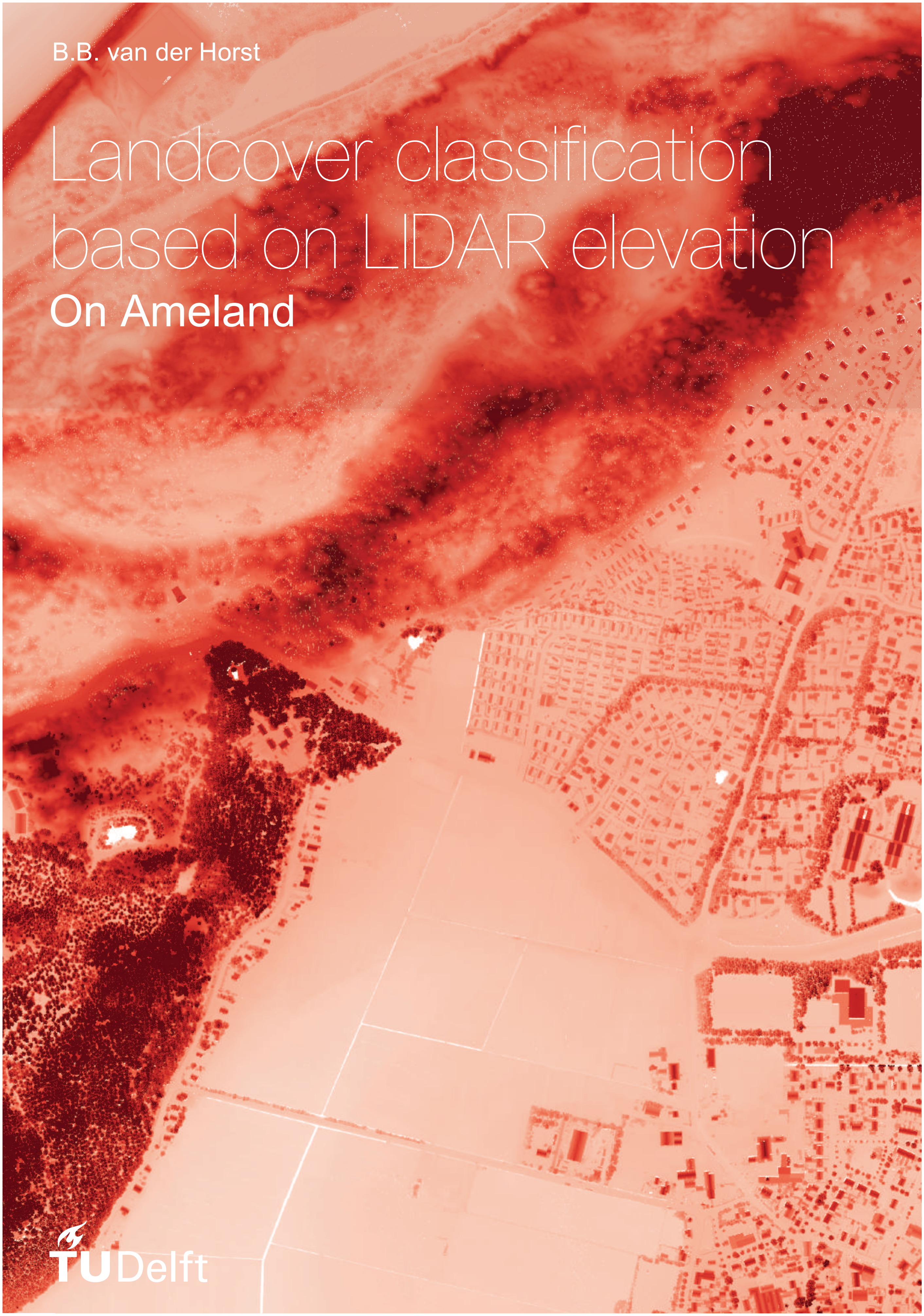


B.B. van der Horst

Landcover classification based on LIDAR elevation

On Ameland

TUDelft

Landcover classification based on elevation

On Ameland

By

B.B. van der Horst

Supervisor:	Dr. R. C. Lindenbergh
2nd supervisor:	Dr. ir. F. van Leijen

Cover: AHN height data of the 0,5 meter raster.

TU Delft

Preface

This report was written as main part of the Bachelor thesis of Applied Earth Science, by Bibi van der Horst at the Delft University of Technology.

I would like to thank dr R.C. Lindenberg for supervising the research and guiding me through the process. I would like to thank A.L. van Natijne for supporting me with his programming skills and for proofreading the report. I would like to thank J. van der Horst, dr. R.C. Lindenberg for proofreading the report too.

B. B. van der Horst
Delft, April 2016

Abstract

This research was done as main part of the Bachelor thesis of Applied Earth Science. In this research the main objective was to study how airborne laser scan data sampling Ameland can be used to produce a land cover map. From airborne laser scan data one can derive different features, which are used to distinguish different land cover categories. In this research the classification classes are:

- Woods;
- Buildings;
- Agriculture;
- Beach.

The classification was performed for 5 meter and 0,5 meter rasters. The classification of the 5 meter raster was done based on six elevation related features. The result of this classification method was coarse. The classification of the 0,5 meter raster was done with added features to improving the classification method. The result of this was a detailed map.

Contents

1 Introduction

At present airborne laser data of the Netherlands is publicly available. The start of collection of this data was in 1996, when there was a demand of this data for water management. It took 7 years to complete the first version of AHN (Actueel Hoogtebestand Nederland, “Current Height file Netherland”). Today, 20 years later, the acquisition of the third version, AHN3, is ongoing.

1.1 Research question

The main objective of this bachelor thesis is to study if and how airborne laser scan data sampling Ameland can be used to produce a land cover map. Part of this objective is to determine which method and which classes are most suitable and how the results depend on the methodology and on the properties of the data. For example, the spatial resolution of different data products will be considered. First, elevation rasters are used for classification, after which point are used to get more detail.

For validation, the results will be compared to other classification products like Bestand Bodemgebruik and aerial photographs.

1.2 Ameland

A small introduction to the area of interest. Ameland is the fourth inhabited West Frisian Island (Waddeneiland) counted from the West. The neighbouring islands are Terschelling to the West and Schiermonnikoog to the East. The area of the land is approximate 59 km^2. With approximately 3610 inhabitants Ameland is the middle largest of the Frisian Islands (Centraal Bureau voor de Statistiek, 2016). The population is scattered over four villages, Hollum, Ballum Nes and Buren, from West to East. As shown in Figure 1 Ameland exists mainly of sand dunes, moorlands and a swampy nature reserve, the Oerd. In this nature reserve the seawater flows in and out by channels. But there are woods too.

Figure 1: Overview of Ameland (OpenStreetMap)

2 Data

In this project different data was used. For the height data AHN will be used. AHN stands for Actueel Hoogtebestand Nederland. There are three types of AHN, in this project AHN3 will be used, the latest version of AHN. The height data is obtained with LiDAR (LIght Detection And Ranging). The data are distributed as 3D-point clouds or as raster files (Digital Surface Models).

2.1 LiDAR

LiDAR is a method which measures the distance to objects. Fixed-wing aircrafts or helicopters do airborne laser scanning by flying over and sending a laser pulse down to the surface. The returned laser signal reflects at the surface and is received at the aircraft. The time between sending and receiving is measured. Because the velocity of light waves is known and the travel time of the signal is measured by the airborne system, the distance from the airborne system to the object can be calculated by:

$$Distance = \frac{speed\ of\ light \times time\ of\ travel}{2}$$

A GNSS (Global Navigation Satellite System) is used to measure the exact position of the system. To know if an aircraft flies horizontal or with an angle, the orientation is measured by an IMU (Inertial Measurement Unit). The IMU measures the rotation in three directions, the roll, pitch and yaw of the aircraft. The combination of GPS and IMU reconstructs the flight path. (www.LiDAR-UK.com, 2016)

The flight paths are normally straight flight lines. Figure 2 illustrates the principle of LiDAR (Airborne and terrestrial laser scanning, 2010).

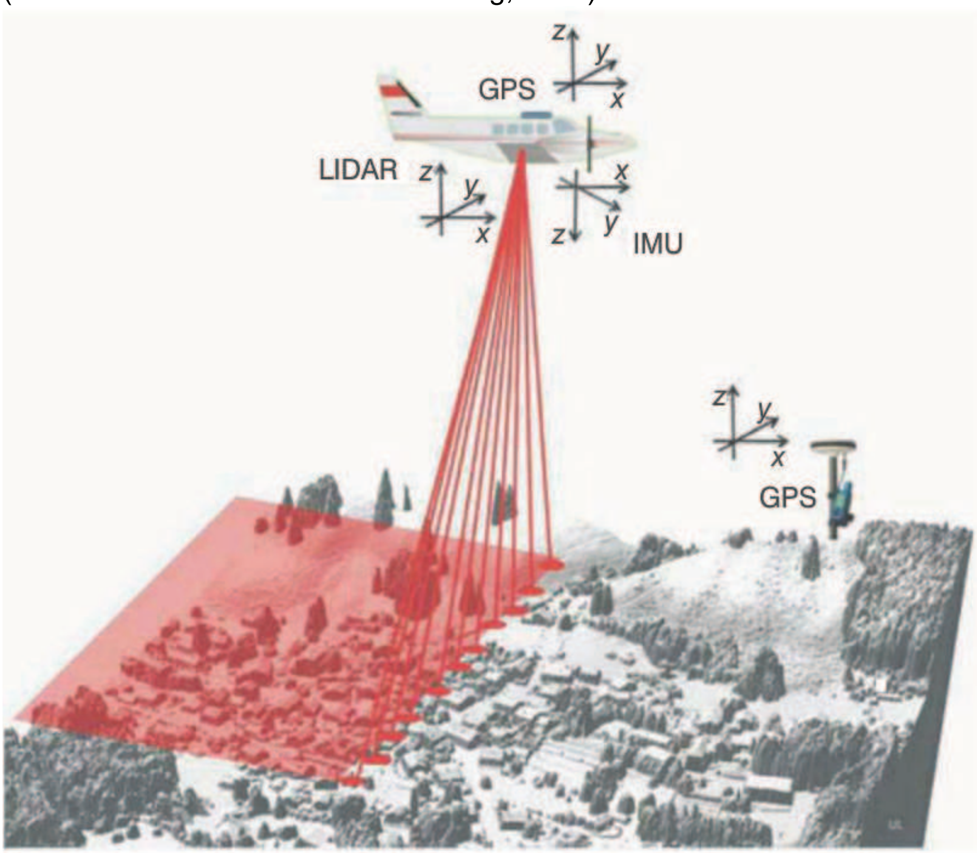


Figure 2: Airborne laser scanning principle (Airborne and terrestrial laser scanning, 2010)

2.2 AHN

The AHN is a file with height data for the whole of The Netherlands. AHN is obtained with LiDAR. The first version of AHN (AHN1) is made between 1996 and 2003. At present the collection of AHN3 is well under way. Between 2014 and 2019 the total area of the Netherlands will be scanned for AHN3. Since September 2015 all AHNs (AHN1, AHN2, and partly AHN3) are freely available via nationaalgeoregister.nl and pdok.nl (Publieke Dienstverlening Op de Kaart). Available AHN3 data is shown in Figure 3. This data was scanned in 2014 as shown in Figure 3.

