position and size by a k-means algorithm to one index between 0 and 1. To simplify the implementation, the measure used in this study did not incorporate such a data reduction method but simply added-up the metrics.

The metrics RA_{SO} and RP_{SO} are included in eCognition (Definiens, 2009). These metrics are calculated for every intersection polygon of segmented object and reference objects with either the segments or the reference polygons as *super* objects. Every objects RA_{SO} and RP_{SO}

to both the reference as the segments are exported to R (R Core Team, 2012).

Within R the RA_{SO} and RP_{SO} are added up and weighted with the polygons area.

$$C = \sum_{i} \frac{A_i(RA_{SO_i} - RP_{SO_i})}{A}$$

where

C	The average comparability
i	The intersected polygon
RA_{SO}	Relative area to super object, with either segment or reference object
RP_{SO}	Relative position to super object
A	The total area

This gives a C of the segmentation intersection to the reference (over segmentation) and to the segmentation (under segmentation). The optimum segmentation parameter is the scale and shape where under and over segmentation are equal and the comparability is highest. This point is selected visually from plots of over and under segmentation.

See Section 4.3 for more information on the concepts over and under segmentation and see Section 4.5 for a description of RA_{SO} *and* RP_{SO}

9 CLASSIFICATION BY RANDOM FOREST

The objects have been classified into ten vegetation structure classes on the basis of 181 variables (Table 4) and 113 observations (Table 3). Classification is done by RF (Breiman, 2001) which is implemented in the statistical language R (R Core Team, 2012) via the the package *randomForest* (Liaw and Wiener, 2002).

9.1 VARIABLES

Table 4 shows all the variables used in classification. The variables have been computed using eCognition for all the objects of the segmentations which have been classified. Gray level co-occurence matrices and gray level difference vectors have been calculated for all the data layers used in the segmentation.

9.2 CLASSES

The classes are defined according to Figure 13. It is similar but not completely to the class definition of the PI map of Rijkswaterstaat. Some classes were left-out because these do not occur in the study area, others have been merged to get to a class definition which is simpler to implement and more compliant to Rijkswaterstaat's pilot vegetation legger. For a list of the classes used in this study and those used by Geerling et al. (2009) and by Rijkswaterstaat in the vegetation legger and the PI map, please see Appendix A.

9.3 TRAINING DATA

The forest is based on 113 training objects. These objects have been manually selected and labeled according to the class definition of Figure 13. Labelling has been done manually on the basis of visual interpretation of the RGB layers of the CIR dataset. The 113 training objects are from a segmentation of the CIR and AHN2 with scale 150 and shape 25%.

The distribution of the different classes within the training set is given in Table 3. The abundance of some classes was low in the study area which made it impossible to get an even amount of training objects for every class.

Table 3: Training object and validation point distribution

CLASS	TRAINING OBJECTS	VALIDATION POINTS
water	18	5
pioneer	7	3
sand	9	4
built-up	22	4
grass	17	9
field	6	1
herbaceous	13	6
forest	10	3
bush	9	6
orchard	2	2

9.4 VALIDATION DATA

The training data is used in conjunction with the OOB cross-validation of RF to estimate the accuracy of the forest, additionally a field validation has been performed. The distribution of the field observations used to perform the validation is also given in Table 3.

Five objects of every class of the classified result depicted in Appendix C were selected randomly. With the exception of *orchard*, since there were only two objects classified as such. The center of gravity of every object has been determined and visited using a conventional Garmin eTrex handheld GPS. One area predicted to be *pioneer*, one to be *sand* and two *built-up* points could not be visited due to access difficulties. These were removed from the dataset.

Table 4: The 181 object features used as variables in the classification. Unless indicated different, the features have been calculated for all the data layers of the color-infrared and the AHN2

CATEGORY	DESCRIPTION
Spectral	Mean, Min, Max, Standard deviation, Ratio and Skewness of the four CIR bands
	Intensity, Hue and Saturation transformation of the RGB values
	Brightness, as in the total of the mean values of CIR
Topographic	Mean and Standard deviation of DEM, DSM and DCM
Textural	GLCM Homogeneity
	GLCM Contrast
	GLCM Dissimilarity
	GLCM Entropy
	GLCM Standard deviation
	GLCM Correlation
	GLDV Angular Second Moment
	GLDV Entropy
	GLDV Contrast
	GLDV Mean
Geometric	Area
	Length
	Width
	Compactness
	Rectangular fit
	Border index
	Border length
	Shape index
	Density
	Main direction asymmetry
	Elliptic fit
	Length/Thickness
	Length/Width
	Relative border to image border
	Roundness

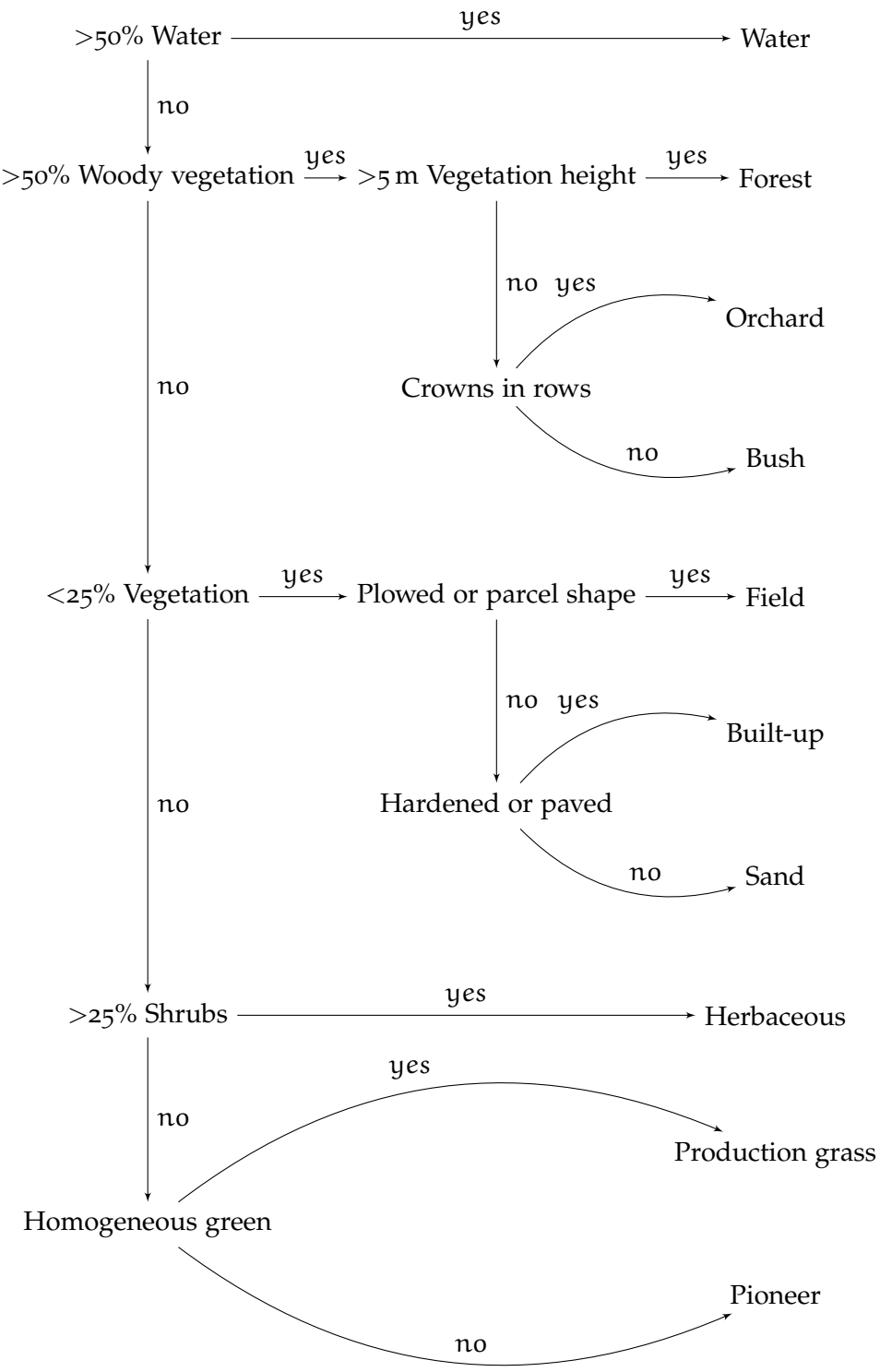


Figure 13: Class definition key adapted from Rijkswaterstaat (1998)

Part III

RESULTS

The automated mapping methodology follows the same two steps a human mapper would. The boundaries of distinct objects are identified and sequentially classified to the most probable vegetation structure class. Since the automated method lacks the flexibility of identifying objects of the right scale and shape straight away, this step is done many times. The resulting boundaries that show the most resemblance to reference boundaries are chosen for classification. Following, manually classified-objects form the basis for many automatically created decision trees which together choose the most probable class to assign. As such the structure of the floodplain vegetation is mapped in a reproducible, time-consistent manner, with maintained precision, flexibility and overall map quality.

10 SEGMENTATION

The segmentation parameter selection method has been tested on the 19 circular reference plots and the CIR 2008 and AHN2 2011 dataset. The results are presented in three levels of abstraction. Two abstract levels in graphs, both the aggregation as the individual values of the 19 plots in Figure 14. To illustrate the behavior of the segmentation and the selection method, one of the 19 circular plots is separately presented in graphs (Figure 15) and concrete graphical form (Figure 16).

The chapter continues by presenting results of the selection with the legger, an alternative reference dataset, both abstract (Figure 17, Figure 18) as graphical (Figure 19) and the results of an alternative source dataset, the GREN, abstract (Figure 20, Figure 21) and graphical (Figure 22).

10.1 MANUALLY DELINEATED REFERENCE PLOTS

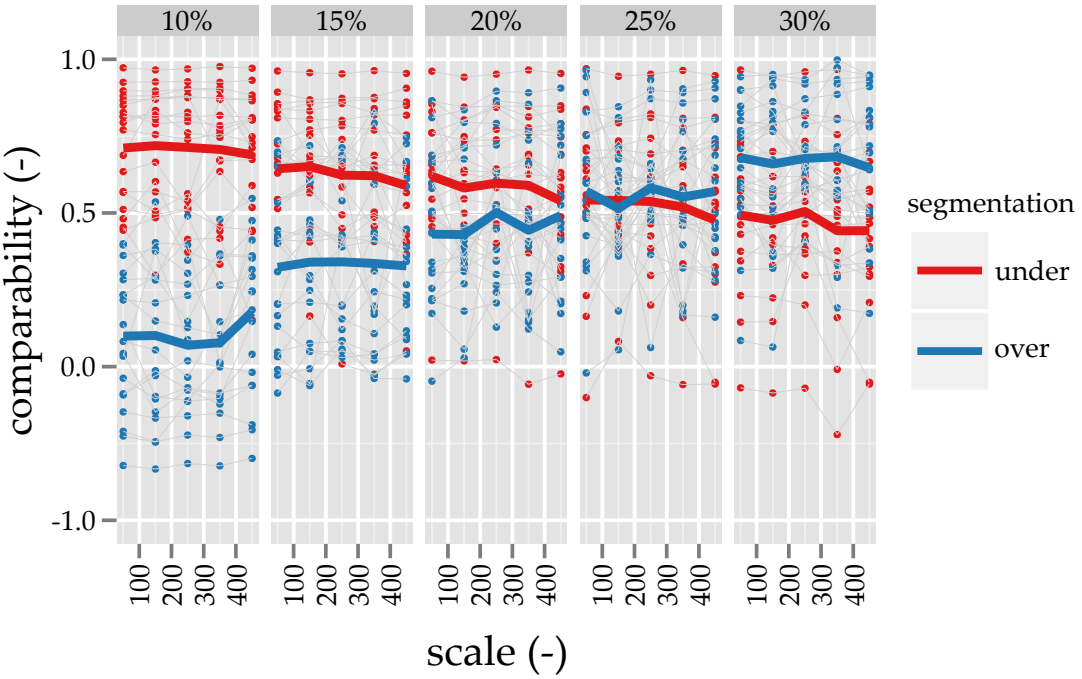


Figure 14: Average comparability between circular reference plots and segmentation result for five shape settings between 10% till 30% (indicated at the top) and scale between 50 and 450 (indicated at the right) of the 2008 color-infrared and 2011 AHN2 dataset

The total allowed heterogeneity of the objects was set between 50 and 450, incremented with 100 for every segmentation. This was done

five times with emphasis on shape of the objects between 10% till 30%. In total 25 segmentations were performed.

To select the segmentation with the most resemblance to the desired size and shape, the objects were compared to the manually digitized reference plots. This resulted in comparability as depicted by Figure 14. The dark bold lines show the average comparability.

Bold red and blue lines

The relative size and position are quantifications of the discrepancy between the reference polygons and the segments. The concept is explained in Section 4.5 and the implementation used in this study in Section 8.2

The segmented objects have been intersected with the 206 reference objects of the 19 plots distributed over the study area. The average relative position of the gravity centers and the average size of the intersections are compared to the segmented objects to get the under segmentation, indicated in red in Figure 14. The blue line in Figure 14 shows over segmentation: the comparability between the intersections and the reference polygons.

