

Supervised machine learning of fullcube hyperspectral data

Extremely randomized trees for classification of tree species

MASTER THESIS

to obtain the degree "Master of Science (MSc)"
at the Department of Geoinformatics, University of Salzburg.

SUPERVISOR:

Dr. Nicole Pinnel
German Aerospace Center (DLR)

AUTHOR:

Yannic Timothy Fetik

Salzburg, May 2017

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Aims	3
2	Materials	4
2.1	Study Area	4
2.2	Data description	6
2.2.1	Ground truth data	6
2.2.2	Hyperspectral Data	8
2.2.3	Vegetation indices	9
2.2.4	LiDAR Data	13
2.2.5	Stand density	13
2.2.6	Elevation data	14
2.2.7	Forest Mask	15
2.3	Validation / evaluation metrics	16
2.3.1	Out-of-bag Error	16
2.3.2	Cross Validation	16
2.3.3	Confusion matrix	17
2.3.4	Precision - Recall - F1 Score	17
2.3.5	Cohens Kappa statistics	19
3	Methodology	20
3.1	Validation of previous study	21
3.2	HySpex Preprocessing	22
3.2.1	Level 1A/1B	22
3.2.2	Level 1C	23
3.2.3	Level 2A	23
3.2.4	BRDF correction using BREFCOR	23
3.2.5	Savitzky Golay Filter	27
3.2.6	Brightness Normalization	28
3.3	Sample Data pre-processing	29
3.3.1	Outlier removal - Isolation Forest	29
4	Supervised Machine Learning	31
4.1	Random Forest	31
4.2	Extremely randomized trees	32
4.3	Choosing one classifier	33

4.4	Features	33
4.5	Feature engineering	34
4.6	Feature selection	35
4.6.1	Feature importance	36
4.6.2	Recursive feature elimination	36
4.7	Test Training split	38
4.8	Hyperparameter optimization	38
4.8.1	Maximum Trees	39
4.8.2	Maximum Features	40
4.8.3	Class weight	41
5	Results	43
5.1	Accuracy assessment	43
5.2	Classifier accuracy	44
5.2.1	Classifier accuracy: Species	44
5.2.2	Classifier accuracy: Species Groups	46
5.2.3	Classifier accuracy: Conifers / Broadleaf	48
5.3	Map accuracy	49
5.3.1	Map accuracy: Species	50
5.3.2	Map accuracy: Species Groups	57
5.3.3	Map accuracy: Conifers / Broadleaf	59
6	Discussion	61
6.1	Best result	61
6.1.1	LiDAR data	61
6.1.2	BRDF problems	63
6.1.3	Tree species classification	64
6.1.4	Species group classification	65
6.1.5	Conifers / Broadleaf	66
6.2	Further studies	67
6.2.1	Sampling design	67
6.2.2	Additional data	68
6.2.3	Savitzky-Golay filter & brightness normalization	68
6.2.4	Feature selection	68
6.2.5	Machine Learning	69
7	Conclusion	70
A	Appendix	71

List of terms	79
----------------------	-----------

B Eidesstattliche Erklärung	100
------------------------------------	------------

List of Figures

1	Study Area	4
2	Height zones	5
3	Tree species according to inventory	5
4	Tree species mixture	6
5	Coordination calculation for forest inventory pixels	7
6	Sampling locations	8
7	Calculated tree density for the northern part	14
8	Canopy Height Model calculation	14
9	Satellite image overlayed with CHM	15
10	Workflow of single processing steps	20
11	Confusion matrix for validation of previous study (Pixels)	22
12	Diagram visualising BRDF problems	24
13	Diagram visualising BRDF problems	25
14	Geometric and volume scattering	25
15	Mosaic of Northern HySpex data	26
16	Mean spectrum of species before and after Savitzky-Golay filter .	28
17	Mean tree species spectrum after Brightness Normalization . . .	29
18	Outliers detected by Isolation Forest	30
19	Single tree showing first two levels of tree species classification using vegetation indices	32
20	Recursive feature elimination	37
21	Selected bands by band selection	37
22	Common splitting	38
23	Error rates for OOB in relation to number of Extremely randomized trees	39
24	Error rates for OOB in relation to size of random subwindow . .	40
25	Pixels per tree species (percentage)	42
26	Training pixels per species group	46
27	Confusion matrix VNIR / SWIR	52
28	Confusion matrix VNIRfull / SWIRfull (no DTM)	53
29	Confusion matrix Spectral / Spectral + VIs	54
30	Image of VNIR and SWIR	55

31	Image of Spectral and Spectral + VIs	55
32	Image of VNIRfull and SWIRfull with DTM	56
33	Image of VNIRfull and SWIRfull without DTM	56
34	Confusion matrix VNIR / SWIR for species groups	58
35	Confusion matrix VNIRfull / SWIRfull for species groups (no DTM)	58
36	Confusion matrix Spectral / Spectral + VIs for species groups . .	59
37	Kappa scores for species classifiers	61
38	Histogram of pixels for the DTM	63
39	Zoom on probabilities for the predicted classes of classifier N14 .	63
40	Probabilities for the predicted classes of classifier N14	64
41	HySpex vegetation spectrum with chosen indices	75
42	Height distribution for training data	76

List of Tables

1	Table showing sensor parameters.	8
2	Pixel frequency for tree species	9
3	Table showing vegetation indices used in this study.	10
4	Example of an confusion matrix for multiclass classification . . .	17
5	Classification errors for binary classification	18
6	Results of validation metrics obtained from validating previous study	21
7	Available data for Machine Learning	34
8	Combinations of features used	34
9	F1 scores for validation of weights using Spectral data	42
10	Overview of all calculated classifier results	43
11	Classifier accuracy for Species	44
12	Classifier accuracy for Species Groups	47
13	Classifier accuracy for Conifers / Broadleaf	48
14	Percentage of tree species used in classification according to forest inventory (2002/2003).	50
15	Predicted results for Species (North)	51
16	Predicted results for Species Groups (North)	57
17	Predicted results for Conifers / Broadleaf (North)	60
18	Best species predictions compared to forest inventory data	65
19	Best species groups predictions compared	66
20	Selected spectral ranges by applied recursive feature elimination .	71
21	Predicted results for Species (South)	72
22	Predicted results for Species Groups (South)	73
23	Predicted results for Conifers / Broadleaf (South)	74
24	Classifier accuracy for species predicted by Random Forest	77
25	Predicted results for Species (North) using Random Forest	78

Abstract

This thesis is part of multiple studies aimed at using remote sensing technologies for monitoring the changes in the Bavarian Forest National Park (48°58' N 13°23' E), which has an overall area of 24.369 hectares. In order to detect changes of biodiversity, additional methods contributing to regular forest inventories were used. One key component of that goal is the tree species classification. The scientific team of the National Park has been working together with the German Aerospace Center conducting multiple airborne campaigns collecting hyperspectral data. The fullcube hyperspectral data (0.4 - 2.498 μm) with a resolution of 3.2 m, acquired in July 2013 is the basis data set for the pixel based supervised machine learning of tree species. Along with two field campaigns and the forest inventory data, 4775 pixels of ground truth data were derived. The spectral data were processed using the established CATENA processing chain developed at DLR, and for further enhancing the predictive capabilities smoothed and brightness normalized. BRDF effects were minimized using the novel approach of BREFCOR included in the ATCOR4 software for correcting atmospheric disturbances of airborne remote sensing data.

The classification was carried out using common open source software.¹ Additional LiDAR data and a set of vegetation indices were also used as input data. Seven classifiers of extremely randomized trees were trained using different feature combinations. Three levels of predictions were made based on I. species, II. species groups, and III. coniferous / broadleaf trees. The classification accuracy was evaluated using Kappa scores, F1-measurements and confusion matrices. Over fitting was detected as a problem, when using LiDAR based DTM data, because of the small size of available training data and the specific behaviour of random forests. The large number of training pixels, which would be needed for representing the multitude of differences in species distribution over the height above zero was not achieved. The greedy behaviour of the used forest of randomized trees lead to a biased learning behaviour. Apart from comparing machine learning metrics retrieved from the ground truth data, the overall tree species composition of the both parts of the Bavarian Forest National Park was calculated and the northern part was evaluated by comparing predicted results to the latest forest inventory. The fullcube hyperspectral spectrum combined with selected vegetation indices showed an overall better suitability for classifying the selected tree species reaching a κ score of 0.589 for the test data set. The highest F1-scores were recorded for the species *Pinus mugo* with 0.88, followed by the species *Fagus sylvatica* (0.80), *Picea abies* (0.65) and *Fraxinus excelsior* (0.64). Difficulties in the classification were observed within the conifers and broadleaved species, rather than between these two groups. The coniferous minority class species *Pseudotsuga menziesii* (0.14) showed low F1-scores based on high misclassification as *Abies alba* and *Picea abies*. While the broadleaved species *Acer pseudoplatanus* (0.29) showed high misclassification as *Fagus sylvatica*.

¹SciPy Stack v1.0, Geospatial Data Abstraction Library v2.1.1, scikit-learn v0.18.1

1 Introduction

1.1 Motivation

In the year of 2015 31% of the global land area is covered with forests and within Europe the area covered by forests is increasing (Keenan et al. 2015). Due to the importance of forests to the world climate, forest ecosystems especially in Europe have to fulfil a wide range of needs. The United Nations Conference on Environment and Development in Rio 1992 established the so called "Rio Forest Principles". This is a legally non binding document which lays out several principles of sustainable forest management (Schlaepfer et al. 2000)

Forest resources and forest lands should be sustainably managed to meet the social economic, ecological, cultural and spiritual needs of present and future generations. These needs are for forest products and services, such as wood and wood products, water, food, fodder, medicine, fuel, shelter, employment, recreation, habitats for wildlife, landscape diversity, carbon sinks and reservoirs, and for other forest products.²

In Germany the "Bundeswaldgesetz" obligates forests to fulfil the needs of economical use, environmental aspects especially the sustainability of the ecosystem, the climate, hydrology's balance, and other factors. For this obligations to be fulfilled an forest management plan is created. The foundation of every forest management plan is the Forest Inventory. The Forest Inventory structures a forest in different classes, usually based on parameters concerning the economic age, species, and height. The needed data acquisition for this inventories is usually done in teams of two persons and is very time consuming, and thereby expensive. Thus inventories are only usually conducted once every decade.

As soon as worldwide remotely sensed data was accessible with the Launch of Landsat 1 in 1972 (originally named "Earth Resources Technology Satellite"³), scientists started to examine the possibilities of these data sources. (Kirvida et al. (1973), Ebtehadj (1973), Lawrence et al. (1975)). While in the beginning of Remote Sensing of forest parameters the focus was at gaining information on forest stands (average tree height, structure and the overall biomass), with the increased capabilities of data acquisition (Hyperspectral, LiDAR) and processing (computation power) the focus shifted towards information about individual

²Report of the United Nations Conference on Environment and Development (CONF.151/26 (Vol. III) Principle 2b

³www Landsat.usgs.gov

trees (J. Hyypä and Inkinen (1999), Straub et al. (2011), Martens (2012)). Using Hyperspectral data for tree species classification has already been proven possible (Gong et al. (1997), Clark et al. (2005), Buddenbaum et al. (2005), Boschetti et al. (2007), Pu (2009), Dalponte, Ørka, et al. (2013)). Other studies have already shown the benefit of combining LiDAR and Hyperspectral data for assessing Forest parameters (Dalponte, Bruzzone, et al. (2008), Holmgren et al. (2008), Naidoo et al. (2012), Dalponte, Bruzzone, et al. (2012)). Forest inventory is especially cost intensive in the rugged mountainous terrain of the Bavarian National Forest. Due to the increasing heterogeneity of a forest where natural development without much interference is allowed, common forest inventory statistics aiming at a homogeneous distribution of species are not appropriate anymore (Heurich, Krzystek, et al. 2015). By law until the year 2027 three quarters of the National park should be in a state where no human interference takes place. Through the natural processes, this will lead to an even greater diversity in structure and species. Especially the changes in species distribution (closer to the possible natural vegetation), but also the changes from a mostly one layered forest (in parts with majority of *Picea abies*) towards a non layered multi-aged and multi-species forest, increase the difficulties in manual monitoring. But due to the need of regular monitoring, which is not possible using common forest inventories, other approaches are being pursued.

The scientific team of the Bavarian National Park has been investigating different methods of remote sensing to contribute to forest inventories for several years (Heurich, T. Schneider, et al. (2003), Tiede et al. (2004), Heurich, Schadeck, et al. (2004), Aulinger et al. (2005), Wei et al. (2012)). Methods for estimating biomass and structure have already been proven possible using LiDAR data (Lim et al. 2003). The key component of any further analysis of forests is the species composition. For achieving a remotely sensed classification of tree species, the National park is closely working together with the DLR to create a database of airborne sensed hyperspectral data to investigate the capabilities of species classification.

1.2 Aims

The basis for this study consists of a previous tree species mapping exercise conducted by Sommer (2015). This previous study used data from the VNIR region of the hyperspectral spectrum ($0.4 - \sim 1.0 \mu m$) only, obtained from an airborne campaign in 2013. In this current study the same data set was used, but the spectral regions of VNIR and SWIR were combined to obtain a hyperspectral full cube data set. The goal was to optimize the classification and to understand which regions of the spectrum contributes the most to an enhanced classification result.

Further data acquisition was successfully completed in 2015 and a continuation of these campaigns is planned in future. The aim of this study is to create a reproducible method of utilizing the reflectance of trees (in addition with LiDAR data) for a pixel based supervised classification of tree species. In order to achieve this goal a second field campaign for additional sampling of ground truth data was carried out (2.2.1). These data were used as training data for the classification after the raw hyperspectral data were preprocessed (3.2) using the established processing chain at the DLR.

This classification could be the base for monitoring and prediction of the development of biodiversity and structure in the Bavarian National Park. In order to obtain measures of biodiversity, several indexes have been developed, such as Shannon index, (Shannon 1948) and the related Evenness index, the Species Profile Index (Pretzsch (1995) , Pretzsch (2009)) and the Simpson index (Simpson 1949). But for the applicability of these indexes a successful classification of tree species, which goes further than a normal point based forest inventory, is essentially needed.

2 Materials

2.1 Study Area

As the first national park in Germany, the Bavarian Forest National Park ($48^{\circ}58'$ N $13^{\circ}23'$ E) was founded in 1970 with an area of 13.000 hectares (Falkenstein-Rachel area (North)) and extended in 1997 (Rachel-Lusen area (South)) to an overall area of 24.369 hectares. Stretching from around 600 meters above standard elevation zero to 1453 meters (Mountain "Großer Rachel"). As the name suggest the most of the area of the national park is densely forested, with exception of the raised bog areas at the mountain tops.

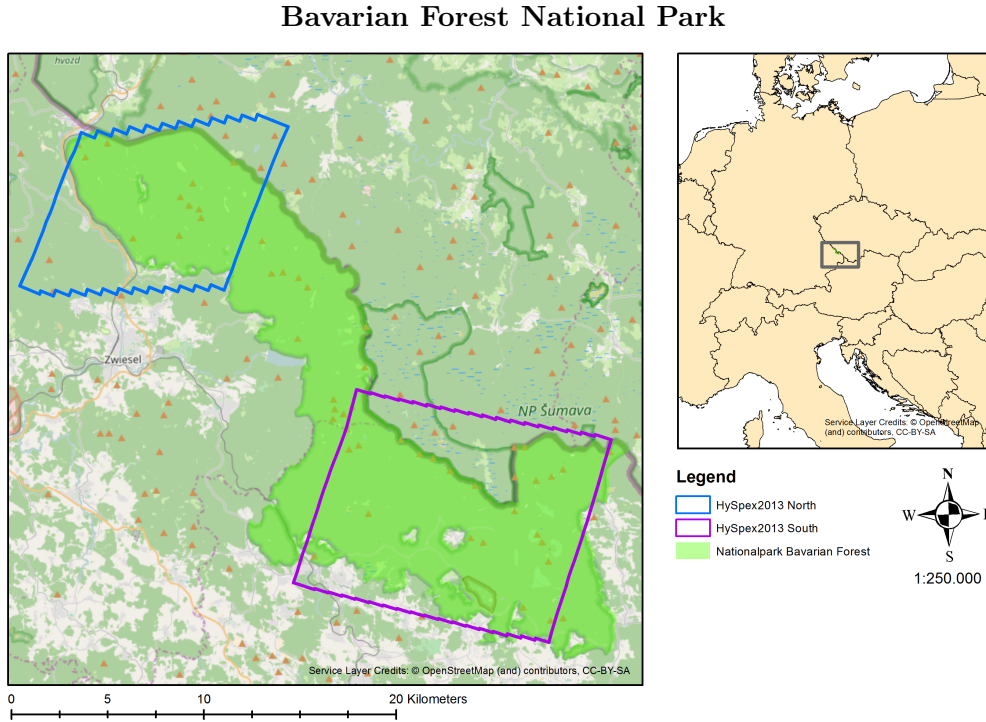


Figure 1: Study Area

The forest area of the national park is classified as mountainous forest, and belongs to the growing area 11 (11.2 & 11.3, see Figure 2, Walentowski et al. 2006). Further more the forest area can be differentiated into three categories, concerning the elevation and slope of the terrain, which results in different natural forest communities:

- High level: 1.150 m up to 1.450 m covering 16 %.
- Hillside: 700 m up to 1.150 m covering 68 %.
- Valley: 650 m up to 900 m covering 16 %.

Each with an individual natural forest vegetation:

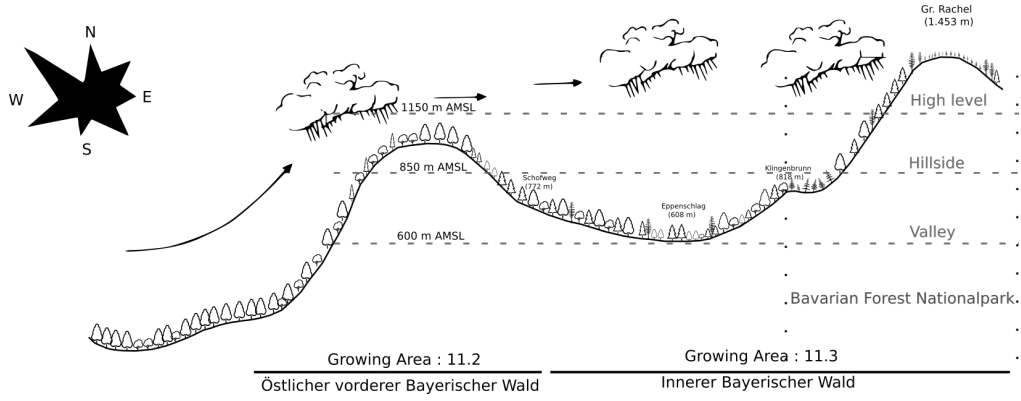


Figure 2: Height zones adapted from Walentowski et al. (2006)

- High level: *Calamagrostio villosae* (*Piceetum barbilophozietosum*)
- Hillside: *Luzulo Fagetum* (*Asperulo Fagetum*)
- Valley: *Calamagrostio vill* (*Piceetum bazzanietosum*)

However due to the management history of the Bavarian Forest National Park, the species *Picea abies* is currently still overrepresented (Heurich and Neufanger 2005).

The concept of the national park is based on having as little human interference as possible. These natural dynamics of a forest ecosystem leads to a declining percentage of *Picea abies*. As the human interference is at present still visible in the frequency of tree species in the national park, the forest communities are not only grouped by their possible natural vegetation (2.1), but also by their current composition of tree species (Fig. 4). For a classification this information is highly interesting, as it determines how pixels have to be separated by the classifier.

In the North (Falkenstein-Rachel area) the area covered only by *Picea abies* (less than 5% other species) is with 25.4% more than twice as big as in the South (Rachel-Lusen area, 10.1%). In addition to that, 17.7% of the area in the northern part is covered by *Fagus sylvatica* only, which results in ~42% of the area covered by only a single tree species. In the South the forest is due to its history much

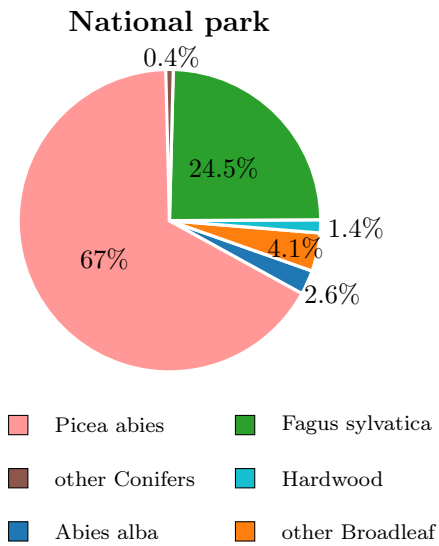


Figure 3: Tree species according to inventory (2002/2003)

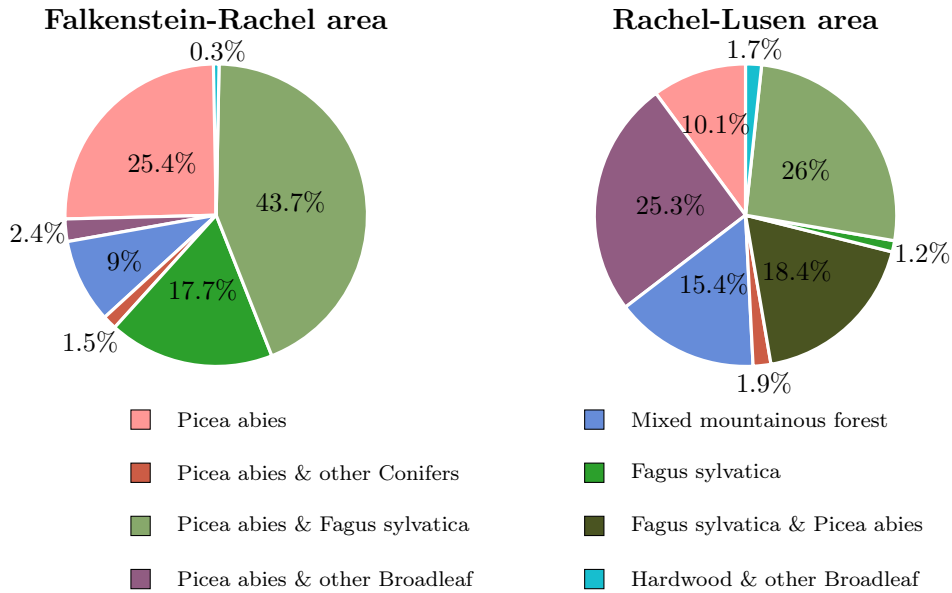


Figure 4: Tree species mixture according to Heurich and Neufanger (2005)

more mixed, and only around $\sim 11\%$ of the area is covered predominantly by one species. This increases the difficulty of the classification task by a even higher possibility of mixed pixels in the South.

It should also be noted, that since the last Inventory (in 2002/2003) natural hazards, like storms (eg. 18/19.1.2007 Kyrill) and the following rise in bark beetle had a great influence on the distribution of the species *Picea abies*.

2.2 Data description

2.2.1 Ground truth data

The term 'ground truth' is used in many fields, from earth sciences to data science. In remote sensing ground truth is used for data which has been verified or collected in field work (Næsset 1997) and is also referred to as reference data. In a supervised classification task, prior knowledge about some data points is needed for training the classifier. In remote sensing this knowledge is represented by pixels where the label of each pixel has been determined. In the case of tree species in an mixed heterogeneous forest this knowledge could only be obtained through a field survey.

First Field Campaign

A first field campaign was conducted in October 2014 (leaf off condition) by Sommer (2015) as part of a previous tree species mapping exercise. For further description see Sommer (2015).

Second Field Campaign

A second field campaign was additionally carried out for this study in December 2015 using an Magellan MobileMapper TM CX with an external antenna. Each point was recorded for 15 minutes. Due to the known accuracy issues under canopies, the data acquisition took place in the first two weeks of December under 'leaf off' conditions. It was taken care that only single trees surrounded by trees of the same species were recorded, in order to reduce the possibility of mixed or mislabeled pixels.

Forest inventory

In 1991 a regular grid with an edge length of 200 meters was calculated and put over the national park. Every node is the center of a variable size plot. After the extension of the national park in 1997 this resulted in 5.841 points. Typically forest inventories are repeated every ten years. In 2002 / 2003 the Bavarian Forest National Park conducted the second forest inventory based on this grid. As for the forest management the knowledge of the exact coordinates of these plots is not relevant (the position in the field is marked by a magnetic material), only some points were therefore measured by GPS (Czaja 2003). In total: 352 points were recorded.

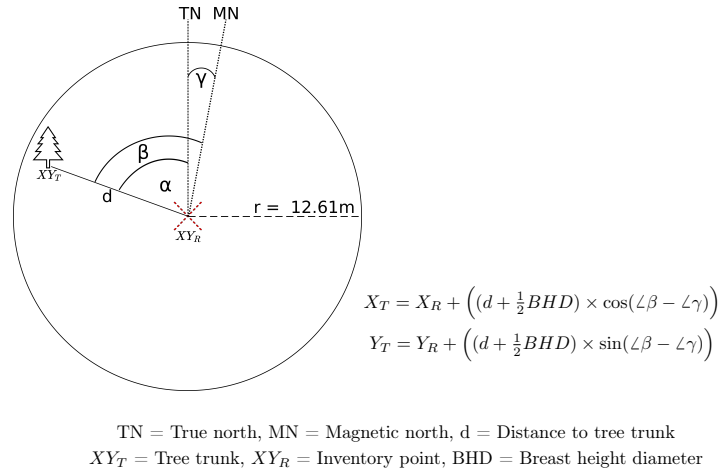


Figure 5: Coordination calculation for EPSG:31468

In order to evaluate the data quality, these points were measured multiple times and the position were calculated using these measurements. Out of these 352 points 132 points were inside the boundaries of the hyperspectral data. For each of these points, the measured trees and their exact positions relative to the sampling point (XY_R) (angle (β) and distance (d)) are known. The tree positions (XY_T) could then easily be calculated (see Figure 5). For the association of the pixel with the corresponding label, the tallest tree was used.