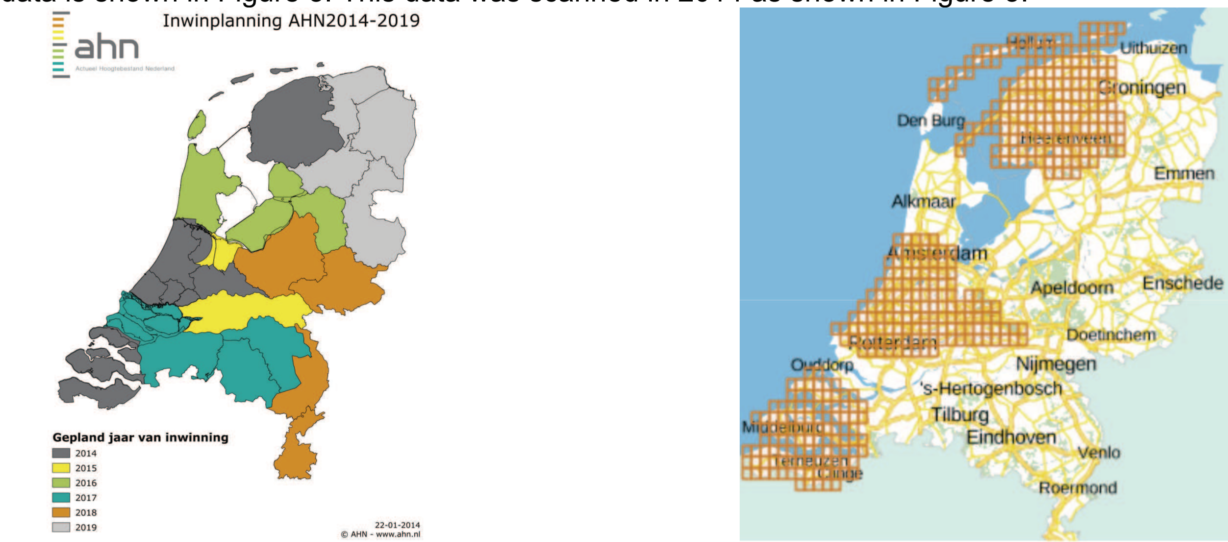


Figure 3: Left: Sampling plan AHN3; Right Available AHN3 data (https://www.pdok.nl/nl/ahn3-downloads)

The tiles that cover Ameland and used in this project are:

- 01hn1
- 01hn2
- 02cn1
- 02cn2
- 02dn1
- 02cz1
- 01hz2
- 01hz1
- 01gn2
- 01gz2

These tiles are shown in Figure 4.

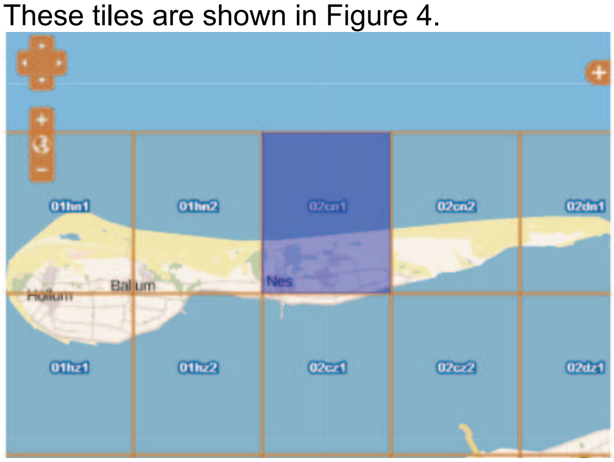


Figure 4: AHN tiles used in this project (https://www.pdok.nl/nl/ahn3-downloads)

The obtained data are point clouds which is transformed in rasters of 0,5X0,5 meter and 5X5 meter for AHN3. The 5 meter raster of AHN3 are shown in Figure 5. The difference between the 0,5 and 5 meter rasters are shown in Figure 6. These rasters are available in GeoTIFF format, while the point clouds are available in LAS files. LAS files are binary files which are commonly used to save and exchange LiDAR data. Those LAS files are compressed to a LAZ file (LASzip).

In the point cloud date there are at least eight measurements every square meter.

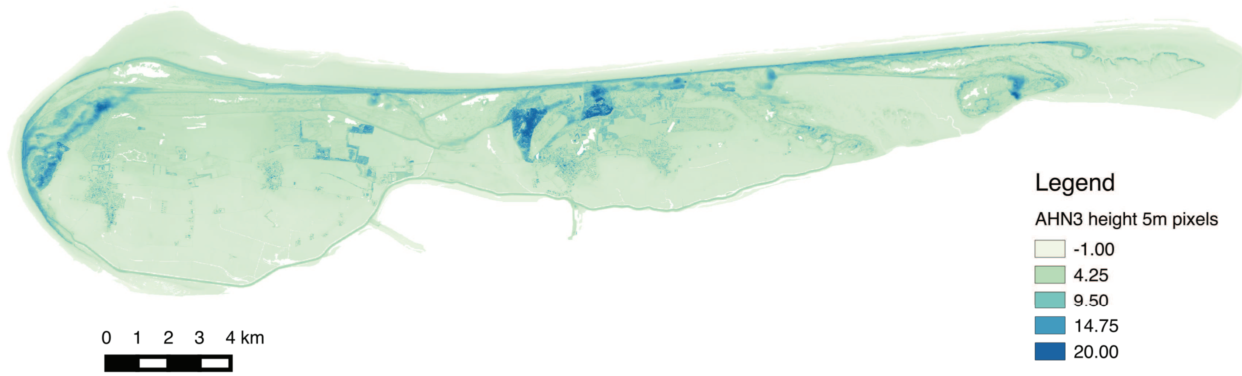


Figure 5: AHN3 height with 5 meter pixels

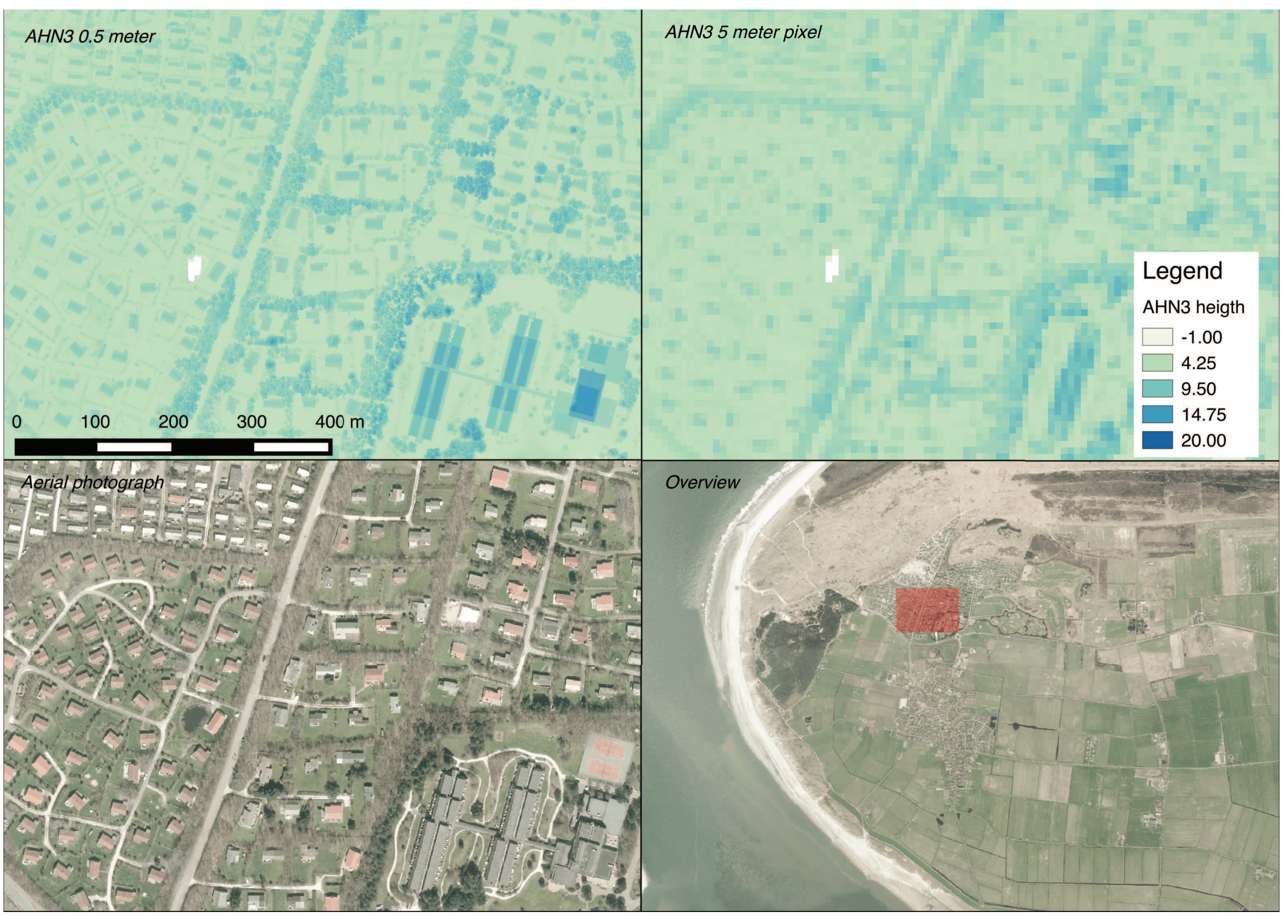


Figure 6: Clockwise from top left: AHN3 0,5 meter pixel; AHN3 5 meter pixel, Aerial photograph of inzoomed area; Aerial photograph overview

2.3 Land use (bestand Bodemgebruik)

Bestand Bodemgebruik (land use file) is a Dutch land cover map. Since 1989 the Centraal Bureau voor de Statistiek, CBS, has published every 3 to 4 year the new boundaries of the land use in the Netherlands. These boundaries are based on TOP10NL, a digital topographic file from the land register (Kadaster). This data can be downloaded from PDOK. (Centraal Bureau voor de Statistiek, 2016)

The classification of this data is in nine main classes and more subclasses. For the classification of Ameland the following classes are used:

- Main roads (Hoofdwegen)
- Airport (Vliegveld)
- Built-on area (Bewoond terrein)
- Industrial area (Bedrijfsterrein)
- Semi built-on area (Semi-bebouwd terrain)
- Recreation area (Recreatieterrein)
- Agriculture area (Landbouw)
- Woods (Bos)
- Open dry area (Open droog terrein)
- Open wet area (Open nat terrein)
- Inland waterway (Binnenwater)
- Sea (Zee)

In Figure 7 the classification of Ameland can be seen.