As the total allowed heterogeneity increases, the segmented objects gradually grow larger such that the intersection resembles more the reference polygons. The same holds the other way around. When the total allowed heterogeneity increases, the segmented objects gradually grow larger, such that the comparability between the intersection and the segmented objects decreases.

As can be seen in Figure 14 the lines intersect around the parameter setting with a scale of 150 and shape 25%. This means that the segmentation of this dataset with scale 150 and shape 25% shows the most resemblance to the reference plots in size, shape and position.

Red and blue dots

19 circular reference plots with a radius of 150 m were digitized. The plots were randomly selected. In total 206 objects were distinguished with an average size of 0.5 ha.

The over and under segmentation of the 19 individual plots is depicted in Figure 14 by red and blue dots, connected by gray lines. Appendix B shows the 19 plots individually. Even with scale and shape settings that result in almost equal over and under segmentation, the deviation between the plots remains large.

Segmentation parameters that work for certain reference plots in certain areas of the study area, do not for others. While the 150 scale and 25% shape setting functions well on average, some areas might consist of objects which are larger or smaller and have a very different shape than the pursued objects of the manually drawn reference plots.

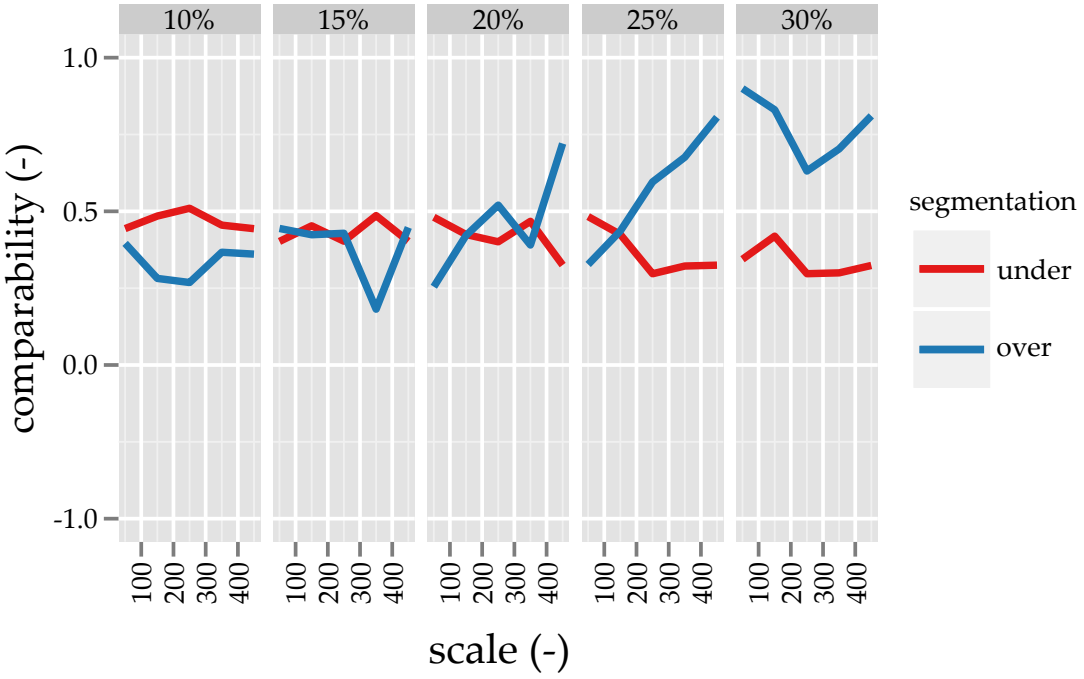


Figure 15: Comparability between circular reference plot seven and segmentation results for five shape settings between 10% till 30% (indicated at the top) and scale between 50 and 450 (indicated at the right) of the 2008 color-infrared and 2011 AHN2 dataset

Reference plot 7

Figure 11 of Section 7.3 shows plot 7 on a false-color background

One of the gray lines connecting red and a blue dots of Figure 14 and in Appendix B represents reference plot 7, these lines are depicted separately in Figure 15. To concretely illustrate the relationship between the segmentation parameters and the comparability, plot number seven is taken as a representative case. As can be seen in Appendix B, the behavior of plot seven resembles that of the averages in Figure 14 best, compared to the other plots depicted in Appendix B.

As can be seen in Figure 15, plot seven follows a clear increase in over segmentation and decrease in under segmentation with higher scale setting for a shape setting of 25%.

Figure 15 can be seen as an abstraction of the discrepancies between the reference polygons and the generated segments as depicted in Figure 16. Every square of Figure 16 shows the result of one combination of segmentation parameters and corresponds to one point on the blue and red lines of Figure 15. As such Figure 16 gives a representative insight in how polygons created with optimal and sub-optimal parameter settings visually compare to the reference plots. With a lower scale setting there are more and smaller objects and with less emphasize on shape, these objects are more oddly shaped.

It can be seen that some objects contained in the reference plot are clearly identified from the source data, while others are not. This is

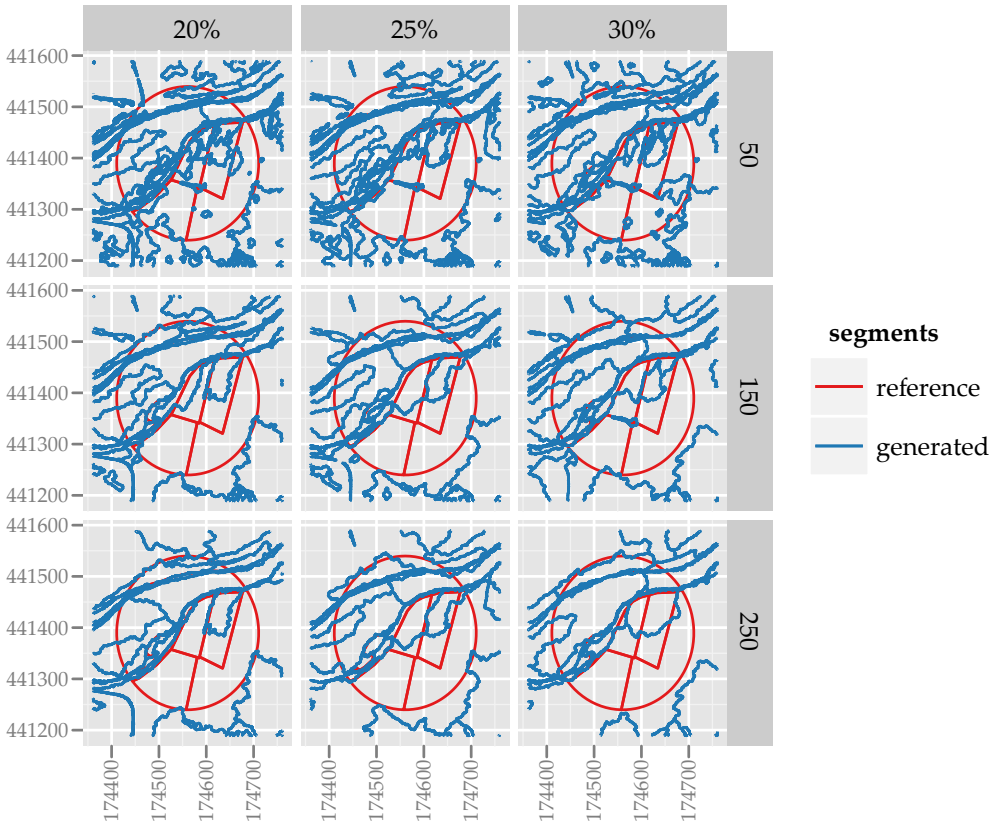


Figure 16: Representation of segmentation results of 2008 color-infrared and 2011 AHN2 dataset with optimal (center, shape 25% and scale 150) and sub-optimal (periphery) parameter settings with the circular reference plot 7 as reference

the case for every shape and scale setting. As in Figure 15, Figure 16, shows that the 150/250 scale and 25% shape gives the most comparable result. Visually this seems due to the two square reference-objects in the middle of the plot, which are reasonably bounded by the segmentation results of 150/250 and shape 25%.

All results contain a lot of small noise objects and all except the 150/250 scale and shape 25% do not segregate between two square reference-objects.

10.2 VEGETATION LEGGER AS REFERENCE

Instead of using specifically digitized objects for use as reference, it might be possible to use an existing map product as reference. This is investigated by substituting the reference plots by Rijkswaterstaats pilot vegetation legger.

As can be seen in Figure 17 the complete Rijkswaterstaat pilot vegetation legger as a reference does not give satisfactory results.

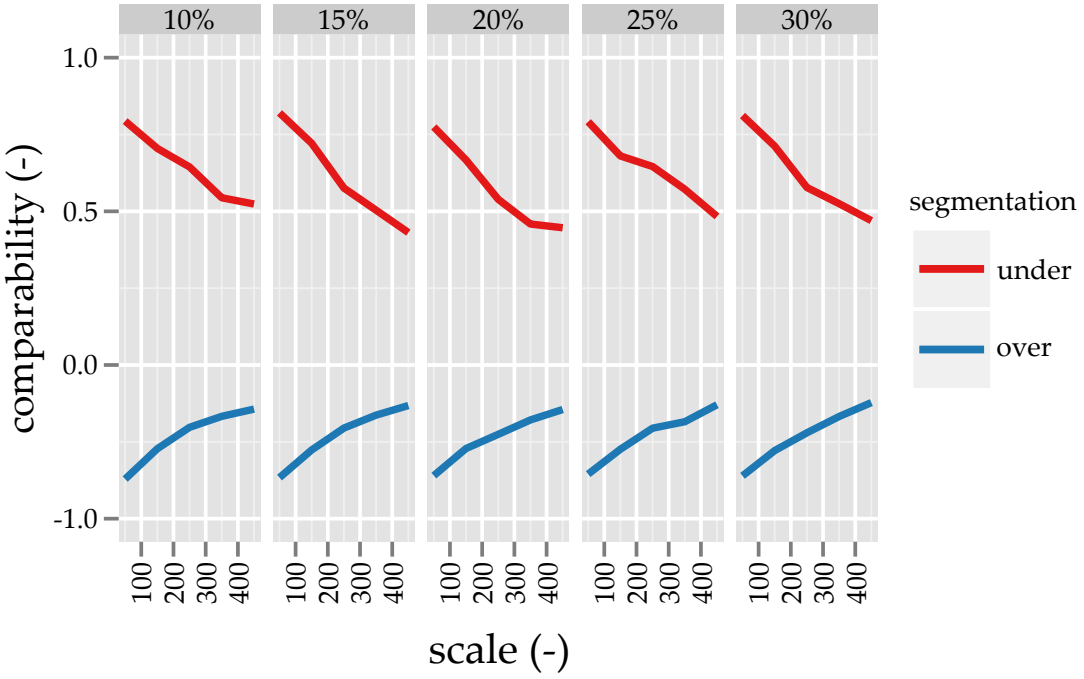


Figure 17: Comparability between the full vegetation legger and segmentation result with shape 10% till 30% and scale between 50 and 450 of the color-infrared 2008 and the AHN2 2011 datasets

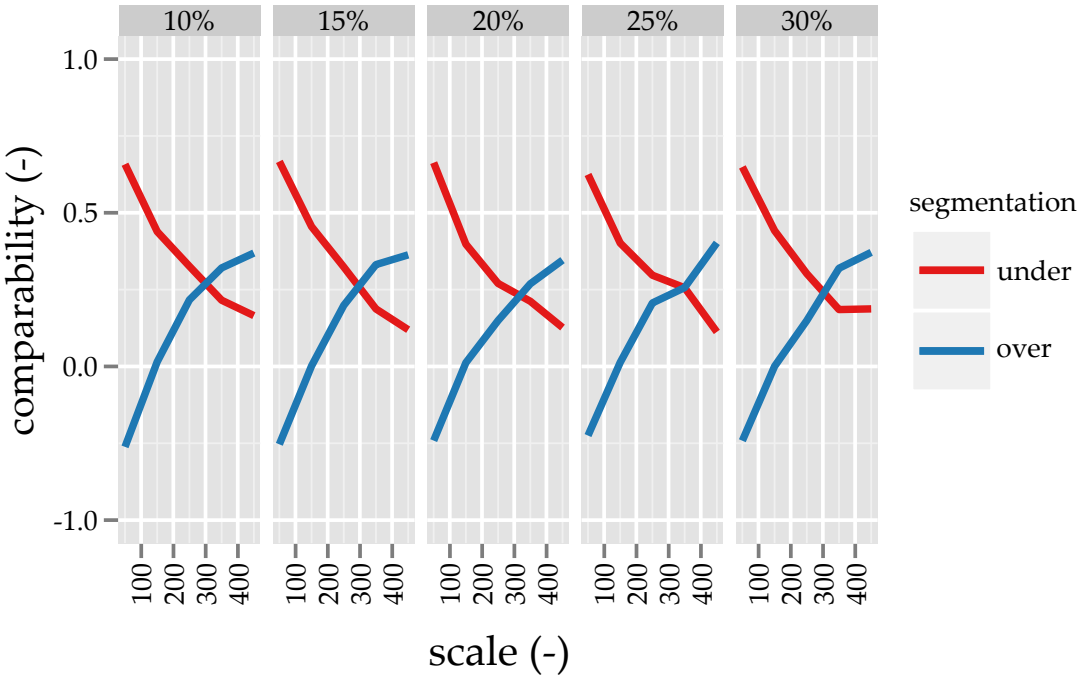


Figure 18: Comparability between the vegetation legger *the largest polygon* and segmentation result with shape 10% till 30% and scale between 50 and 450 of the color-infrared 2008 and the AHN2 2011 datasets

Figure 18 shows the same reference plots excluding the large natural grass polygon. Figure 18 shows that the legger without this poly-

gon gives clearly increasing over segmentation and decreasing under segmentation with increasing scale. This trend is observable for all the five shape settings. Although the sensitivity to shape is low, a low shape setting seems to be favorable. The segmentation of this dataset with scale 300 and shape 10% shows the most resemblance to Rijkswaterstaats pilot vegetation legger in size, shape and position (Figure 18). Note that this holds true for the area which is mapped in the pilot project, objects in the rest of the study area might be less comparable to the size and scale of the vegetation legger.