Spatial distribution of sample sites

Due to the heterogeneous distribution and low frequency of non dominant tree species, the sample design could not be random. As explained in 2.2.1 the sites

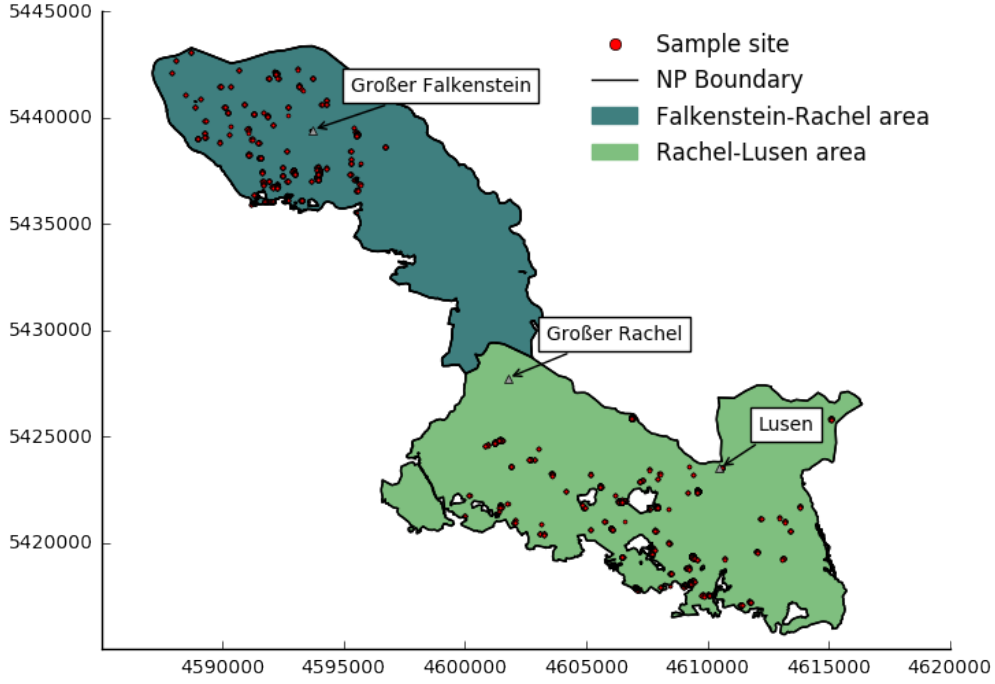


Figure 6: Sampling locations in EPSG:31468

were chosen by gathering information of specific sites rich in species diversity.

2.2.2 Hyperspectral Data

The hyperspectral data (DLR Remote Sensing Technology Institute 2016) used in this study was acquired on the 22th and 27th of July 2013 using hyperspectral sensors (HySpex) developed by Norsk Elektro Optikk⁴ and covering a spectral range of 0.4-0.992 μm (HySpex VNIR-1600) and 0.968-2.498 μm (HySpex SWIR-320m-e), respectively (see Table 1).












⁴<http://www.hyspex.no/>

Table 1: Table showing Sensor parameters. (Köhler and M. Schneider 2015)

Sensor	Spectral range (μm)	Bands	Bandwidth (μm)
HySpex VNIR-1600	0.4–0.992	160	0.5
HySpex SWIR- 320m-e	0.968–2.498	256	0.6

In contrary to the HySpex VNIR-1600 data (with a spatial resolution of 1.6 m) the Full Cube hyperspectral data used in this study was resampled to a ground resolution of 3.2 m (SWIR sensor spatial pixels) in order not to introduce artificial reflectance data. The coarser geometric resolution resulted in less ground truth pixels. This results in the amount of Pixels, shown in Table 2.

Table 2: Tree species and respective number of associated pixels (after outlier removal)

Scientific name	English	Abbreviation	Number of Pixels	Color
<i>Abies alba</i>	European silver fir	AA	543	
<i>Acer pseudoplatanus</i>	Sycamore maple	AP	371	
<i>Alnus glutinosa</i>	European alder	AG	204	
<i>Betula pendula</i>	Silver birch	BP	329	
<i>Fagus sylvatica</i>	European beech	FS	1408	
<i>Fraxinus excelsior</i>	European ash	FE	386	
<i>Larix decidua</i>	European larch	LD	346	
<i>Picea abies</i>	Norway spruce	PA	725	
<i>Pinus mugo</i>	Mountain pine	PMu	237	
<i>Populus tremula</i>	European aspen	PT	120	
<i>Pseudotsuga menziesii</i>	Douglas fir	PM	106	
Total Pixels $\Sigma = 4775$				

While a typical tree crown in the upper tree layer can have a diameter of approximately 5–15 m (20 - 180 m^2) (Fassnacht et al. 2014), mixed pixels (two or more species) will occur at all spatial resolutions.

2.2.3 Vegetation indices

Vegetation indices (VIs) are combinations of ground reflectance data of two or more bands. Due to their design they are more stable than single bands (Asner et al. 2003), as indices are less sensitive to atmospheric effects and to soil brightness (Bannari et al. 1995). With vegetation indices it is possible to measure specific physiological differences of leaves, due to the difference in spectral reflectance (Asner 1998, Sims et al. 2002, Schlerf et al. 2005, Ollinger 2010).

As this study is a continuation of a previous study which used the VNIR region only (Sommer et al. 2016), the same vegetation indices with additional two indices in the SWIR region were used. An overview of all indices applied in this study, including their used center wavelength is listed in Table 3. A graphical presentation of the used indices and their exact position on a vegetation spectrum is shown in Figure 41.

Table 3: Table showing vegetation indices used in this study.

Indices	Abbreviation	Center wavelength (μm)
Cellulose Absorption Index	CAI	2.202, 2.028, 2.106
Chlorophyll Index Green	CIG	0.545, 0.858
Enhanced Vegetation Index	EVI	0.660, 0.800
Normalized Difference Lignin Index	NDLI	1.681, 1.753
Normalized Difference Leaf Mass (per area)	NDLma	1.495, 2.262
Normalized Difference Vegetation Index	NDVI	0.649, 0.858
Red Edge NDVI	RENDVI	0.700, 0.750
Red Edge Inflection Point	REIP	0.671, 0.700, 0.739, 0.779
Modified NDVI	mNDVI	0.444, 0.707, 0.750

Cellulose Absorption Index (CAI)

The CAI was originally developed to quantify exposed surfaces that contain dried plant material (Daughtry et al. 2004). It is usually used in applications such as crop monitoring, fire fuel conditions, and grazing management. As a strong absorption in the used wavelengths indicate a strong presence of cellulose, a differentiation between at least conifers and broadleaf seems possible (Nagler et al. 2003). The value range of this index ranges from -3 to more than 4. The common range for green vegetation is -2 to 4.

$$0.5 \times (\lambda_{2005\mu m} + \lambda_{2203\mu m}) - \lambda_{2106\mu m} \quad (1)$$

Chlorophyll Index Green (CIG)

Originally proposed by Gitelson, Viña, et al. (2003), the Chlorophyll Index Green (CIG) was developed to remotely estimate the chlorophyll content of crops. After proven to be successful in estimating the chlorophyll content (E. Raymond Hunt, Jr. et al. 2013) and also having an close relationship to the LAI, the CIG was also used for estimation of canopy chlorophyll content in a coniferous forest (Wu et al. 2012).

$$\frac{\lambda_{860\mu m}}{\lambda_{545\mu m}} - 1 \quad (2)$$

Enhanced Vegetation Index (EVI)

The Enhanced Vegetation Index (EVI) is an vegetation index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and a reduction in atmosphere influences. The range of values for the EVI is -1 to 1,

where vegetation usually reaches values of 0.20 to 0.80. The coefficients used in this study were adopted from the MODIS-EVI algorithm (Huete et al. 2002).

$$2.5 \times \frac{(\lambda_{800\mu m} - \lambda_{660\mu m})}{(\lambda_{800\mu m} + 6 \times \lambda_{660\mu m} - 7.5 \times \lambda_{480\mu m} + 1)} \quad (3)$$

Normalized Difference Lignin Index (NDLI)

Normalized Difference Lignin Index (NDLI) estimates the relative amounts of lignin contained in canopies. The lignin concentration in canopies is one key of estimating the growth and decomposition capabilities of a ecosystem (Serrano et al. 2002). But as the leaves of tree species contain significant different amounts of lignin (Melillo et al. 1982), this index was also used in the effort of achieving a better classification result. The value of this index ranges from 0 to 1.

$$\frac{\left(\log \frac{1}{\lambda_{1754\mu m}}\right) - \left(\log \frac{1}{\lambda_{1680\mu m}}\right)}{\left(\log \frac{1}{\lambda_{1754\mu m}}\right) + \left(\log \frac{1}{\lambda_{1680\mu m}}\right)} \quad (4)$$

Normalized Difference Leaf Mass (per area) (NDLma)

In an experimental study Maire et al. (2008) found out that the Normalized Difference Leaf Mass (per area) (NDLma) is suitable for estimating leaf mass in data derived from spectroscopy measurements. Despite the suitability with laboratory measurements, for remotely sensed data the index was only significant for one out of two study sites. However because of it's correlation with the LAI (Maire et al. 2008) and therefore the relation to biophysical and structural tree parameters, using the shortwave infrared spectrum, it was included as an experimental VI in this study.

$$\frac{\lambda_{2260\mu m} - \lambda_{1490\mu m}}{\lambda_{2260\mu m} + \lambda_{1490\mu m}} \quad (5)$$

NDVI

The Normalized Difference Vegetation Index (NDVI) is the most basic vegetation index in remote sensing, and is widely used for distinguishing between vegetated area and no vegetated area. It was first proposed by Rouse et al. (1974) for the detection and monitoring of vegetation. It is calculated from the visible and near-infrared wavelengths. The chlorophyll in the plant leaves absorbs the visible light for photosynthesis. It results in an index ranging from -1 to +1, while it is assumed that values below 0.3 are non vegetated areas (clouds, water, bare

soil Pettoirelli (2013)). Studies have shown that the canopy structure of trees has an influence on the NDVI values (Gamon et al. 1995), as well as that there is a relationship between the NDVI and the Leaf area index (LAI) (Carlson et al. (1997), Q. Wang et al. (2005)).

$$\frac{\lambda_{858\mu m} - \lambda_{649\mu m}}{\lambda_{858\mu m} + \lambda_{649\mu m}} \quad (6)$$

REDNDVI

The Red Edge NDVI (REDNDVI) was developed in order to adapt the NDVI for being more sensitive in changes of chlorophyll. Gitelson and Merzlyak (1994) found that using this adapted index there is a close correlation between chlorophyll concentration of leaves and this index, even for high Chlorophyll concentrations. The value of this index ranges from -1 to 1.

$$\frac{\lambda_{750\mu m} - \lambda_{705\mu m}}{\lambda_{750\mu m} + \lambda_{705\mu m}} \quad (7)$$

REIP

The Red Edge Inflection Point (REIP) utilizes the sharp change in leaf reflectance between 680 nm and 750 nm (Horler et al. 1983) in order to detect the chlorophyll content (Moss et al. (1991), Vogelmann et al. (1993)).

$$700 + 40 \left(\frac{\frac{(\lambda_{670\mu m} + \lambda_{780\mu m})}{2} - \lambda_{700\mu m}}{\lambda_{740\mu m} - \lambda_{700\mu m}} \right) \quad (8)$$

mNDVI

The Modified NDVI (mNDVI) is another index based on the NDVI, altered in the manner of being more sensitive to the vegetation condition (Jurgens 1997). The values range from -1 to 1.

$$\frac{\lambda_{800\mu m} - \lambda_{680\mu m}}{\lambda_{800\mu m} + \lambda_{680\mu m} - \lambda_{2445\mu m}} \quad (9)$$

2.2.4 LiDAR Data

The initial use of Light Detection And Ranging (LiDAR) data in forest inventory began around year 2000, before that LiDAR was, as described by Means et al. (2000), limited to: *'... geo-technical applications such as creation of digital terrain models for layout of roads or logging systems.'* With the availability of full waveform LiDAR data at a high density scale (~ 30 points per m^2) scientific research started to evaluate the chances of LiDAR for forest inventory (Reitberger et al. (2008), J. Hyypä, H. Hyypä, et al. (2008)). However classifications based on the shape derived from detected single trees have limited accuracy only, and are mostly used in homogeneous forests with a small number of different tree species (Vauhkonen et al. 2013). Therefore LiDAR data are to date mostly used for biomass estimations (Koch 2010).

The data for this study was provided by the Bavarian Forest National Park and is the outcome of an airborne campaign in July 2012. The full waveform LiDAR data was acquired with a RIEGL LMS-Q 680i, with a last pulse point density of approximately 30-40 points (Latifi et al. 2015) per m^2 . In this study different LiDAR products were used:

- First as Digital Terrain Model (DTM) (height above standard elevation zero for every pixel)
- Second as an indicator of the stand density (Gadow 2003), which was derived from the single trees detected in a previous study (Latifi et al. 2015).

2.2.5 Stand density

For every hyperspectral pixel the count of detected trees within a radius of 12.61 meters ($\sim 500m^2$) was calculated (see Figure 7). As the number of stems in a forest correlates with the age of the forest stand (Assmann 1961) as well as with the mean diameter (Clutter et al. 1983), the density is an appropriate indicator of forest structure.

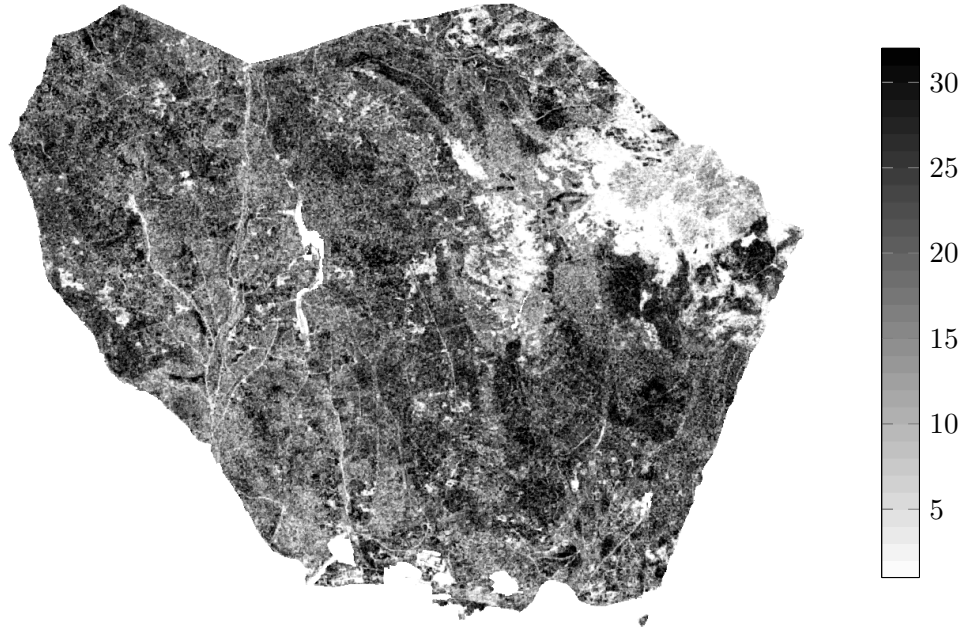


Figure 7: Calculated tree density for the northern part

2.2.6 Elevation data

Digital Elevation Model (DEM) & Digital Terrain Model (DTM)

As explained in section 2.2.4 the basis for the DEM (and DTM) used in this study was a DEM (DTM) derived from the LiDAR data with a spatial resolution of 1x1 m. The elevation data were re-sampled (using bilinear re-sampling) to fit the resolution of the hyperspectral data ($3.2m \times 3.2m$) using the Geospatial Data Abstraction Library (GDAL) (GDAL Development Team 2016).

Canopy Height Model (CHM)

The Canopy Height Model (CHM) was computed as a difference between DSM and DTM, in addition all non forest area was removed from the datasets.



Figure 8: Canopy Height Model calculation

2.2.7 Forest Mask

In order to ensure that anonymous pixels consist of forest area only, several data sources were combined for the creation of a forest mask.

Boundaries

The majority of the road network and small settlements within the national park were excluded using a vector data-set representing the official boundaries of the national park.

NDVI

As for this study only pixels of actual trees were needed, non vegetated pixels were excluded using a threshold of 0.6 on the calculated Normalized Difference Vegetation Index (NDVI) value of the pixels.

$$Forest = NDVI > 0.6 \quad (10)$$

CHM

Finally, the calculated Canopy Height Model (CHM) was used to exclude deforested areas (Barkbeetle, Storms), otherwise not excluded through the NDVI filter, using a height threshold of 2 meters. However, it should be noted that the "Tree layer" as commonly defined in ecology (Dierschke 1994) starts at 5 meters. Nevertheless the threshold of 2 meters was chosen in order to remove areas where due to natural hazards, plants like *Rubus vulgaris* are at the top of the vegetation layer, but areas where pioneer species like *Betula pendula*, *Alnus glutinosa* have established a canopy layer are included.

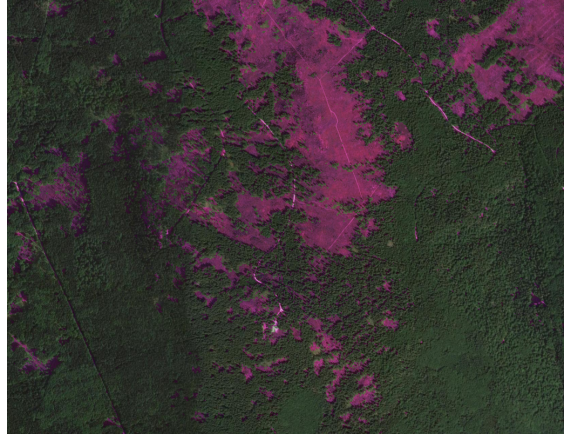


Figure 9: Sat. image overlaid with CHM
Imagery ©2017 Google

This should prevent non tree species from interfering with the prediction result. Additionally it should be mentioned that the species *Sorbus aucuparia* which is also common pioneer species, especially in the higher regions, has not been included in the analysis, and thus cannot be accounted for.

2.3 Validation / evaluation metrics

Validation in remote sensing means to create measures of the correctness of the classification result (Foody 2002). In Machine Learning there is a difference between validation and evaluation. Whereas **validation** refers to the process of validating the results while training the model, **evaluation** refers to the measuring of performance of the finally trained model.

As a result of this differentiation, two types of metrics were applied. First during the training of the classifier those metrics for assessing the effect of different parameters on the result (OOB, CV) were used. These measures are obtained using the training dataset only. Second those measures meant to evaluate the predictions of the final model utilizing the test dataset were used. As each statistic measure has its limitations, different metrics were calculated for evaluating the classifier performance (Kappa score, Confusion matrix, Precision, Recall, F1-score). A detailed description of the calculated metrics is given in the following paragraphs (2.3.1 - 2.3.5). All metrics were used as implemented in scikit-learn 0.18 (Pedregosa et al. 2011).

2.3.1 Out-of-bag Error

Originally proposed by Breiman (1996b), every decision tree classifier where bagging is employed, is able to use the samples not included in the construction of a tree for measuring the prediction error of a forest. The Out-of-bag error (OOB) is the misclassification error averaged over all samples. The benefits of the OOB error rate is mainly, that there is no additional computation needed to achieve that measure. Also there is no need for splitting the Training data (see 4.7) into three parts (James et al. 2013) as the OOB error eliminates the need of a separate validation set.

2.3.2 Cross Validation

Cross-validation (CV) is one of the most used method for estimating prediction error (Hastie et al. 2009). In this study a stratified k-fold cross validation with 5 folds was used in addition to the OOB error in order to evaluate the tuning of hyperparameters. The training set is split into 5 (k) smaller sets, and the model is trained using k-1 of the folds. The trained model is then validated on the remaining part of the data. This is repeated for every fold. The resulting measurement is the average of all of the resulted accuracy scores. The additional benefit to only using the OOB error is that there is a second independent measure

for the reliability of the predictions of the classifier. This measure can then be used to compare the classifier with a non bagging classifier.

2.3.3 Confusion matrix

A confusion matrix creates a two dimensional representation of the predicted data against test data. It visualizes the performance of the classifier (Sammut et al. 2011).

Actual	Class 1	<div>n</div>	<div>n</div>	<div>n</div>
	Class 2	<div>n</div>	<div>n</div>	<div>n</div>
	Class 3	<div>n</div>	<div>n</div>	<div>n</div>
		Class 1	Class 2	Class 3
		Predicted		

Table 4: Example of an confusion matrix for multiclass classification

For every class the count of actual (true) pixels is listed according to its belonging to the predicted class. This generates an easy overview on how the classifier is performing in predicting the correct classes, and between which classes misclassifications happen.

2.3.4 Precision - Recall - F1 Score

In order to evaluate the performance of a classifier, one should consider what kind of results a classifier can produce. If the classifier predicts correctly it is called true positive (Fawcett 2006). The higher the true positive rate the better the classifier predicts for one class correctly. However a precision score of 1.0 for a Class 1 only means that every pixel labeled as Class 1 belongs indeed to Class 1. So precision says nothing about the pixels from Class 1 which were labeled as another class. In order to account for that error the Recall Score is used. A Recall of 1.0 shows that every pixel belonging to Class 1 was indeed labeled as Class 1. As a result in a way of combining these two measures, the F1-score is used.

Actual	Class 1	True positive (tp)	False positive (fp)
	Class 2	False negative (fn)	True negative (tn)
		Class 1	Class 2
		Predicted	

Table 5: Classification errors for binary classification

Precision is the number of correctly classified (tp) over the number true positives plus the number of false positives.

$$\text{precision} = \frac{tp}{tp + fp} \quad (11)$$

Recall is the number of correctly classified (tp) over the number true positives plus the number of false negatives.

$$\text{recall} = \frac{tp}{tp + fn} \quad (12)$$

The F1 score is the weighted average of Recall and Precision. The maximum F1 score is at value 1 and the minimum is 0.

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (13)$$

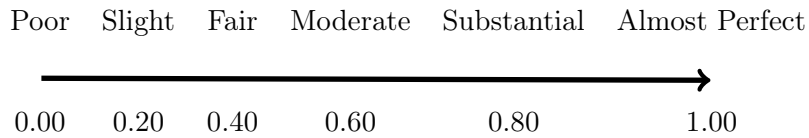
2.3.5 Cohens Kappa statistics

Originally developed by Cohen (1960) in order to rate the agreement of two humans classifying binary data, it has also been used in land classification problems (Foody 2002) for the comparison of different classification results. Most recently it has also been used as a metric in Machine learning to compare the performance of different classifiers.

$$\kappa = (p_o - p_e)/(1 - p_e) \quad (14)$$

p_o is the observed agreement ratio, and p_e is the expected agreement.