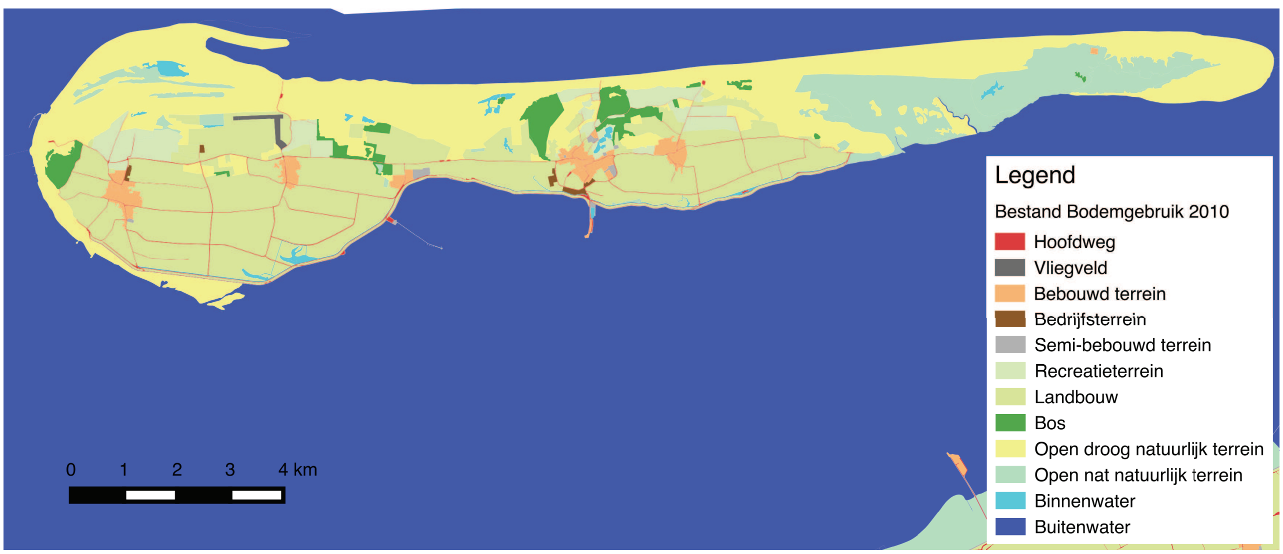


Figure 7: Landcover of Ameland in 2010. (Centraal Bureau voor de Statistiek, 2016)

2.4 Aerial photograph

Aerial photographs of The Netherlands can also be provided by PDOK. Figure 8 shows an overview of Ameland with this data. This data is originating from 2014, except for the West Frisian Islands. Through inadequate flight weather the islands are photographed during 2013. The resolution of the data is 50 cm. The green fields are agriculture and the dark green area are the woods. The white boundary of the island are the beaches. The villages in this figure are difficult to see, but if zoomed in, they appear. The grey green area are dunes.

Figure 8: PDOK aerial photograph overview of Ameland

3 Methodology

3.1 Data preparation

In this research the AHN data is used. But classification is not possible with only the height of the surface. To create more classification features the gradient and the roughness are determined.

3.1.1 Gradient

To calculate the gradient of the height matrix, a script is written. This script calculates difference in height in all directions. This is done by making a new zero matrix with the size of the AHN height matrix. This empty matrix is filled up by moving the height matrix one place in the desired direction and another empty matrix is filled up by doing this in the opposite direction. Calculating the gradient is finalized by taking the central difference. The formula of central difference is given by:

$$Q_c(h) = \frac{f(x+h) - f(x-h)}{2h}$$

where:

- f(x) = the height on position x;
- f(x+h) = the height on position x+h;
- f(x-h) = the height on position x-h;
- h = the step size.

(Numerical Methods for Ordinary Differentioal Equations, 2015)

The step size for the diagonals, like the NorthEast and SouthEast directions, is calculated by:

$$h_{diagonal} = \sqrt{h^2 + h^2} = \sqrt{2h^2} = |h|\sqrt{2}$$

To calculate the correlation between two directions, the standard Matlab function was used. If there is a linear relationship between two directions the correlation is high, around 1 or -1. If there is no relationship between two directions the correlation is around zero. When all correlations between every cardinal direction are calculated, which is done in Table 1, you can see that the opposite side of the direction has a negative correlation. In Table 1 the values with the same colour are from the same direction or negative correlated. For example, the direction south east, this is coloured in light green. The horizontal values are the same as the vertical values but may be of opposite sign. And the values of the north west direction, the opposite part of the compass rose, is negative correlated. That's why it was chosen to used only the gradient of the east part of the compass rose; North, Northeast, East, Southeast.

Table 1: Correlation matrix of different gradient directions of Ameland 5 meter rasters

	N	NE	E	SE	S	SW	W	NW
N	1,000	0,628	0,050	-0,657	-1,000	-0,628	-0,050	0,657
NE	0,628	1,000	-0,438	-0,223	-0,628	-1,000	0,438	0,223
E	0,050	-0,438	1,000	-0,491	-0,050	0,438	-1,000	0,491
SE	-0,657	-0,223	-0,491	1,000	0,657	0,223	0,491	-1,000
S	-1,000	-0,628	-0,050	0,657	1,000	0,628	0,050	-0,657
SW	-0,628	-1,000	0,438	0,223	0,628	1,000	-0,438	-0,223
W	-0,050	0,438	-1,000	0,491	0,050	-0,438	1,000	-0,491
NW	0,657	0,223	0,491	-1,000	-0,657	-0,223	-0,491	1,000

In Figure 9 and Figure 10 the gradient in south-east and east direction are shown. The largest difference between these two figures is the gradient along the north coast of the island. In the south-east direction there is a large difference between the gradient, while in the east direction almost no elevation difference can be seen. This is probably caused by the dunes, which are descending steeper to the beach than to the adjacent dunes.

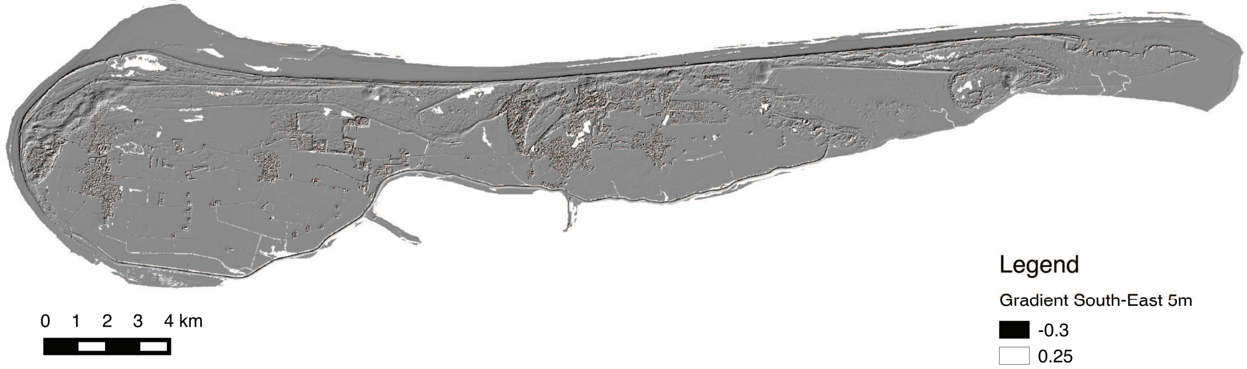


Figure 9: Gradient in south-east direction at 5 meter

Figure 10: Gradient in east direction at 5 meter

3.1.2 Surface roughness

Another feature for the classification is the surface roughness. This is obtained with the roughness index tool from the 'raster terrain'-analysis plugin in QGIS. This tool calculates the ruggedness, by summarizing the difference in elevation within a 3x3 pixel raster. If there is a clear difference in height between the pixels, the sum of the absolute differences is greater, thus the roughness indexes too. When the roughness index is small, the surface is smooth. (Nicolaides, 2013)
Figure 11 shows the roughness index of Ameland calculated with the 5 meter pixels. It shows that the woods have a higher roughness than the flat grass-lands.

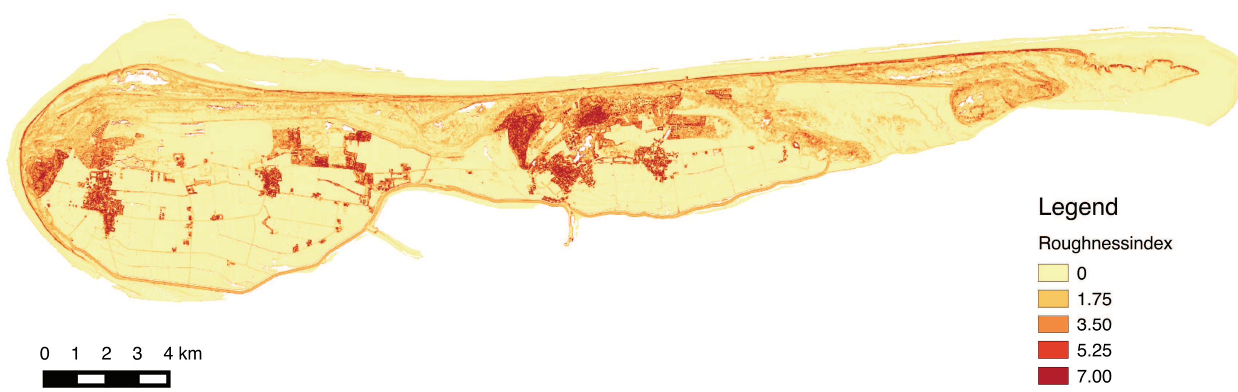


Figure 11: Roughness index of Ameland with 5 meter rasters

3.1.3 Intensity

In the LAS files of AHN3 there are more parameters inside the data, like intensity and first and last laser pulse. The intensity is the amount of reflected light. From the LAS data rasters of 0,5 meter are made with the intensity values inside the custom made rasters. In Figure 12 the intensity of tile 01HN1 is shown. The intensity for buildings and woods are low, while the intensity of grasslands is high. The figure shows also kind of stripes over the island. Probably these stripes have to do with the flights of the airplane.