Representative area around plot 7

Nine segmentation results of the same area around plot seven are depicted in Figure 19. Figure 19 clearly shows that a higher scale setting gives more comparable results than low scale settings. It is also visible that a low shape setting leads to a better identification of the small square areas in the middle of this subset.

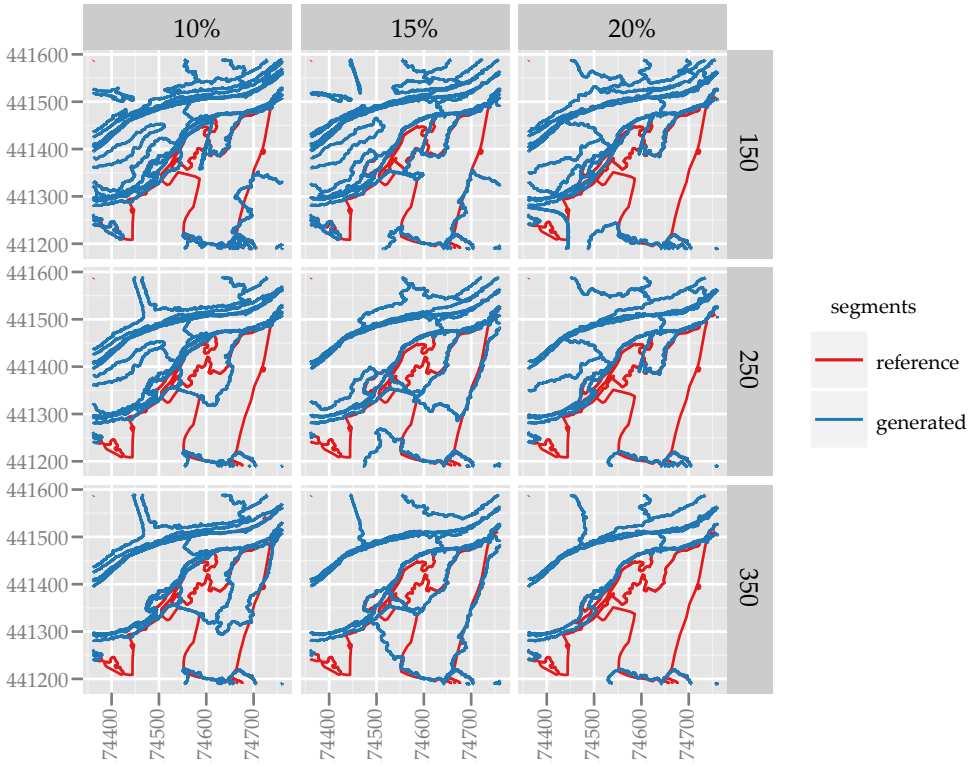


Figure 19: Representation of segmentation results of 2008 color-infrared and 2011 AHN2 dataset with optimal (shape 10% and scale 250/350) and sub-optimal parameter settings with the Rijkswaterstaat pilot vegetation legger as reference

Differences between the legger and the manual reference

The Rijkswaterstaat pilot vegetation legger consist of very connected objects. This gives large and oddly shaped areas. The generated segments are generally a lot smaller, even with a high scale setting. Figure 14 and Figure 18 show that comparability at the intersection point is lower for the legger than the manually delineated reference plots.

10.3 AN ALTERNATIVE DATASET

The previous section showed results of the comparison between segmentation results with varying parameter values and two different maps used as reference. This section shows results of the comparison between reference plots and two different datasets, the CIR from 2008 and the Unmanned Aerial Vehicle (UAV) GREN image of 2012. Both the CIR as the GREN were used in conjunction with the AHN2 dataset since there was no substitute for structural information of 2012.

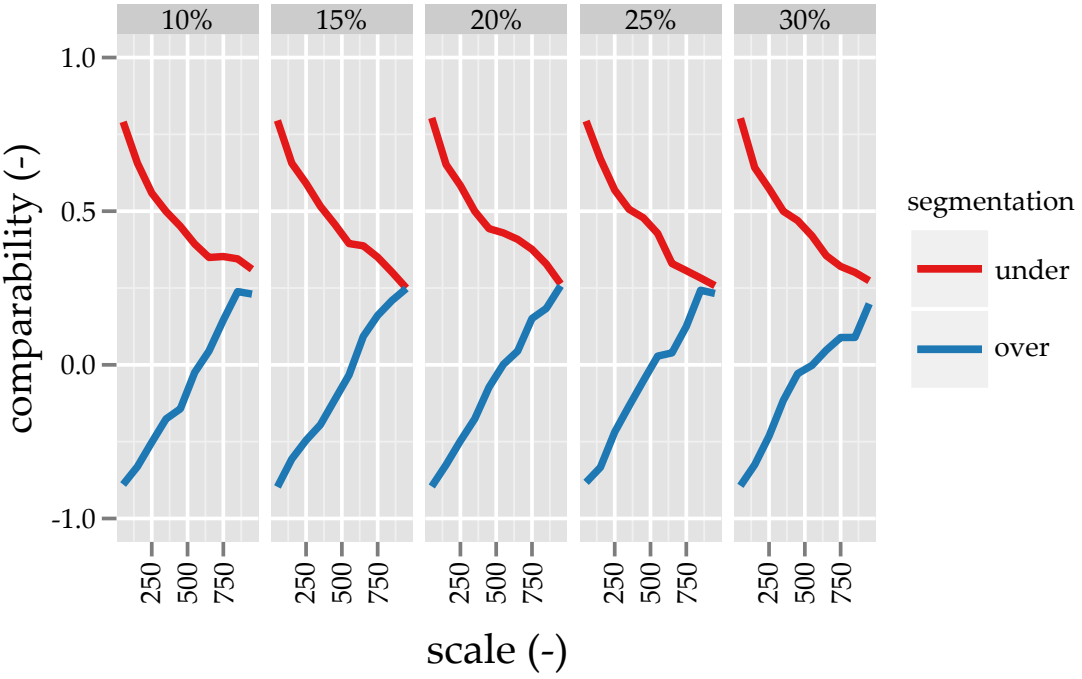


Figure 20: Comparability between the vegetation legger *excluding* the largest polygon and segmentation results with shape 10% till 30% and scale between 50 and 950 of the AHN2 and the UAV image

Segmentation results of the GREN image and the AHN2 data, with varying scale and shape settings are compared to 170 objects of the pilot vegetation legger which lie within the area covered by the GREN image.

As can be seen in Figure 20 the over and under segmentation of the GREN image in comparison to the vegetation legger equalize at a scale setting of 950 for the 15%, 20% and 25% shape. This is a higher

scale value than the optimum value of 250, deducted from Figure 18 for the CIR image.

The same insensitivity to the shape parameter for the GREN image is observed in Figure 20 as for the CIR in Figure 18.

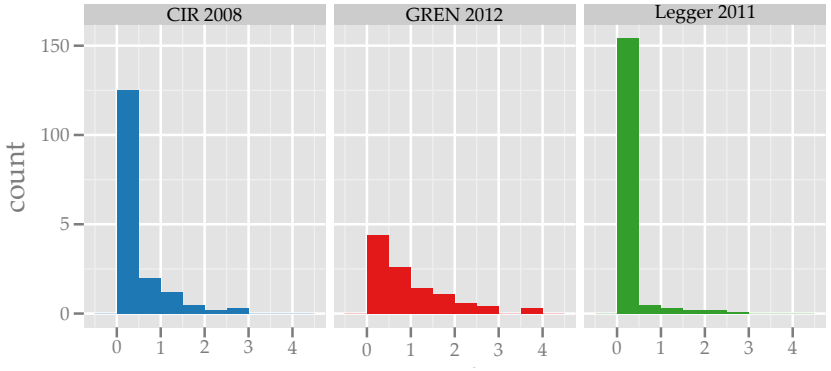


Figure 21: Histogram of object size for the segmentation of color-infrared 2008 with scale 250 and shape 10%, green red red-edge 2012 with scale 950, shape 15% and the manually delineated Rijkswaterstaat vegetation Legger *excluding* the largest polygon

Figure 22 shows the segmentation results of the parameter values showing the optimum comparability to the reference plots for the area of the GREN 2012 image. The lowest plot shows the 170 manually drawn vegetation legger objects which are used as reference to find the optimum segmentation parameter values used to segment the GREN 2012 and CIR 2008 image, which are depicted in the above two plots of Figure 22. Ideally, all three plots would consists of objects of similar size and shape, differences between the top two would only be due to physical changes between 2008 and 2012. Visually it seems that there are less small objects present in the segmentation of the GREN image than the CIR image. Figure 21 confirms this.

Section 7.2 includes images (Figure 10) and a short description of the change in the part of the study area covered by the GREN image

Figure 21 shows the density of objects grouped on the size of the objects. It shows that size of the objects of the different datasets is reasonably comparable, with more small objects for the CIR image of 2008 than the GREN of 2012. Although the segmentation of GREN did not segregate the small objects of Legger well, the comparability at the intersection of the under and over segmentation of the GREN in Figure 20 and the CIR in Figure 19 are both around 0.25. This means that the scale, shape and position of the segmented objects of both images equally compare to the reference plot *on average*. Small objects on the GREN image are not identified well, but medium sized objects are identified better which *on average* gives an equal comparability to the objects of the Legger.

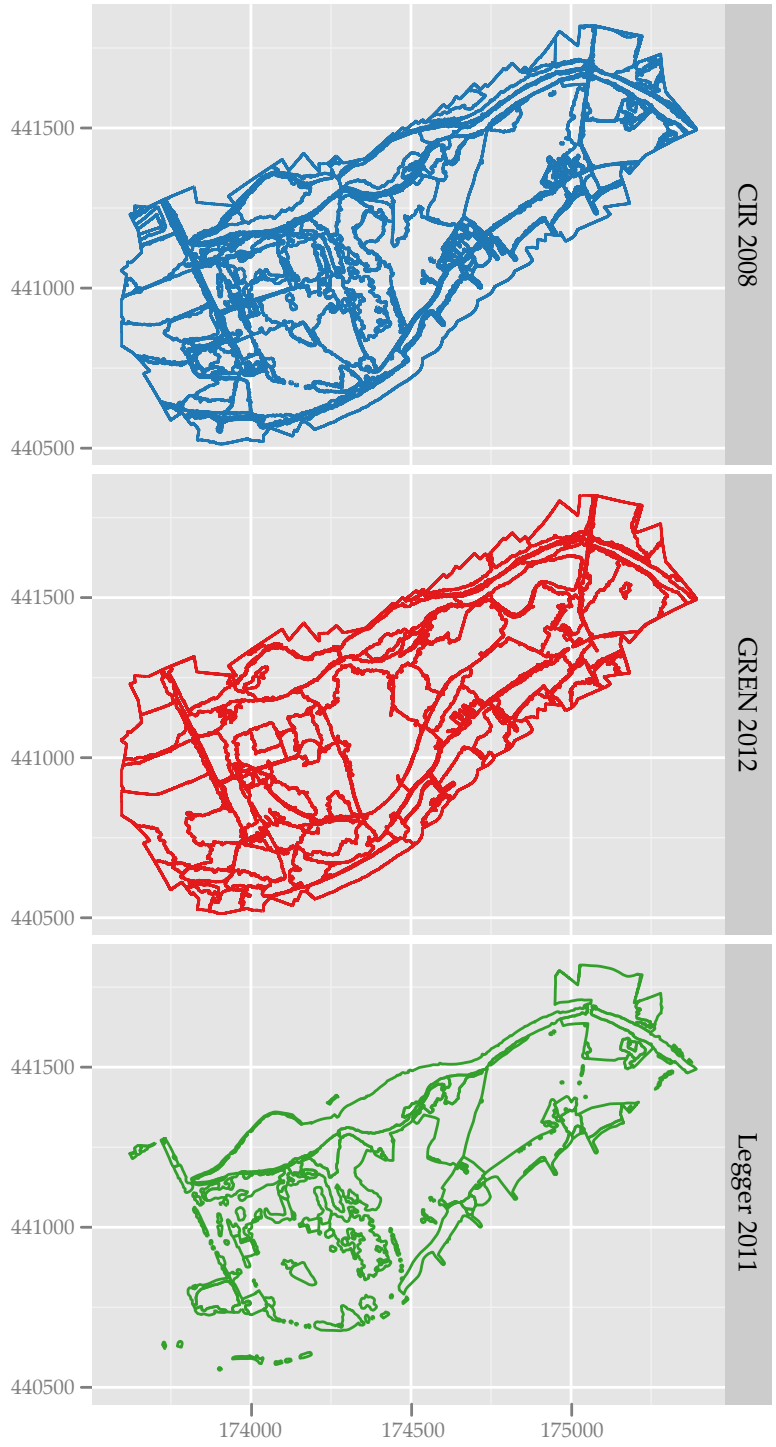


Figure 22: Visual representation of optimal segmentation of the CIR2008 with scale 250 and shape 10%, the GREN2012 dataset with scale 950 and shape 15% and the reference objects of the legger *excluding* the largest polygon

11 CLASSIFICATION

To investigate the effects of segmentations with different parameter settings and different datasets on the final map product, RF was applied on the objects. Additional information from RF is analyzed to further understand the function of the datasets in the classification and the workings of RF.