Cohen's kappa coefficient (Kappa) is a metric that compares an Observed Accuracy with an Expected Accuracy (accuracy resulting out of random chance). Minimum value is -1 which means complete disagreement and maximum is 1 representing complete agreement. A Kappa value of 0 is equal to a result predicted by chance. Between -1 and 1, there is no standardized measurement which defines the scale of κ scores, as it always depends on the data used. But there is the generally accepted range of values proposed by Landis et al. (1977), which resembles the agreement:



3 Methodology

The following Chapter describes the methodologies used in treating the data and obtaining the results. The flowchart in Figure 10 provides an overview of single processing steps and visualizes the main work flow :

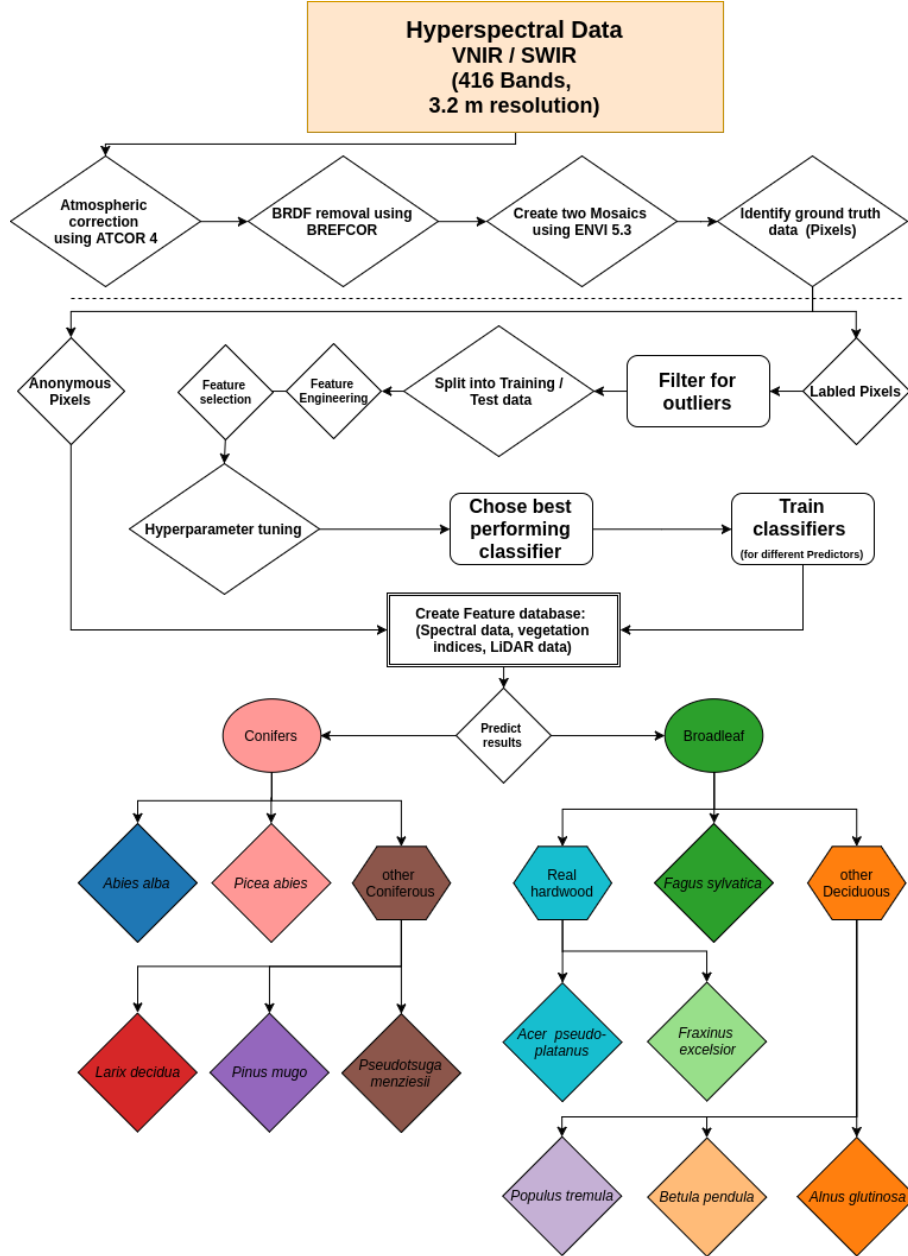


Figure 10: Workflow of single processing steps

3.1 Validation of previous study

The previous classification result by Sommer (2015) was validated using the combined ground truth data from the second field campaign and the forest inventory data. The focus hereby lies on the thematic accuracy, which is the correspondence between the label of the pixel assigned by the ground truth data, and the predicted label assigned by the classifier.

Table 6: Results of validation metrics obtained from validating previous study

Scientific Name	Precision	Recall	F1-score	Number of Pixels
<i>Abies alba</i>	0.22	0.28	0.25	99
<i>Acer pseudoplatanus</i>	0.12	0.35	0.18	65
<i>Alnus glutinosa</i>	0.34	0.53	0.41	34
<i>Betula pendula</i>	0.00	0.00	0.00	22
<i>Fagus sylvatica</i>	0.62	0.60	0.61	227
<i>Fraxinus excelsior</i>	0.30	0.07	0.12	180
<i>Larix decidua</i>	0.64	0.82	0.72	67
<i>Picea abies</i>	0.64	0.68	0.66	372
<i>Populus tremula</i>	0.54	0.16	0.24	45
<i>Pseudotsuga menziesii</i>	0.00	0.00	0.00	29
	0.47	0.47	0.45	$\Sigma = 1140$

Table 6 shows Precision, Recall and F1-score (see equations (11), (12) and (13)) for the matching pixels of the previous classification. Overall 1140 pixels were matched with the additional data gathered during the second field trip and combined with the data derived from forest inventory points. For some tree species (*Pseudotsuga menziesii*, *Betula pendula*) very limited ground data were available (less than 30 pixels), thus the metrics are not reliable for these species. But also for some other species with more pixels, the table shows relatively low scores (F1-score below 0.20) (*Fraxinus excelsior* and *Acer pseudoplatanus*). This is also visualized in the confusion matrix (see Fig. 11). The best predictions occurred for the species *Fagus sylvatica*, *Picea abies* and *Larix decidua* (F1-score above 0.60). The coniferous species *Abies alba* however was mainly misclassified as *Picea abies* and *Alnus glutinosa* and consequently obtained very low prediction values (F1-score below 0.3). This result was unexpected as the affiliation of *Picea abies* and *Alnus glutinosa* belongs to different Orders (Conifers / Broadleaf). The overall F1-score of 0.45 suggests limited abilities of the classifier to predict

AA	28	19	5	3	2	12	1	28	0	1
AG	2	18	14	0	0	0	0	0	0	0
AP	1	0	23	1	1	26	0	13	0	0
BP	8	8	2	0	1	1	0	2	0	0
FE	7	0	102	0	13	25	2	26	3	2
FS	12	1	26	4	14	136	6	26	2	0
LD	0	0	0	0	1	3	55	7	0	1
PA	55	1	13	2	7	16	22	253	1	2
PM	0	0	0	0	0	0	0	29	0	0
PT	13	6	1	1	5	0	0	12	0	7
	AA	AG	AP	BP	FE	FS	LD	PA	PM	PT

Figure 11: Confusion matrix for validation of previous study (Pixels)

correctly for this data set. The Kappa score of 0.348 for the whole classification reflects the low scores retrieved from validation result.

3.2 HySpex Preprocessing

The data used in this study was processed internally by DLR using the established processing chain CATENA (Krauß et al. 2013). The products of CATENA are categorized in different levels, according to the applied processing steps :

- Level 0: Raw data recorded by the HySpex sensor system
- Level 1A: Raw data with synchronised navigation data (3.2.1)
- Level 1B: At-sensor radiance including navigation data for georeferencing (3.2.1)
- Level 1C: Orthorectified at-sensor radiance (3.2.2)
- Level 2A: Orthorectified surface reflectance (3.2.3)

For a detailed description see (Köhler 2016) and (Köhler and M. Schneider 2015). After the final product (Level 2A) processed and delivered by CATENA, additional processing steps were employed in order to remove remaining artefacts and prepare the data for the classification task:

- BRDF Correction (3.2.4)
- Spectral polishing (3.2.5)
- Brightness normalization (3.2.6)

3.2.1 Level 1A/1B

During the Level 1A and Level 1B processing, the raw Digital Number recorded by the sensor get converted to radiance using laboratory measurements. This

includes the removal of artifacts (nonlinearity, stray light, spectral response). Bad pixels are corrected by interpolating neighbouring signals. Smile and keystone effects are removed based on camera models derived in the laboratory. The VNIR data is also additionally corrected for a radiometric nonlinearity and stray light (Lenhard et al. 2015).

3.2.2 Level 1C

The individual images (VNIR & SWIR) are matched using the BRISK algorithm (Leutenegger et al. 2011) and are then orthorectified, using the provided DEM (2.2.6), with the DLR's in house software ORTHO (Müller et al. 2005).

3.2.3 Level 2A

This step aims at correcting the signal for the atmospheric absorption and scattering mainly due to water vapor and aerosols (atmospheric correction). In order to reduce the effect of gases, aerosols and water (clouds), the image data is corrected using the atmospheric correction software ATCOR-4. ATCOR-4 is a software for atmospheric/topographic correction of airborne scanner data. ATCOR-4 correction functions are compiled using the radiative transfer model MODTRAN-5⁵. For the right algorithm to be applied either the user defines the atmospheric parameters or the software automatically calculates the appropriate parameters (aerosol type, visibility, water vapor content). For an comprehensive description of the process see Richter et al. (2016). The main steps of the atmospheric corrections are:

- masking of haze, cloud, water, and clear pixels
- haze removal
- de-shadowing

3.2.4 BRDF correction using BREFCOR

The previously described processes were conducted in the purpose of removing noise and artefacts, caused by internal (sensor) or external (atmosphere) disturbances, from the remotely sensed hyperspectral signal. However, even when no distortions happen, the signal (observed reflectance) heavily depends on the surface it was reflected from. Reflectance is heavily influenced by the viewing angle, solar illumination, time and location (Peddle et al. 2001). With a change

⁵<http://modtran.spectral.com/>

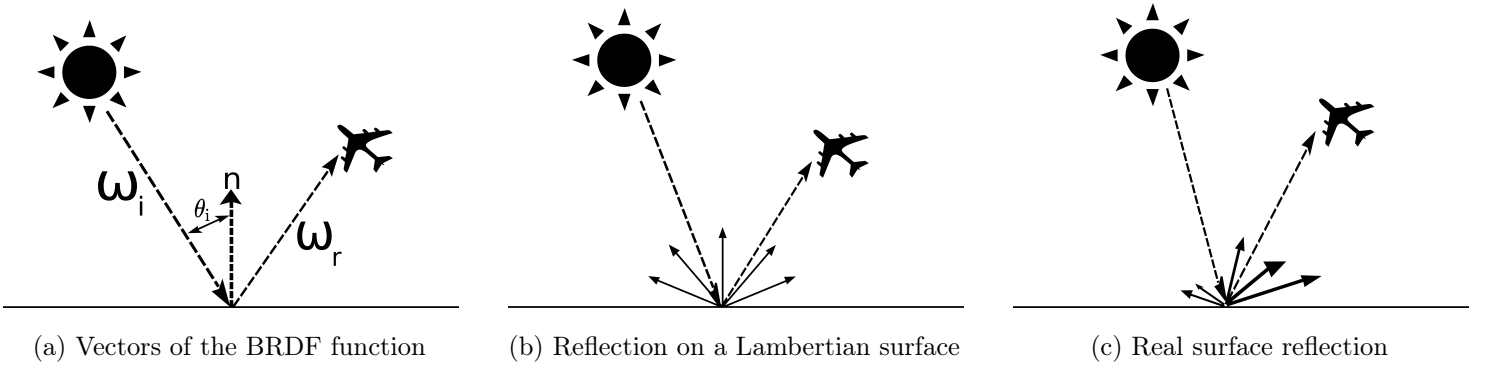


Figure 12: Diagram visualising BRDF effects (simplified)

in angle along the flight line from nadir, the signal measured at the edge of the flight lines is reflected differently than the signal at nadir (see Fig. 13c) resulting in brightness gradients across the flight line.

The illumination effects differ also during the acquisition of the data for each flight line (time, angle) as well as because of different topography (mountains, valleys). In order to be able to classify the image data almost seamlessly, the differences in brightness (reflectance) should be minimized before mosaicking the single lines. The Bidirectional Reflectance Distribution Function (BRDF) as first described in F. Nicodemus et al. (1977) (see Eq. 15) describes the amount of light reflected (wavelength dependent) depending on the observation angle for a given incident light direction (illumination angle) (See Fig. 12a), with further description of the application to remote sensing, and the caveats as summarized in Schaepman-Strub et al. (2006).

$$f_r(\omega_i, \omega_r) = \frac{d L_r(\omega_r)}{d E_i(\omega_i)} = \frac{d L_r(\omega_r)}{L_i(\omega_i) \cos \theta_i d \omega_i} \quad (15)$$

L = Radiance, E = Irradiance, i = Incident light, r = Reflected light, d = Perfectly diffuse

ω_i = Illumination angle, ω_r = Observation angle

As this problem is not exclusive to remote sensing, several mathematical models were developed for correcting these differences (Blinn-Phong, Lafortune, Torrance-Sparrow). However these models are not designed for vegetation surfaces, but for rather generic surfaces. In remote sensing specific models were proposed, in order to compensate for the BRDF of different natural surfaces (Dymond, Roujean, Ross-Thick/Li-Sparse, Walthall). These models take the directional reflectance properties including vegetation and forest canopies into account (T. Koukal et al. 2010), as the distribution of leaves has an significant impact on reflectance properties (see Fig. 13a).

The ATCOR-4 software includes an semi automatic correction algorithm based

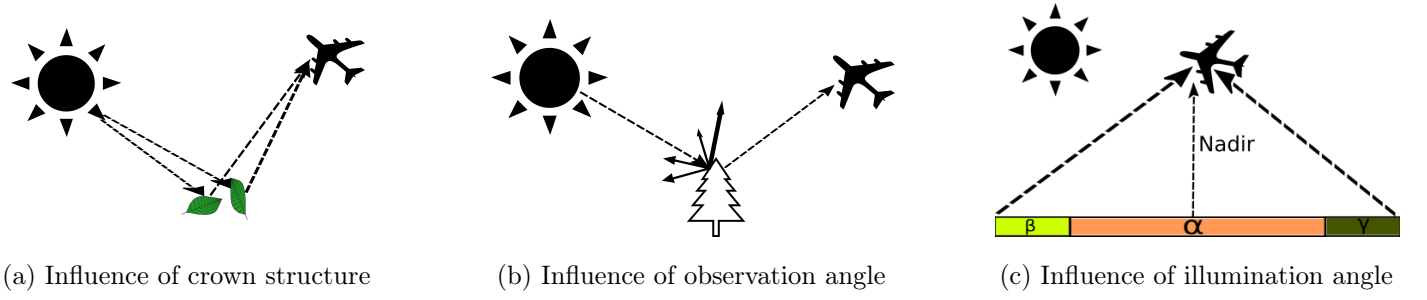


Figure 13: Diagram visualising BRDF problems (simplified)

on the Ross-Thick/Li-Sparse (RTLS) kernel, with the ability to remove hot spot disturbances, as explained in Maignan et al. (2004). The approach is called BREFCOR (BRDF effects correction). BREFCOR uses the surface cover type and the pixel observation angle to find appropriate factors for correcting the BRDF effects. It was tested for suitability with HySpex data (Schläpfer and Richter 2014). The BRF for every pixel and spectral band is given as:

$$\rho_{BRF} = \rho_{iso} + f_{vol}K_{vol} + f_{geo}K_{geo} \quad (16)$$

ρ_{iso} = Isotropic reflectance (defined at nadir)

f_{vol}, f_{geo} = Weighting coefficients, K_{vol} = Ross-Thick kernel function

K_{geo} = Li-Sparse kernel function

The Ross-Thick kernel is used for correcting the volume scattering (Fig. 14a), while the Li-Sparse kernel is used for correcting the geometric scattering (Fig. 14b). For a detailed description of the Kernels see Schläpfer, Richter, and Feingersh (2015) and Richter et al. (2016).

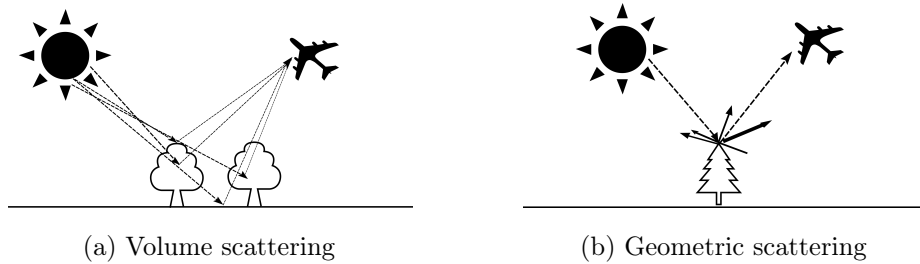


Figure 14: Geometric and volume scattering

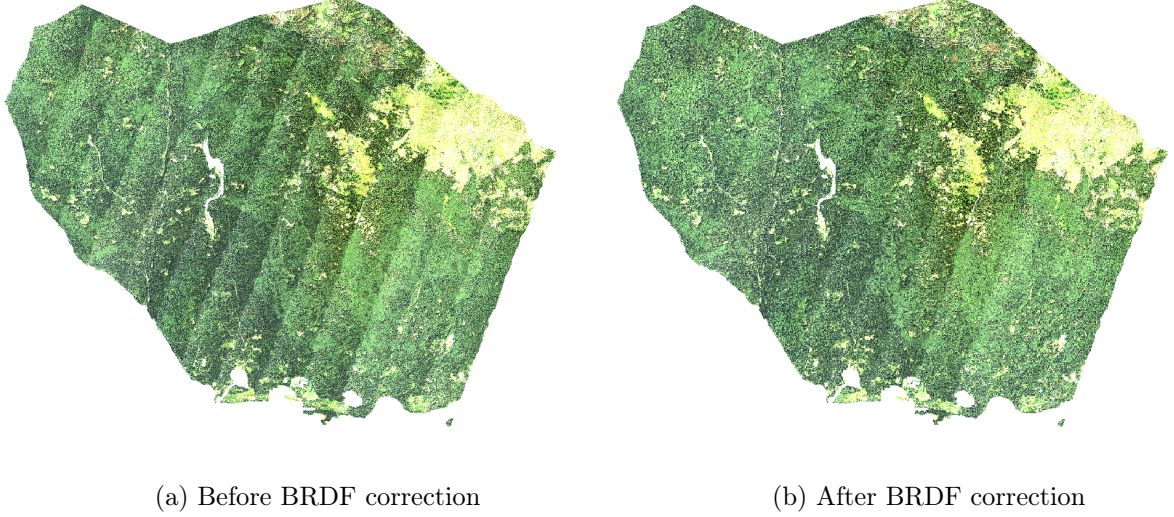


Figure 15: Mosaic of Northern part (RGB)

The user chooses 2-4 flight lines, which should be an optimal representation of the overall image data. These flight lines are classified by the characterization of the surface (BRDF cover index (BCI)). The BCI is based on the BRDF properties of four wavelengths (460, 550, 670, 840 nm) and separates the surface into different classes (water, artificial materials, soils, sparse vegetation, grassland, and forests). The threshold for the classes can be set manually, on an empirical basis, for a better fit to the specific image data. For each BCI class, the kernel weights are computed and averaged over the sample images. For the final correction of the whole image data, the BCI has to be calculated from each image and is used to get a continuous correction function (BRDF model). Based on the BCI values, the observation and illumination angle, and the BRDF model, an anisotropy map is calculated for each wavelength. For every pixel the anisotropy factor (ρ_{iso}) is used in equation 16 for the final image correction.

The result of the BREFCOR correction can be seen in Fig. 15. Figure 15a shows an mosaic created from the flight lines without correction. The single lines are visually distinct due to BRDF effects. Figure 15b shows a mosaic of corrected images, in the west and in the east the correction worked well, and BRDF effects are removed. However in the middle part of the mosaic, the single flight lines are still visible, thus the correction did not work well enough to remove the BRDF effects over the whole image data. As the BREFCOR method is still under development and BRDF removal, especially on full cube data (VNIR and SWIR) is not a trivial task, BRDF correction should further be investigated as results are

crucial for the classification performance. The visually best BREFCOR output was used in this study of for further analysis.

3.2.5 Savitzky Golay Filter

Hyperspectral data consists of many narrow bands, and the noise within the data is also increasing with the number of bands, as the narrow bandwidth can only capture very little energy which gets polluted by self-generated noise of the sensor (Vaiphasa 2006). In addition, every remotely sensed data is disturbed by the atmosphere and various illumination conditions. As these effects can never be corrected completely, the data will always be affected by some noise. In order to increase the signal to noise ratio, in spectroscopy it is common to use polishing algorithms smoothing the reflectance spectra in order to eliminate this noise. However it is important, while noise is reduced, to preserve the characteristics of the spectral data. Different filters are currently used, e.g. Minimum-Noise-Fraction, Mean-Smooth-Filter, and Savitzky-Golay filter (Vaiphasa (2006), Miglani et al. (2011)). Savitzky-Golay is a low-pass filter, which assumes that close data points have significant redundancy, that can be used to reduce the level of noise (Press et al. 2007). It has proven to be a good fit with hyperspectral data (Tsai et al. (1998), King et al. (1999), Schläpfer and Richter (2011)).

$$Y_j = \sum_{i=-\frac{m-1}{2}}^{\frac{m-1}{2}} C_i y_{j+i} \quad (17)$$

(Savitzky et al. 1964)

The parameters to be set is the size of the moving window and the degree of the polynom used. The window size is very important in order to maintain the integrity of the spectrum (Ruffin et al. 2008). A smaller window results in less smoothing, while a wider windows results in changes of the spectral shape. In this case a window size of 3 and the first degree polynom was used.

Other methods exist for not truncating the edges of the spectrum (Hui et al. 1996), for the end of the spectrum the interpolated values were used. Therefore no padded values had to be interpolated. As this was not possible for the beginning of the spectrum, the used library⁶ fits a polynom to the values of the window lengths of the edge. This polynom is used to evaluate the values left of the beginning of the data. The result of the applied filter can be seen in Figure 16.

⁶Scipy 0.18.1

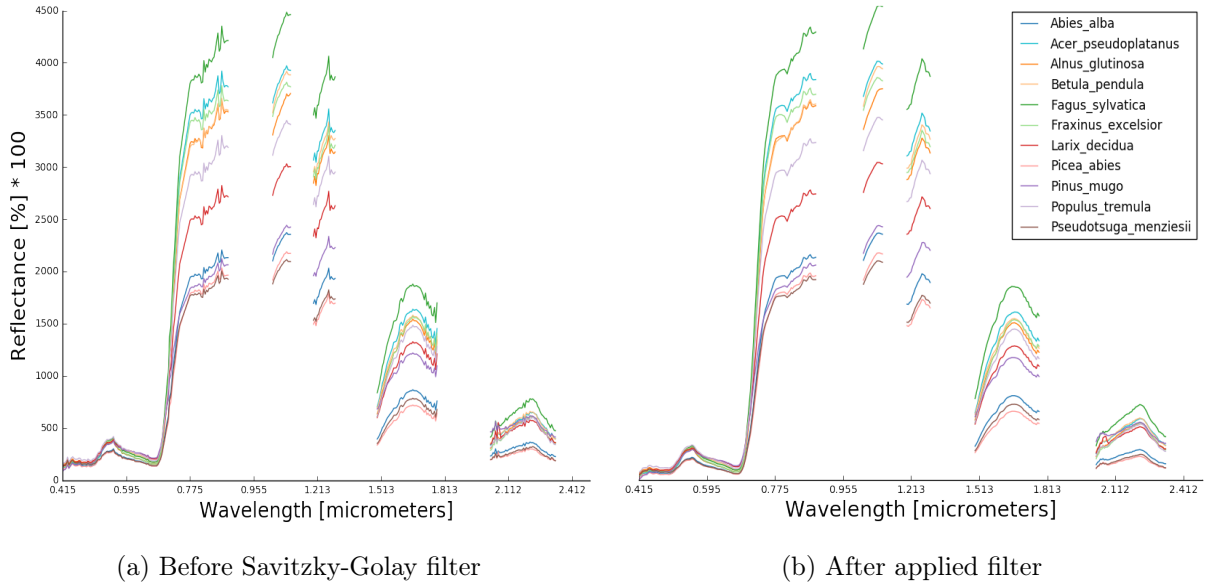


Figure 16: Mean spectrum of species before and after Savitzky-Golay filter

3.2.6 Brightness Normalization

The hyperspectral data were processed in order to I. remove atmospheric effects (3.2.3), II. to correct for the remaining Bidirectional Reflectance Distribution Function (BRDF) (3.2.4) effects and III. to remove noise by polishing the spectrum using Savitzky-Golay filtering (3.2.5). However a small difference in brightness always remains. The sensor detects the radiation reflected from the trees, whereas the amount of photons absorbed by the surface depends on multiple factors such as illumination angle and reflectance angle. The more light is absorbed, the lower (darker) the pixel value becomes (Cervone et al. 2014). The BRDF correction was performed in order to remove difference in across-track brightness. But another important reason for the difference in brightness of single pixels are the different tree species, which reflect light differently due to their habitus (crown shape). Other reason is the large area covered through multiple flightlines where changes in sun and viewing angle effect the illumination geometry. These effects can only be removed to limited extent. Also remaining BRDF artifacts (shaded areas) are common in forests resulting from different reflectance angles of the crown surface (vertical structure). These artifacts can not be corrected by a pixel-based BRDF correction (Schläpfer, Richter, and Feingersh 2015). Thus, the proportional differences of reflectance from the tree crowns is not only dependent on species, but also on the solar and viewing angle. This results in so called 'shadow pixels' in the data. As these pixels, if not represented equally in the training data, would lead to misclassification, each pixel was normalized to a uniform brightness of one.