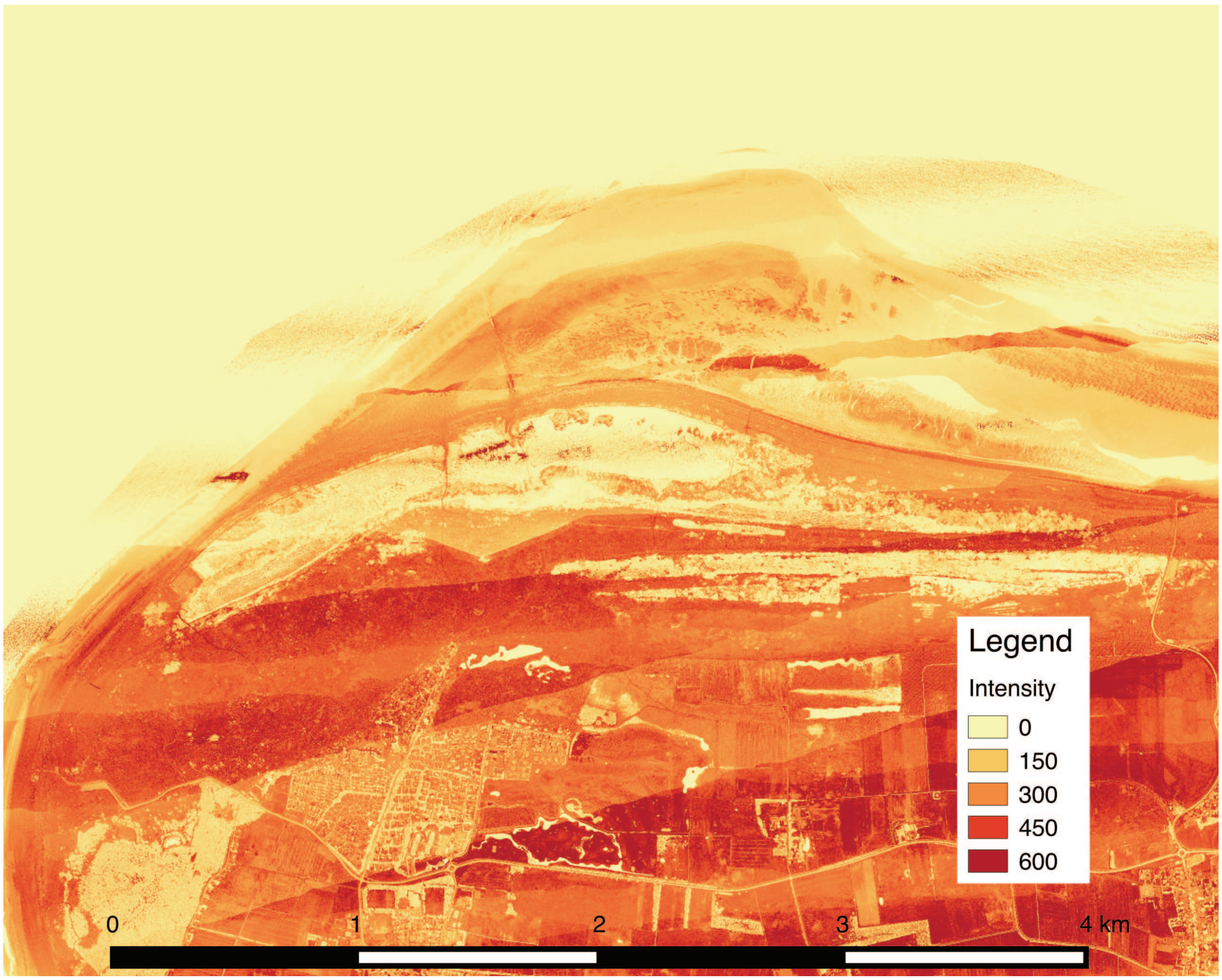


Figure 12: Intensity of tile 01HN1 in 0,5 meter

3.2 Classification

When combining all this self-made features for the 5 meter pixels with the AHN data, there are 6 bands for classification; AHN height, surface roughness and the gradients in the following directions: north, northeast, east and southeast.

The classification is done with the semi-automatic classification plugin, simply SCP from QGIS. This plugin is a free open source plugin written by Luca Congedo. As this is supervised classification, training data will be used in the classification algorithm. The training areas, Regions of Interest (ROIs), are created easily with integrated tools. Shapes are drawn over a homogeneous area of which it is known that that area belongs to the same land cover class.

Table 2: Percentages of the different classes at Ameland

Classes	Area of Ameland in [%]
Agriculture	34
Woods	3
Buildings	7
Dry area	39
Wet area	12
Recreation area	1
Airport	0,2
Water	1
Roads	2

From the BBG 2010 data the amount of area of each class were calculated. Table 2 shows the results of these calculations. Most of Amelands area is dry area and agriculture, respectively 39% and 34% of the area. Woods and buildings have together an area of 10% of the total area of Ameland. These four classes are the main and largest classes at Ameland and were chosen as classes for this project. Thus the classification classes are:

- Woods
- Buildings
- Agriculture
- Dunes/beach

From the training data the signatures are calculated automatically. There are three classification algorithms available with this plugin:

- Minimum Distance
- Maximum Likelihood
- Spectral Angle Mapping

Because there do not work with spectral data in this project and because for the computer minimum distance is easier to calculate than maximum likelihood, minimum distance was chosen as the classification method.

The classification will be done for the 5 meter pixels and for the 0,5 meter pixels. This are the available sizes of AHN. Shown in Figure 6 are differences between 0,5 and 5 meter. For example, in the building areas trees are distinguished in 0,5 meter rasters, while in the 5 meter raster the overall pixels will be a little bit higher.

3.2.1 Minimum distance

The minimum distance algorithm calculates the Euclidean distance d(x,y) between the signatures of image pixels and training signatures, according to the following equation:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

where:

- x = signature vector of an image pixel;
- y = signature vector of a training area;
- n = number of features.

For every pixel of the input picture, in this case the 6 features derived from the elevation raster, the distance to the signature is calculated. Every pixel is assigned to the class it is closest to (smallest Euclidean distance). This is according to the following discriminant function:

$$x \in C_k \iff d(x, y_k) < d(x, y_j) \forall k \neq j$$

where:

- C_k = land cover class k;
- y_k = signature of class k;
- y_j = signature of class j.

(Congedo, 2015)

3.2.2 Training data

For the selection of the trainings data the PDOK aerial photograph is used. In the aerial photograph the different classification classes can be distinguished easily. As mentioned before four classes were chosen:

- Woods;
- Buildings;
- Agriculture;
- Dunes/beach.

In Figure 13 the selected classification shapes are shown in green and blue. The green shapes, classification shape 1, were drawn roughly. The blue shapes were drawn more carefully, and attention was paid to outliers. Outliers are for example buildings in meadows or open area, like gardens or streets in the building area.

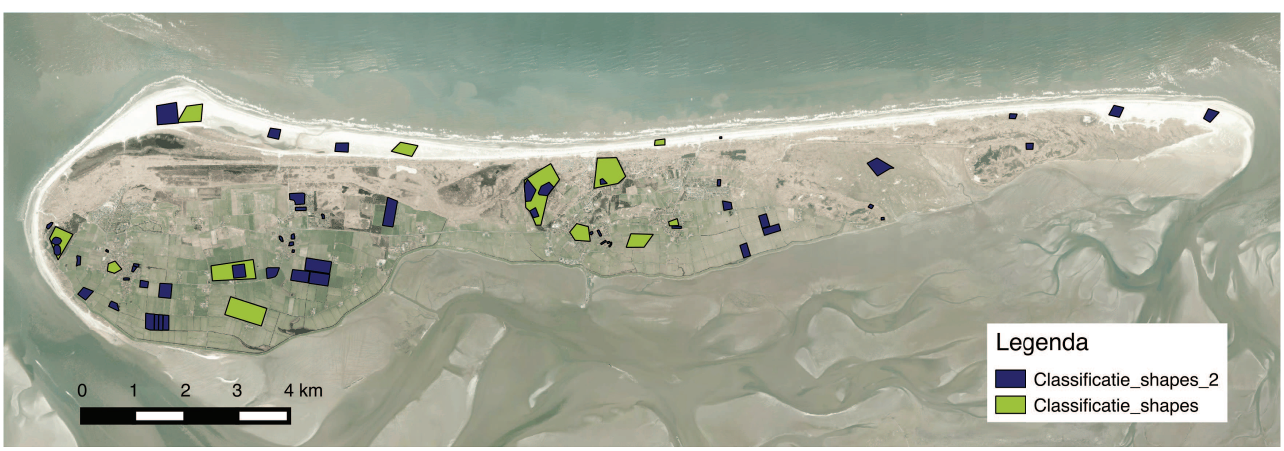


Figure 13: Selected classification shapes Source: PDOK aerial photograph

3.2.3 Signatures

After making the shapes the plugin calculated the signatures of every class. These values are distinguished into the different classes by the plugin. In the next chapter the signatures of the different classes will be discussed.

4 Results

In this project the classification is done for different size rasters: 5 meter and 0,5 meter raster. First the results for the 5 meter raster are discussed followed by the 0,5 meter raster result.

4.1 Results of 5 meter rasters

Figure 14 shows the result after running the semi-automatic classification plugin with shape 1. Figure 15 shows the result of shape 2. The difference of these two classifications are calculated, and is shown in Figure 16. It is the subtraction of the classification values of shape 1 and shape 2: Classification1-Classification2. In Figure 16 is seen that the most difference between the two classification shapes are the boundaries of the same area. Also in Figure 16 there are three closer views of the areas where the most difference was.
The yellow area is on the beach, but was in both cases classified as agriculture. Shape 1 classified here better than shape 2, because in shape 1 more pixels are classified as beach as they should be.
The blue area are meadows. Here the big difference in both classifications is that shape 1 classified this as beach while shape 2 classified it as agriculture. In this case shape 2 fits more as if we look at the aerial photograph.
The purple area is a salt marsh (kwelder), where water of the Waddenzee flows in and out. The classification with shape 1 classified this area as beach and agriculture. On the other hand, shape 2 classified this area as well as beach and agriculture, but with more agriculture than shape 1. In this case it's difficult to agree with one shape, because a salt marsh is nor agriculture nor beach.