For a study wide classification result, please see Appendix C.

11.1 CLASSIFIER PERFORMANCE

A forest of 1000 trees with 15 of the 181 variables per tree has been constructed on the basis of 113 training objects. RF includes an estimate of the error of the forest based on a cross-validation with the OOB objects. The estimated error of the forest used to classify the objects in this study was 14.16%. This value showed no sensitivity to number of trees and number of variables of each tree.

Section 5.2 describes that the trees of RF are constructed on subsets and objects not part of a tree are considered OOB

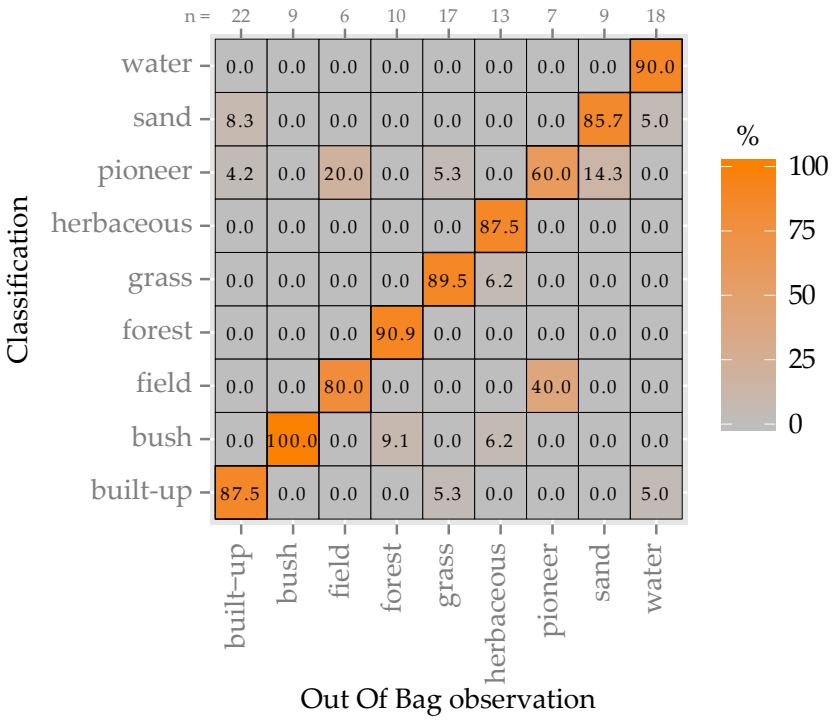


Figure 23: Heatmap of the confusion matrix of the random forest and the OOB objects of a the forest constructed with 1000 trees and 15 of the 181 variables deduced from the 2008 color-infrared and 2011 AHN2 data

A heatmap of the confusion matrix is given in Figure 23. Figure 23 shows that except for the classes *field* and *pioneer* all the classes are well distinguishable and the cross-reference of the OOB objects show that the performance of the forest is high. On average 85.84% of the OOB objects are predicted correctly.

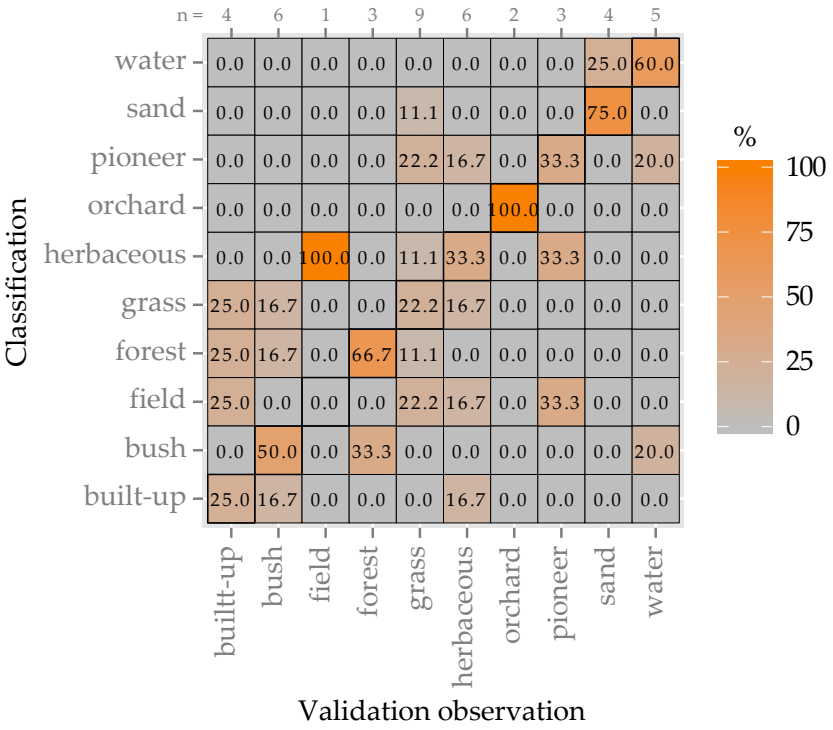


Figure 24: Heatmap of the confusion matrix of the random forest prediction and the field validation observations

The results of the field validation show much lower accuracies. On the basis of 44 randomly stratified validation object center points with classes as strata, the overall accuracy was estimated to be 43%.

Figure 24 shows the confusion matrix of the predictions and the actual class. The validation data is of September 2012 while the source and training data is of 2008 and 2010.

11.2 CLASSIFICATION OF OPTIMAL AND SUBOPTIMAL SEGMENTATION RESULTS

Figure 25 shows the result of a prediction of the forest on the area around plot number seven. The area is the same as in Figure 16, though slightly larger.

Figure 25 clearly shows that the same dataset may be used to map the area for different map scales. At a scale of 50, there are a lot of small patches of grass within the herbaceous area. These areas are all mapped as herbaceous for a higher scale setting. The same is observed for patches of forest within the bush areas.

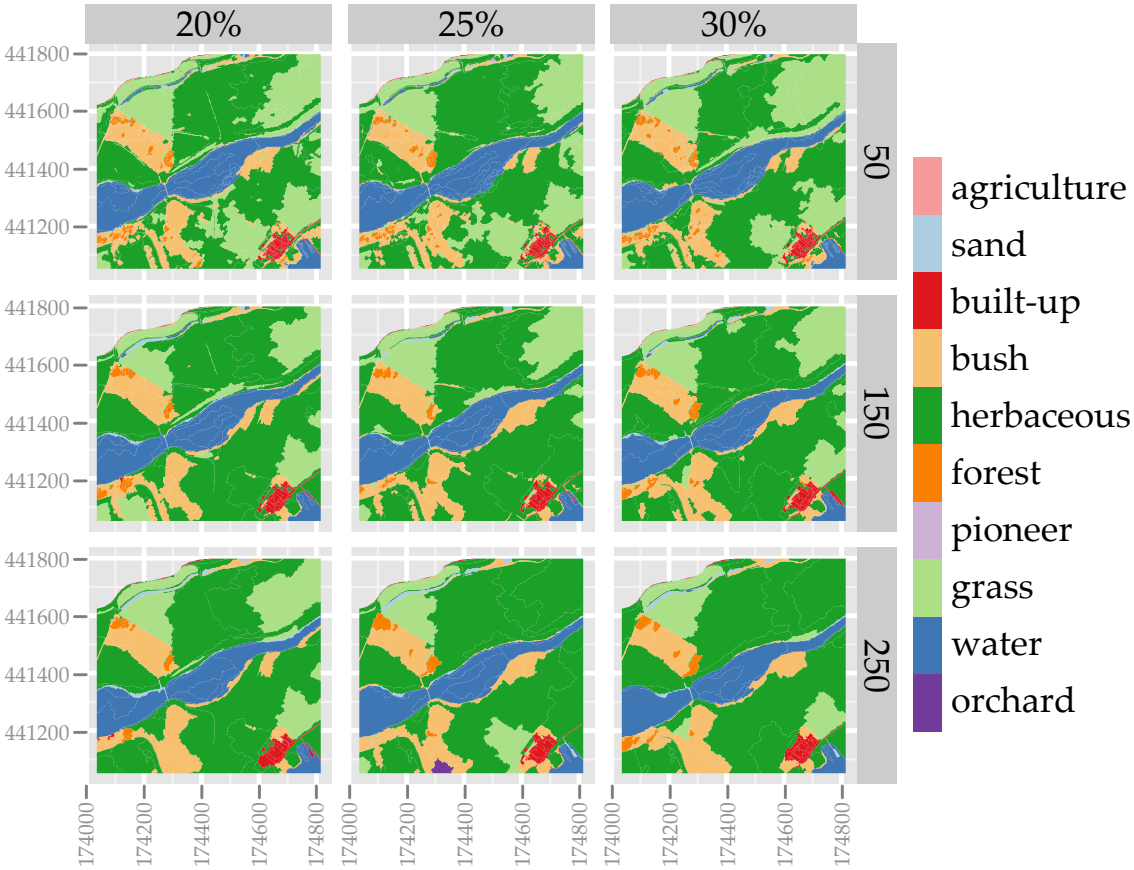


Figure 25: Representation of classification results of 2008 color-infrared and 2011 AHN2 dataset with optimal (center, shape 25% and scale 150) and sub-optimal (periphery) segmentation parameter settings

For the low scale setting it is also clearly visible that more emphasis to shape leads to more compact patches while a lower shape setting results in more irregularly shaped patches.

11.3 CLASSIFICATION OF AN ALTERNATIVE DATASET WITH OPTIMAL SEGMENTATION PARAMETERS

Figure 18 and Figure 20 showed that the CIR with scale 250 and shape 10% and the GREN with scale 950 and shape 15% are both on average the most comparable to the vegetation legger. This should result in map results with similar mapping scale. Figure 26 shows the classified results of two datasets in conjunction with AHN2 and the reference objects.