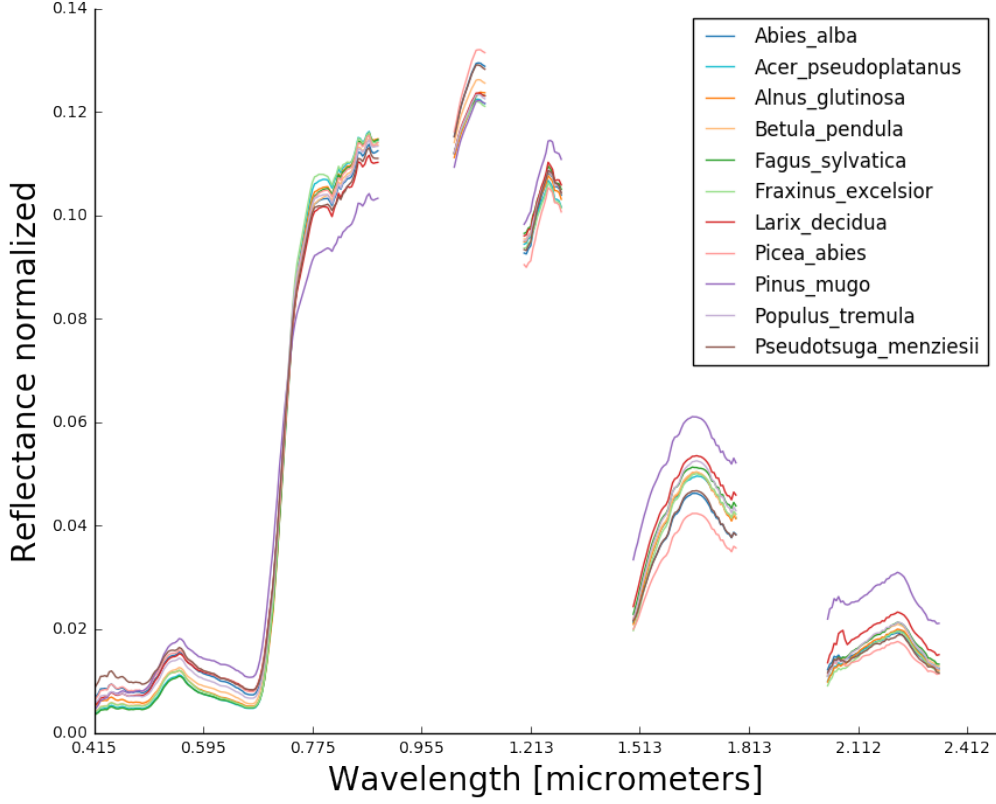


Figure 17: Mean tree species spectrum after Brightness Normalization

The brightness normalization was applied as described in Feilhauer et al. (2010).

$$x_{bn} = \frac{\vec{x}}{\|\vec{x}\|} = \frac{x_i}{\left(\sum_{i=1}^n x_i^2\right)^{1/2}} \quad (18)$$

x = Pixel, $\|x\|$ = Length of vector (bands)

As a result, the intra species brightness difference is removed, while the spectral shape is preserved. As can be seen in Figure 17, it accounts for almost all species, with the exception of *Pinus mugo*, which still shows significant differences in brightness to the other species.

3.3 Sample Data pre-processing

3.3.1 Outlier removal - Isolation Forest

As the ground truth data were collected by different people and also consists of multiple sources (Field work, Inventory data), various errors (mislabeling) could be present. In order to reduce the chance of polluted training data, and thereby the possibility of biased parameters (Ben-Gal 2005), a way of outlier detection seemed necessary. An outlier itself is defined as:

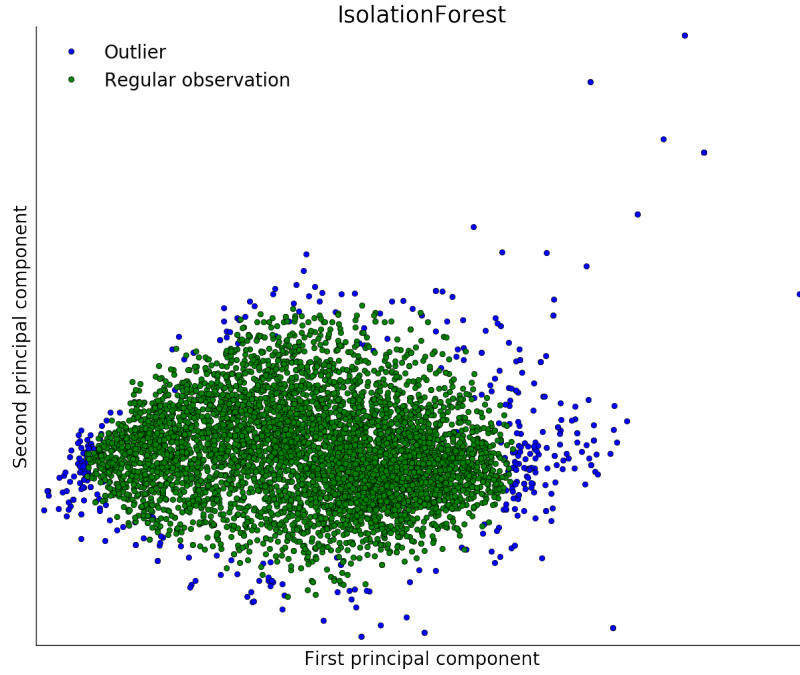


Figure 18: PCA reduced visualization of outliers detected using Isolation Forest

An outlying observation, or “outlier,” is one that appears to deviate markedly from other members of the sample in which it occurs. (Grubbs 1969)

In this context it means a pixel of eg. *Picea abies* is either mislabeled as another species, or not representing forest at all. As hyperspectral data are high dimensional data, and also highly correlated consisting of redundant features, outlier detection methods using density or distance are not a good fit (Aggarwal et al. 2001). As this problem increases with availability of high dimensional data (Curse of dimensionality (Zimek et al. 2012)), other methods than using distance were proposed by (Kriegel et al. 2009, F. T. Liu et al. 2008).

For example the Isolation Forest is a method which is able to work with data where the existence of outliers is not known (F. T. Liu et al. 2012). It works by utilizing the abilities of a random tree to classify data, as outliers (or anomalies) are supposed to be "few and different" from the other sample data. This leads to the fact that outliers are classified closer to the root of the tree. The algorithm measures the path length, averaged over a forest of random trees for every sample. The path length is then treated as a measure of abnormality, detecting outliers. For this study the ground truth data (spectral values of the pixels) was checked for outliers using the IsoForest module⁷ provided in scikit-learn. In order to

⁷<http://scikit-learn.org/stable/modules/generated/sklearn.ensemble>.

visualize the result of the outlier detection using the Isolation Forest algorithm the dimensionality of the dataset was reduced using a Principal Component Analysis (PCA) and is shown in Figure 18.

4 Supervised Machine Learning

Supervised learning refers to a process that learns a function from a tuple $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, of *inputs* (also called predictors) and *outputs* (also called label), in order to predict the label of inputs where the output is not known. Every single input variable (*instance*, eg x_1) consists of multiple *attributes* (also called *features*). Two typical examples of supervised learning are classification and regression tasks (Sammur et al. 2011). Given a data-set X , for some instances of the data-set (x_i) , a corresponding label is known (y_i) . This data is used in order to train an algorithm to predict the corresponding label of the instances of X , where (y_i) is unknown. So the algorithm defines a function (f) , which predicts \hat{y} as denoted in $\hat{y}_i = f(x_i)$. Supervised classification is one of the most frequently used procedures for classifying remotely sensed data (Richards 2012b). There are many algorithms available, designed for this task, such as Support Vector Machines (SVM), Neural networks, Nearest neighbor methods, Forests of randomized trees. However, because some of these algorithms were primarily designed to handle low-dimensional data, problems can arise when dealing with high dimensional data (Fuan et al. 2002).

4.1 Random Forest

One of the most prominent supervised classifiers at this time is the Random Forest algorithm proposed by Breiman (2001). It is a widely used implementation of random decision forests, and is based on an ensemble of decision trees (see Fig. 19). It is therefore categorized as an ensemble classifier. The Random Forest algorithm combines decision trees and bagging. Bagging (Bootstrap aggregating) is used in order to reduce the variance (Breiman 1996a), as a side effect it enables the estimation of the generalization error (OOB error). The samples are drawn at random with replacement from the training instances, the unused part of the samples is utilized to calculate the out-of-bag error (see 2.3.1). Each tree is grown with a randomly sampled feature set (random subspace method (Ho 1998)) selected for each split (4.8.2). This way the correlation between the single trees is decreased, which improves the accuracy of the algorithm (Hastie et al. 2009).

[IsolationForest.html](#)

This method has shown to perform better than using bagging only, especially when many redundant features are present (Archer et al. 2008). In the original implementation of Random Forest, all trees take a vote on the predicted class (majority voting). The implementation applied in this study (Pedregosa et al. 2011) uses an averaged probability prediction (as seen in Fig. 40). However the differences should be negligible (Breiman 1996a).

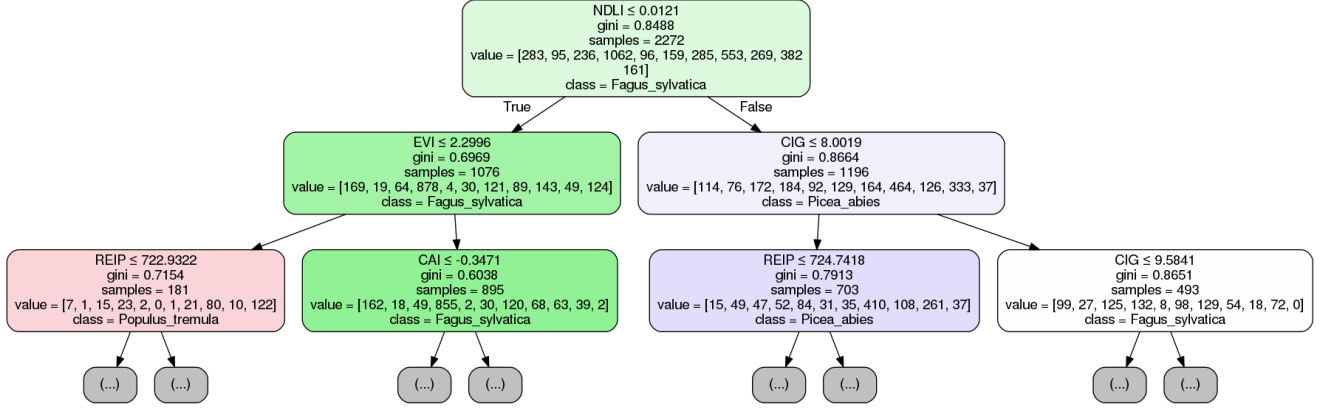


Figure 19: Single tree showing first two levels of tree species classification using vegetation indices

The main advantages of random forests are:

- Non parametric
- Robust against outliers
- Embarrassingly parallel
- Achieves a high prediction accuracy with only tuning a few parameters (4.8)

(Hastie et al. 2009, Cutler et al. 2012, Qi 2012)

Random forests are widely used in the field of environmental remote sensing for surface classification (Pal 2005, Gislason et al. 2006), in particular in combination with hyperspectral data (Crawford et al. 2003, Ham et al. 2005, Chan et al. 2008). It is also known as robust supervised learning method in general (Caruana et al. 2006, Fernández-Delgado et al. 2014).

4.2 Extremely randomized trees

Extremely randomized trees (ET) were developed by Geurts et al. (2006). The forest of randomized trees is based on RF, but the algorithm takes the randomization to the next level. The trees are also grown using bagging for the samples

and the random subspace method, as described in section 4.1. But in addition cut points are set at random (chosen randomly among the training data) and the split point with the best result is selected in order to compute the split for all features present in the random subspace of the maximal features considered (4.8.2). In the default settings each tree is built from the complete training set (no bootstrapping) as a measure to further reduce bias, but it should be noted that in order to obtain the OOB score, bootstrapping was used in this study.

4.3 Choosing one classifier

The decision which classifier to use was made based on several reasons. While both classifiers, Random Forest (RF) and Extremely randomized trees (ET), showed high stability after enough trees, Random Forest (RF) showed a minimally better performance regarding the error metrics (Fig. 24). On the other hand during evaluation of the classifiers and test runs, Extremely randomized trees (ET) classifier showed a better avoidance of over-fitting, especially when the DTM was used. But one of the biggest advantages regarding ET is the computational benefit (Geurts et al. 2006).

Intel(R) Core(TM) i7-6700K CPU @ 4.00GHz

Fitting: 5 Fold cross validation:

RF: 3min 37s 10min 30s

ET: 0min 22.7s 1min 28s

Especially during the tuning of the classifier parameters (section 4.8) and cross-validation, this performance difference becomes apparent. As ET performs around ~8 times faster, with the benefit of better avoidance of over-fitting (Geurts et al. 2006), this algorithm was finally applied in this study.

4.4 Features

The previously describe available data sets (2.2.2 - 2.2.5), are combined for further processing to one single database (Table 7). Each individual measured or calculated property of an instance is called *feature*, eg. each band (λ) represents one feature. In total 275 features ($||X||$) are available.

Table 7: Available data for Machine Learning

Spectral Data			LiDAR data	
VNIR	SWIR	Vegetation indices	DTM	Treecount
131 Bands	133 Bands	9 Indices	Meters above MSL	Tree density

From all available data various combinations of features can be employed as input for classification and compared later on for their classification performance. All optional combinations used in this analysis are listed in Table 8 with an abbreviation.

Table 8: Combinations of features used

	Spectral Data			LiDAR data	
Abbrv.	VNIR	SWIR	VI	DTM	Treecount
All	X	X	X	X	X
SpecInd	X	X	X		
Spectral	X	X			
SWIR		X			
SWIRFull		X	X	X	X
VNIR	X				
VNIRFull	X		X	X	X

For visualizing the effect of the DTM on the map accuracy (possible over-fitting), all feature combinations were also calculated with and without the DTM input. Also an additional run was conducted with a reduced number of bands (band selection (BS), see section 4.6). It should be mentioned that all vegetation indices (VIs) were calculated using both VNIR and SWIR bands (see 2.2.3). Thus the VIs are not supposed to replace the spectral bands, but to detect possible advantages by using them.

4.5 Feature engineering

In Machine Learning it is common to extend the feature space through applied domain knowledge. This happens by aggregating, combining or splitting features to create new features. This process is also called feature construction.

Feature construction [...] is a process that discovers missing relations among features and augments the space of features by inferring or creating additional compound features, and thus could expand the feature space.

H. Liu et al. (1998)

As an example, if the relationship of the label of the two instances (x_1 and x_2) depends mainly on the angle of two features, it can increase the classification result, if that relationship is added to the feature space. This approach of including the relationship of features in an analysis is not new in the field of applied remote sensing. Vegetation indices (2.2.3) have been in use since the 1970s (Bannari et al. 1995), and can be considered a feature construction method in the machine learning sense, as the aim is to construct new representations of the data by combining multiple bands.

4.6 Feature selection

In hyperspectral data exploitation it is common to apply band selection to the data in order to remove highly correlated bands and noise. Due to the large number of spectral bands the difference in information contained in bands close to each other is often minimal (Spectral redundancy, see Foody et al. 2004 and Richards 2012a). In Machine learning such an approach can be considered as Feature selection, as the goal is to reduce the numbers of features necessary for achieving a good classification. The main outcome of a reduction in features is the reduced computation time, while also the risk of over-fitting can be reduced (Bolón-Canedo et al. 2015), and the accuracy can be increased (Janecek et al. 2008). In addition Hänsch et al. (2016) have shown that due to the redundancy of information in correlated spectral bands, random forests are able to achieve an already very good classification result with less bands (see Figure 20). However Forests of randomized trees automatically determine which features are the most useful for class separation during the training phase (see 4.6.1). The amount of features used to construct a forest is one of the parameters tuned while optimizing the classifier (4.8.2). This study should not only output results of a classifier, but also aims at investigating how the prediction results are influenced by different approaches. Therefore, the decision was made to include some kind of band selection and perform the classification with and without it, in order to evaluate the influence of the removal of certain spectral bands on the classification result.

4.6.1 Feature importance

The Gini importance (mean decrease in impurity) measure can be used to rank the features according to their importance in the models ability for class separation. Alternatively, the mean decrease in accuracy could be used as well, however it has been proposed to use the Gini importance for studies with smaller sample sizes (Archer et al. 2008). The Gini importance is based on the Gini impurity and measures how often a feature was selected for a split. As splits based on Gini impurity try to produce pure nodes (Breiman 1996c), the features selected for a higher amount of splits resulting in purer splits, is considered more important than the other features. Or in other words the Gini importance is based on the weighted impurity decrease at all nodes the variable was used, averaged over all trees (Louppe et al. 2013). But studies have shown that using the Gini criterion for feature importance can be biased (Strobl, Boulesteix, Zeileis, et al. 2007). However this is only true if features are not of the same measurement scale. If only continuous features are used (eg. spectral data), feature importance based on the Gini criterion can be used for feature selection. But that still leaves a problem, because even so the algorithm scores features without bias, one still doesn't know how many features are relevant. Especially as the Gini importance tends to rank the first features very high while for remaining features a sharp drop in importance is observed. Another problem is that the algorithm can't differentiate between correlated features (Archer et al. 2008, Menze et al. 2009), and has problems scoring them correctly (Strobl, Boulesteix, Kneib, et al. 2008, Tolosi et al. 2011, K. Nicodemus 2011). The usage of feature importance is a highly controversial topic (Gregorutti et al. 2016) when correlated predictors are included. This does not change the prediction results, but from an analysis point of view the feature importance can't be reliably used to distinguish important features from non important ones if the features are highly correlated. This makes feature importance unsuitable for feature selection when working with hyperspectral data.

4.6.2 Recursive feature elimination

Recursive feature elimination is a method where an estimator is trained repeatedly with a given set of features, and during every run one feature is left out. In a cross-validation loop the accuracy (the ratio of correctly predicted observations to the total amount of observations) of the classifier is evaluated to find the optimal number of features. That increases the stability of the selection results

by reducing the effect of the correlation on the importance measure (Gregorutti et al. 2016).

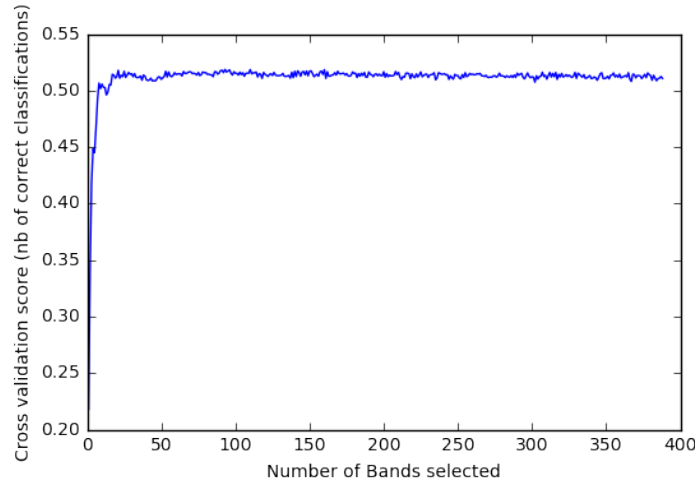


Figure 20: Recursive feature elimination for hyperspectral data. Showing an optimal number of 93 Bands

The recursive feature elimination clearly shows that the error rate for predictions with spectral data stabilizes after around 50 features (Figure 20). The algorithm itself determines 93 features as an optimal cut off point. But on the other hand the plot also shows that the error rate does not increase even if all features are used. The applied recursive feature elimination resulted in the selected bands shown in Figure 21 in red. For an overview, see Table 20

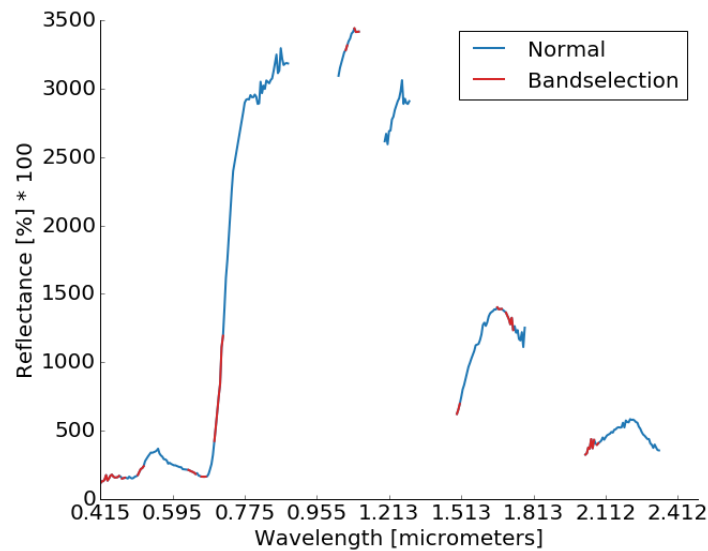


Figure 21: Selected bands in red on a sample vegetation spectrum (in blue, interpolated bands not shown).

4.7 Test Training split

It is best practice to split the data-set into three parts, one for training, one for testing and one for validation (Hastie et al. 2009). However splitting a data set into three parts limits the size of the data available for training. This problem is even greater in this study, as the data available for some species is already very limited. A solution to this problem is to split the data only into training and test samples and use a cross validation (2.3.2) (and also OOB error 2.3.1) for validation and estimating the hyper-parameters (Louppe 2014).

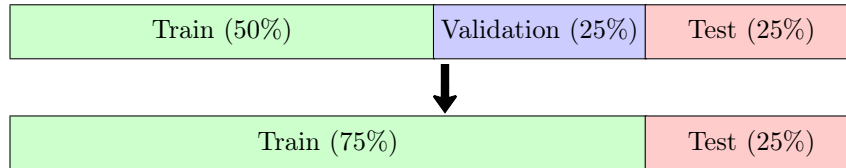


Figure 22: Common splitting

This way, over-fitting can be avoided (through leaking knowledge of the Testset into the tuning process), and an untouched Testset is available for evaluating the model generalization.

4.8 Hyperparameter optimization

For using a learning algorithm, in this case Extremely randomized trees, some preferences need to be set in advance, which can't be learned from the training data. These parameters are estimated by evaluating the predictive performance of each model trained with these parameters (Bishop 2006). The common way to approach hyperparameter tuning in a machine learning task is either manually, or conducting a grid search, or because of the high computational requirements, do a random search (Bergstra and Bengio 2012).

As mentioned before one of the main advantages of classifiers based on forests of randomized trees is the minimal amount of parameters to evaluate to obtain a high accuracy. Other learning algorithms such as Neural networks consists of more than ten parameters (Bergstra, Bardenet, et al. 2011) which have to be evaluated (e.g learning rate, batch size, number of hidden units, regularization, training iterations (Bengio 2012)). For a forest of randomized trees the main parameters are:

- Number of trees in the forest
- Number of features in random subwindows
- Size of the trees (maximum depth, number of samples for splitting)

While the maximum depth of the trees is desired to be infinite (until all leaves are pure, no pruning (Breiman 2001)), there is no need for optimization of that parameter. On the other hand the number of trees (4.8.1) and the number of features considered for splitting (4.8.2), are the main parameters with the greatest influence on the randomization of the forest, and therefore should be evaluated in detail.

4.8.1 Maximum Trees

The main goal is to grow enough trees so that every attribute is used multiple times. As there is only the computation cost as a downside of using a larger forest (the bigger the better for prediction, with a saturation of decreasing the error rate), the aim is using as many trees as needed for the best result possible. To achieve this goal, the OOB error is obtained for the classifiers, each with an increasing amount of trees. In order to visualize it is common to plot the error rate against the number of trees (as seen in Figure 23). The error rate drastically decreases with the usage of 100 up to 250 trees, after around 500 trees the error rate has been stabilized.

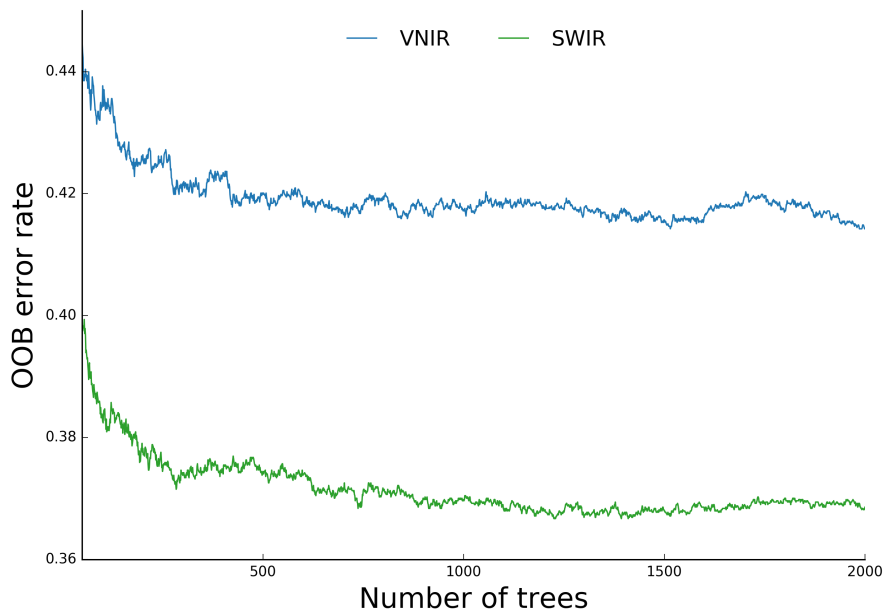


Figure 23: Error rates for OOB in relation to number of Extremely randomized trees

4.8.2 Maximum Features

The size of the random subspace, defines how many features are considered at each split (maximum features). The less features the greater the randomization, but on the other hand that also reduces the possibility of including relevant features at each split. In a high dimensional dataset with expected high correlation between the features, it is important to evaluate this parameter. The error rates for decimal steps of this parameter are shown in Figure 24. It shows a steep decline in the error rate at the beginning, with a stabilization when 70% to 80% of the features are included. This is reflected in the findings of Geurts et al. 2006, as stated in the original paper that an increase of the maximum features considered at splitting leads to a decrease of the error rate when a high percentage of the features is irrelevant. As more features are considered for splitting, that increases the possibility of filtering out the irrelevant ones. Of course if one should set the maximum features to 1 (all) that would mitigate the effect of decorrelating the single trees using random subspaces, increasing the possibility of over-fitting (Bernard et al. 2009).