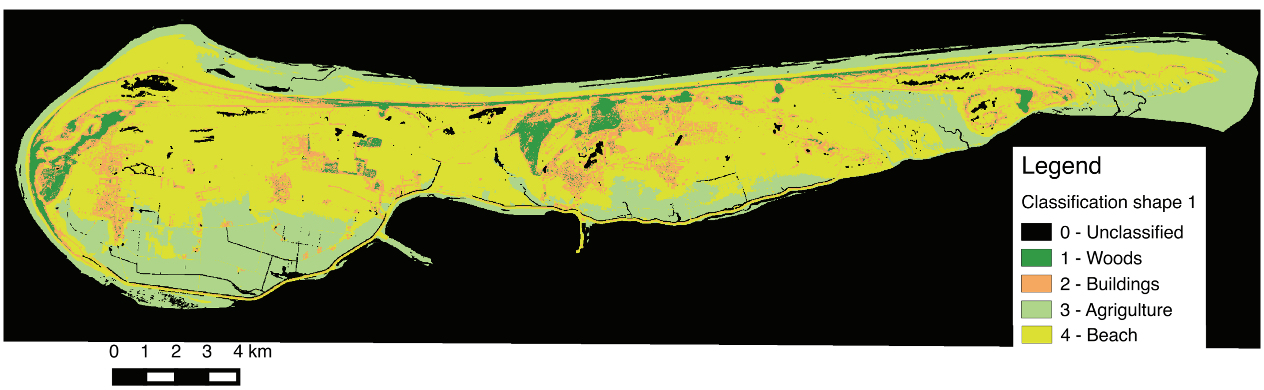


Figure 14: Shape 1 classification with minimum distance

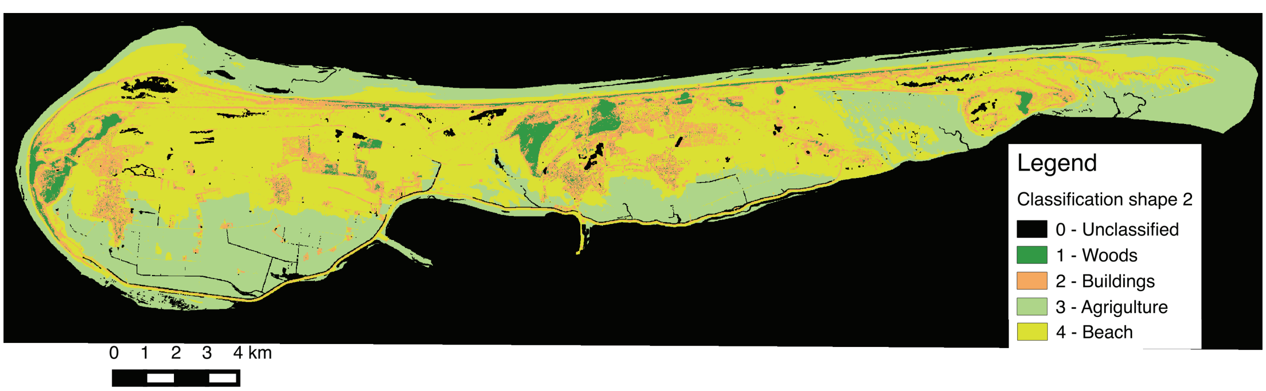


Figure 15: Shape 2 classification with minimum distance

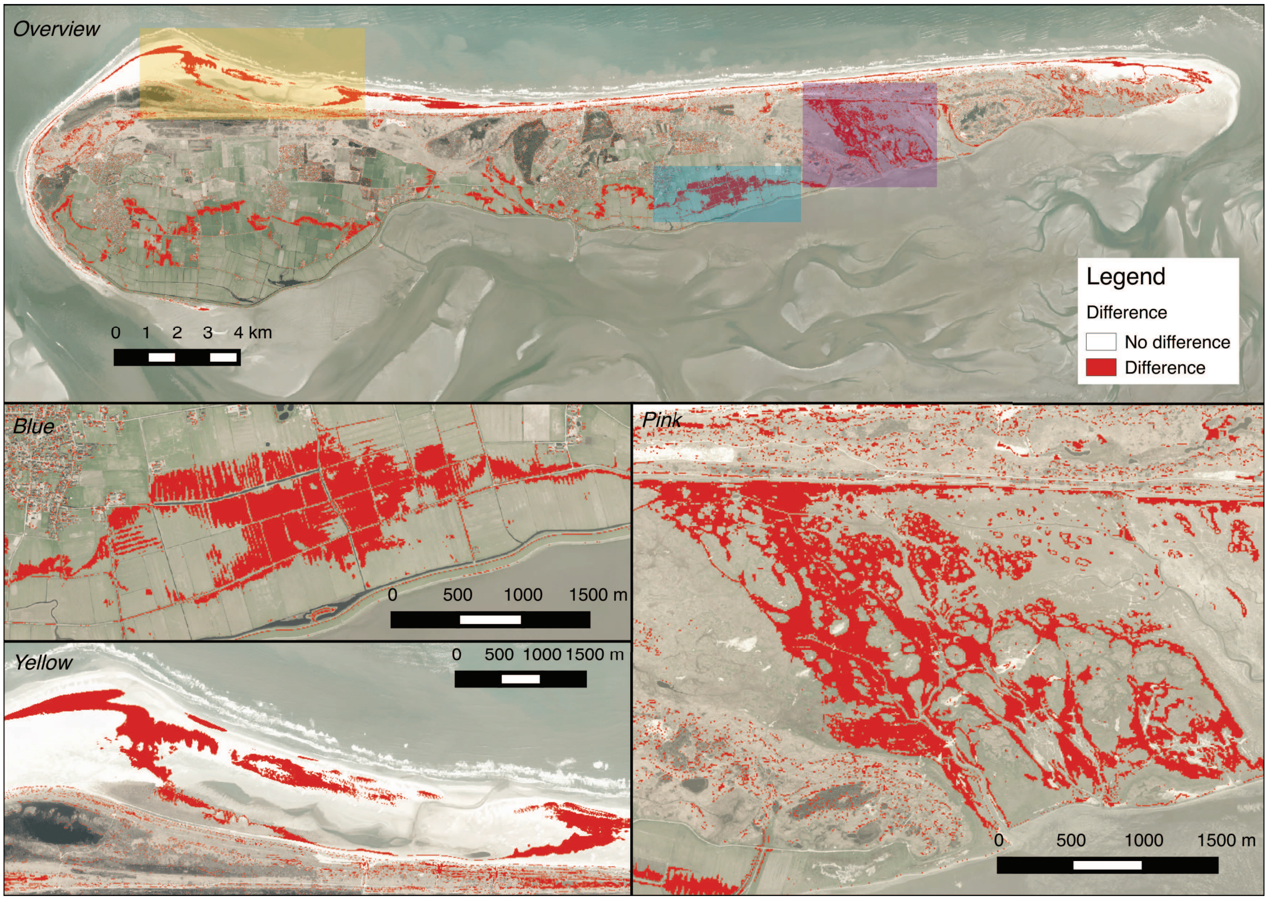


Figure 16: Differences between shape 1 and shape 2 in classification and background PDOK aerial photograph

For validation of the classification, shape 2 is compared to the Bestand Bodemgebruik from 2010. In this file are more classes than the four classes used in this project. To compare this two files, it was necessary to convert the classes of the Bestand Bodemgebruik to our own classes. Table 3 shows the classes of BBG subdivided in the our classification classes.

Table 3: Convert table from BBG to own classes

	Own classes		BBG 2010
1.	Woods	60.	Woods
2.	Buildings	20-24	Buildings
		30-35	Semi-Buildings
		44.	Place of residence
3.	Agriculture	40.	Park
		41.	Sport area
		42.	Public gardens
		51.	Agriculture
4.	Beach	61.	Open dry area
		62.	Open wet area
5.	Other (not in classification)	70-78	Inside waters
		80-83	Outside waters
		10.	Track
		11.	Roads
		12.	Airport

In Figure 17 the subdivided classes of BBG 2010 are shown. With this file the validation of the classification will be done. Of course is this not the best subdivision, because BBG is used for whole Netherlands and has more general categories than there are on Ameland. For example, the salt marsh (kwelder), in Figure 16 the purple area, is very specific for this part of The Netherlands. According to the classification shows in Figure 17 it is now beach, but actually it's a salt marsh. This is happening too for al the dune and coast area of Ameland.

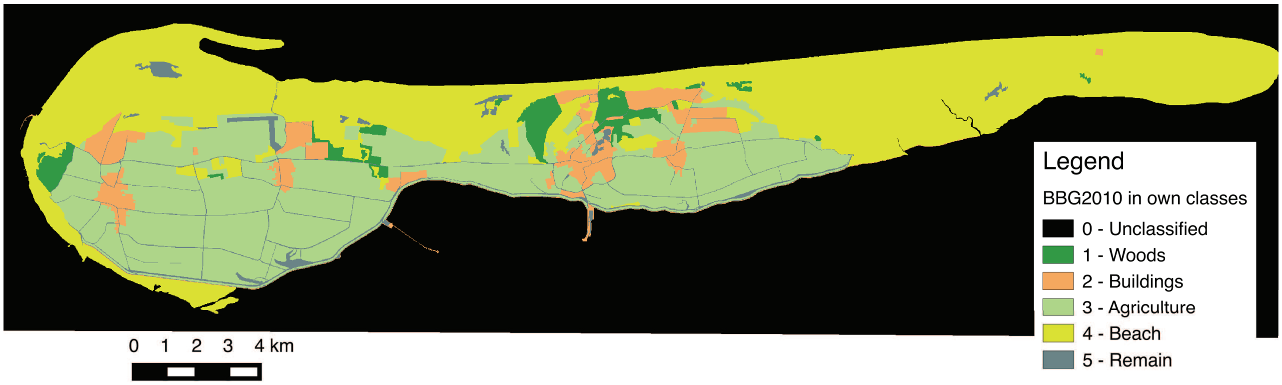


Figure 17: BBG2010 in the four classes used in this report

4.1.1 Validation

As mentioned above, the validation was done with the modified BBG file, shown in Figure 17. To compare the validation with the own classification a confusion matrix was made. This was done with Matlab. Every pixel is compared with the (validation) pixel of the BBG file. Pixels of the same class were counted. For example, there are 38814 pixels where the BBG pixel is woods and the own classification pixel is wood too. After this is done for all classes, the confusion matrix is made, and shown in Table 4.

Table 4: Confusion matrix of shape 2 at 5 meter pixels

5 meter	BBG 2010					
Classification data	1 - Woods	2 - Buildings	3 - Agriculture	4 - Beach	5 - Remain	Total
1 - Woods	38814	7137	1437	42418	351	90157
2 - Buildings	32271	83467	32890	169585	5216	323429
3 - Agriculture	1	4244	393892	280120	8302	686559
4 - Beach	5109	63124	396356	685730	34549	1184868
5 - Remain	0	0	0	0	0	0
Total	76195	157972	824575	1177853	48418	2285013

From the values of Table 4 the overall accuracy can be calculated:

$$\frac{sum\ of\ diagonal}{total} = \frac{1201903}{2285013} = 53\%$$

That means that half of the pixels are classified correct. Which class is best classified, the "user's accuracy" can be calculate. The user's accuracy is an indication of the probability that a pixel classified actually represents that category on the ground. The user's accuracy is calculated by the total number of correct pixels in a category divided by the total number of pixels that where classified in that category (Congalton, 1991). For example, the user's accuracy of woods is 38814/90157=43%. This calculation is done for every class, and is shown in Table 5.

Table 5: User's accuracy of the classification of shape 2

User's accuracy	Percent
1 - Woods	43%
2 - Buildings	26%
3 - Agriculture	57%
4 - Beach	58%
5 - Remain	0%

When looking at Table 4 the classification has most problems with distinguishing agriculture and beach. This can be seen through the agriculture of the validation data is classified a lot as beach and visa versa. The amount of pixels classified as beach, but actual agriculture and visa versa is really high. The distinction between woods and buildings is better, the false positive and false negative are for these classes lower than for the different between agriculture and beach.

4.1.2 Differences (between BBG and the own classification)

The confusion matrix contains lot of values. To show the difference between the classification in this project and the BBG classification for the four classes, the difference in outline of the difference data was calculated. Figure 18 shows the difference of Ameland of the AHN data and the BBG data in red. The biggest difference of this two outlines is the beach part, like on the north coast. This is caused by the dynamics of the island and because airborne data is made with ebb, low tide. In the following comparison this part will be disregard.