Scale and shape of the patches

Irrespective to the labels, the patches of the GREN image of Figure 26 seem larger, less detailed than those of the CIR image. The old brick-

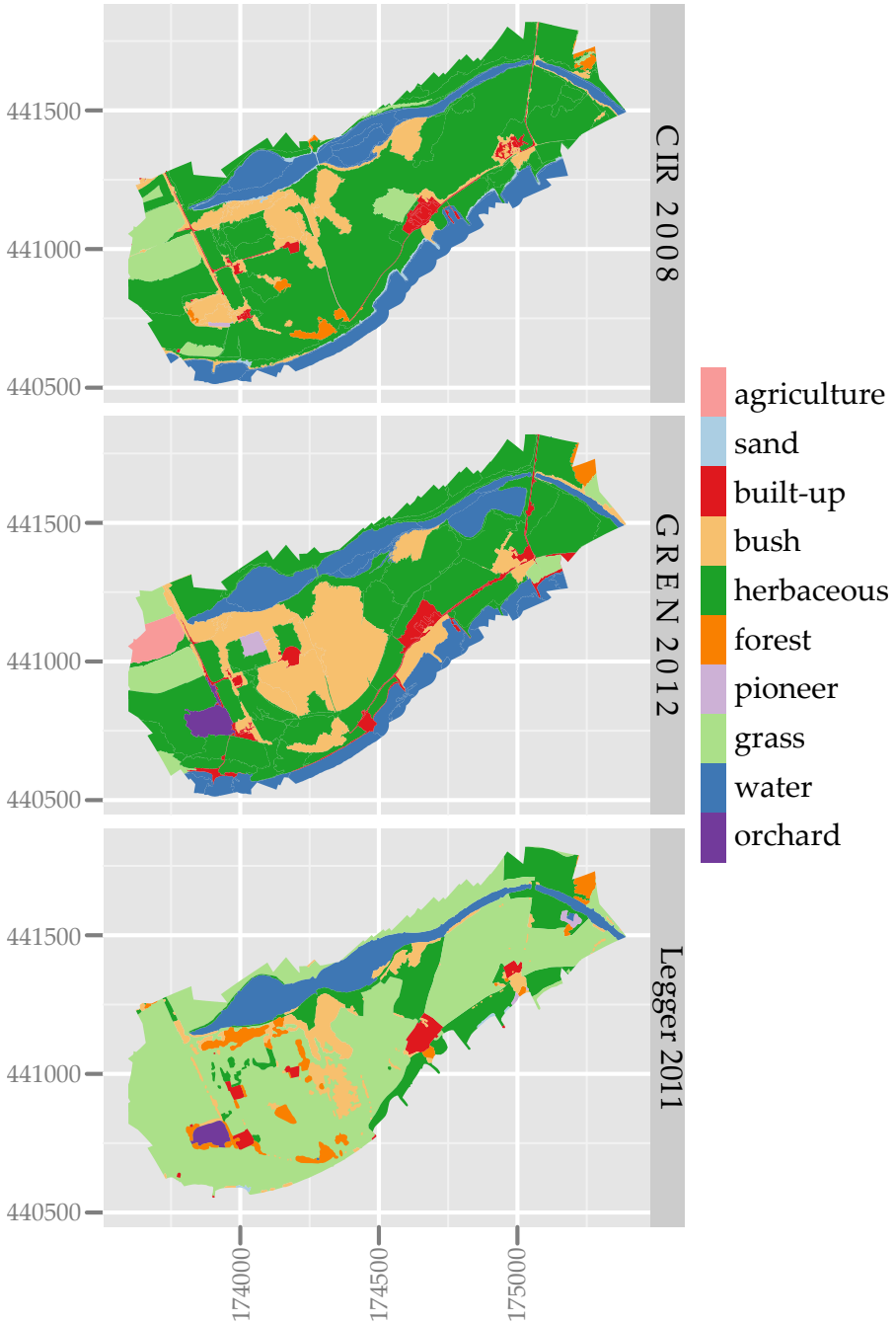


Figure 26: Classification results of the CIR 2008 dataset, the GREN 2012 and the rijkswaterstaat vegetation legger which has been used as reference in the segmentation

works, in the middle of the image near the river is labeled correctly as built-up in all three maps. Though, in the GREN image it has become one big object, while in the reference and the CIR image it is a built-up patch surrounded by bush and grass patches. On the contrary, small features such as houses in the west of the image and roads, are identified correctly in all three images. Overall it might be that both images

compare to the reference equally on average, concerning scale and shape. This is backed up by Figure 19 of Section 10.3 which showed that the GREN image contains less small objects, but more medium sized objects.

Class distribution

Since the GREN image contains different data (other spectral bands and a different bit depth), a new forest had to be constructed. The quality of this forest is limited by the small amount of objects and the thereby low amount of training data of 26 of the 110 objects in total.

It is clear from Figure 26 that Rijkswaterstaat definition of grass is wider than the definition used in this study. This is also illustrated by Figure 27 which shows the total area per class for every dataset. The biggest difference between the legger and the classified images are due to this broader definition of grass. Next, it is clear that the area has shrub encroachment, the bush area has become bigger and woodier. The Rhine has not been included in the legger, this explains the higher area of water for the classified images. Between 2008 and 2012, the water area has increased which is visible in the North-East area of Figure 26 and also in a slightly higher bar in Figure 27.

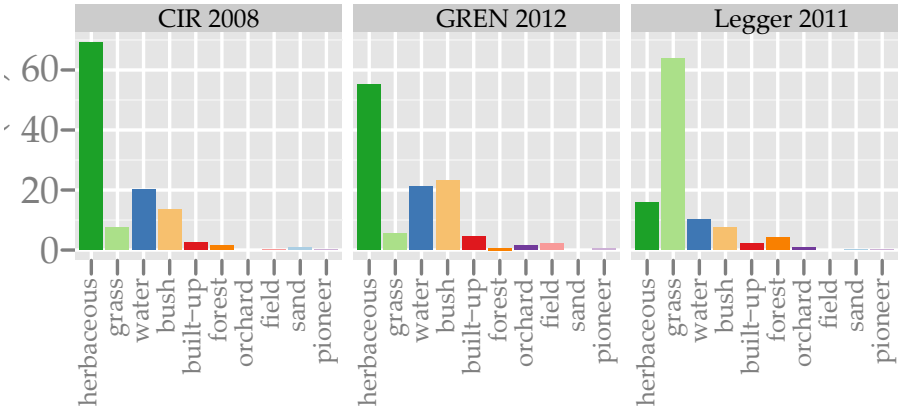


Figure 27: Area per class of the CIR 2008 dataset, the GREN 2012 and the rijkswaterstaat vegetation legger which has been used as reference in the segmentation

11.4 VARIABLE IMPORTANCE

To further investigate the value of all the variables for the classification, Figure 28 gives the 28 most important variables in respect to the increase of the Out of Bag of object classification error as calculated by random forest.

The dots in Figure 28 indicate the average for all classes. As can be expected from an image with many vegetated features, infrared is the most important variable in the classification. After around the tenth

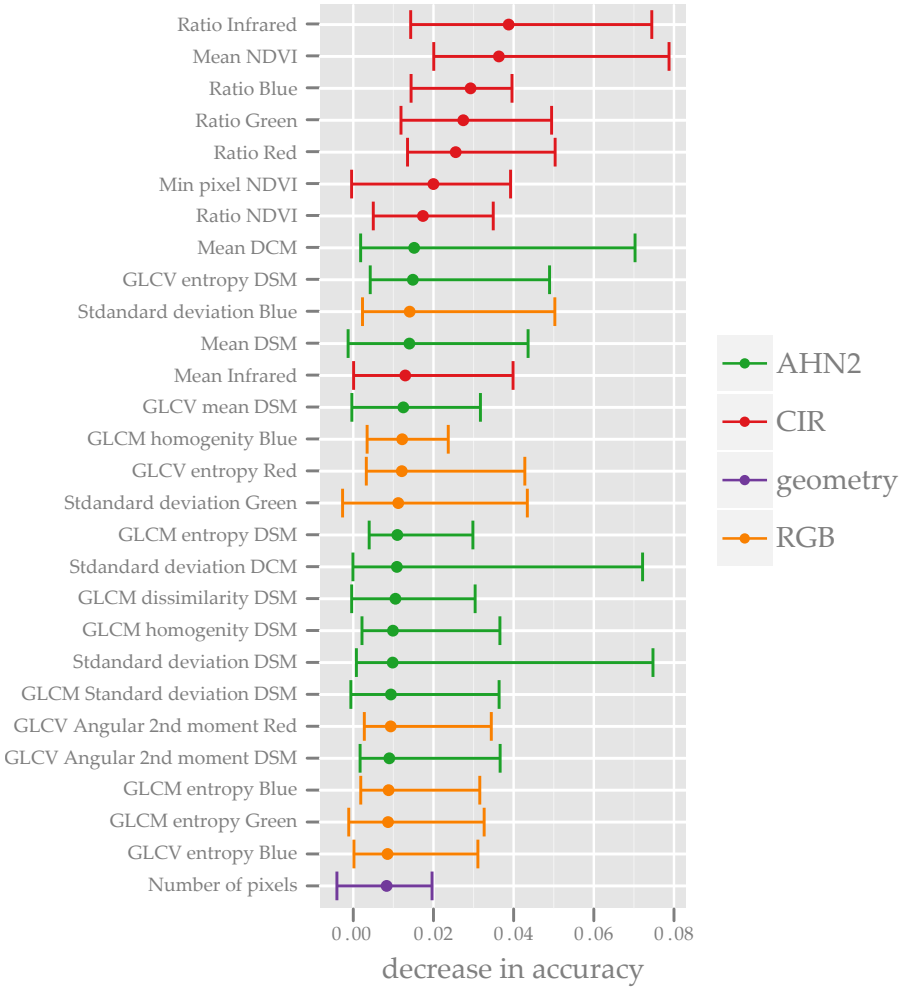


Figure 28: Decrease in Out of Bag object classification accuracy for the 280 most important variables colored according to the source data layer

most important variable, the next variable has very low effect on the OOB decrease in accuracy. This indicates that the variables are very correlative.

The bars indicates the minimal and maximal importance for a class. the bars indicate that many features of the AHN2 are not important for the distinction of all classes but are for some. For example the majority of the difference between forest and bush is captured by the DCM and the standard deviation of the DSM.

The colors of Figure 28 indicate the source of the variable. The top seven variables are all deducted from the infrared or a combination of the infrared and the color data.

This is also illustrated by which shows the decrease in accuracy aggregated per data source. This shows that CIR is the most valuable information for classifying the study area in vegetation structure classes. It also shows that the AHN2 is a valuable dataset to include and that geometry of objects contains very little information.

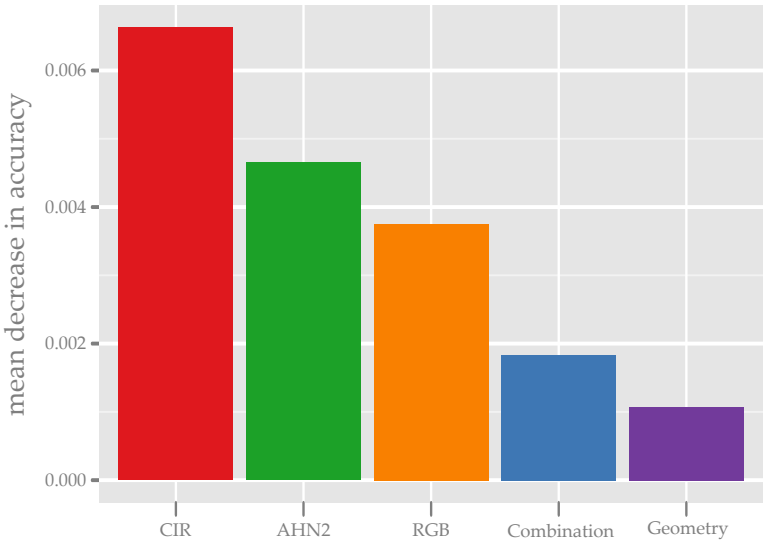


Figure 29: Mean decrease in Out of Bag object classification accuracy for all the variables grouped per data source

12 DISCUSSION

The results of Chapter 10 show that manually delineated reference plots may serve to parameterize a region growing segmentation algorithm. The resulting parameter settings are, as expected, sensitive to the used reference areas. The more detailed and more realistically shaped vegetation legger of Rijkswaterstaat, covering a portion of the study area, results in lower shape settings and a higher scale setting compared to the coarser reference plots. Synchronously, a different dataset, with a higher spatial resolution, results in a higher scale setting and a comparable shape setting when used in conjunction with the vegetation legger as reference.

Chapter 11 presented results concernign the classification of the objects. It showed that RF is well able to distinguish the 9 out of 10 vegetation structure classes used in this study. Both with the CIR as the GREN dataset with structural information of AHN2. The variable importance measures from RF showed that vegetation indicators such as NDVI or Red-Edge, and structural information has a great influence on the accuracy of the classification. Additionally, including textural transformations seem to be beneficiary to the classification result.

12.1 SEGMENTATION PARAMETER SELECTION

The parameter selection method was tested on two sets of reference objects and on two datasets. The method tested on the CIR data and the manually digitized reference plots gave an optimal scale parameter of around 150 and a shape parameter of 25%, as shown by Figure 14 and scale 950, shape around 20% for the GREN dataset (Figure 20).

THE SCALE SETTINGS FOUND IN THIS STUDY ARE HIGH in comparison to FNEA scale settings found in most other studies. Möller et al. (2007) found a scale setting of 56 with a similar method on a different dataset. Liu et al. (2012) found optimum scales in the range of 50-200 with different methods based on reference polygons. Johnson (2011) found an optimum of 70 based on a global intra- and inter-segment *empirical goodness measure*.

The absolute difference can be explained by the differences in the used datasets. Möller et al. (2007) tested their method on 18 landsat multi-spectral and multi-temporal bands with a resolution of 30 m. Johnson (2011) used four CIR bands with a 30 cm resolution. And

Liu et al. (2012) used a 2.5 meter panchromatic band and four 10 m spectral bands.

The data in this study includes six high-resolution spectral bands and additional high resolution elevation data. Such a high amount of complex data is also used in FNEA segmentation by Geerling et al. (2009) who used 10 2 m spectral bands and four 2 m LiDAR layers. Their adapted version of the Möller et al. (2007) comparison index, selected scales between 200 and 600 for different areas of a river floodplain. The high sensitivity of the scale to the datasets used to segment, is also observed in results of varying datasets in Figure 18 and Figure 20 of this study.

THE IRREGULAR BEHAVIOR of the average under and over segmentation observed in Figure 14 is also observed in similar figures of Möller et al. (2007). Both Figure 14 as Möller et al. (2007) presents results obtained by comparing manually digitized circular reference plots with a low amount of relative rectangular objects (400 (Möller et al., 2007) and 206 in this study).