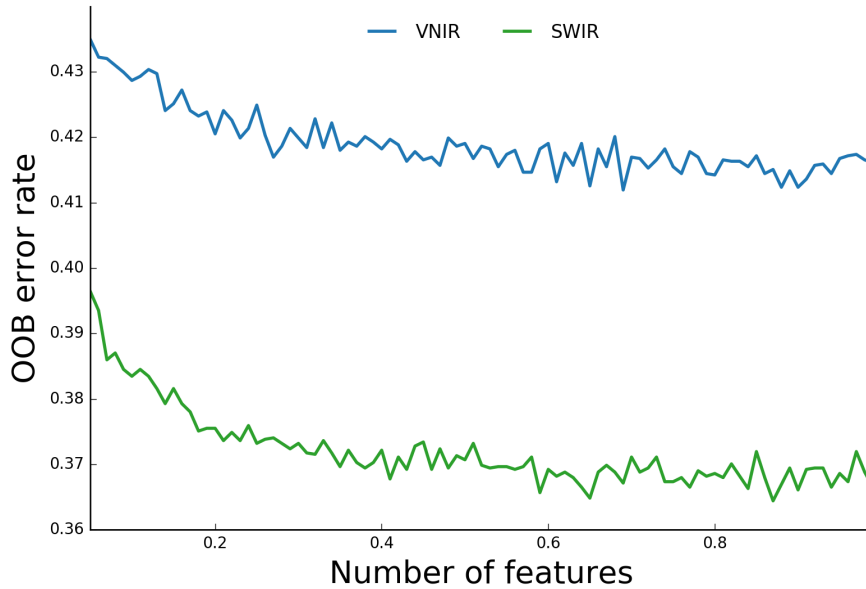


Figure 24: Error rates for OOB in relation to size of random subwindow

4.8.3 Class weight

Despite the fact that the problem of in between class imbalance is common in multiclass classification tasks, because of the rare occasion that natural processes produce multiple classes (Laurikkala 2001) which are rarely balanced, the focus usually lies on data imbalance of binary classification tasks (S. Wang et al. 2012). In these cases class imbalance lies in the order of 100:1, 1.000:1, or sometimes even 10.000:1 (Chawla et al. 2004, He et al. 2009). Techniques accounting for that problem, can be grouped in two categories, either cost sensitive learning, or different sampling techniques (over- undersampling, SMOTE, overbagging). For the field of remote sensing the oil spill study by Kubat et al. 1998 is one of the most prominent examples for an imbalanced binary classification task, dealing with an imbalance ratio of 22:1 (96% belonging to one class). The main problem of the usage of imbalanced data in an forest of randomized trees is, while drawing a bootstrap sample, the high possibility that a minority class is underrepresented in that sample. That results in a classifier focused towards correct classification of the majority class, which can decrease the accuracy of the minority classes (Chen et al. 2004). The methods developed for binary classification problems are also used for balancing multiple classes. For random forests the proposed way is an approach at the algorithmic level, instead of stratifying the sampling and thus changing the data to use weights (see Equation 19) incorporated into the classifier. These weights (class and/or sample weights) are used in order to balance the number of samples used for the calculation of the probability estimates (Chen et al. 2004). This approach offers the possibility of cost sensitive learning by using a higher weight for a minority class. That can lead to a better result of classifying minority classes, which shows that a balanced dataset must not always lead to the best possible results (Mellor et al. 2015).

The used implementation offers different possibilities for calculating the weights, as well as accepting an array with pre defined weights for each sample / class. The two different methods for automatic calculation are 'balanced' and 'balanced subsample' where the heuristic weighs the classes proportionally to the inverse of their frequency.

$$\text{weight} = \frac{\text{Number of total instances}}{\text{Number of classes} \times \text{Number of samples}} \quad (19)$$

Either for the overall sample or for every bootstrapped sample. The results of the different approaches are shown in Table 9. For this study, the various degrees of

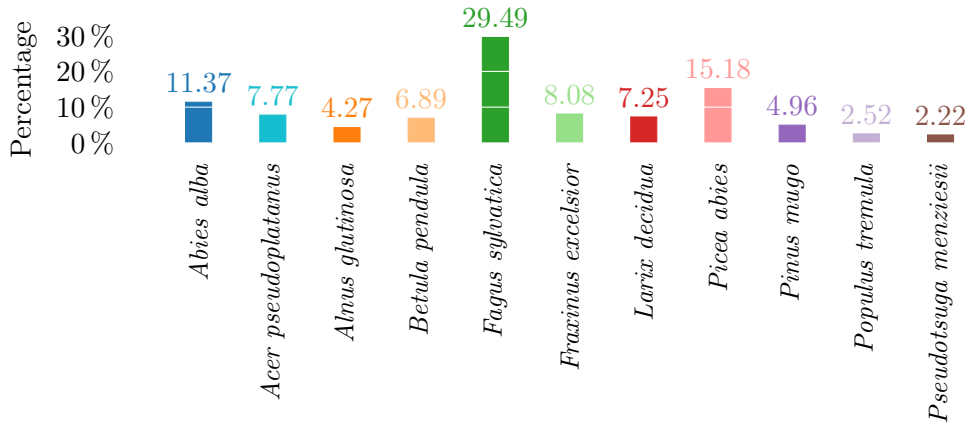


Figure 25: Pixels per tree species (percentage)

Table 9: F1 scores for validation of weights using Spectral data

Scientific Name	No weights	Balanced	Sample balanced
<i>Abies alba</i>	0.45	0.46	0.47
<i>Acer pseudoplatanus</i>	0.23	0.20	0.19
<i>Alnus glutinosa</i>	0.44	0.48	0.48
<i>Betula pendula</i>	0.44	0.42	0.46
<i>Fagus sylvatica</i>	0.79	0.78	0.78
<i>Fraxinus excelsior</i>	0.71	0.72	0.73
<i>Larix decidua</i>	0.53	0.55	0.57
<i>Picea abies</i>	0.69	0.70	0.69
<i>Pinus mugo</i>	0.94	0.93	0.93
<i>Populus tremula</i>	0.26	0.28	0.28
<i>Pseudotsuga menziesii</i>	0.08	0.16	0.15
Weighted F1	0.614	0.609	0.611
Kappa	0.562	0.559	0.560
OOB	0.639	0.639	0.640

imbalance can be seen in Figure 25. The majority class is *Fagus sylvatica* which has a ratio of 13:1 to the minority class of *Pseudotsuga menziesii*. This is a fairly low imbalance compared to other studies.

As can be seen the changes of the F1 score for the different tree species are only minimal. As well as the Kappa and OOB statistics. The reason is, that the overall between class imbalance is not as big as in usual binary classification tasks, and therefore no major improvements using balanced classes could be achieved.

5 Results

As the performed band selection did not provide any significant results, and also due to the large amount of various combinations, all following normalized confusion matrices and corresponding images showing the spatial distributions, display the results of the classifiers without the applied band selection. However the effect of the algorithm for selecting relevant bands, can still be studied in the respective tables.

Table 10: Overview of all calculated classifier results

Group	Classifier accuracy	Predicted results	Confusion matrix	Image
Species	Link	Link	Link	Link
Species Groups	Link	Link	Link	
Conifers / Broadleaved	Link	Link		

5.1 Accuracy assessment

It is not only necessary to evaluate the model performance (Classifier accuracy), but also to evaluate the overall prediction results (Map accuracy). As Shi et al. (2016) have shown, even the classifier parameters can have an affect on the map accuracy up to 16%, even when the classifier accuracy is not really affected. This is even more true when different features are combined. In addition, even when the obtained evaluation metrics show a good fit of the classifier, it does not automatically imply a good generalization performance (possible overfitting). This is even more important when the size of the training data is relatively small compared to the prediction data (Geurts 2002). In order of being able to compare the single species to each other and how the results have changed due to the combination of features, the previously described metrics are listed in the Tables 11, 12, 13.

$$c_v = \frac{\sigma}{\mu} \quad (20)$$

The Coefficient of variation (c_v) was additionally utilized as a measure of spread to describe the amount of variability relative to the mean. This way conclusions regarding the stability of the prediction results for a tree species can be drawn. The main differences for the classifications are presented in the following two sections, first for the classifier accuracy and second for the predicted results. Due to the vast possible combinations only the most important classifiers are described in detail.

5.2 Classifier accuracy

All metrics are calculated as the mean of 10 random iterations of each feature set, with an random Split and Seed used at each time. Standard deviations are not listed for each group, as it was insignificant for all three groups. For the group "Species" the standard deviation for the three metrics used is listed below:

• CV: 0.0014 • Kappa: 0.0028 • OOB: 0.0018

The highest value for each metric / species, is marked (■).

5.2.1 Classifier accuracy: Species

The classification results are obtained using all available classes (eleven tree species listed in Table 2).

Table 11: Classifier accuracy for Species

Classifier Data							F1-Score										
ID	Data	DTM	BS	OOB	CV	Kappa	AA	AP	AG	BP	FS	FE	LD	PA	PMu	PT	PM
1	all	True	True	0.809	0.796	0.762	0.64	0.77	0.84	0.79	0.88	0.79	0.83	0.76	0.93	0.63	0.52
2	all	True	False	0.812	0.797	0.774	0.67	0.74	0.82	0.75	0.9	0.82	0.81	0.77	0.95	0.67	0.55
3	all	False	True	0.704	0.692	0.624	0.52	0.47	0.66	0.63	0.84	0.7	0.6	0.66	0.91	0.49	0.11
4	all	False	False	0.699	0.688	0.649	0.55	0.56	0.58	0.6	0.86	0.69	0.67	0.73	0.86	0.31	0.07
5	specind	False	True	0.661	0.653	0.558	0.44	0.26	0.57	0.5	0.79	0.7	0.51	0.61	0.88	0.4	0.12
6	specind	False	False	0.655	0.646	0.589	0.56	0.29	0.48	0.55	0.8	0.64	0.63	0.65	0.88	0.59	0.14
7	spectral	False	True	0.626	0.622	0.579	0.47	0.34	0.56	0.36	0.81	0.73	0.63	0.66	0.9	0.3	0.12
8	spectral	False	False	0.642	0.632	0.555	0.46	0.24	0.47	0.36	0.77	0.64	0.59	0.69	0.92	0.36	0.08
9	swir	False	True	0.613	0.604	0.519	0.39	0.26	0.33	0.26	0.74	0.73	0.45	0.68	0.85	0.24	0.08
10	swir	False	False	0.609	0.598	0.539	0.52	0.32	0.36	0.23	0.74	0.69	0.52	0.71	0.82	0.06	0
11	swirfull	True	True	0.837	0.826	0.829	0.74	0.83	0.89	0.77	0.91	0.91	0.85	0.84	0.96	0.81	0.27
12	swirfull	True	False	0.845	0.834	0.8	0.77	0.74	0.85	0.81	0.88	0.88	0.85	0.83	0.96	0.74	0.45
13	swirfull	False	True	0.716	0.705	0.648	0.48	0.59	0.7	0.62	0.84	0.72	0.66	0.69	0.86	0.62	0
14	swirfull	False	False	0.709	0.696	0.639	0.49	0.54	0.67	0.64	0.85	0.74	0.64	0.66	0.91	0.47	0.07
15	vnir	False	True	0.569	0.561	0.495	0.38	0.26	0.37	0.36	0.77	0.49	0.59	0.57	0.85	0.27	0.06
16	vnir	False	False	0.565	0.56	0.468	0.37	0.27	0.44	0.29	0.76	0.48	0.53	0.52	0.86	0.28	0.07
17	vnirfull	True	True	0.811	0.801	0.784	0.69	0.81	0.92	0.77	0.89	0.84	0.85	0.76	0.94	0.69	0.44
18	vnirfull	True	False	0.811	0.798	0.778	0.71	0.82	0.82	0.74	0.9	0.81	0.8	0.76	0.95	0.72	0.65
19	vnirfull	False	True	0.672	0.666	0.64	0.48	0.59	0.62	0.65	0.86	0.75	0.62	0.67	0.86	0.25	0.12
20	vnirfull	False	False	0.678	0.664	0.622	0.47	0.55	0.56	0.65	0.83	0.65	0.63	0.7	0.88	0.53	0.19
C_V							0.245	0.414	0.289	0.333	0.067	0.155	0.192	0.117	0.048	0.438	0.94

Highest scores

The highest overall scores (OOB, CV, Kappa) were achieved using the SWIR spectrum combined with the vegetation indices and the LiDAR data (SWIRfull, ID 11 & 12). Differences between the classifiers with reduced bands (band

selection) and those with all bands are small (e.g. Kappa 0.800 to 0.829). This is also true for the F1-scores of the single tree species, only *Pseudotsuga menziesii* showed an significant increase for the classifier ID 18. However this can partially be attributed to the usage of the DTM (generally lower scores, if the DTM is not used) as well as the limited support of pixels available for testing. If the DTM is not used the Kappa scores for the feature combination of SWIRfull drops to 0.639 (ID 14), this is almost the same score as using all the available data except the DTM (ID 4).

Comparison VNIR / SWIR

When comparing the results for the spectral regions of VNIR and SWIR without a band selection (ID 16 & 10), it shows an overall increased result in favour of the SWIR region (0.468 to 0.539 κ). Regarding the single tree species (F1 score), the biggest difference can be seen in the classification of *Fraxinus excelsior* and *Picea abies*, which have an increased score from 0.48 to 0.69 and 0.53 to 0.71, respectively. Whereas the usage of the SWIR region decreased the results for *Populus tremula* from 0.28 to 0.06.

Comparison Spectral / Spectral & VIs

The usage of the VIs showed only an very limited increase in κ scores (0.555 (ID 8) to 0.589 (ID 6)). This increase is mostly due to the better classification (as seen in an increase in F1-scores) for *Abies alba* (0.46 to 0.56), *Betula pendula* (0.36 to 0.55) *Populus tremula* (0.36 to 0.59). For the other species, the usage of the chosen VIs did not add a significant change in classification, neither positive or negative.

Tree species classification

The most robust results were obtained for *Pinus mugo* (c_v 0.048) and *Fagus sylvatica* (c_v 0.067). This can be attributed in the case of *Fagus sylvatica* to the size of available training data for this species. In the case of *Pinus mugo* to the habitus of this species, which is unique within the classified species. For both species, robust predictions can be made by using a multitude of features, the difference for prediction based on SWIR or VNIR is only 0.02. The highest variation was recorded for *Pseudotsuga menziesii* (c_v 0.94), which can partially be explained by the very limited amount of pixels. The overall correct classification for *Pseudotsuga menziesii* is also very limited, especially for those classifiers where no LiDAR data was used.

5.2.2 Classifier accuracy: Species Groups

The species groups represent a grouping of tree species according to common German forestry standards. The idea behind using these groups is that species with similarities in habit and ecological niche, share spectral similarities. This could lead to an increased accuracy, by reducing the number of classes from 11 to 6, and thereby reducing the complexity of the classification task. Also the amount of training data for each group is increased by combining the tree species with a small amount of pixels available for training (see Figure 26).

Applied groupings:

Larix decidua, *Pinus mugo*, *Pseudotsuga menziesii* as 'other coniferous' (**Sndh**).

Acer pseudoplatanus, *Fraxinus excelsior* to 'real hardwood' (**Elbh**), and

Alnus glutinosa, *Betula pendula*, *Populus tremula* to 'other deciduous' (**Slbh**).

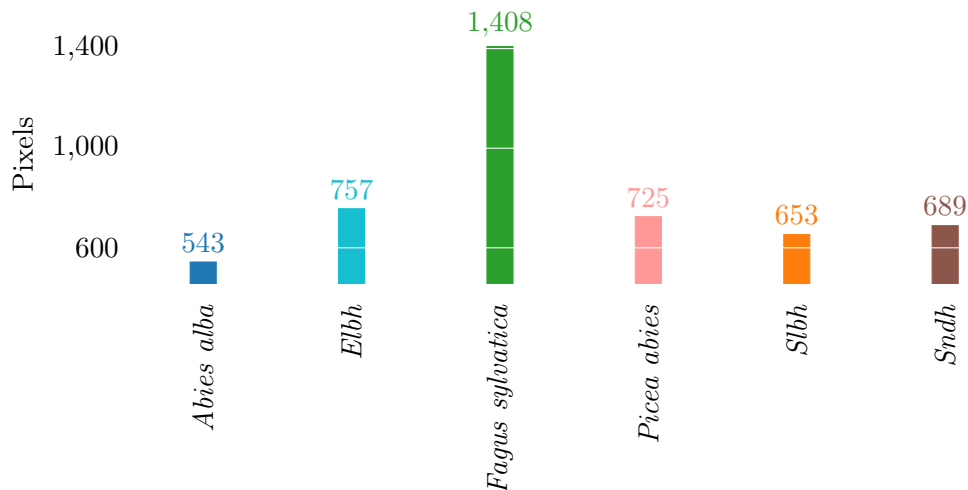


Figure 26: Training pixels per species group

Table 12: Classifier accuracy for Species Groups

Classifier Data							F1-Score					
ID	Data	DTM	BS	OOB	CV	Kappa	AA	Elbh	FS	PA	Slbh	Sndh
21	all	True	True	0.802	0.793	0.774	0.69	0.86	0.9	0.76	0.84	0.76
22	all	True	False	0.804	0.797	0.768	0.66	0.83	0.89	0.76	0.82	0.78
23	all	False	True	0.717	0.707	0.684	0.55	0.73	0.86	0.7	0.67	0.75
24	all	False	False	0.725	0.716	0.662	0.49	0.72	0.86	0.71	0.67	0.73
25	specind	False	True	0.694	0.683	0.635	0.51	0.65	0.84	0.71	0.62	0.7
26	specind	False	False	0.701	0.691	0.627	0.49	0.65	0.83	0.68	0.64	0.69
27	spectral	False	True	0.668	0.654	0.636	0.52	0.72	0.82	0.71	0.61	0.7
28	spectral	False	False	0.675	0.666	0.605	0.43	0.66	0.81	0.68	0.57	0.71
29	swir	False	True	0.664	0.65	0.566	0.47	0.6	0.76	0.67	0.49	0.67
30	swir	False	False	0.662	0.653	0.577	0.44	0.64	0.76	0.68	0.6	0.61
31	swirfull	True	True	0.825	0.818	0.819	0.77	0.87	0.91	0.8	0.85	0.84
32	swirfull	True	False	0.832	0.823	0.805	0.78	0.82	0.9	0.82	0.83	0.84
33	swirfull	False	True	0.733	0.727	0.664	0.49	0.7	0.86	0.71	0.67	0.73
34	swirfull	False	False	0.729	0.724	0.693	0.54	0.75	0.87	0.73	0.71	0.74
35	vnir	False	True	0.613	0.607	0.534	0.47	0.57	0.8	0.59	0.44	0.59
36	vnir	False	False	0.617	0.614	0.542	0.41	0.62	0.79	0.59	0.5	0.65
37	vnirfull	True	True	0.815	0.81	0.744	0.67	0.8	0.87	0.7	0.78	0.82
38	vnirfull	True	False	0.813	0.801	0.772	0.7	0.78	0.89	0.73	0.86	0.82
39	vnirfull	False	True	0.695	0.691	0.641	0.52	0.68	0.85	0.68	0.61	0.72
40	vnirfull	False	False	0.684	0.681	0.67	0.54	0.7	0.87	0.69	0.66	0.74
\mathcal{C}_V							0.211	0.119	0.052	0.081	0.187	0.096

Highest scores

The 'best' result for species groups was also achieved using the SWIRfull combination (ID 31 & 32). With a κ score of 0.805 (32) for the classifier not using the DTM, which is very similar to the κ score of the single tree species classification using the same features (0.800)(12).

Comparison VNIR / SWIR

The differences between VNIR and SWIR are comparatively small between species groups, the biggest difference is visible between other broadleaved trees (Slbh) with a difference of +0.1 in F1-scores, and *Picea abies* (+0.09), favouring the SWIR spectrum.

Comparison Spectral / Spectral & VIs

The maximum change from adding VIs appeared for real hardwoods (Elbh) with a small increase of 0.07. This is only a small change, which can also be seen in the minimal overall change of κ scores (0.026).

Species groups classification

The main difference is an increased κ score for the classifier not using the DTM (ID 34) compared to the classifier on species level (ID 14) from 0.639 to 0.693, while the results for the classifiers using the DTM did not benefit from the reduction in classes (see 5.2.2). Comparing the results for the classifiers without the DTM, *Abies alba* decreased from 0.52 to 0.44 compared to the SWIR region. However for the VNIR region an increase from 0.37 to 0.41 was recorded. For *Fagus sylvatica* no significant changes occurred by reducing the amount of classes. This can also be seen in the c_v results: *Fagus sylvatica* (0.052) and *Picea abies* (0.081) show a small benefit from a reduction in classes, while the grouping of *Pseudotsuga menziesii*, *Pinus mugo*, and *Larix decidua* slightly increased the variation of scores between the classifiers (0.096), making *Fagus sylvatica* again a very robust class to predict.

5.2.3 Classifier accuracy: Conifers / Broadleaf

The lowest level of classification is the differentiation between conifers and broadleaved trees, where the expected biggest difference in spectral shape is based on different habitus. As opposed to single tree species, where species like *Fraxinus excelsior* and *Acer pseudoplatanus*, show similar spectral behaviour.

Table 13: Classifier accuracy for Conifers / Broadleaf

Classifier Data							F1-Score	
ID	Data	DTM	BS	OOB	CV	Kappa	Conifer	Broadleaf
41	all	True	True	0.911	0.91	0.801	0.89	0.92
42	all	True	False	0.907	0.904	0.824	0.9	0.93
43	all	False	True	0.908	0.905	0.774	0.87	0.91
44	all	False	False	0.904	0.899	0.807	0.89	0.92
45	specind	False	True	0.908	0.908	0.778	0.87	0.9
46	specind	False	False	0.903	0.9	0.803	0.88	0.92
47	spectral	False	True	0.902	0.898	0.807	0.89	0.92
48	spectral	False	False	0.901	0.902	0.782	0.87	0.91
49	swir	False	True	0.9	0.899	0.795	0.88	0.92
50	swir	False	False	0.902	0.9	0.786	0.87	0.92
51	swirfull	True	True	0.919	0.915	0.828	0.9	0.93
52	swirfull	True	False	0.917	0.914	0.859	0.92	0.94
53	swirfull	False	True	0.915	0.915	0.809	0.88	0.92
54	swirfull	False	False	0.911	0.909	0.844	0.91	0.94
55	vnir	False	True	0.882	0.88	0.715	0.83	0.89
56	vnir	False	False	0.883	0.881	0.713	0.84	0.87
57	vnirfull	True	True	0.9	0.894	0.784	0.88	0.91
58	vnirfull	True	False	0.899	0.893	0.8	0.88	0.92
59	vnirfull	False	True	0.893	0.893	0.785	0.87	0.91
60	vnirfull	False	False	0.891	0.889	0.804	0.89	0.92
C_V							0.024	0.017

Highest scores

While the highest scores (0.859 κ) were again obtained for the SWIRfull classifier (ID 52), the overall differences are minimal, with the minimum κ score of 0.713 for the VNIR region (ID 56).

Comparison VNIR / SWIR

The difference of κ between ID VNIR (56) and SWIR (50) is only 0.073, with a better outcome for the SWIR region. This is illustrated by the higher F1 scores (0.84, 0.87 to 0.87, 0.92) for the classification using the SWIR spectrum.

Comparison Spectral / Spectral & VIs

A small benefit could be achieved by adding the vegetation indices (VIs) as attributes to the classifier, increasing the κ from 0.782 (ID 48) to 0.803 (ID 46).

Binary classification

The classifications of only two classes is very robust which can be seen on the low values for the coefficient of variation (0.024 and 0.017). With a difference between the minimum F1 scores of 0.83 & 0.87 to the maximum of 0.92 & 0.94, reaching the highest κ score (0.859) for the classifier using the SWIR spectrum combined with the VIs and the LiDAR derived treecount (ID 52).

5.3 Map accuracy

The benefit of working with remote sensing data, is the ability for plausibility checks of the predicted results, based on the overall percentages of classes (Tables 15, 16, 17). As the time necessary for predicting a whole dataset is too long for also using 10 random iterations for each prediction, the data was predicted only once. But as illustrated before, the standard deviation for the metrics measured was rather small. So the differences are also expected to be minimal as the classifier has been proven to be stable. Using only one iteration is also the reason why the classifier with the same ID has a slightly different error rate listed as in the section 5.2, as those metrics are the mean of 10 iterations. In order of being able to compare the different feature combinations, all runs were conducted using the same seed and split.

The results of the predictions can be compared to the latest forest inventory which was conducted in 2002/2003. Table 14 shows the overall species distribution for the (whole) national park for the eleven species used in the classification. The predicted results are however only predicted for the northern part of the national park, which leads to minor changes in the species distribution (see Figure 4).

However it should be noted that the assessment of the predicted results is also heavily dependent on the properties of the anonymous pixels, which is not known. Therefore no final assumptions can be made regarding the quality of the predicted results based on the predictions of tree species.

Table 14: Percentage of tree species used in classification according to forest inventory (2002/2003).

<i>Species</i>										
AA	PA	PM	PMu	LD	FS	FE	AP	AG	BP	PT
2.6	67	0.2	0.1	0.1	24.5	0.1	1.2	0.0*	0.7	0.1
<i>Species groups</i>										
AA	PA	Sndh			FS	Elbh [†]		Slbh [‡]		
2.6	67	0.4			24.5	1.4		4.1		
<i>Conifers - Broadleaf</i>										
Conifers					Broadleaf					
71					29					
Missing: [†] <i>Acer platanoides</i> 0.1 [‡] <i>Sorbus aucuparia</i> 3.1 <i>Salix</i> 0.2, *below 0.1										

Missing: [†] *Acer platanoides* 0.1 [‡] *Sorbus aucuparia* 3.1 *Salix* 0.2, *below 0.1

5.3.1 Map accuracy: Species

For the classifiers used in predicting the tree species the spatial distribution of the classes is shown in Figures 30 - 33. The normalized confusion matrices (Figures 27 - 28) visualize the errors in classification, and the change of errors over the different classes, when other features are used.