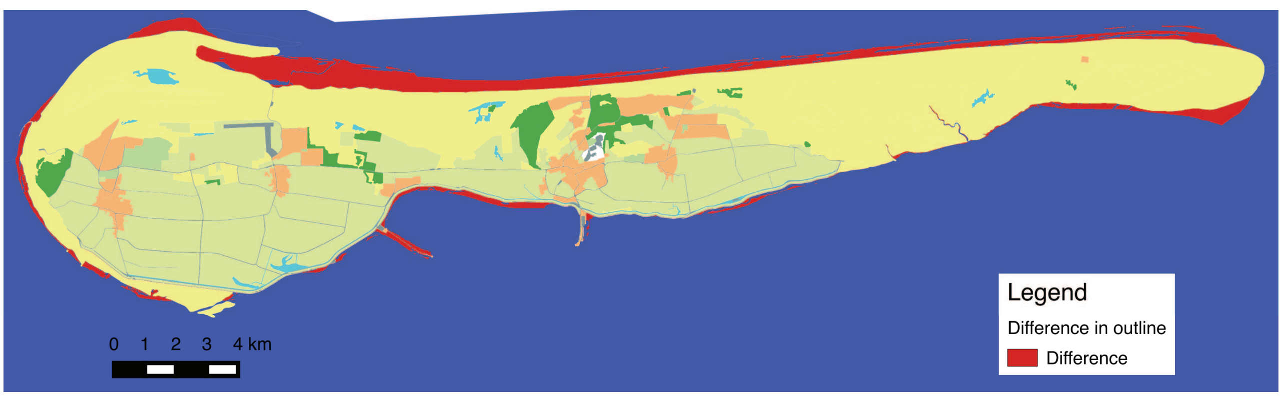


Figure 18: Difference between in outline AHN data and BBG

Figure 19 shows the differences between the BBG and the classification with shape 2. There are four areas, purple, blue, yellow and red, which will be discussed separately. With the help of the PDOK aerial photograph it is decided which classification is more correct.
The purple area are meadows. As shown in the overview of Figure 19, there is a lot of differences between the BBG classification and the classification of this project. Most of the purple areas are meadows and this is classified with the BBG too. But in this project these areas are classified as beach. In this case the classification of the BBG was correct.
The yellow area is a salt marsh (kwelder). Originally the BBG classified this part as open wet area, as shown in Figure 7. But because the number of categories is reduced in this project, this specific part of Ameland was not classified as open wet area but as beach. This part is classified as beach and agriculture in the project classification and actually it's not the same as a salt marsh, but this classification comes near. Both classifications are here not correct but this part is so specific that it's hard to classify this.
The red area is dunes and beach. In the BBG modified classification this area is classified as beach. The classification of this project classified this part as agriculture. The BBG classification is more correct.

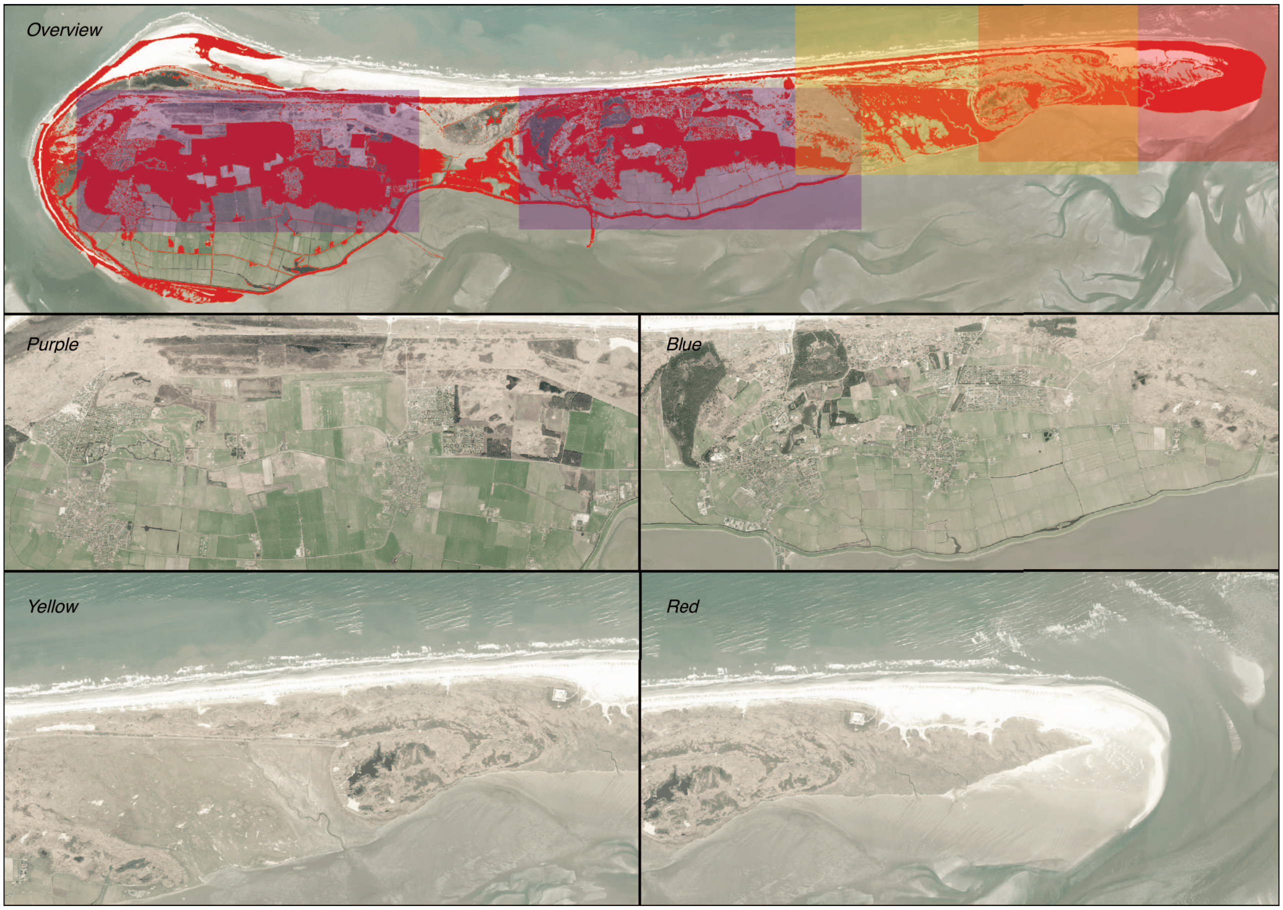


Figure 19: Differences between BBG and own classification with PDOK aerial photograph background

4.1.3 Signatures

To have an idea on how the semi-automatic classification plugin classifies, the underlying signatures are discussed in this part. The values of signatures are used to distinguish between categories. To have an idea of the difference in the values of the signatures, they are shown in Table 7 and Table 8. Table 8 shows the signature values for every feature of each category. There is a large difference between the values. The elevation and roughness are relativelly large numbers, while the gradients have values close to zero. That's why Table 7 didn't shown the real values, but a ratio. The values are scaled between 0 and 1 for every band, and plotted in a bar graph. The scaling was calculated according:

$$\frac{value - \min{(values)}}{\max(values) - \min{(values)}}$$

When the result is 1, this was the maximal value and 0 the minimal value of the feature. For example, the ratio for the AHN height at shape 2 will calculated. The heights of the four classes are shown in Table 6. As an example the calculation is done for the buildings:

$$\frac{5{,}01 - 1{,}44}{14{,}15 - 1{,}44} = 0{,}28$$

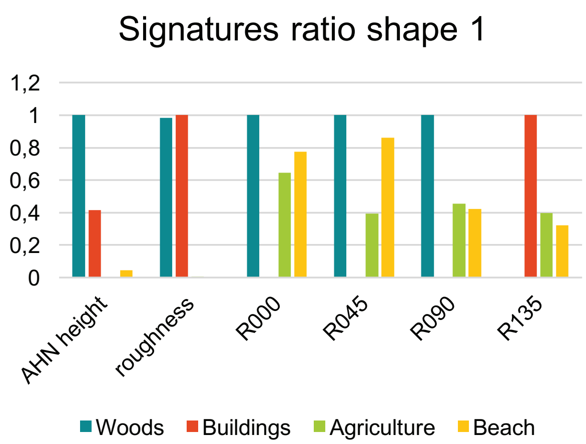
Table 6: AHN heights for the categories of shape 2

Classes	**AHN height**
Woods	14,15
Buildings	5,01
Agriculture	1,44
Beach	1,74

The different bands on the x-axis are the six features:

- AHN height;
- Roughness index;
- R000, the gradient in the north direction;
- R045, the gradient in the north east direction;
- R090, the gradient in the east direction;
- R135, the gradient in the south east direction.

Table 7: Signature ratio of shape 1 and 2



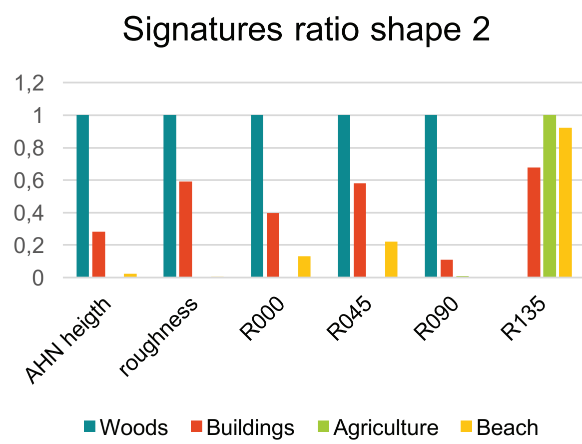
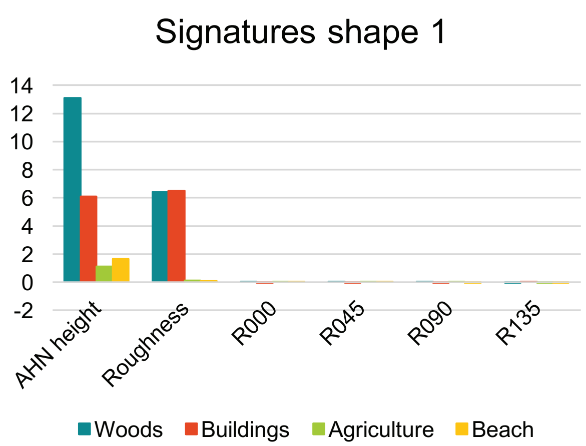
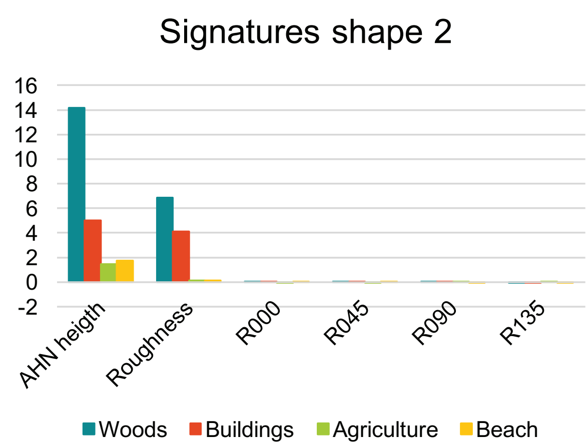


Table 8: Signature plots of shape 1 and 2





To compare the results of the two training areas each feature will be discussed below.
AHN height; its clear that there is difference in height between the categories. For Ameland, trees are between the 13-14 meter high. Therewith woods have the same height as a trees. There is no significant difference between the two classification shapes. In shape 2 the woods are a little bit higher than for shape 1. For buildings the height is between 5 and 6 meters high. There is less height difference between agriculture and beach.

Roughness; the roughness for woods and buildings at shape 1 is more or less the same, while there is a clear difference at shape 2. The roughness for agriculture and beach are in both shapes difficult to distinguish.
Gradients; the gradients are calculated in four different cardinal directions. The values of the gradients are small, but Table 7 shows that the values of the wood are for three of this cardinal direction (north, north east and east direction) are high in ratio. It is remarkeble that woods don't have the highest gradient in the south east direction, R135, but the lowest. The difference between the two shapes is that for shape 1, the inaccurately drawn shapes, the gradients of the building is the lowest for the north, north east and east directions. While for shape 2, the carefully drawn shapes, the gradient values are just in the middle.