A much more linear behavior can be seen when the vegetation legger is used as a reference in Figure 18 and Figure 20.This more linear behavior gives a much clearer intersection between over and under segmentation and is therefore possibly better in selecting the optimal parameters. Though the linear result may also be due to the smaller portion of the study area covered by the legger. As such the segmentation parameters found with the legger as reference may not be applicable to the rest of the study area.

BY PLOTTING THE 19 REFERENCE PLOTS INDIVIDUALLY in Figure 14 and Appendix B, more insight is gained on the variability of the comparison. The linear results of Figure 18 and Figure 20 are from the segmentation results and the vegetation legger as reference. The vegetation legger covers only the most western part of the study area. Contrary, the circular reference plots are well distributed over the full 1000 ha and the small dots, connected by grey lines in Figure 14 show that optimal settings in one area are far from optimal in other areas. This partially explains the irregular behavior of the average under and over segmentation, of both Figure 14 as results of Möller et al. (2007).

EMPHASIS ON SHAPE IS HIGH for the circular reference plots (Figure 14), while much lower when the vegetation legger is used as reference (Figure 18). The CIR dataset segmentations compared to the circular plots result in a shape setting of 25% while it results in 10% with the vegetation legger.

The relative high emphasize on shape (40%) is also observed by the method with circular reference plots of Möller et al. (2007). This

is contradicted by other authors who kept the shape parameter low and constant since the result was not sensitive to it (Geerling et al., 2009; Johnson, 2011).

Liu et al. (2012) looked into the shape factor after finding the optimal scale setting. This is in compliance with Figure 18, which showed little susceptibility to shape. Though, Figure 14 showed that it is important to consider both parameters integrally. This even suggests that it is worthwhile to include the parameter compactness into the analysis, at the cost of tripling the complexity of the analysis.

Reference objects

Liu et al. (2012) used individual reference polygons and compared these with the *corresponding* segmented object. A corresponding segment has more than half of its area within the reference polygon or contains more than half of the reference polygon within itself. They created 30 reference polygons for every landcover class in the study area, totalling 120 polygons. These polygons had a total of 2106 corresponding segments which belonged to one of the 120 reference polygons. Plots of the comparability metrics resulted in scatter plots which depicted the scale and the metric value with a high amount of spread between the values of individual segments.

This approach of using reference digitization is different from those of circular plots used for Figure 14 in this study and used by Möller et al. (2007). And both approaches differ from the use of a complete map used for Figure 18 and Figure 20. And in between is Geerling et al. (2009) who manually selected representative polygons from an existing complete map and calculated the Möller et al. (2007) comparison index based on the relative size only for the *corresponding* segments.

The method of non-touching reference objects (Liu et al., 2012; Geerling et al., 2009) seems to be favorable to the method of this study and the method of Möller et al. (2007). With the non-touching method, *corresponding* segments belong to one reference object. In other words, there is a one-to-many relationship between the reference and the segments. With the method of random circular plots of Möller et al. (2007), the reference plots are more objective since there is no choice involved in which areas are selected. But there is a many-to-many relationship between segments and reference polygons. This requires a middle layer. Which is created by intersecting the reference polygons with the segments. Then both the segments as the reference have a one-to-many relationship with the intersection. This is the source of the two metrics describing segmentation quality: over and under segmentation. With the method of Möller et al. (2007) these metrics cannot be considered separately, with the non-touching polygons of Liu et al. (2012) they can.

When making or selecting reference polygons, it seems to be important to distribute these well over the full area. Figure 14 showed that spatial variability in *optimum* parameter setting is high and to get an accurate average, a high amount of reference polygons are required.

Figure 18 and Figure 20 in comparison to Figure 14 shows that detailed and realistically shaped polygons are preferred above manually drawn compact and simplified objects. Figure 18 and Figure 20 and results from Geerling et al. (2009) combined, suggests that it is beneficiary to the selection, to use a stratified random sample of non-touching polygons from a high quality manually drawn map as a reference. With strata of the classes in the map. As such a both spatially as thematically well-distributed set of high quality reference polygons may reliably select an optimum mapping scale.

Different datasets and the issue of scale

Figure 20 showed that is possible to find an optimum parameter combination to segment a different dataset with the same reference information. The hypothesis was that this would result in objects of similar mapping scale. To test this hypothesis Figure 22 and Figure 21 have been created.

From these results it is difficult to conclude whether the hypothesis holds true. The CIR 2008 as the GREN 2012, segmented with FNEA and the selected parameter settings, *theoretically* both result in objects which best match the polygons of the legger reference given the properties of the data and the the algorithm. Additionally, the almost equal comparability at the intersection of under and oversegmentation of Figure 18 and Figure 20 mean that the match is both equally good.

Though, Figure 22 shows that the resulting objects are very different. This is better illustrated by the size of the objects as shown by Figure 21. On average both segmentations are equally comparable to the reference, but still, this does result in very different objects.

Comparability measure

This study as the study of Möller et al. (2007) compared the intersection of the segmentations and the references to the references and the segmentations. This results in two metrics which show the similarity between the intersection and both the reference as the segmentation. Both metrics are a combination of relative size and relative position. As such both the topological as the geometric differences are taken into account.

Geerling et al. (2009) stated that calculating the geometric errors did not provide additional value to the selection for ecotope delineation but provided no evidence of this. Intuitively it seems important

to include both metrics in order to get a realistic delineation, not one with the lowest amount of objects within a reference area. Though the combination of the metrics in this study and that of Möller et al. (2007) obscure information about the functioning of the separate metrics and provide no clues to the value of the individual metrics.

Liu et al. (2012) evaluated seven metrics to find the best match between individual reference plots and segmented objects, though none of these included geometric errors such as the relative distance between gravity centers. They concluded that a combination of topological and arithmetic metrics are best used. Though this is focused on preventing under-segmentation as this error definitively propagates to the classification result (Liu et al., 2012). The combination of metrics of Möller et al. (2007) used in this study is more complex to interpret, but does make the workings transparent. Metrics of Liu et al. (2012) are very reliant on the assumption that the optimum segmentation parameters are there where the under and over segmentation intersect. The method of (Möller et al., 2007) is also relying on this assumption, but by plotting under and over segmentation as separate metrics, the effects of this assumption are made transparent. For instance, Figure 14 shows that for shape 15% under segmentation goes down with increasing scale, instead of over segmentation going up. With a very high scale setting, the line eventually will intersect, at a very low comparability. A combined metric, such as those of Liu et al. (2012), will obscure that this comes at the cost of under segmentation, making objects too large.

12.2 CLASSIFICATION RESULTS

Appendix C shows the possibility to map river floodplain vegetation structure on a wide scale with AHN2 and CIR aerial images. These datasets are available for all of the Dutch river floodplains and are frequently updated.

A validation with 44 random stratified field observation points in 2012 of this map showed a relative low overall accuracy of 43%. This value is in stark contrast to the validation accuracy of 74% found by Geerling et al. (2009) with similar data and 348 randomly selected validation points deducted from aerial images and expert knowledge. The value of 43% is similar to the field validation of the ecotope map (Knotters et al., 2008). The full ecotope map consists of more classes and covers a much wider area. Though they found an accuracy of 48%.

The relative low accuracy has to do with inaccuracies in the validation set due to a high temporal offset of four years and a relative high spatial offset due to inaccuracies in the gps device used to locate the points. The four years difference between image acquisition and point observation is especially influential in the estimate of classes

which generally change rapidly, such as agricultural fields and pioneer vegetation. This is reflected in the confusion matrix of Figure 24.

The relative high spatial detail of the map and low GPS accuracy caused difficulties in finding the exact center points of the stratified randomly sampled objects. The inaccuracy of the device often exceeded the dimensions of the objects.

Liaw and Wiener (2002) experienced accurate accuracy estimates of RF OOB estimates. Gislason et al. (2006) found that the OOB estimate of the accuracy is generally *lower* than the accuracy estimated based on a separate validation set. This is logically due to the fact that each tree is constructed on a subset of training data and the accuracy estimate is therefore based on only a small amount of the total trees in the forest. As expected a subset of the forest has a lower performance the the complete forest. The RF OOB estimate of the object classification accuracy was 86%. This constitutes to the average of the correctly classified objects.

The high difference between the low field validation and the high OOB accuracy is believed to mainly show the high degree of change in the study area between 2008 and 2012 and that river floodplain vegetation is highly dynamic.

Scale and the Modifiable Aerial Unit Problem

Figure 25 gives a representative insight in the function of scale on the resulting map. With a finer analysis scale especially grass objects emerge within areas classified as herbaceous at coarser scales. This behavior is also observed with multi-scale OBIA of IKONOS data by Hall et al. (2004). Floodplain landscape is complex and composed of objects at different scales. As with other vegetation parameters (Addink et al., 2007), the roughness deducted from maps with a coarse analysis scale will differ from maps with a finer analysis.

OBIA, contrary to pixel based analysis, is believed to be a partial way out of the MAUP (Marceau, 1999). Hay et al. (2003) acknowledge that this is true when there is an objective way to determine segmentation parameters and consistently create objects of similar map-scale. Figure 26 shows that it is possible by selecting appropriate segmentation parameters by using reference polygons to control the analysis scale, independent of the data. Figure 27 shows that the effect in area distribution between the classes is limited to increase in bush area. This is compliant with expert knowledge about the shrub encroachment of the area.

Data layers and classification variables

The use of RF as a classifier supplied additional information on the importance of the used data layers and computed classification variables. This has been shown in Figure 29 and Figure 28 respectively.

As expected Figure 29 shows that CIR data contains a high distinction potential for different vegetation classes. It clearly shows that RGB alone will have a negative impact on classification accuracy. Additionally the AHN2 data ranks second and contains valuable information especially to distinct specific classes such as bush vegetation from forests and herbaceous as is visible in Figure 28. This is acknowledged by Geerling et al. (2009) who found that including LiDAR with hyperspectral data has beneficiary effects on classification accuracy of river floodplain vegetation. Though it is also widely recognized that LiDAR data cannot be used solely and is of a lower importance than spectral information when it comes to classifying floodplain vegetation (Forzieri et al., 2010; Gislason et al., 2006; Geerling et al., 2009).

It is widely ackowledged that RF is not negatively affected by a high number of highly correlating and possibly noisy variables (Friedman et al., 2001; Liaw and Wiener, 2002). Though Figure 29 showed that there is little value in incorporating geometric features in the classification.

The color-space transformation of RGB into IHS did not showed up in the top 28 variables of Figure 28. Although successfully used with remote sensing data to for instance map burned areas with Landsat data (Koutsias et al., 2000), such a transformation did not showed beneficiary effects for mapping of floodplain vegetation. Though, these features might contain valuable information during the segmentation stage of OBIA. This has not been done in this study, but has been proved to be a successful method when segmenting scenes containing vegetated objects (Trias-Sanz et al., 2008).

Haralick's texture features, proved to be beneficiary to the classification accuracy of Figure 28. The rank of these features was similar to the *sub-top* ranks of these features found by Yu et al. (2006). The total decrease in classification error due to the inclusion of texture has not been investigated, but a similar study of riparian vegetation found a decrease of 2-19% by including texture depending on the vegetation class (Johansen et al., 2007).

13 CONCLUSION AND RECOMMENDATIONS

The CIR aerial images with the LiDAR dataset: AHN2, will both have national coverage for the Netherlands in the near future. This research has shown that these novel datasets may serve as input into an object-based procedure to map river floodplains into vegetation structure classes and that a high map accuracy is feasible.

FIGURES OF MAP RESULTS with different scales and datasets (Figure 25, Figure 26 and Figure 27), emphasize that it is important to choose an appropriate analysis scale and be able to consistently apply that scale to different datasets. Without being able to do so, the analysis becomes arbitrary, loses validity and the resulting map products become unfit to use in hydrological models.

The empirical goodness measure used in this study has shown to be a suitable method to select segmentation parameters. By quantifying the discrepancies between reference polygons and segmentation results, the optimal scale and shape parameter may be selected. As such it is possible to objectively select appropriate segmentation parameters on the basis of reference polygons.

Comparing different datasets with the same reference (Figure 26) has shown that this is a generic method. It results in a similar mapping scale with aerial images of a different sensor and with different specifications.

A FIELD VALIDATION OF 44 OBSERVATIONS four years after the data acquisition date resulted in an overall accuracy of 43%. The study wide map was created by classifying segments created with FNEA scale 15, shape 25% and compactness 50%.

Although the overall accuracy is low, RF in combination with the datasets used in this study has shown to be a capable classifier. The internal validation measure of RF showed a producers accuracy of 86% on the basis of the OOB training samples. Previous research suggests that this value is generally lower than the actual object-classification accuracy (Gislason et al., 2006). The great difference is therefore explained to be due to the low amount of validation observations, high spatial uncertainties during the acquisition of the validation set and high degree of change in the four years between image acquisition and validation.