Table 15: Predicted results for Species (North)

Classifier Data							Percentage										
ID	Data	DTM	BS	OOB	CV	Kappa	AA	AP	AG	BP	FS	FE	LD	PA	PMu	PT	PM
N1	all	True	True	0.815	0.799	0.788	16.29	4.31	2.21	0.91	31.13	6.39	3.21	33.41	1.6	0.08	0.45
N2	all	True	False	0.803	0.787	0.772	15.56	3.87	2.29	0.87	30.56	6.82	2.8	35.22	1.57	0.07	0.35
N3	all	False	True	0.709	0.698	0.648	17.1	4.63	1.57	1.34	28.77	7.75	5.9	31.11	1.55	0.08	0.19
N4	all	False	False	0.7	0.686	0.646	13.86	3.33	1.15	1.11	29.27	8.19	4.77	36.86	1.31	0.09	0.07
N5	specind	False	True	0.657	0.651	0.587	19.62	1.87	1.53	1.06	36	3.1	4.37	30.69	1.45	0.13	0.18
N6	specind	False	False	0.659	0.644	0.583	14.34	0.91	0.84	0.96	36.6	3.09	3.65	38.17	1.27	0.12	0.05
N7	spectral	False	True	0.641	0.626	0.565	20.55	3.31	2.8	1.12	35.74	2.71	4.38	27.7	1.18	0.24	0.28
N8	spectral	False	False	0.638	0.623	0.559	15.29	1.39	1.41	0.84	36.15	3.17	3.5	36.76	1.22	0.21	0.06
N9	swir	False	True	0.593	0.584	0.515	19	2.93	0.5	1.92	30.36	1.44	6.22	34.96	2.15	0.14	0.38
N10	swir	False	False	0.606	0.591	0.533	17.44	2.76	0.48	1.43	30.86	1.56	5.78	37.3	2.12	0.08	0.2
N11	swirfull	True	True	0.849	0.841	0.838	9.82	4.55	1.5	0.45	27.75	5.8	3.98	42.68	2.08	0.08	1.3
N12	swirfull	True	False	0.835	0.827	0.829	9.37	5.17	1.61	0.45	27.11	5.66	4.07	43.07	2.26	0.08	1.13
N13	swirfull	False	True	0.719	0.713	0.679	9.14	4.74	0.98	0.64	24.64	5.71	8.48	43.37	2.1	0.06	0.16
N14	swirfull	False	False	0.713	0.701	0.671	9.24	5.39	0.87	0.67	24.61	5.24	8.7	43.05	2.07	0.06	0.09
N15	vnir	False	True	0.557	0.545	0.461	18.07	1.88	3.03	1.4	32.46	3.75	9.93	25.39	2.89	0.75	0.45
N16	vnir	False	False	0.566	0.562	0.465	17.39	1.54	2.59	1.54	31.88	3	11.18	25.96	3.64	1.03	0.24
N17	vnirfull	True	True	0.828	0.818	0.791	14.93	3.44	2.21	2.02	27.33	8.67	5.16	34.6	0.88	0.21	0.55
N18	vnirfull	True	False	0.812	0.803	0.774	15.21	2.82	2.17	2.64	26.42	8.57	5.44	34.79	1.16	0.29	0.5
N19	vnirfull	False	True	0.691	0.684	0.626	13.24	3.36	2.04	3.81	21.24	11.42	10.03	31.71	2.85	0.18	0.13
N20	vnirfull	False	False	0.674	0.674	0.612	12.83	3.36	2.39	3.92	19.97	11.29	10.68	32.18	3.05	0.2	0.12

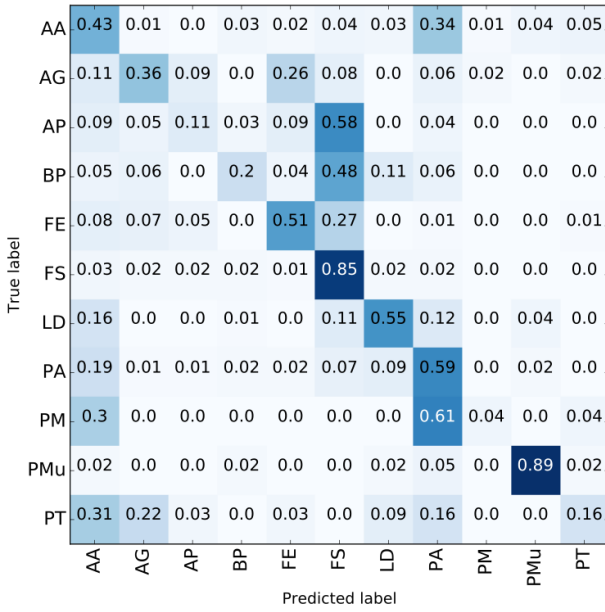
Highest scores

The highest scores were as well retrieved for the classifier SWIRfull using the DTM (5.2.1), but taking a look at the classification images (Figure 32b and 33b) it is evident that the spatial distribution of species is heavily influenced by the usage of the DTM (overfitting). Even though the overall classification results (Kappa, F1-scores) show a good result, the generalization of that classifier is not valid.

Comparison VNIR / SWIR

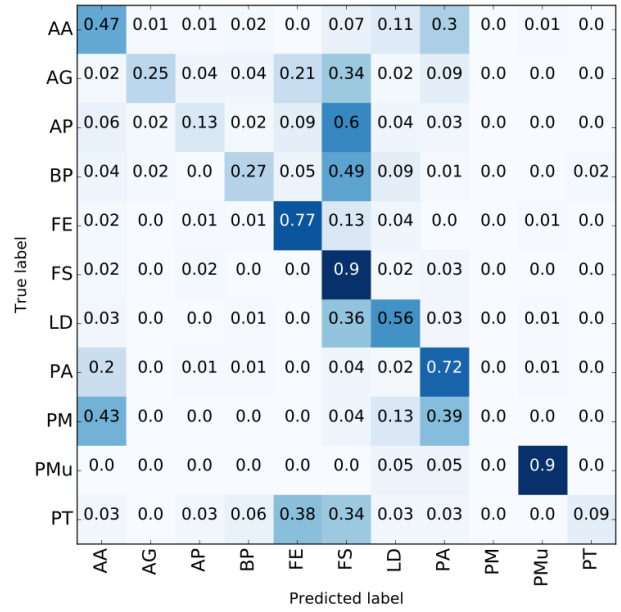
The difference for *Abies alba* (ID 16, 17.39%) is as high as the classifier using only the SWIR region (ID N10, 17.44%), the percentage is even higher when the band selection is used. Both percentages are far too high for being plausible. The usage of the feature treecount (ID N14) reduces the percentage to 9.24%, while the differences in misclassifications based on the spectral regions are minimal with a small overall plus for the SWIR (Figure 27). *Abies alba* is mostly misclassified as *Picea abies*.

Misclassifications for *Populus tremula*, have shifted from *Abies alba*, *Alnus glutinosa* and *Picea abies*, to mostly *Fraxinus excelsior* and *Fagus sylvatica*. Showing a better fit of the SWIR region for separating broadleaved and conifers.



(a) VNIR (ID N16)

Image



(b) SWIR (ID N10)

Image

Figure 27: Comparison of VNIR and SWIR

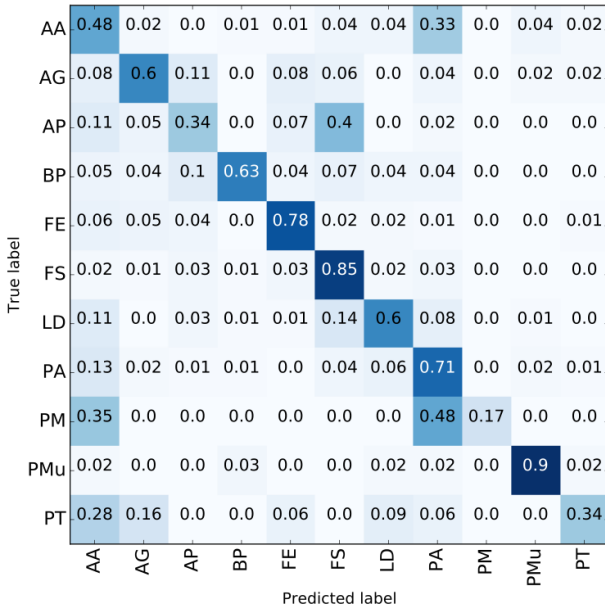
For *Picea abies* the situation is the opposite, with a far higher percentage for the SWIR classifier (37.3%) compared to the VNIR (25.96%), which is also confirmed by the higher accuracy seen in the confusion matrix (Figure 27b), and in the F1 scores showed above (Table 11).

Fraxinus excelsior also has an increased accuracy using the SWIR, resulting in a decrease of percentages compared to the VNIR, based on a reduced false classification as *Fagus sylvatica*.

While the F1-scores for *Larix decidua* do not differ much, there is a big difference in predicted percentages (11.18% to 5.78%), a change can also be seen in the confusion matrices (Figure 27) where *Larix decidua* is mostly misclassified as *Fagus sylvatica* for the SWIR, while for the VNIR misclassifications are evenly distributed over *Abies alba*, *Fagus sylvatica* and *Picea abies*.

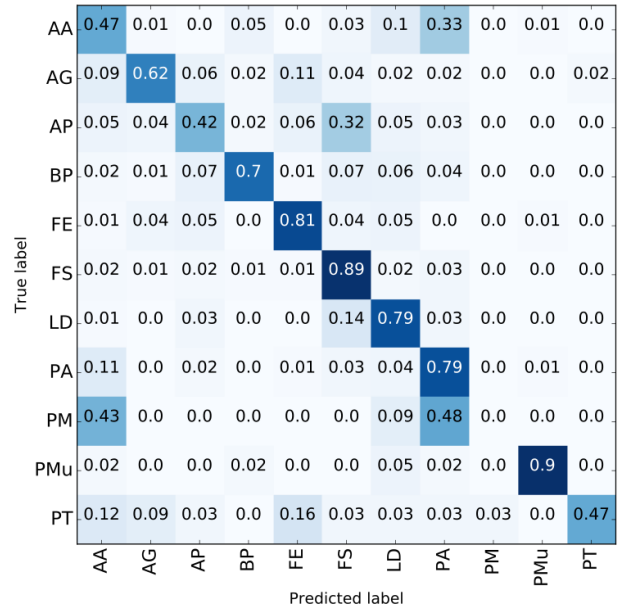
Fagus sylvatica shows minimal differences (1%) in percentage between the classifiers using the VNIR spectrum (N16) and the SWIR (N10).

For *Pinus mugo* no real difference can be seen in the confusion matrices, while the percentage predicted is higher (~1.4%) for the VNIR.



(a) VNIRfull (ID N20)

Image



(b) SWIRfull (ID N14)

Image

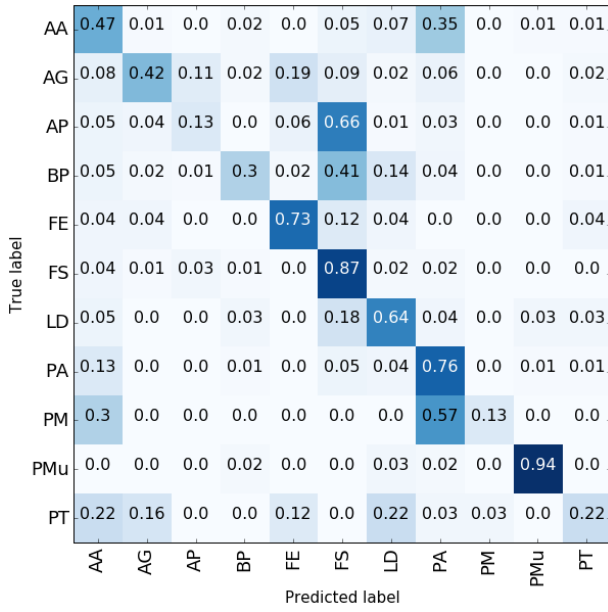
Figure 28: Comparison of VNIRfull and SWIRfull without DTM

Comparison VNIRfull / SWIRfull

Both classifiers showed a high reduction in percentages for *Abies alba* compared to the single spectral regions. For the SWIRfull the percentage is the lowest overall score (9.24%). While the usage of the DTM has no significant effect on the percentages of *Abies alba* for the SWIRfull (see N14 vs N12), the classifier VNIRfull (N20) shows an increase in percentage when the DTM is used. However as mentioned before, the spatial distribution (Figure 32), reveals the over-fitting of the classifier using the DTM as a feature, and thus the results based on this feature are not valid.

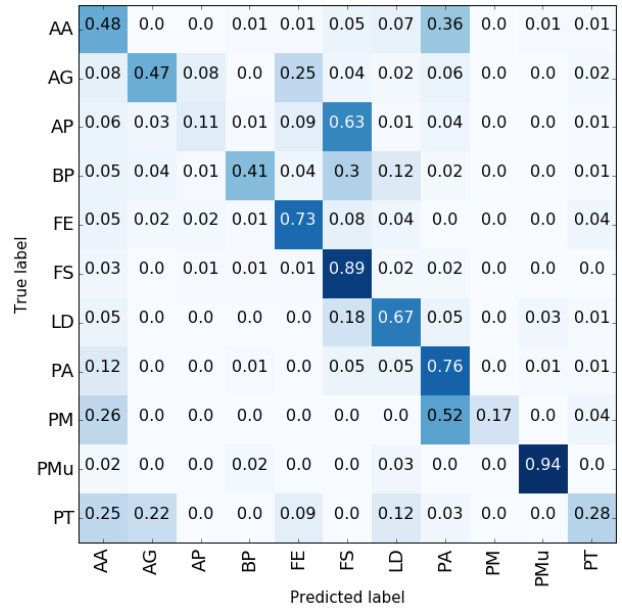
For *Fraxinus excelsior*, percentages for both classifiers are considerably higher than using only the spectral data.

Fagus sylvatica shows the same behaviour as with VNIR / SWIR, but with an overall reduced share in percentages.



(a) Spectral (ID N8)

Image



(b) Spectral + VIs (ID N6)

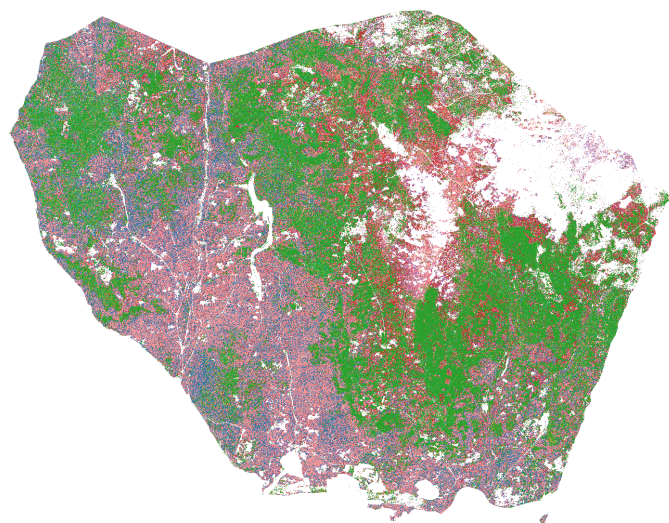
Image

Figure 29: Comparison of Spectral and Spectral + VIs

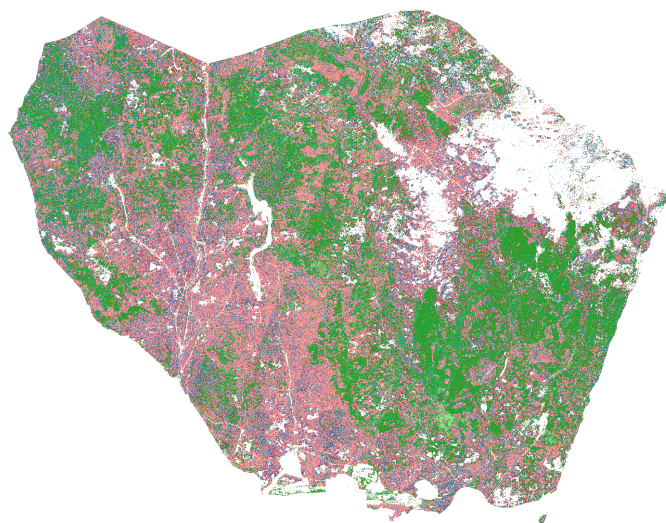
Comparison Spectral / Spectral & VIs

The impact of adding the selected VIs to the classifier has a minimal impact on the predicted percentages of species, while the κ scores show a small increase of 0.024, the main difference is a decrease of *Abies alba* of only 0.95%. Compared to the results obtained from the two single spectral regions, *Abies alba* shows an reduced percentage when the full spectrum is used for classification.

These minor changes can also be seen when comparing the confusion matrices for both classifiers (Figure 29).

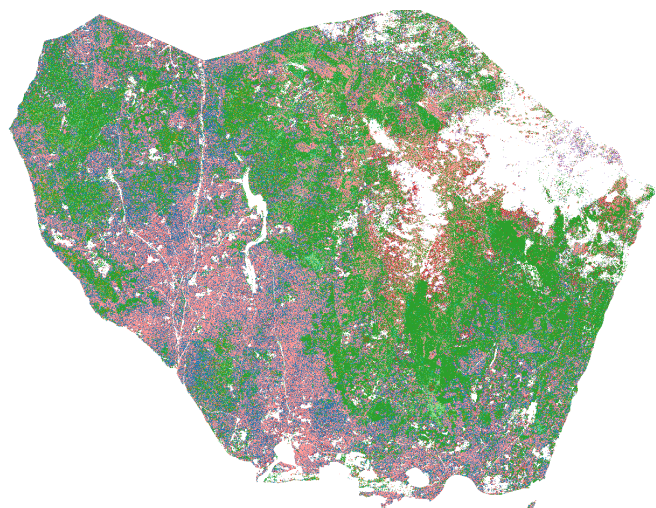


(a) VNIR (ID N16)
Confusion matrix

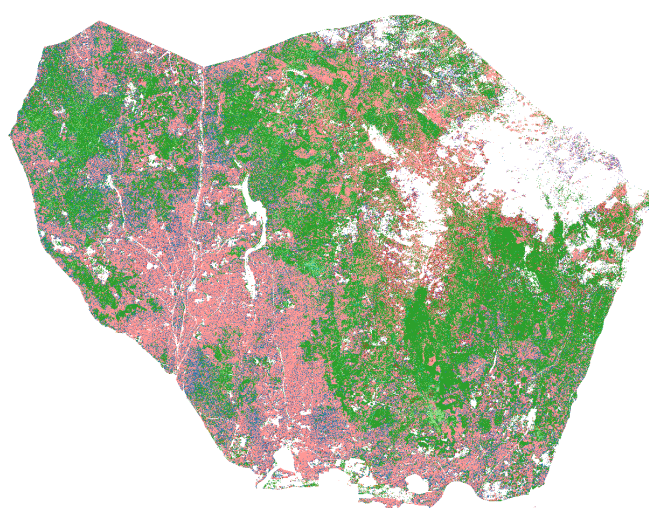


(b) SWIR (ID N10)
Confusion matrix

Figure 30: Predictions for VNIR and SWIR

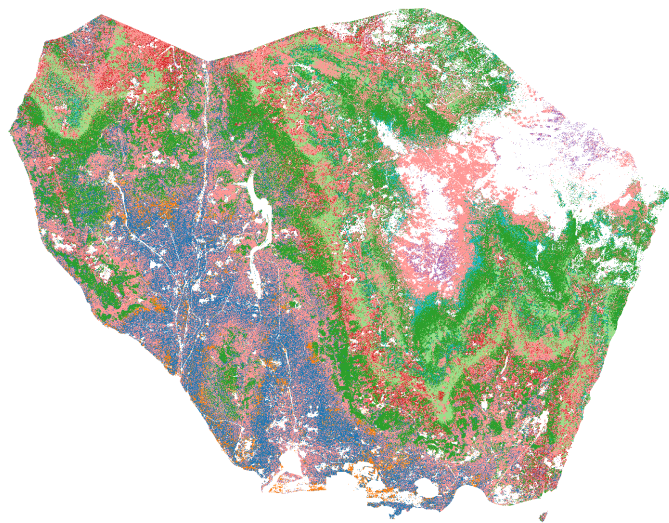


(a) Spectral (ID N8)
Confusion matrix

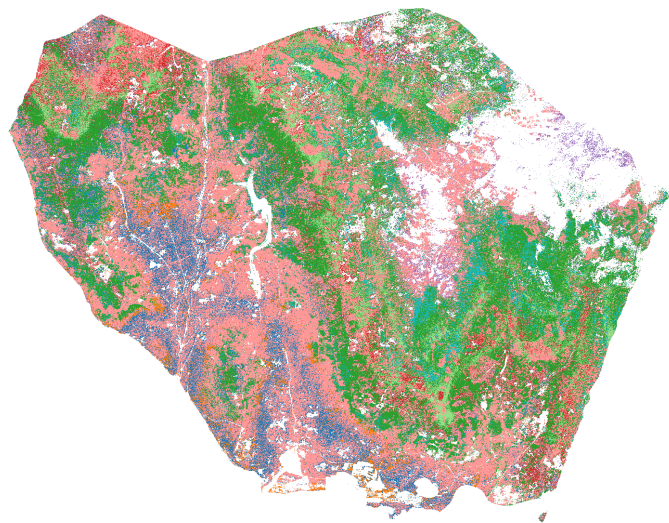


(b) Spectral + VIs (ID N6)
Confusion matrix

Figure 31: Predictions for Spectral and Spectral + VIs

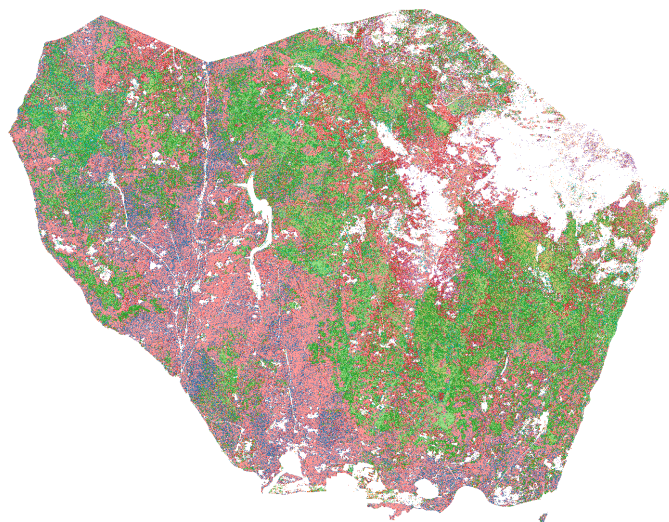


(a) VNIRfull (ID N18)



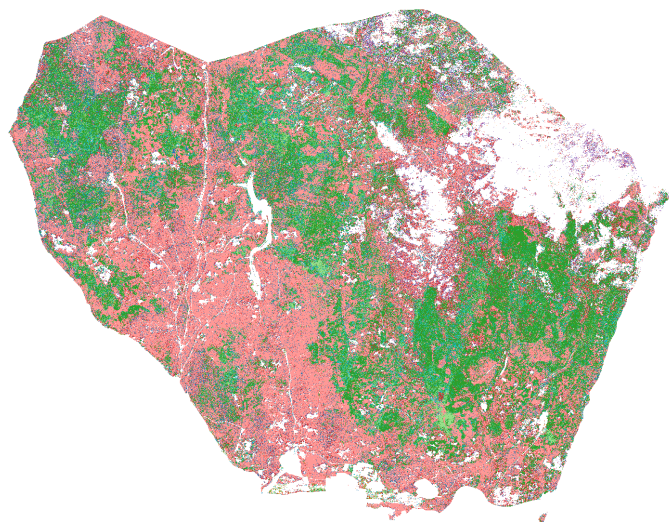
(b) SWIRfull (ID N12)

Figure 32: Predictions for VNIRfull and SWIRfull with DTM



(a) VNIRfull (ID N20)

Confusion matrix



(b) SWIRfull (ID N14)

Confusion matrix

Figure 33: Predictions for VNIRfull and SWIRfull without DTM

5.3.2 Map accuracy: Species Groups

Table 16: Predicted results for Species Groups (North)

Classifier Data							Percentage					
ID	Data	DTM	BS	OOB	CV	Kappa	AA	Elbh	FS	PA	Slbh	Sndh
N21	all	True	True	0.816	0.806	0.764	15.07	15.55	26.26	30.86	5.52	6.74
N22	all	True	False	0.806	0.798	0.748	12.75	15.9	24.63	35.38	5.71	5.63
N23	all	False	True	0.729	0.724	0.655	15.31	16.32	25.03	30.46	4.76	8.12
N24	all	False	False	0.725	0.716	0.652	10.98	15.71	24.01	37.46	4.96	6.89
N25	specind	False	False	0.698	0.689	0.612	11.62	6.44	33.15	38.26	5.22	5.31
N26	specind	False	True	0.704	0.692	0.622	17.7	8.48	33.1	29.63	4.44	6.65
N27	spectral	False	True	0.685	0.675	0.581	19.13	8.67	34.34	25.8	5.37	6.69
N28	spectral	False	False	0.682	0.67	0.591	12.69	5.85	32.44	36.51	7.04	5.46
N29	swir	False	True	0.649	0.642	0.562	18.29	5.63	27.44	34.43	3.97	10.25
N30	swir	False	False	0.66	0.648	0.588	16.63	5.53	27.67	36.4	4.04	9.73
N31	swirfull	True	True	0.839	0.829	0.807	7.52	15.5	20.72	42.41	3.75	10.11
N32	swirfull	True	False	0.83	0.82	0.79	7.72	16.09	19.54	42.03	4.29	10.32
N33	swirfull	False	True	0.741	0.736	0.667	7.67	14.43	20.2	40.5	3.73	13.45
N34	swirfull	False	False	0.736	0.728	0.674	8.15	14.88	19.54	40.31	4.05	13.06
N35	vnir	False	True	0.606	0.595	0.49	15.49	8.41	28.7	25.14	7.8	14.46
N36	vnir	False	False	0.62	0.612	0.515	14.99	7.16	27.16	25.97	8.28	16.45
N37	vnirfull	True	True	0.824	0.809	0.763	12.65	18.02	20.51	33.24	6.32	9.27
N38	vnirfull	True	False	0.809	0.799	0.76	12.81	18.36	17.86	33.47	7.43	10.07
N39	vnirfull	False	True	0.714	0.712	0.628	10.88	20.63	14.8	33.75	7.22	12.71
N40	vnirfull	False	False	0.703	0.699	0.624	10.41	21.55	12.44	34.34	7.83	13.43

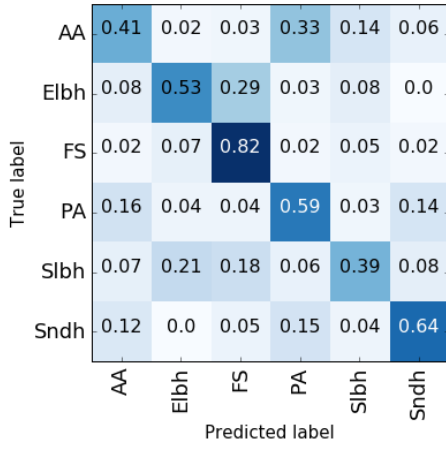
Comparison VNIR / SWIR

Similar to the single tree species classification, the greatest change occurs in the comparison of the rates for *Picea abies* with 36.4% for the SWIR (N30), which is only 0.90% lower than the species classification, and 25.97% for the VNIR (N36), which is the same rate as before. Thus for the VNIR region there is no changes in percentages of *Abies alba* for the classification of species groups. For the SWIR region the percentage of the group 'other conifers' was increased by the small reduction of *Picea abies*.