Overall are the values of the signatures of woods the highest in both trainings shapes, except for the gradient in the south east direction. For this is no explanation yet. Beach and agriculture are more or less the same, they have more or less the same values.

4.2 Results of 0,5 meter rasters

First the classification of total Ameland was done with the same signatures which have been used for the 5 meter pixels. This was done by merging all the feature layers of the 0,5 met raster together. Then the classification with the signatures of the 5 meter rasters was used. The result is shown in Figure 20. It is clear that this result is not correct. The error was found in the merged layers. In the third layer half of the file was empty and following layers where completely empty. This might have been caused by the lack of working memory (RAM) on the laptop used.

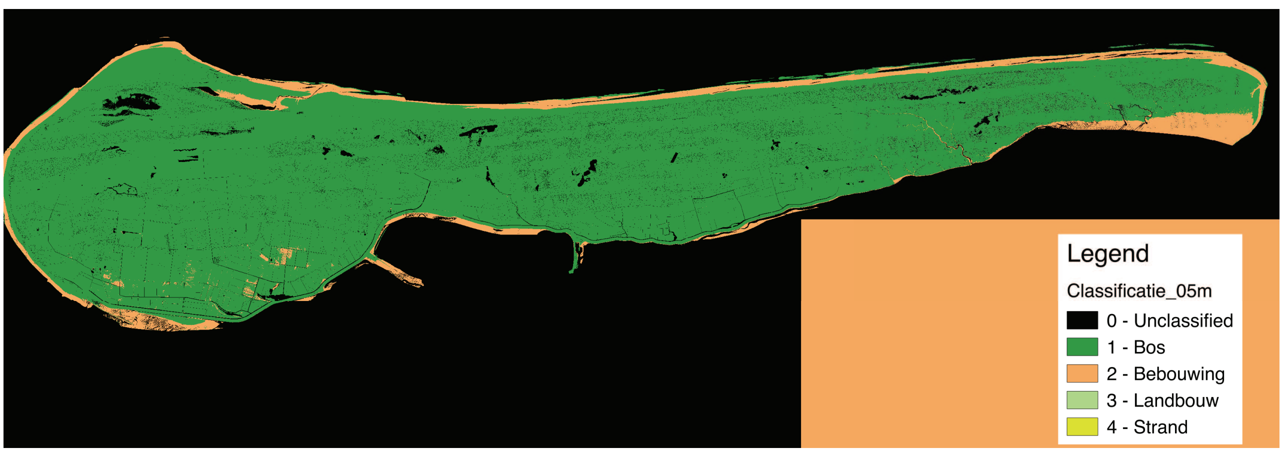


Figure 20: First result classification 0,5 m rasters

When the rasters are finer, the amount of data is larger. That's why it was chosen to work on smaller areas, to work and test faster. In Figure 21 the study area is shown, it is part of tile 01HN1. Inside the study area there are woods, building area, agriculture and beaches.

Figure 21: Study area for 0,5 meter pixels

Figure 22 gives the classification results of tile 01HN1 for 0,5 meter when classified with the same six features as in the classification of the 5 meter raster. Of course these results are not correct, every pixel is classified as beach.

Figure 22: Classification result with 6 bands of 0,5 meter

4.2.1 Extra features

When the classification of the 0,5 meter raster didn't work with the six features, a test is done with extra created features for tile 01HN1. For this tile 15 layers were calculated from the LAS data. Al the points are organized in 0,5 meter pixels. The layers are as follows:

- AHN height;
- Roughness index;
- R000, the gradient in the north direction;
- R045, the gradient in the north east direction;
- R090, the gradient in the east direction;
- R135, the gradient in the south east direction;
- Average height;
- Minimal height;
- Maximal height;
- Second percentile of height;
- 98th percentile of height;
- Range between 2-98 percentile of height;
- Count of measurements;
- Intensity;
- Standard deviation of intensity.

The first six features are the same as previous discussed features. The average height is more or less the same as the AHN height. The minimal height and the second percentile of the height, the maximal height and the 98th percentile of the height are very similar too. The range is the difference between the 2nd and the 98th percentile of height. The count of measurements is the amount of points in the 0,5 meter pixel, this is shown in Figure 23. There are stripes with more measurements in the figure shown. The intensity, the amount of reflected light.

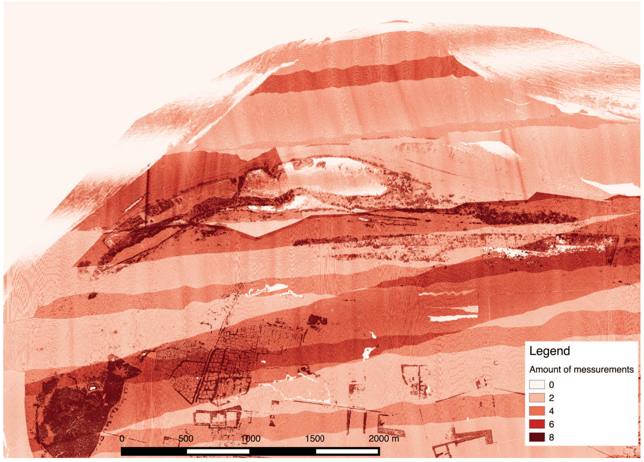


Figure 23: Measurement count on tile 01HN1

The classification of tile 01HN1 with the extra layers was done with the trainings shapes as shown in Figure 24. The result of this classification is shown in Figure 25 and Figure 26. Figure 25 shows the classification of the whole tile, while Figure 26 gives a closer view. On first thought it seems disappointing, but zoomed in the results are sufficient. In the overview there are stripes from west to east classified as buildings. But in the closer view is seen that the classification makes distinguish between the trees and the houses.

Figure 24:classification shapes for tile 01HN1

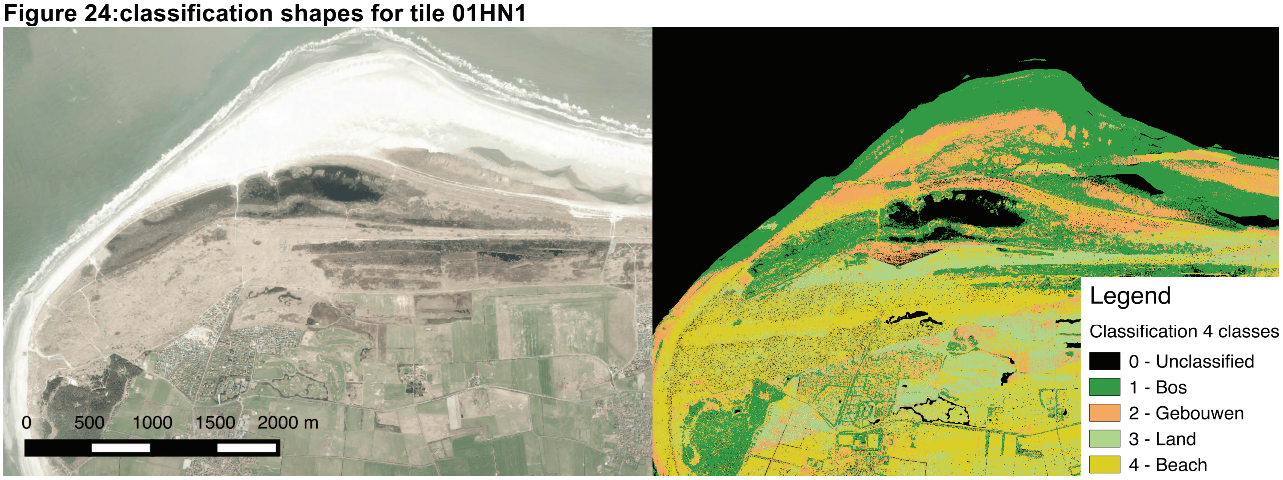


Figure 25: Left PDOK aerial photograph, Right the classification with extra features and 4 classes

Figure 26: Left PDOK aerial photograph, Right the classification with extra features and zoomed in

Table 9: Confusion matrix of four catergories,5 meter pixels

0,5 meter	BBG 2010					
Classification data	1 - Woods	2 - Buildings	3 - Agriculture	4 - Beach	5 - Remain	Total
1 - Woods	1026574	707423	411571	7679727	265514	10090809
2 - Buildings	185220	828450	811767	5905849	185673	7916959
3 - Agriculture	64287	525331	5172065	4525228	384020	10670931
4 - Beach	70862	862414	6828203	9610742	369159	17741380
5 - Remain	0	0	0	0	0	0
Total	1346943	2923618	13223606	27721546	1204366	46420079

The overall accuracy of the 0,5 meter pixels with four categories is 36%. The user's accuracy, shown in Table 10, were overall low. The user's accuracy of agriculture and beach are close together.

Table 10: User's accuracy of the classification of 0,5 meter rasters

User's accuracy	**Percent**
1 - Woods	10,2%
2 - Buildings	2,3%
3 - Agriculture	0,6%
4 - Beach	0,4%
5 - Remain	0,0%

4.2.2 Signatures

The signature plot of the classification of tile 01HN1 is shown in Figure 27. The different features will be discussed separately. Because average height, minimal height, maximal height, second percentile of height and 98th percentile of height are more or less the same, they are not being discussed.

Elevation; there is a difference between the elevation of the 4 categories. For woods the elevation would be high, around 12 meters. For buildings the elevation is just below the 5 meters. For agriculture and beach, the elevations are respectively 2,5 and 2 meter. There is probably enough difference between the first three categories to distinguish them. For the categories agriculture and beach, the differences in these features are minimal. These features didn't distinguish the two categories.

Roughness; there is a clear difference between the values of the roughness. Woods are the roughest of the four categories. There after the category agriculture followed by beach and buildings. This is remarkable, one would like to expect that after woods the buildings are the most roughness.

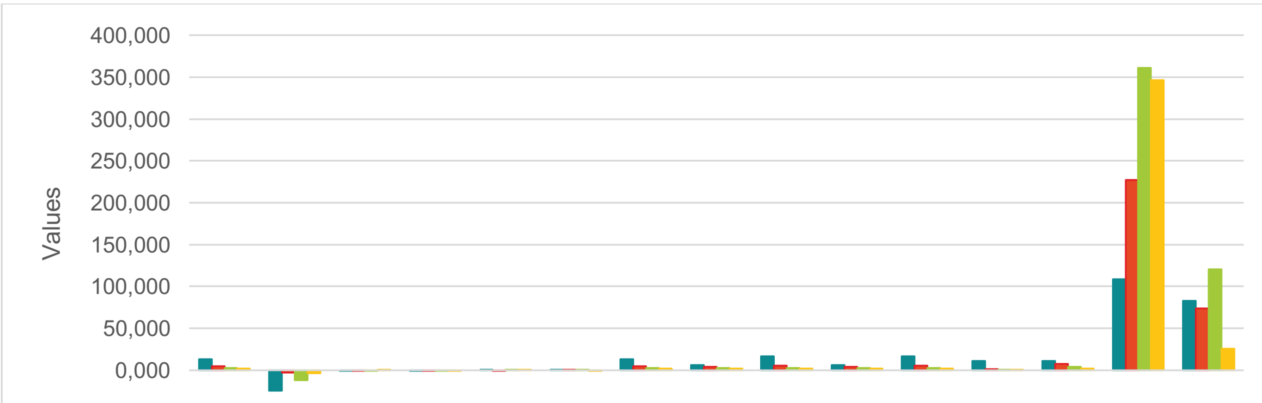
Gradients; overall the gradients are really small numbers. The difference between the categories are small too, but there is a minimal difference. The most difference in the values is at the gradient in the south east direction.