THE VARIABLE IMPORTANCE MEASURE of RF (Figure 28 and Figure 29) has shown that CIR is by far the most important source of in-

formation to distinct different vegetation objects. Including structural information of a LiDAR dataset has proven to be beneficiary, especially in the classification of high and woody vegetation such as forest and bush. Vegetation types which exert a high amount of friction on water flow.

The importance measure depicted that including an IHS color transformation or geometric object features does not increase classification accuracy. While including GLCM and GLDV derived texture metrics did show to have slight positive effect on the classification accuracy. This suggests that texture of vegetated objects is a valuable part of the information contained in high resolution data.

IT IS RECOMMENDED TO include other parameters such as compactness and the degree of spectral difference merging. Though it is also recommended to consider segmentation parameters simultaneous as has been done in this study. Doing both will drastically increase the computational burden since all possible combinations have to be tested.

When using reference polygons it is recommended to use individual non-touching polygons as done by Liu and Xia (2010). This makes it possible to directly compare reference to corresponding segments, without having to intersect both. As such the analysis becomes better understandable and enables visualizations (maps) of the discrepancies and its spatial distribution. Additionally, the difference between the two reference datasets used in this study emphasize that these polygons should be of high quality and well distributed over the area.

When operationalizing the mapping process on a national scale, there is a trade-off between analysis complexity and the quality and labeling of the segments. Keeping some parameters constant and using a simpler classifier such as a k-nearest neighbor classifier will slightly lower the overall accuracy but will keep the analysis understandable and the results reliable. This might be of priority to maps used in models predicting the discharge capacity of floodplains and the required flood defenses.

Part IV

APPENDIX

A

APPENDIX CLASSES

Classes distinguished by Rijkswaterstaat Photo Interpretation map, Rijkswaterstaat pilot vegetation legger, Geerling et al. (2009) and this study

PI MAP	LEGGER	GEERLING ET AL. 2009	THIS STUDY
Summer bed [1]	Water	Water	Water
Pioneer vegetation	*not present*	*not present*	Pioneer
Gravel[2]	Bare	Bare	Sand
Built-up/hardened	Built-up/hardened	Built-up/hardened	Built-up
Production grassland	Grass	Production grassland	Grass
Structure rich grassland	Grass	Structure rich grassland	Grass
Agricultural field	*not present*	Agricultural field	Field
Bulrush	*not present*	Bulrush	Herbaceous
Reed and other helophytes	Reed	Reed and other helophytes	Herbaceous
Herbaceous vegetation	Herbaceous vegetation	Herbaceous vegetation	Herbaceous
Natural forest	Natural forest	Natural forest	Forest
Production forest	*not present*	Production forest	Forest
Tidal forest	*not present*	Tidal forest	Forest
Bush	Bush	Bush	Bush
Orchard and tree nursery	Orchard and tree nursery	*not present*	Orchard

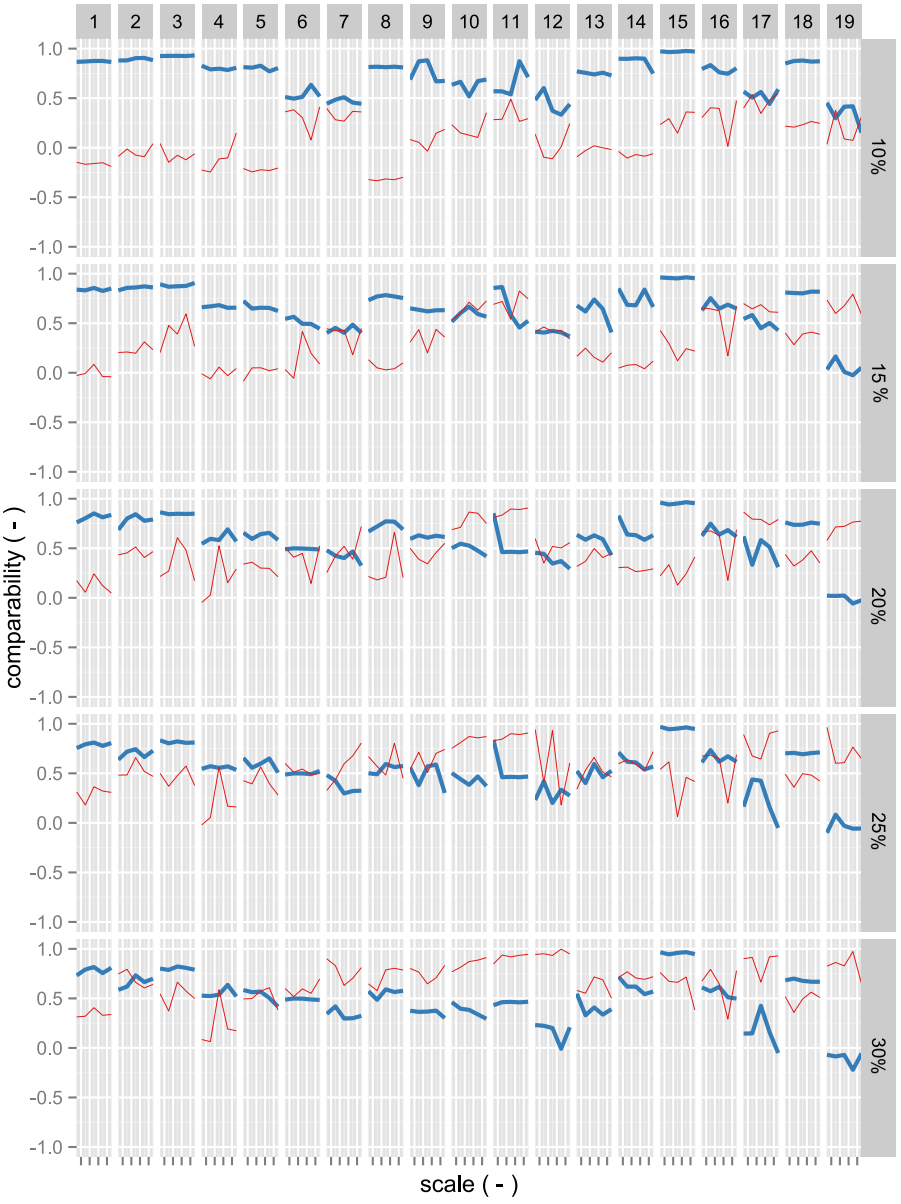
1 Along summer bed, the PI map consisted of other water classes: Side channel and Artificial or oxbow lake.

2 Along gravel, the PI map consisted of other sand and shore classes: Natural shell bank; Hard clay or peat bank; Beach or shelf; Bare shore levee and Rest (often temporarily bare soil).

B

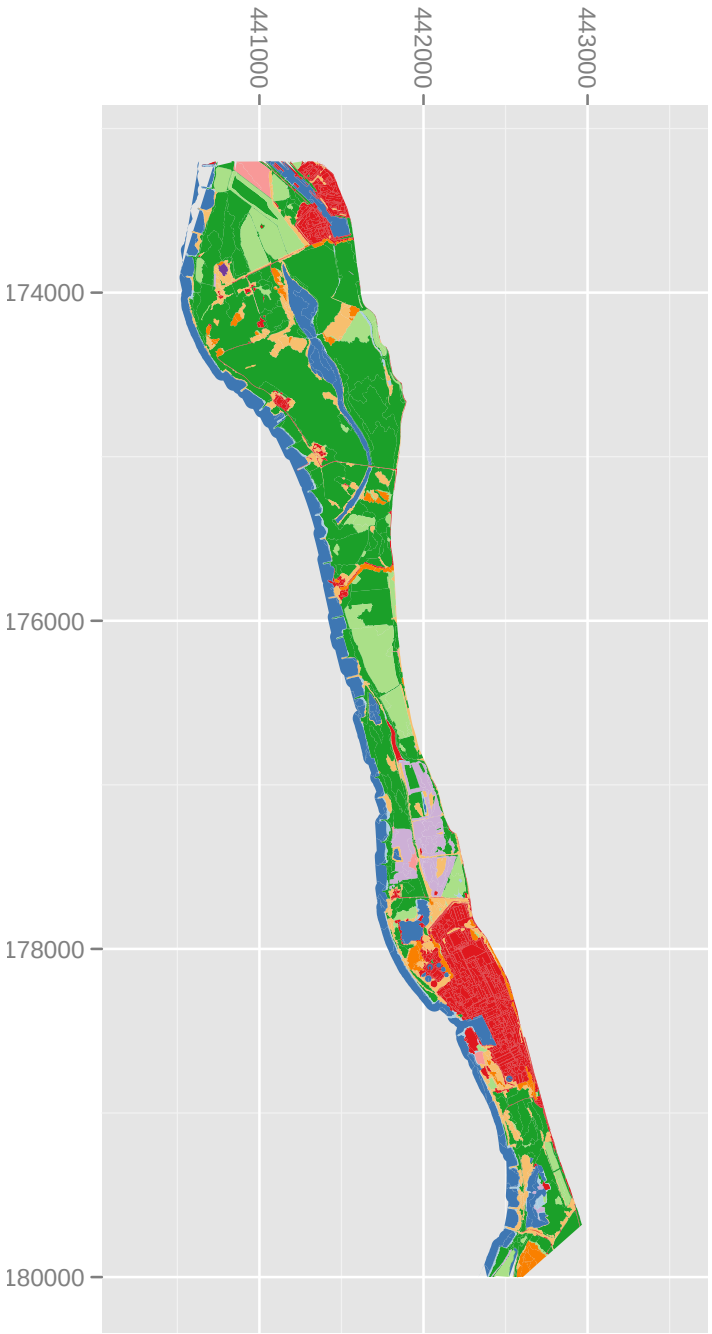
APPENDIX COMPARABILITY OF 19 INDIVIDUAL REFERENCE PLOTS

Average comparability between the reference objects of the 19 individual circular manually digitized reference plots (indicated at the top) and segmentation results of the 2008 CIR and the 2011 AHN2 dataset. Scale is set between 50 and 450 (horizontal axis) and repeated for the five scales settings between 10% till 30% (indicated at the right).



C

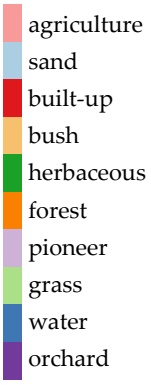
APPENDIX CLASSIFICATION

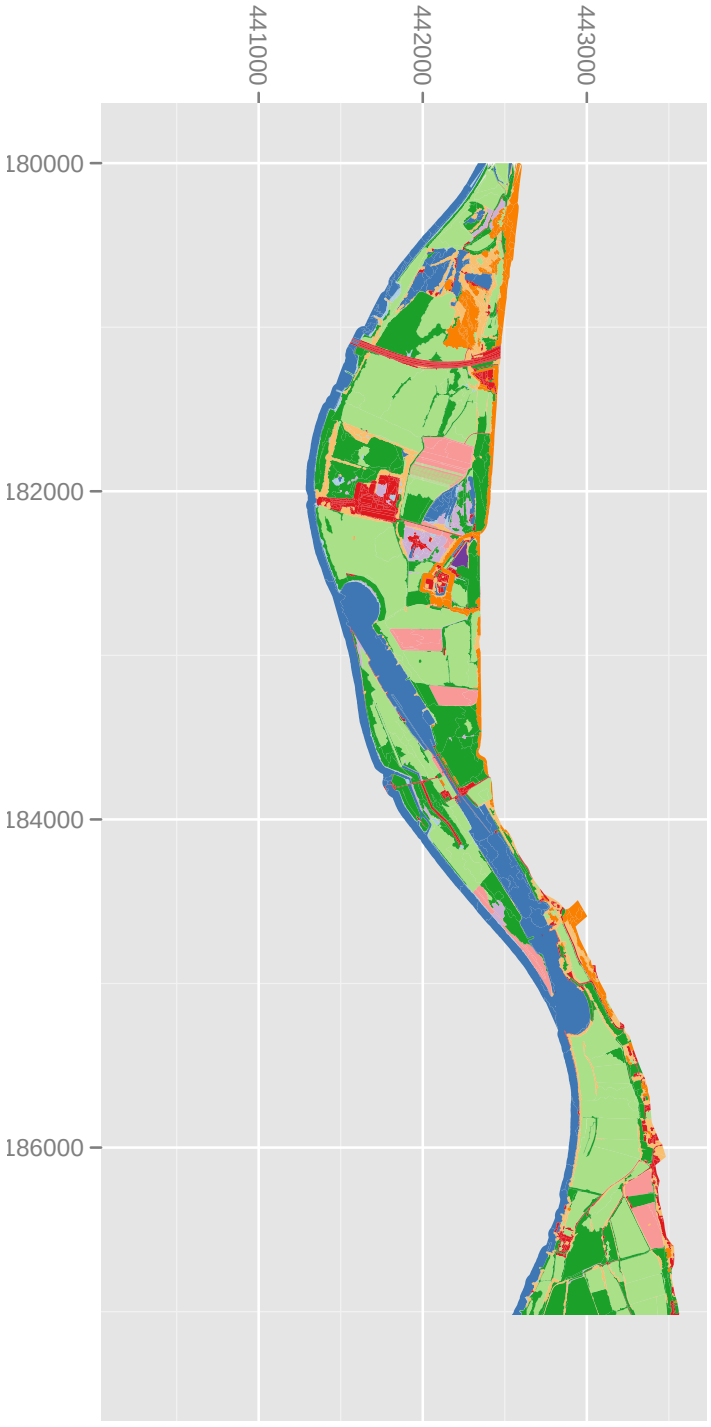


Western part of the study area, North is to the right of the paper. Floodplains around Wageningen and Renkum classified into floodplain vegetation structure classes. Objects created by segmentation in eCognition's multiresolution segmentation with scale 150 and shape 25%, spectral difference segmentation 2 and classified by random forest with 1000 trees and 15 variables each

- agriculture
- sand
- built-up
- bush
- herbaceous
- forest
- pioneer
- grass
- water
- orchard

Eastern part of the study area, North is to the right of the paper. Floodplains to the east of Renkum and west of Arnhem classified into floodplain vegetation structure classes. Objects created by segmentation in eCognition's multiresolution segmentation with scale 150 and shape 25%, spectral difference segmentation 2 and classified by random forest with 1000 trees and 15 variables each





BIBLIOGRAPHY

E.A. Addink, S.M. de Jong, and E.J. Pebesma. The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. *Photogrammetric Engineering and Remote Sensing*, 73(8): 905, 2007.