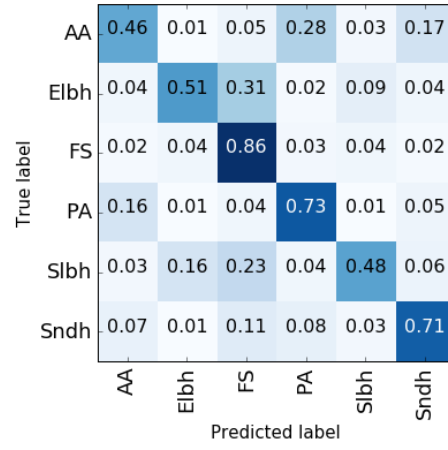
Fagus sylvatica shows only marginal differences in percentage (-0.51%) between the two spectral regions. However compared to the species level classification, numbers are reaching the percentages of single tree species classification. This is a reduction of almost 5% for the VNIR and SWIR.

Percentages of *Abies alba* were reduced by 2.4% for the VNIR, and (-0.81%) for the SWIR as compared to single species classification.

The group of real hardwoods (Elbh) shows a difference of 1.63%, between the two spectral regions. The grouping had only a small effect on the percentages for the



(a) VNIR (ID N36)

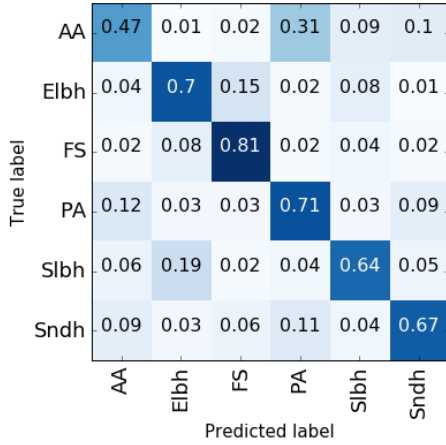


(b) SWIR (ID N30)

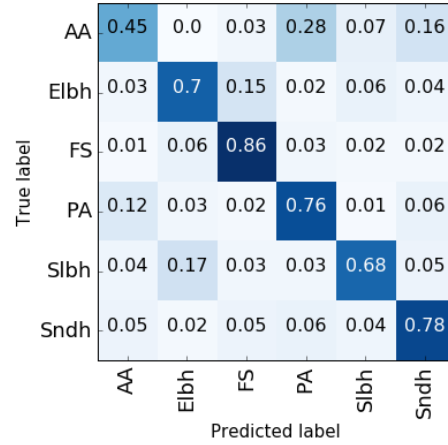
Figure 34: Comparison of VNIR and SWIR

SWIR region (4.32% to 5.53%), the VNIR region showed a significant increase in percentages (4.54% to 7.16%).

Comparison VNIRfull / SWIRfull



(a) VNIRfull (ID N40)



(b) SWIRfull (ID N34)

Figure 35: Comparison of VNIRfull and SWIRfull (no DTM)

Both classifiers show a greatly reduced percentage for *Fagus sylvatica* compared to the VNIR/SWIR classifiers. The decrease for the VNIRfull (N40) to only (12.44%) is even higher than for SWIRfull (19.54%). Even though both classifiers show similar F1-scores (see 5.2.2) for this tree species. Comparing the results with the species level classification, the major tree species (*Fagus sylvatica* and *Picea abies*) show a decrease in percentage after the grouping of minority species. The big discrepancy when comparing the percentages of these classifications to

the results of the latest forest inventory is raising the question of validity of the results.

Comparison Spectral / Spectral & VIs

AA	0.43	0.0	0.04	0.33	0.07	0.13
Elbh	0.04	0.56	0.28	0.01	0.08	0.02
FS	0.04	0.05	0.83	0.01	0.04	0.03
PA	0.14	0.0	0.04	0.73	0.01	0.07
Slbh	0.04	0.2	0.16	0.03	0.46	0.11
Sndh	0.08	0.0	0.06	0.06	0.06	0.74
	AA	Elbh	FS	PA	Slbh	Sndh

(a) VNIRfull (ID N28)

AA	0.43	0.01	0.02	0.33	0.07	0.14
Elbh	0.04	0.58	0.27	0.02	0.08	0.02
FS	0.03	0.06	0.85	0.02	0.02	0.02
PA	0.12	0.0	0.04	0.74	0.01	0.09
Slbh	0.04	0.18	0.1	0.03	0.54	0.11
Sndh	0.09	0.0	0.06	0.05	0.05	0.74
	AA	Elbh	FS	PA	Slbh	Sndh

(b) SWIRfull (ID N26)

Figure 36: Comparison of Spectral and Spectral + VIs

The results for the species groups classification based on the spectral data and spectral + VIs are very similar. The main difference occurs for the species *Fagus sylvatica* with an increase of 1.75% when the VIs are used. This effect is also visible for the tree species classification, where the usage of the VIs also had only a limited influence on the prediction results. The minimal changes of the confusion matrices for both classifiers also reflect these results. However the influence of the grouping on the overall percentages for both classifiers can be seen on the percentages for *Abies alba* with an increase in of 4.79% and 2.41% as compared to the single species classification.

5.3.3 Map accuracy: Conifers / Broadleaf

The maximum difference of the binary classification is 14.23% between all classifiers, with a maximum difference in κ scores of only 0.083. When the algorithms using the band selection are excluded, the difference drops to only 9.51% and difference in κ is restricted to 0.075. The percentages for conifers are the highest using the SWIR region and the lowest using all available spectral data.

Comparison VNIR / SWIR

As it was already shown in the classifications of other levels, the percentages of

Table 17: Predicted results for Conifers / Broadleaf (North)

<i>Classifier Data</i>							<i>Percentage</i>	
ID	Data	DTM	BS	OOB	CV	Kappa	Conifers	Broadleaf
N41	all	True	True	0.91	0.908	0.825	50.54	49.46
N42	all	True	False	0.906	0.904	0.82	52.94	47.06
N43	all	False	True	0.907	0.906	0.817	50.77	49.23
N44	all	False	False	0.904	0.899	0.804	53.44	46.56
N45	specind	False	True	0.908	0.902	0.808	50.76	49.24
N46	specind	False	False	0.901	0.898	0.801	53.19	46.81
N47	spectral	False	True	0.906	0.901	0.806	46.92	53.08
N48	spectral	False	False	0.901	0.897	0.801	51.56	48.44
N49	swir	False	True	0.895	0.894	0.806	61.15	38.85
N50	swir	False	False	0.899	0.898	0.815	61.07	38.93
N51	swirfull	True	True	0.92	0.919	0.841	59.88	40.12
N52	swirfull	True	False	0.915	0.912	0.839	59.84	40.16
N53	swirfull	False	True	0.915	0.91	0.825	60.16	39.84
N54	swirfull	False	False	0.911	0.91	0.818	60.1	39.9
N55	vnir	False	True	0.871	0.868	0.758	56.16	43.84
N56	vnir	False	False	0.872	0.868	0.764	57.6	42.4
N57	vnirfull	True	True	0.904	0.903	0.806	58.05	41.95
N58	vnirfull	True	False	0.898	0.894	0.804	58.43	41.57
N59	vnirfull	False	True	0.898	0.897	0.797	58.15	41.85
N60	vnirfull	False	False	0.891	0.888	0.793	58.32	41.68

conifers are slightly lower when the VNIR region is used for classification. This automatically leads to a higher percentage of broadleaved trees.

Comparison VNIRfull / SWIRfull

Adding the VIs and the 'treecount' to the features, it leads to a balancing of the percentages with a difference of only 1.78%, while the κ scores still favour the SWIR region. The use of the DTM has only a minor influence on the classifiers, e.g. comparing N52 and N54 shows only a difference of 0.26%.

Comparison Spectral / Spectral & VIs

Adding only the VIs to the whole spectrum increases the percentages of conifers by 1.63% to 53.19%, which is still considerably lower than the predictions using only one spectral region.

All data

When all available data except the DTM are used, the percentages are almost equal, and also using the DTM results in a 50:50 split of the data, which is highly unlikely given the latest forest inventory results.

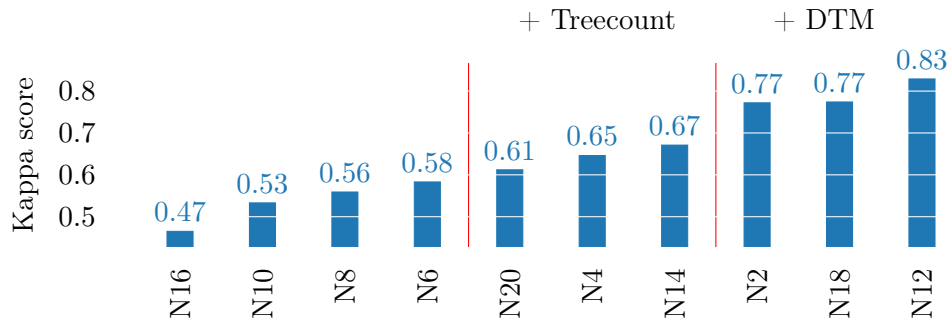


Figure 37: Kappa scores for species classifiers

6 Discussion

6.1 Best result

Three different levels of complexity (species, species groups, conifers / broadleaf) were used as basis for the classification, the highest scores (Kappa, F1) were consistently obtained using the SWIR spectrum + VIs + LiDAR data (12). However in contrast to the latest forest inventory and also compared to spatial distribution of the predicted pixels, the highest score for the model does not necessarily imply a good representation of the reality. Therefore the highest scores are not always the 'best' results and should be interpreted with caution. The reason for this behaviour can be diverse (e.g. shadow pixels, remaining BRDF effects), but one of the main factors is the integration of the features derived from LiDAR data, which in this case lead to over-fitting, especially when the DTM was used as a feature. In summary the classifiers including the SWIR spectrum showed a higher ability of classifying specific species, such as *Abies alba*, *Fraxinus excelsior* and *Picea abies* as opposed to those classifiers that used the VNIR spectrum only. On the other hand, the combination of both spectral regions did not improve the F1-scores for all species. The integration of VIs in the classifiers however increased F1-scores again, especially for minority classes such as *Populus tremula* and *Betula pendula*.

6.1.1 LiDAR data

Two features ('DTM' and 'treecount') based on LiDAR data were used for classification. Whenever the DTM was used, the scores, especially for rare tree species, increased significantly, and thereby the percentages for the predicted results as well. However, as discussed earlier, when looking at the spatial distribution of the pixels, a pattern of topographical relief becomes visible. The explanation for this behaviour lies in the algorithm of random forests, combined with the properties

of the available training data.

The algorithm aims at creating pure nodes, splits are favoured on attributes which enable a better class separation than others. The vast majority of features are spectral bands, where the class separation based on a single wavelength for eleven classes is very hard to achieve due to multiple intersection between classes. On top of that the hyperspectral data set consists of 4507359 pixels for the northern part, while the training data set is only comprised of 4775 pixels for the same area ($\sim 0.001\%$ of RS data).

An acquisition of ground truth data, which would represent the real distribution of the different levels of elevation would be preferable but is rather unfeasible, especially when the field study is not specifically designed for it. As can be seen in Figure 38, the distribution of pixels over the elevation differs greatly between training and anonymous pixels. This effect is even stronger when considering the distribution for each single tree species (Figure 42). The basic idea of including the elevation as a feature for the classification is based on adding the knowledge of natural height limits for some species, or the preferred habitat. But given the sub-optimal sample design, the chances are extremely high that especially rare tree species are only sampled for specific elevation levels, which does not reflect the real distribution as found in nature.

Even though the features considered for splitting are limited by the size of the random sub-window, and the value of the feature considered for splitting is chosen at random, the Extremely randomized trees (ET) algorithm still chooses the attribute with the purest possible split. This leads to a situation where splitting on the feature 'DTM' is favoured, as it achieves a far greater class separation. This leads to a situation where possible class separations based on other attributes are not learned as well by the classifier. Due to this greedy behaviour of the algorithm, the probability of prediction for e.g. *Abies alba* in a specific elevation level is far higher than it would be when considering the real distribution.

The same problems are present when the 'treecount' feature is used, however the influence is much smaller than for the 'DTM'.

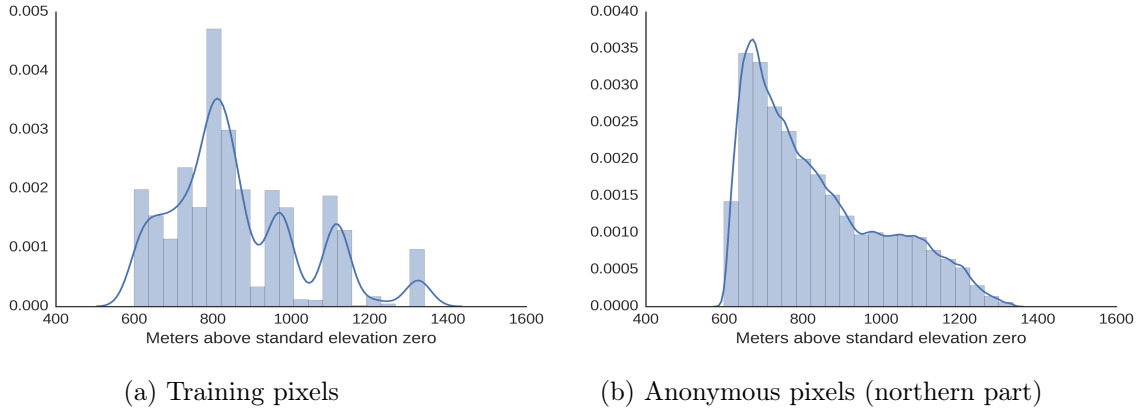


Figure 38: Histogram of pixels for the DTM

6.1.2 BRDF problems

Despite extensive BRDF corrections (3.2.4), BRDF effects are still influencing the classification, as can clearly be seen in the images of the predicted results (Figures 31). Even though these effects are not as noticeable as in the other images for the predictions, which are based on single spectral regions of VNIR and SWIR only (Figure 30), the BRDF effects were observed for all obtained predictions. The effect has only limited influence on the accuracy metrics due to the spatial distribution of the training data, but an even greater impact on the overall classification is to be expected.

The influence can also be visualized through the calculation of probabilities with which a class is predicted for every single pixel. The prediction probabilities are computed as the mean predicted class probabilities of all the trees in a single forest. For each pixel the selected class probability (highest scoring) is recorded and displayed in Figure 40. A high probability (green) shows that the majority of the trees predicted the same class, while a low

probability (red) shows that a majority of the trees predicted different classes for the same pixel. Remaining BRDF effects still have an influence on probabilities of the classification, as can clearly be seen when looking at the overlapping areas of the single flight lines in the lower right corner of the image. There are pattern

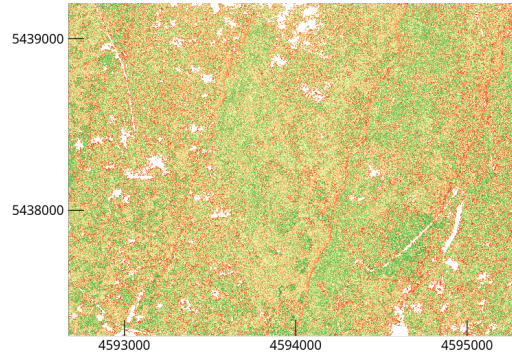


Figure 39: BRDF effects still visible in probabilities of predicted classes. (For a colorbar see Figure 40.)

of breaks in probability values between the flight lines, which is already visible in Figure 40, and can be seen in detail in Figure 39.

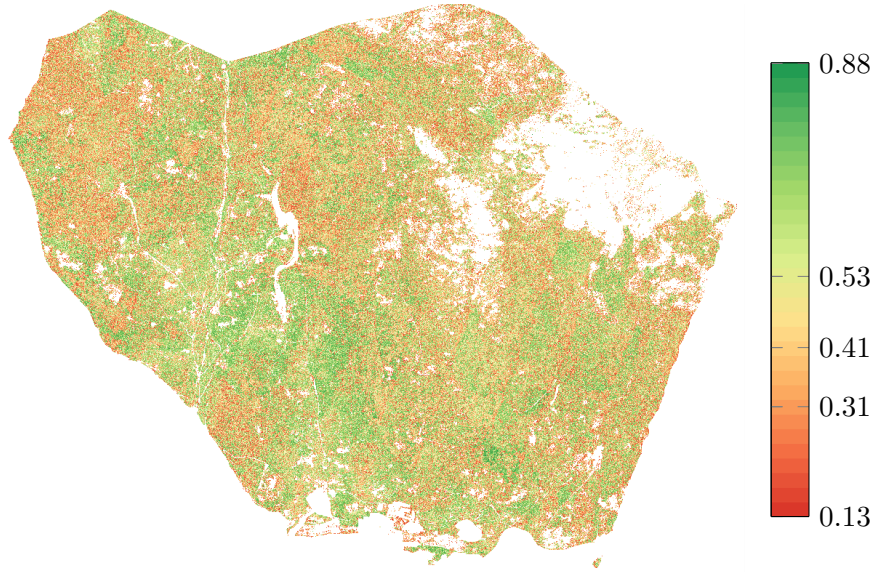


Figure 40: Probabilities for the predicted classes of classifier N14

6.1.3 Tree species classification

The main goal of this study was to classify eleven tree species. The classifiers with the highest scores (ID 11, 12) were impaired by the problems described above, the results can thus not be considered as fully valid. It can be argued whether the classifier with second best metrics (0.671 κ , SWIRfull without DTM (ID N14) is considered as valid, or the reduced kappa score (0.583 κ) of the classifier (specind ID N6), not using any LiDAR derived features, is best accepted. As discussed in 6.1.1, using the 'treecount' feature suffers from the same behaviour based on the greedy nature of the algorithm. Even though the percentages of *Abies alba* (Table 18), for the classifier (ID N14) shows percentages closer to the results from the latest forest inventory. The differences in F1-scores between both classifiers show an increased score for *Abies alba* for classifier N6. Also the minority groups of 'Elbh' and 'Sndh' show significant reduced percentages when compared to the classifier N14.

These results are based on the difference percentages of *Acer pseudoplatanus* and *Fraxinus excelsior*, which are predicted far more often when the LiDAR based 'treecount' feature was applied.

The percentage of single tree species is an important metric for measuring the behaviour of the different classifiers. The example of *Abies alba* shows, that even with similar F1-scores (0.52, 0.49) for this tree species (see classifiers ID 10,

Table 18: Best species predictions compared to forest inventory data

<i>ID</i>	<i>Data</i>	<i>Kappa</i>	<i>F1-score</i>										
6	specind	0.589	0.56	0.29	0.48	0.55	0.8	0.64	0.63	0.65	0.88	0.59	0.14
10	swir	0.539	0.52	0.32	0.36	0.23	0.74	0.69	0.52	0.71	0.82	0.06	0.00
14	swirfull	0.639	0.49	0.54	0.67	0.64	0.85	0.74	0.64	0.66	0.91	0.47	0.07
			<i>Percentage</i>										
Species			AA	AP	AG	BP	FS	FE	LD	PA	PMu	PT	PM
Forest inventory			2.6	1.2	0.0	0.7	24.5	0.1	0.1	67	0.1	0.1	0.2
N6	specind	0.583	14.34	0.91	0.84	0.96	36.6	3.09	3.65	38.17	1.27	0.12	0.05
N10	swir	0.515	17.44	2.76	0.48	1.43	30.86	1.56	5.78	37.3	2.12	0.08	0.2
N14	swirfull	0.671	9.24	5.39	0.87	0.67	24.61	5.24	8.7	43.05	2.07	0.06	0.09
Species groups			AA	Elbh		FS		PA		Slbh		Sndh	
Forest inventory			2.6	1.4		24.5		67		4.1		0.4	
N6	specind	0.583	14.34	4		36.6		38.17		1.92		4.97	
N10	swir	0.515	17.44	4.32		30.86		37.3		1.99		8.1	
N14	swirfull	0.671	9.24	10.63		24.61		43.05		1.6		10.86	
Conifers / Broadleaf			Conifers					Broadleaf					
Forest inventory			71					29					
N6	specind	0.515	57.48					42.52					
N10	swir	0.515	62.84					37.16					
N14	swirfull	0.671	63.15					36.85					

14), the result of percentages can differ considerably between multiple classifiers (17,44%, 9.24%). The comparison between N10 and N14 clearly shows that the properties of the anonymous pixels are not completely captured in the training data. An additional classification using a reduced proportion of the SWIR spectrum ($1.3\mu\text{m} - 2.4\mu\text{m}$) showed a κ score of 0.515 with a F1-score of 0.45 for *Abies alba* resulting in 8.7% predicted percentages. The highest scores with the most valid result based on the spatial distribution was obtained for the full spectrum combined with the vegetation indices (specind) (N6) reaching a κ score of 0.583. However this classifier still has problems separating *Abies alba* from *Picea abies*, hence overestimating *Abies alba* considerably.

6.1.4 Species group classification

When aggregating the results from the single tree classification, the biggest differences between the three classifiers show up for the minority classes (see Table 18). Also the percentages of the classifiers for the aggregated results,

show an additional increase in difference for the minority classes. This result does not contribute to the validity of the predictions obtained by the classifiers using the reduced classes. In addition when evaluating the performance of the classifiers using the reduced classes (Table 19), the comparison of the F1-scores for the remaining species, e.g. *Abies alba* or *Fagus sylvatica* clearly shows that a reduction in classes does not necessarily imply an increase in F1-score. Thus the remaining tree species (*Abies alba*, *Fagus sylvatica* and *Picea abies*) did not profit from a simplified classification scheme. Conclusively the reduction of classes to species groups is not essential. When grouping the single species classification and comparing it to results obtained from the grouped classification, the accuracy between N6 and N26 remains 86.40%.

Table 19: Best species groups predictions compared

<i>ID</i>	<i>Data</i>	<i>Kappa</i>	AA	Elbh	FS	PA	Slbh	Sndh
			<i>F1-Score</i>					
N26	specind	0.627	0.49	0.65	0.83	0.68	0.64	0.69
N30	swir	0.561	0.44	0.64	0.76	0.68	0.60	0.61
N34	swirfull	0.674	0.54	0.75	0.87	0.73	0.71	0.74
			<i>Percentage</i>					
N26	specind	0.612	11.62	6.44	33.15	38.26	5.22	5.31
N30	swir	0.588	16.63	5.53	27.67	36.04	4.04	13.06
N34	swirfull	0.671	8.15	14.88	19.54	40.31	4.05	13.06

6.1.5 Conifers / Broadleaf

Comparing the aggregated results of the species level classification (agg. results for N6) with the predicted results of the classifier N46 (53.19%, 46.81%), it shows a difference of only 4.29%. The F1-scores are rather high, which is to be expected for a binary classification based on groups with the biggest in between class differences in spectral signature. However due to the fact that the results of the classifiers show no major difference to the aggregated results, the benefit from conducting a binary classification is rather questionable. The high accuracy of the comparison of the classifiers N6 and N46 with 93.17% confirms this opinion. As most misclassifications happen within the groups of conifers or broadleaved trees.