Range between 2nd and 98th percentile; the range is the difference between the values of the 2nd and the 98th percentile. For woods is this value around the 10 meters, and therewith the biggest range of the categories. For buildings the range is low, but not so low as the category agriculture. The difference in range of agriculture and beach were nil.

Count of measurements; this is a strange feature, because it has noting to do with height. But there is a big difference in values between the categories. There are lot measurements at woods and this will be smaller for buildings. For agriculture and beach, the amount is small, and there is not a big difference between them.

Intensity; the intensity is the amount of reflected light. There is a clear difference between the categories. Here is the opposite as the other features, the woods have the smallest value and the agriculture the highest. Buildings is just in the middle. Beach is close to the Agriculture. It is difficult to distinguish between agriculture and beach.

Standard deviation of intensity; for this feature there is a difference between the categories. For agriculture the standard deviation is relative large. The standard deviation of the category woods is in the middle and buildings has the lowest standard deviation. Beach had a small value. There is thus a big difference of values for agriculture and beach. This would be a good feature to distinguish agriculture and beach.



	Elevation	Roughness	R000	R045	R090	R135	z_avg	z_min	z_max	z_2	z_98	z_r	Count	I	I_std
Woods	12,65	-24,6	-0,05	-0,04	0,016	0,042	12,69	6,090	16,79	6,274	16,72	10,63	10,66	108,5	83,01
Buildings	4,736	-3,26	-0,08	-0,05	-0,00	0,060	4,770	4,055	5,542	4,069	5,511	1,456	7,668	227,3	73,71
Agriculture	2,516	-11,8	-0,00	-0,00	0,001	0,001	2,518	2,506	2,529	2,507	2,528	0,022	3,707	361,3	120,4
Beach	1,935	-3,85	0,009	-0,00	0,016	-0,01	1,930	1,925	1,935	1,925	1,935	0,010	2,039	346,4	25,60

Figure 27: Signatures of tile 01HN1 with 15 layers and four categories

5 Discussion

The results of the different classifications are given in the previous chapter. In this chapter the results will be discussed which problems were identified and are the results according reality or more or less wrong.

5.1 Scale

The classification was done with two difference raster sizes of the AHN height, the 5 meter and the 0,5 meter. The 5 meter raster was classified with two different shapes. The shapes of shape 1 were drawn roughly without taking in to account the outliers, like farm houses in meadows. This kind of outliers were taking in to account by drawing the shapes of shape 2. The effect of this kind scale difference was nil. Only at boundaries of classes difference was seen. The difficulty distinguishing categories like agriculture and beach was not better classified with the carefully drawn shapes.

The 0,5 meter rasters are 100 times as small as the 5 meter raster. A lot more information is there in this data. Because the rasters are smaller more details appear. This was shown in Figure 6. Therefore the result of this classification had more details, like the trees in villages shown in Figure 26. That these small details appeared was surprising. But when compared with the BBG, the results were bad. Probably this caused through the scale of the BBG. The BBG document didn’t have a lot details in the shapes of the categories. When there is a city, the whole city is classified as buildings, the trees along a street too. When comparing the classified 0,5 meter raster with the BBG these trees where not shown in the BBG, so the user’s accuracy will be low. Actually for the comparison of the results of the 0,5 meter raster requires a more detailed comparison.

5.2 Classes

Ameland is a West Frisian Island (Waddeneiland) and has special nature, like salt marsh and dunes. The BBG has a lot global classes, for example the dunes and beach belong to the category open dry area. But beach and dunes have a lot of difference in elevation. Dunes have more elevation difference and more vegetation like bushes and beach grasses than beaches. Still in this project there didn’t make distinction between these two. These is overall not really a problem, this own classification takes the beach and dunes together, what the BBG does too. This is a problem in the salt marsh, this is classified in the two rasters as agriculture, in the BBG as open wet area, but after converting it was beach.

The classes woods and buildings are better chosen classes. Although for the finer raster the trees between houses are not woods. A better name for this category was ‘trees’.

The name agriculture is fine. Classification categories can select on how they behave in the data. For example, the categories beach and agriculture, both flat, named as one category “bare land”.

5.3 Oddities

In this project a few oddities were seen. First, the amount of measurements was not equal everywhere. Probably the amount of measurements had a negative influence on the classification. In the signatures is shown that there is a difference between the amount of measurement between the categories. At a more roughly face, like the woods and buildings, the amount of measurements is higher than at the relative flat surfaces of agriculture and beach. An explanation for these stripes are the flight lines of the airplane. Figure 23 shows that the stripes are from west to east and one stripe is diagonal. The stripes from west to east are less wide than the diagonal stripe. Probably the less wide stripes are caused through the overlapping of the flight lines, the sides of the flight lines. The diagonal line is probably caused because the airplane flew over again.

5.4 Features

In this project there were used different features for classification the elevation data. For the 5 meter raster 6 features were used. For the 0,5 meter raster the same 6 features and 9 extra features were used. Are all features necessary? As shown in the signatures plots the gradients in all directions are low. The difference of the gradients between the categories are low too. This feature is probably not necessary for classification with height data. This was not tested in this project, thus certainly unknown.
Features like average height, minimal height, maximal height, second percentile of height and 98th percentile of height are more or less equal. These features are not necessary for the classification, because they carry the similar information.
An extra feature is the RGB colours of the aerial photograph. This feature would help distinguish for example agriculture and beach. In our case, beaches are white in the aerial photograph, while agriculture is green.

5.5 Challenges

One of the challenges was the dune area. The appearance of dunes can vary, wet or dry, with vegetation or without. Because of this variation the dune area was confused with agriculture. This was a problem for both rasters used.
The second challenge was to handle the size of data. Soon Geotiffs with a size of 20 GB or larger were generated. Of course there was a limitation on the amount of storage on the hard disc. But the biggest problem was the amount of RAM on my laptop. To load and calculate the Geotiffs in QGIS and Matlab took a long time. Sometimes the files were too large to load in Matlab. Then another laptop was used. Processing the data took a lot time and patience.
An advice is to check the data when a calculation is finished. For example, the merge file of Ameland of the 0,5 meter raster did not have all layers after merging. I checked only the first three layers, but not the last three. They weren't complete.
An advice for handling big data is to cut the tiles in smaller parts. The less work is to use whole tile, if this through the size not possible cutting tiles is an option. When in one tile not all the classes are, don't cut and merge tiles. This cost a lot time. Figure 28 shows this option in blue. Use small tiles like the green one. It is faster and when there is a good result more tiles can be done.

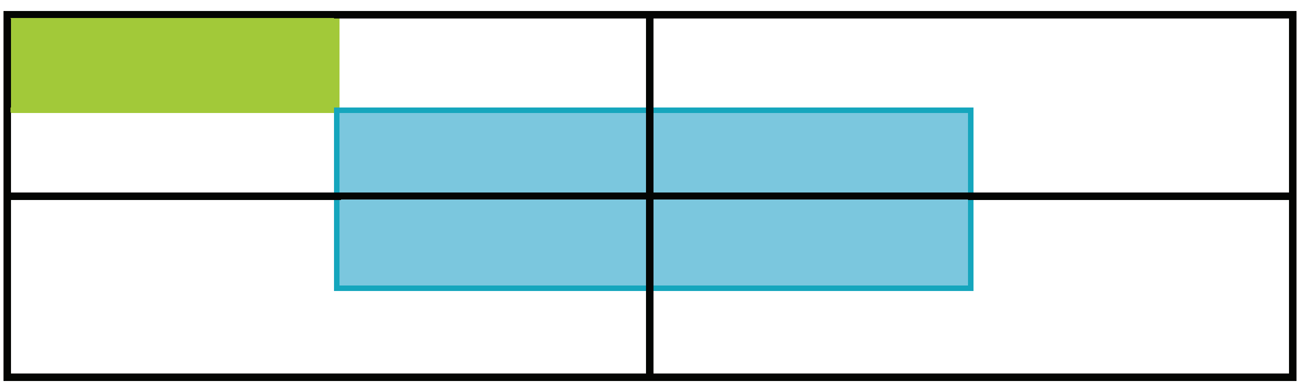


Figure 28: Cutting tiles

6 Conclusion

The research question of this bachelor thesis was to study how airborne laser scan data sampling Ameland can be used to produce a land cover map. From airborne laser scan data one can make different features, which are used to distinguish different land cover categories. For the area of interest, the West Frisian Island Ameland was used. For such islands the characteristics are beach, dunes, agriculture, woods and small villages. The classes used for classification were:

- Woods;
- Buildings;
- Agriculture;
- Beach.

After classification the class beach was not the correct name, it contains the dunes and beach.
The classification was done for 5 meter and 0,5 meter rasters. The classification of the 5 meter raster was done with six features. The result of this classification method was rough. The woods and villages were overall good classified. This method had problems to distinguish agriculture and beach. The classification of the 0,5 meter raster was first done with the same six features as for the 5 meter raster. The result of this was unsuitable. Then more features were used to improve the classification method. This resulted in a detailed map, trees were classified in villages. This classification method had difficulties with distinguishing beach and agriculture too.

Overall the classification method had difficulties to distinguish beach and agriculture. Woods and buildings were classified good enough. To have a better classification the category “beach” can split in the two categories beach and dunes. Or a category “bare land” can used to cover the flat surfaces of beach and agriculture.
When a rough land cover map is made with classification with elevation the classification method of the 5 meter raster is useful. This method was relatively fast and the result was detailed enough to distinguish villages, woods and flat surface like agriculture.
The 0,5 meter raster method is more detailed, but a powerful computer is required. Also there need to be more tests, which features are really necessary for classification the 0,5 meter raster and or there are more features to distinguish between beach and agriculture, like RGB colours.

Bibliography

(2010). In G. Vosselman, & H.-G. Maas, *Airborne and terrestrial laser scanning.* Whittles Publishing.

(2015). In C. Vuik, F. Vermolen, M. v. Gijzen, & M. Vuik, *Numerical Methods for Ordinary Differentioal Equations* (pp. 25-27). Delft: Delft Academic Press.

Centraal Bureau voor de Statistiek. (2016, 02 15). *CBS*. Retrieved 02 15, 2016, from Statline: http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=37230NED&D1=17-18&D2=101-650&D3=l&LA=EN&HDR=T&STB=G1,G2&VW=T

Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of the Environment , 37:35-46*.

Congedo, L. (2015, October 29). *semiautomaticclassificationmanual-v4.readthedocs.* Retrieved March 7, 2016, from http://fromgistors.blogspot.com/p/theinterface-2.html: http://semiautomaticclassificationmanual-v4.readthedocs.org/en/latest/remote_sensing.html#basic-definitions

Nicolaides, L. (2013, May 20). *Qgis planet*. Retrieved from qgis.org: https://planet.qgis.org/planet/tag/terrain%20analysis/#top

www.LiDAR-UK.com. (2016). Retrieved 02 17, 2016, from lidar-uk: http://www.lidar-uk.com/how-lidar-works/