O. Alata and L. Quintard. Is there a best color space for color image characterization or representation based on multivariate gaussian mixture model? *Computer Vision and Image Understanding*, 113(8): 867–877, 2009.

J.P. Ardila, W. Bijker, V.A. Tolpekin, and A. Stein. Context-sensitive extraction of tree crown objects in urban areas using vhr satellite images. *International Journal of Applied Earth Observation and Geoinformation*, 15:57–69, 2012.

M. Baatz and A. Schäpe. *Multiresolution Segmentation - an optimization approach for high quality multi-scale image segmentation*, pages 12–23. Wichmann, Heidelberg, 2000.

M. J. Baptist, W. E. Penning, H. Duel, A. J. M. Smits, G. W. Geerling, G. E. M. van der Lee, and J. S. L. Van Alphen. Assessment of the effects of cyclic floodplain rejuvenation on flood levels and biodiversity along the rhine river. *River Research and Applications*, 20(3): 285–297, 2004.

U.C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3):239–258, 2004.

P.L. Bernstein. *Against the gods: The remarkable story of risk*, volume 1998. John Wiley & Sons New York, 1996.

T. Blaschke. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1):2–16, 2010.

T. Blaschke and J. Strobl. What s wrong with pixels? some recent developments interfacing remote sensing and gis. *GeoBIT/GIS*, 6: 12–17, 2001.

T. Blaschke, C. Burnett, and A. Pekkarinen. *Image segmentation methods for object-based analysis and classification*, pages 211–236. Kluwer acadamic publishers, Dordrecht, 2004.

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

R. Bringhurst. *The elements of typographic style*. Hartley & Marks, 1992.

C. Burnett and T. Blaschke. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological modelling*, 168(3):233–249, 2003.

J.C.W. Chan and D. Paelinckx. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6):2999–3011, 2008.

G. Chen and G.J. Hay. An airborne lidar sampling strategy to model forest canopy height from quickbird imagery and geobia. *Remote Sensing of Environment*, 115(6):1532–1542, 2011.

D.R. Cutler, T.C. Edwards Jr, K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler. Random forests for classification in ecology. *Ecology*, 88(11):2783–2792, 2007.

AG Definiens. Definiens ecognition developer 8 reference book. *Definiens AG, München*, 2009.

Department of LNV. Nature for people, people for nature; policy document for nature, forest and landscape in the 21st century, 2000.

D.C. Duro, S.E. Franklin, and M.G. Dubé. Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *International Journal of Remote Sensing*, 33(14):4502–4526, 2012.

G.M. Espindola, G. Camara, I.A. Reis, L.S. Bins, and A.M. Monteiro. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing*, 27(14):3035–3040, 2006.

European Union. Council directive 92/43/eec on the conservation of natural habitats and of wild fauna and flora, 1992.

G. Forzieri, G. Moser, E. R. Vivoni, F. Castelli, and F. Canovaro. Riparian vegetation mapping for hydraulic roughness estimation using very high resolution remote sensing data fusion. *Journal of Hydraulic Engineering-Asce*, 136(11):855–867, 2010.

G. Forzieri, L. Guarnieri, E. R. Vivoni, F. Castelli, and F. Preti. Spectral-als data fusion for different roughness parameterizations of forested floodplains. *River Research and Applications*, 27(7):826–840, 2011.

J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer Series in Statistics, 2001.

K.S. Fu and JK Mui. A survey on image segmentation. *Pattern recognition*, 13(1):3–16, 1981.

Y. Gao, J. F. Mas, N. Kerle, and J. A. N. Pacheco. Optimal region growing segmentation and its effect on classification accuracy. *International Journal of Remote Sensing*, 32(13):3747–3763, 2011.

G. W. Geerling, M. J. Vreeken-Buijs, P. Jesse, A. M. J. Ragas, and A. J. M. Smits. Mapping river floodplain ecotopes by segmentation of spectral (casi) and structural (lidar) remote sensing data. *River Research and Applications*, 25(7):795–813, 2009.

CE Gehlke and K. Biehl. Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, 29(185):169–170, 1934.

P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson. Random forests for land cover classification. *Pattern Recognition Letters*, 27(4):294–300, 2006.

O. Hall, G.J. Hay, A. Bouchard, and D.J. Marceau. Detecting dominant landscape objects through multiple scales: An integration of object-specific methods and watershed segmentation. *Landscape Ecology*, 19(1):59–76, 2004.

R. M. Haralick, Shanmuga.K, and I. Dinstein. Textural features for image classification. *Ieee Transactions on Systems Man and Cybernetics*, Smc3(6):610–621, 1973.

R.M. Haralick and L.G. Shapiro. Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1):100–132, 1985.

G.J. Hay, T. Blaschke, D.J. Marceau, and A. Bouchard. A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(5):327–345, 2003.

Txomin Hermosilla, Luis A. Ruiz, Jorge A. Recio, and Javier Estornell. Evaluation of automatic building detection approaches combining high resolution images and lidar data. *Remote Sensing*, 3(6):1188–1210, 2011.

J.A.M. Janssen. Monitoring of salt-marsh vegetation by sequential mapping, 2001. PhD dissertation.

J. Järvelä. Flow resistance of flexible and stiff vegetation: a flume study with natural plants. *Journal of Hydrology*, 269:44–54, 2002.

K. Johansen and S. Phinn. Mapping structural parameters and species composition of riparian vegetation using ikonos and landsat etm+ data in australian tropical savannahs. *Photogrammetric Engineering and Remote Sensing*, 72(1):71, 2006.

K. Johansen, N.C. Coops, S.E. Gergel, and Y. Stange. Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. *Remote Sensing of Environment*, 110(1):29–44, 2007.

Z. Johnson, B. Xie. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011.

B.N. Jyothi, G.R. Babu, and I.V.M. Krishna. Object oriented and multi-scale image analysis: strengths, weaknesses, opportunities and threats-a review. *Journal of Computer Science*, 4(9):706–712, 2008.

Y. Ke, L. J. Quackenbush, and J. Im. Synergistic use of quickbird multispectral imagery and lidar data for object-based forest species classification. *Remote Sensing of Environment*, 114(6):1141–1154, 2010.

A. Knotters. Beschrijving methode en resultaten pilot detaillering. Rijkswaterstaat, September 2011.

M. Knotters, B.J. Brus, and A.H. Heidema. Validatie van ecotopenkaarten van de rijkswateren. Technical Report 1656, Alterra, 2008.

N. Koutsias, M. Karteris, and E. Chuvico. The use of intensity-hue-saturation transformation of landsat-5 thematic mapper data for burned land mapping. *Photogrammetric engineering and remote sensing*, 66(7):829–840, 2000.

N.S.N. Lam and D.A. Quattrochi. On the issues of scale, resolution, and fractal analysis in the mapping sciences*. *The Professional Geographer*, 44(1):88–98, 1992.

A. Liaw and M. Wiener. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

D. Liu and F. Xia. Assessing object-based classification: advantages and limitations. *Remote Sensing Letters*, 1(4):187–194, 2010.

Y. Liu, L. Bian, Y. Meng, H. Wang, S. Zhang, Y. Yang, X. Shao, and B. Wang. Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68:144–156, 2012.

L. Lusa et al. Class prediction for high-dimensional class-imbalanced data. *BMC bioinformatics*, 11(1):523, 2010.

J. Maiers and YS Sherif. Applications of fuzzy set theory. *IEEE Transactions on Systems, Man, and Cybernetics*, 15:175–186, 1985.

B. Makaske, G. J. Maas, C. van den Brink, and H. P. Wolfert. The influence of floodplain vegetation succession on hydraulic roughness: Is ecosystem rehabilitation in dutch embanked floodplains compatible with flood safety standards? *Ambio*, 40(4):370–376, 2011.

D.J. Marceau. The scale issue in social and natural sciences. *Canadian Journal of Remote Sensing*, 25(4):347–356, 1999.

D.J. Marceau and G.J. Hay. Remote sensing contributions to the scale issue. *Canadian journal of remote sensing*, 25(4):357–366, 1999.

G. Meinel and M. Neuber. a comparison of segmentation programs for high resolution remote sensing data. *International Archives of the ISPRS*, 35 part B(Commision 4):1097–1105, 2004.

D. Meyer, F. Leisch, and K. Hornik. The support vector machine under test. *Neurocomputing*, 55(1-2):169–186, 2003.

M. Möller, L. Lymburner, and M. Volk. The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation and Geoinformation*, 9(3):311–321, 2007.

O. Mutanga, E. Adam, and M.A. Cho. High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 18:399–406, 2012.

S.W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 2011.

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

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL http://www.R-project.org. ISBN 3-900051-07-0.

Rijksoverheid. Onderhandelingsakkoord tussen rijk en provincies over natuurbeleid, September 2011.

Rijkswaterstaat. Toelichting bij de ecotopenkartering rijntakken-oost 1997, 1998.

Rijkswaterstaat. Ecotopenkartering rijntakken-oost 2005, 2008.

J. Schiewe. Segmentation of high-resolution remotely sensed data-concepts, applications and problems. *International Archives of Photogrammetry Remote Sensing And Spatial Information Sciences*, 34(4): 380–385, 2002.

M.R. Segal. Machine learning benchmarks and random forest regression. http://repositories.cdlib.org/cbmb/bench_rf_regn, 2004. eScholarship Repository. University of California.

W. Silva, F. Klijn, and J. Dijkman. Room for the rhine branches in the netherlands - what the research has taught us, 2001.

M. Straatsma and F. Huthoff. Uncertainty in 2d hydrodynamic models from errors in roughness parameterization based on aerial images. *Physics and Chemistry of the Earth*, 36(7-8):324–334, 2011.

M. Straatsma and H. Middelkoop. Extracting structural characteristics of herbaceous floodplain vegetation under leaf-off conditions using airborne laser scanner data. *International Journal of Remote Sensing*, 28(11):2447–2467, 2007.

M. W. Straatsma and M. Baptist. Floodplain roughness parameterization using airborne laser scanning and spectral remote sensing. *Remote Sensing of Environment*, 112(3):1062–1080, 2008.

J. Stuckens, P. R. Coppin, and M. E. Bauer. Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sensing of Environment*, 71(3):282–296, 2000.

R. Trias-Sanz, G. Stamon, and J. Louchet. Using colour, texture, and hierarchial segmentation for high-resolution remote sensing. *ISPRS Journal of Photogrammetry and remote sensing*, 63(2):156–168, 2008.

C. van der Sande, S. Soudarissanane, and K. Khoshelham. Assessment of relative accuracy of ahn-2 laser scanning data using planar features. *Sensors*, 10(9):8198–8214, 2010.

M. Verleysen and D. François. The curse of dimensionality in data mining and time series prediction. *Computational Intelligence and Bioinspired Systems*, pages 85–125, 2005.

Booij M.J. Van der Klis H. Hulscher S.J.M.H. Warmink, J.J. Quantification of uncertainty in design water levels due to uncertain bed form roughness in the dutch river waal. *Hydrological Processes*, 2012.

Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL http://had.co.nz/ggplot2/book.

Q. Yu, P. Gong, N. Clinton, G. Biging, M. Kelly, and D. Schirokauer. Object-based detailed vegetation classification. with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing*, 72(7):799–811, 2006.

Q. Zhan, M. Molenaar, K. Tempfli, and W. Shi. Quality assessment for geo-spatial objects derived from remotely sensed data. *International Journal of Remote Sensing*, 26(14):2953–2974, 2005.

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

L. Zhang and X. Huang. Object-oriented subspace analysis for airborne hyperspectral remote sensing imagery. *Neurocomputing*, 73 (4-6):927–936, 2010.

Y.J. Zhang. A survey on evaluation methods for image segmentation. *Pattern recognition*, 29(8):1335–1346, 1996.