6.2 Further studies

Generally the results of the classifications, especially for some of the rare tree species (*Abies alba*, *Acer pseudoplatanus*, *Pseudotsuga menziesii*, *Alnus glutinosa* and *Betula pendula*) are not fully satisfying. However before improved classification methods shall be proposed, or additional ground truth data shall be gathered, problems concerning the BRDF effects need to be investigated. A sub-optimal preprocessed data set will always heavily influence the predictions (Tatjana Koukal et al. 2014) regardless of any classification approach. As long as this problem is not solved, a valid classification is rather questionable. In this study the 'best' possible results were obtained using the full spectrum. Comparing the SWIR and the VNIR spectrum, the VNIR showed only a significant higher F1-score (+0.22) for the species *Populus tremula* while the SWIR spectrum showed higher scores for *Abies alba* (+0.15), *Fraxinus excelsior* (+0.21) and *Picea abies* (+0.19). Nonetheless the combination of the SWIR spectrum and only the VIs, which mostly are calculated based on the VNIR region, is something which has not been tested in this study. The robust information gained from the VNIR spectrum together with the overall better classification of the SWIR could achieve better results. Further possible improvements are discussed in the following sections.

6.2.1 Sampling design

The main goal of a tree species classification is not to classify mostly majority tree species, but to achieve a robust classification of rare tree species. Consequently there is a need for enough training data also for those rare tree species. As these species are in fact rare, the sampling design for a field study cannot be random, or a regular grid, as this would lead to mostly gathering ground truth data for majority tree species. However sampling data based solely on expert knowledge has also its flaws, as bias is added through this process. A solution could be to use an unsupervised classification of the whole image data, and based on this segmentation possible sampling locations could be chosen (Appice et al. 2016). Additionally a sampling strategy using the best practice approach established for forest inventories⁸, measuring azimuth and distance for each tree from a fixed point, is highly recommended. Also the collection of precise spatial information of the location of each forest inventory plot is encouraged, as this would create a huge resource of data based on a systematic sampling design.

⁸https://bwi.info/Download/de/Methodik/Aufnahmeanweisung_BWI2.pdf

6.2.2 Additional data

A second airborne campaign has been conducted in 2015 covering the same area. If this data could be co-registered correctly, this data can be added as second layer of hyperspectral data. Based on this, multiple opportunities for feature engineering, e.g. more robust spectral signals, arise which would support class separation, especially between spectrally very similar classes.

Another source of data, which has not been integrated in this study, but is of course always present, is the spatial relationship between pixels. Exploiting these relationships would add information regarding the spectral behaviour of the surroundings, as well as texture (Hänsch et al. 2016). As tree species by nature are growing in relationships dependent on the habitat, natural clusters emerge e.g. *Acer pseudoplatanus* and *Fraxinus excelsior*. Including this information could also improved the classification result.

6.2.3 Savitzky-Golay filter & brightness normalization

Smoothing and normalization of the hyperspectral data lead to a better performance of the classifiers. However the noise polluting the data, which should be reduced by the Savitzky-Golay filter, is random. Without the removal of this noise the classification results can be altered. It is thus important to remove this noise as best as possible. The question if the applied filter is the best suitable approach to this issue arises. The possibilities of other filters (Minimum-Noise-Fraction, Kawata–Minami) and the influence of the parameters of the applied Savitzky-Golay filter for a better classification should be considered for further evaluation. Different approaches for brightness normalization should also be investigated, as in this study only one specific algorithm was applied.

6.2.4 Feature selection

Feature selection can be a vital part of a supervised classification, as a feature does not necessarily contribute towards a better prediction by its single use, but by its combination with other features (Guyon et al. 2003). This makes feature selection for hyperspectral data a complex task, as the correlations between single wavelengths and spectral redundancy are caused by the high dimension of the data. The performed feature selection was based in this study on a recursive feature elimination and has not contributed towards a valid classification. The main reason for this behaviour is the sub-optimal choice of metrics for evaluating the performance. When the overall accuracy was chosen, the evaluation favors those

features which tend to predict majority classes accurately, instead of minority classes. A better choice might have been a metric such as a weighted F1-score, which takes the size of the classes into account. Additional VIs added to this process could have also enabled a feature selection based on more stable spectral signatures. Especially as the vegetation indices (VIs) used in this study were not studied in detail regarding their impact for class separation. Only 9 VIs were used, out of more than 50 indices which have been developed over time. Adding additional VIs as features and including all of them in the process of feature selection could lead to a better selection of spectral features containing important and more stable information.

6.2.5 Machine Learning

One way of reducing the possibility of over fitting, when additional data e.g. a elevation data is used, is to shrink the size of the random sub window from which features are selected. On the other hand this would reduce the information gained from spectral bands. Alternatively a forest of randomized trees algorithm which splits nodes at random could be used. Biau et al. 2008 have proposed such an approach called 'purely random forest'. However such a splitting behaviour also leads to a reduced chance of using crucial spectral information. Apart from the approach of using forests of randomized trees, the field of machine learning is an evolving topic, and especially in recent years big advances have been made regarding image classification. One of the major developments is the adaption of neural networks for image recognition (classification tasks). As the step of feature learning is critical to the classification of high dimensional data, algorithms have been proposed, which include an automated feature learning process, eliminating the need of 'handcrafted' domain knowledge based feature engineering. One of this type of classifiers are convolutional neural networks, which add automatic feature learning and spatial information to the classification through layers of convolutions and spatial pooling. Classifications using this type of neural networks have shown significant capabilities in classifying common hyperspectral data sets (Yu et al. 2017, Li et al. 2017). If the sources of errors could be removed from the data set, such an approach could even further improve the classification result and should be investigated.

7 Conclusion

The aim of this study was to create a repeatable supervised tree species classification based on fullcube hyperspectral data. This was achieved by using common Python libraries for machine learning and data analytics, combined with the usage of GDAL. The ground truth data used in this study relied on three different sources, a fact that adds possible errors, as well as a layer of randomness (forest inventory). Spectral data were combined with LiDAR derived elevation and stem density features. Adding those features lead to an over fitting based on these features, which occurred even stronger for the elevation data. A reason for this behaviour is the greedy nature of the algorithm of forest of randomized trees, combined with the small training sample size, which does in no case reflect all varieties of anonymous pixels present in the image. As a result the highest metrics (kappa, F1-scores) were obtained using the features based on data fusion. However the analysis of the predicted class distribution and the overall percentages of tree species showed that only those classifiers based on the full spectrum offered the most plausible classification.

But even for most valid models, the separation of the species *Abies alba* and *Picea abies* could not sufficiently be solved, leading to a higher percentage of *Abies alba* as potentially meaningful. The benefit of combining the VNIR and SWIR spectrum for a better classification was distinct, but future research should focus even more on combining the optimal spectral information from both sensors to improve discrimination of specific classes. Adding vegetation indices (VIs) as features to the classification had only a limited impact towards higher F1-scores, but in general it was always a positive one. The reduction of classes in order to decrease the complexity of the classification task was not necessary as most classification problems arose within the conifers or broadleaved species, and not between those two classes. This was demonstrated through a comparison between the aggregated results and the results obtained from the predictions based on the reduced classes (species groups). In conclusion the fullcube spectrum showed significant advantages in classifying tree species. However with the increased number of bands, the complexity of the classification task is also increased. Further studies focusing on the separation of specific species and the capabilities of specific spectral ranges should be investigated in more detail.

A Appendix

Table 20: Selected spectral ranges by applied recursive feature elimination

VNIR			SWIR		
0.415 - 0.476	0.476	0.484	1.033, 1.039	1.057 - 1.099	1.423
0.494	0.509 - 0.523	0.728	1.489 - 1.507	1.663 - 1.681	1.699, 1.705
0.624	0.635 - 0.721	0.948 - 0.970	1.711, 1.717	1.723, 1.729	1.777, 1.783
			1.789, 1.795	1.897, 1.903	2.016 - 2.064
			2.076, 2.082	2.154, 2.298	2.310, 2.334
43 Bands selected			50 Bands selected		

Table 21: Predicted results for Species (South)

Classifier Data										Percentage							
ID	Data	DTM	BS	OOB	CV	Kappa	AA	AP	AG	BP	FS	FE	LD	PA	PMu	PT	PM
SS1	all	True	True	0.815	0.799	0.788	13.78	2.8	0.03	3.29	57.67	0.18	2.27	18.34	1.53	0.04	0.06
SS2	all	True	False	0.803	0.787	0.772	12.62	2.07	0.03	3.07	58	0.14	2.16	20.11	1.71	0.02	0.06
SS3	all	False	True	0.709	0.698	0.648	14.63	3.67	0.61	2.83	55.76	0.19	2.97	17.52	1.75	0.03	0.02
SS4	all	False	False	0.7	0.686	0.646	11.77	2.43	0.44	2.2	56.28	0.19	2.63	22.35	1.67	0.02	0.02
SS5	specind	False	True	0.657	0.651	0.587	13.68	2.25	0.73	2.21	55.57	2.62	4.34	16.5	2.02	0.04	0.03
SS6	specind	False	False	0.659	0.644	0.583	10.86	1.34	0.36	1.93	55.52	2.42	3.82	21.81	1.88	0.04	0.02
SS7	spectral	False	True	0.641	0.626	0.565	13.45	2.69	1.44	2.25	55.65	2.96	4.58	15.11	1.73	0.09	0.05
SS8	spectral	False	False	0.638	0.623	0.559	11.48	1.66	0.62	1.77	54.48	3.02	3.82	21.14	1.92	0.08	0.02
SS9	swir	False	True	0.564	0.55	0.493	7.78	1.2	0.56	2.4	38.5	0.73	6.35	37.46	4.87	0.03	0.13
SS10	swir	False	False	0.576	0.567	0.515	7.89	1.06	0.37	3.21	38.51	0.81	6.96	36.55	4.62	0.01	0.01
SS11	swirfull	True	True	0.849	0.841	0.838	11.49	3.05	0.03	2.44	51.24	0.16	3.81	25.67	1.93	0.05	0.13
SS12	swirfull	True	False	0.835	0.827	0.829	10.87	2.46	0.03	2.27	51.73	0.16	3.86	26.26	2.21	0.05	0.11
SS13	swirfull	False	True	0.719	0.713	0.679	11.42	4.28	0.41	1.66	47.02	0.09	5.81	27.05	2.19	0.02	0.05
SS14	swirfull	False	False	0.713	0.701	0.671	10.54	3.46	0.47	1.51	47.52	0.15	6.23	27.87	2.22	0.01	0.03
SS15	vnir	False	True	0.557	0.545	0.461	10.6	3.43	1.9	4.85	48.86	3.97	9.35	14.1	2.45	0.37	0.13
SS16	vnir	False	False	0.566	0.562	0.465	10.57	2.67	1.6	4.68	48.6	2.95	9.74	13.88	4.57	0.68	0.08
SS17	vnirfull	True	True	0.828	0.818	0.791	14.22	4.24	0.03	4.99	54.46	0.2	2.48	17.75	1.49	0.06	0.08
SS18	vnirfull	True	False	0.812	0.803	0.774	14.19	4.06	0.02	5.16	53.5	0.18	2.42	18.39	1.91	0.06	0.11
SS19	vnirfull	False	True	0.691	0.684	0.626	13.54	7.85	0.24	5.37	48.69	0.17	2.65	17.7	3.68	0.09	0.02
SS20	vnirfull	False	False	0.674	0.674	0.612	13.16	8.29	0.2	5.67	46.89	0.16	2.98	18.45	4.06	0.1	0.03

Table 22: Predicted results for Species Groups (South)

<i>Classifier Data</i>							<i>Percentage</i>					
ID	Data	DTM	BS	OOB	CV	Kappa	AA	Elbh	FS	PA	Slbh	Sndh
S21	all	True	True	0.816	0.806	0.764	12.16	4.62	55.99	16.89	4.09	6.25
S22	all	True	False	0.806	0.798	0.748	10.14	3.66	55.95	20.67	4	5.57
S23	all	False	True	0.729	0.724	0.655	12.25	6.57	52.98	17.08	4.9	6.23
S24	all	False	False	0.725	0.716	0.652	8.85	4.85	53.05	22.6	5.08	5.57
S25	specind	False	False	0.698	0.689	0.612	8.66	5.89	52.25	21.95	4.89	6.35
S26	specind	False	True	0.704	0.692	0.622	11.9	7.21	53.3	15.84	4.47	7.27
S27	spectral	False	True	0.685	0.675	0.581	12.29	7.4	54.43	13.77	4.93	7.18
S28	spectral	False	False	0.682	0.67	0.591	9.47	5.4	50.99	21.47	6.17	6.51
S29	swir	False	True	0.618	0.597	0.551	7.69	2.09	37.39	35.78	4.24	12.82
S30	swir	False	False	0.636	0.624	0.561	7.91	2.07	36.92	35.07	5.11	12.92
S31	swirfull	True	True	0.839	0.829	0.807	9.58	5.1	48.08	25.24	2.6	9.4
S32	swirfull	True	False	0.83	0.82	0.79	9.04	4.89	47.99	25.78	2.52	9.78
S33	swirfull	False	True	0.741	0.736	0.667	10.01	5.9	44.83	25.69	3.22	10.36
S34	swirfull	False	False	0.736	0.728	0.674	9.49	5.32	44.99	26.03	3.43	10.74
S35	vnir	False	True	0.606	0.595	0.49	8.89	10.23	44.5	14.4	9.37	12.61
S36	vnir	False	False	0.62	0.612	0.515	9.05	8.79	42.03	14.53	9.89	15.71
S37	vnirfull	True	True	0.824	0.809	0.763	12.16	6.62	51.11	17.21	5.25	7.65
S38	vnirfull	True	False	0.809	0.799	0.76	11.84	6.56	49.91	17.82	5.38	8.5
S39	vnirfull	False	True	0.714	0.712	0.628	10.88	10.99	44.79	19.12	5.99	8.23
S40	vnirfull	False	False	0.703	0.699	0.624	10.77	11.53	42.91	19.37	6.07	9.35

Table 23: Predicted results for Conifers / Broadleaf (South)

<i>Classifier Data</i>							<i>Percentage</i>	
Algorithm	Data	DTM	BS	OOB	CV	Kappa	Conifers	Broadleaved
S41	all	True	True	0.91	0.908	0.825	31.99	68.01
S42	all	True	False	0.906	0.904	0.82	34.86	65.14
S43	all	False	True	0.907	0.906	0.817	32.11	67.89
S44	all	False	False	0.904	0.899	0.804	35.05	64.95
S45	specind	False	True	0.908	0.902	0.808	31.85	68.15
S46	specind	False	False	0.901	0.898	0.801	34.84	65.16
S47	spectral	False	True	0.906	0.901	0.806	28.66	71.34
S48	spectral	False	False	0.901	0.897	0.801	34.19	65.81
S49	swir	False	True	0.882	0.881	0.78	56.85	43.15
S50	swir	False	False	0.889	0.89	0.787	55.67	44.33
S51	swirfull	True	True	0.92	0.919	0.841	43.96	56.04
S52	swirfull	True	False	0.915	0.912	0.839	44.11	55.89
S53	swirfull	False	True	0.915	0.91	0.825	44.32	55.68
S54	swirfull	False	False	0.911	0.91	0.818	44.47	55.53
S55	vnir	False	True	0.871	0.868	0.758	36.77	63.23
S56	vnir	False	False	0.872	0.868	0.764	39.36	60.64
S57	vnirfull	True	True	0.904	0.903	0.806	40.66	59.34
S58	vnirfull	True	False	0.898	0.894	0.804	41.64	58.36
S59	vnirfull	False	True	0.898	0.897	0.797	40.61	59.39
S60	vnirfull	False	False	0.891	0.888	0.793	41.33	58.67

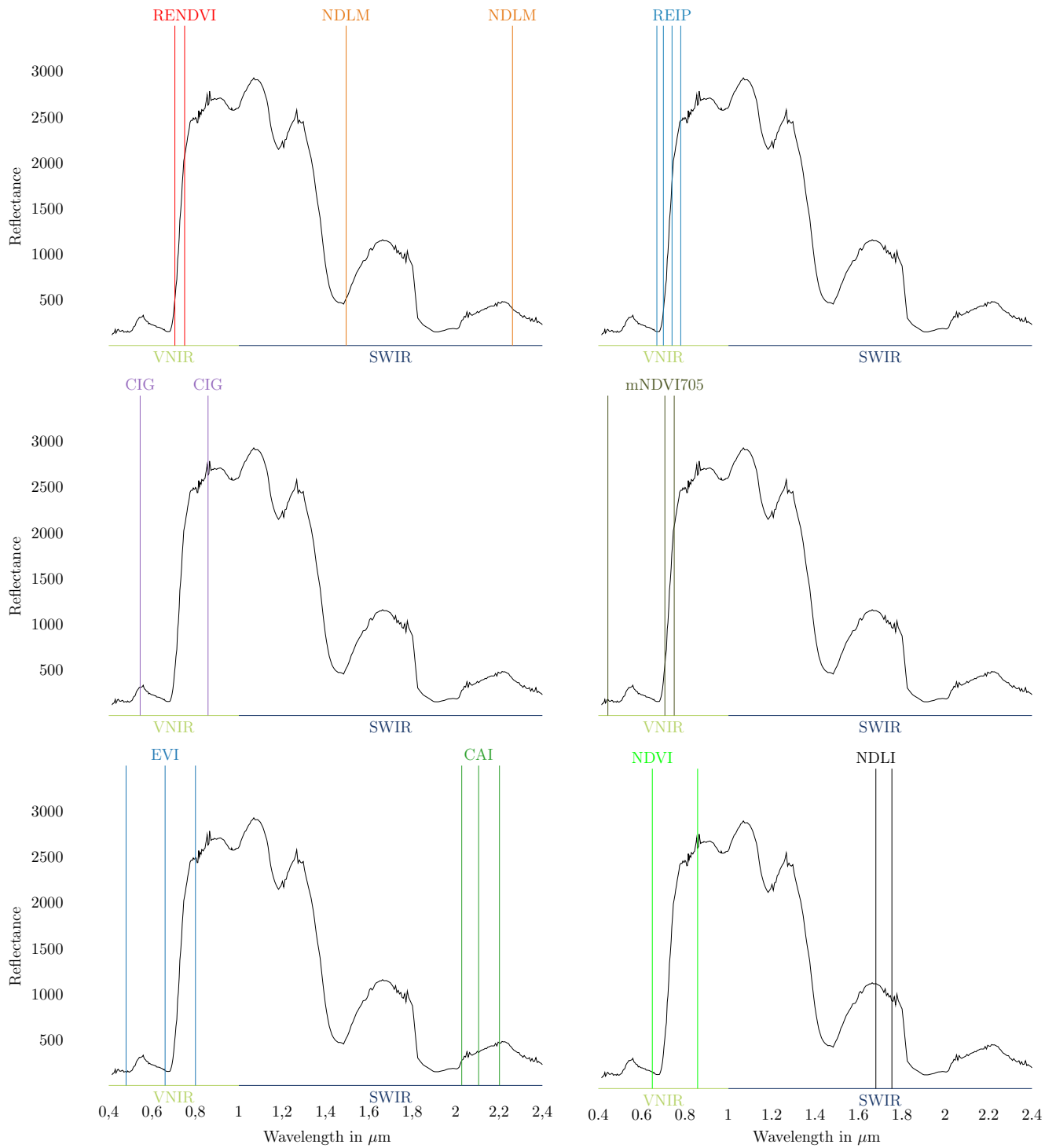


Figure 41: HySpex vegetation spectrum with chosen indices

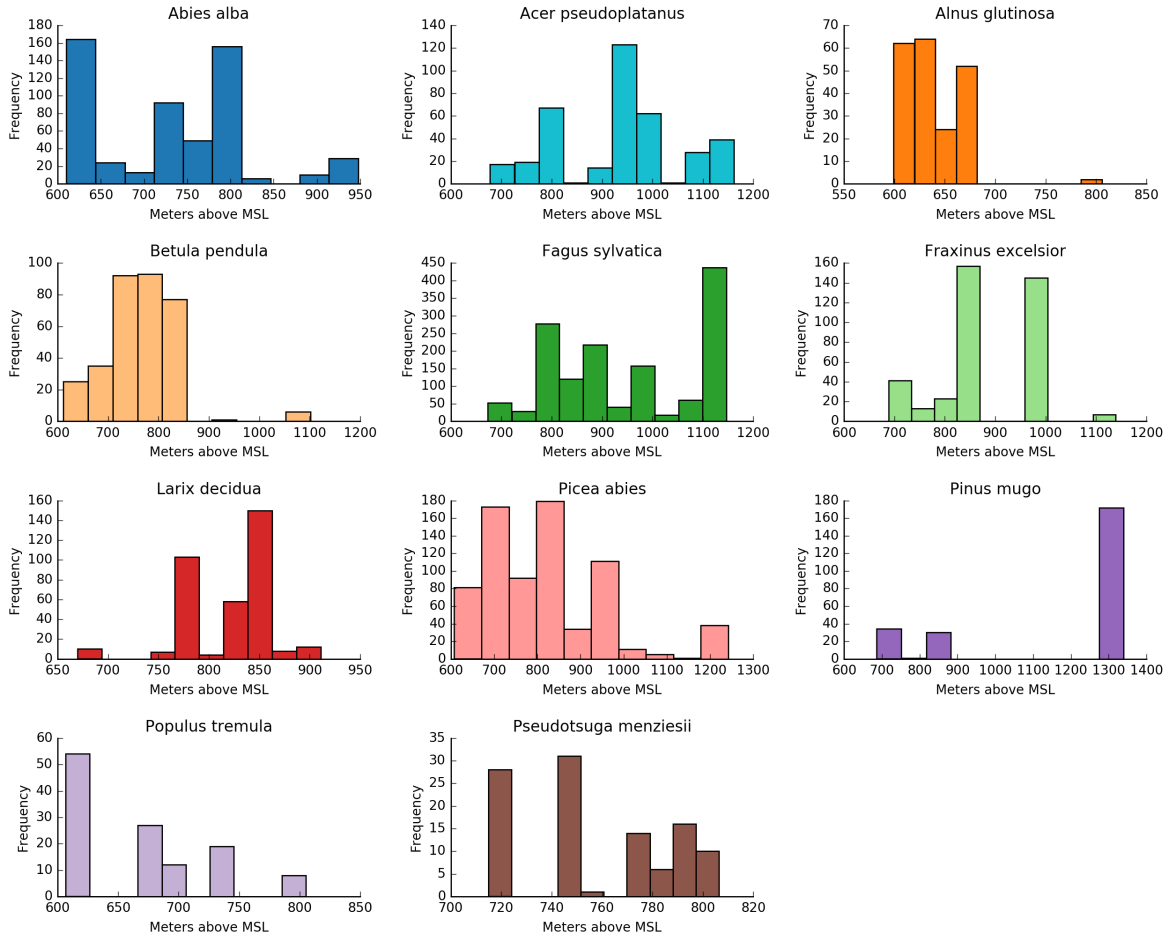


Figure 42: Height distribution for training data

Table 24: Classifier accuracy for species predicted by Random Forest

Classifier Data										F1-Score								
ID	Data	DTM	BS	OOB	CV	Kappa	AA	AP	AG	BP	FS	FE	LD	PA	PMu	PT	PM	
1	all	True	True	0.836	0.829	0.819	0.71	0.79	0.91	0.77	0.91	0.87	0.9	0.83	0.97	0.74	0.69	
2	all	True	False	0.838	0.83	0.807	0.78	0.8	0.82	0.76	0.88	0.87	0.88	0.8	0.95	0.83	0.72	
3	all	False	True	0.712	0.698	0.675	0.59	0.58	0.63	0.69	0.86	0.68	0.72	0.73	0.91	0.55	0.16	
4	all	False	False	0.712	0.702	0.656	0.51	0.62	0.58	0.63	0.85	0.76	0.69	0.7	0.94	0.38	0.38	
5	specind	False	True	0.659	0.649	0.576	0.54	0.27	0.48	0.52	0.83	0.59	0.65	0.66	0.9	0.34	0.14	
6	specind	False	False	0.652	0.65	0.599	0.5	0.34	0.55	0.59	0.84	0.65	0.66	0.64	0.9	0.23	0.32	
7	spectral	False	True	0.643	0.63	0.523	0.47	0.22	0.44	0.39	0.75	0.63	0.62	0.61	0.88	0.36	0	
8	spectral	False	False	0.628	0.619	0.562	0.49	0.29	0.4	0.38	0.79	0.68	0.59	0.66	0.92	0.44	0.22	
9	swir	False	True	0.608	0.598	0.492	0.42	0.32	0.23	0.34	0.72	0.68	0.5	0.65	0.83	0.13	0	
10	swir	False	False	0.602	0.6	0.511	0.35	0.35	0.32	0.34	0.75	0.64	0.5	0.65	0.86	0	0	
11	swirfull	True	True	0.844	0.828	0.84	0.82	0.8	0.85	0.84	0.92	0.89	0.92	0.79	0.96	0.82	0.61	
12	swirfull	True	False	0.847	0.831	0.811	0.78	0.82	0.88	0.78	0.89	0.86	0.93	0.76	0.96	0.81	0.55	
13	swirfull	False	True	0.712	0.7	0.658	0.6	0.55	0.57	0.66	0.84	0.76	0.73	0.68	0.86	0.62	0.33	
14	swirfull	False	False	0.709	0.697	0.669	0.55	0.7	0.66	0.67	0.87	0.71	0.71	0.69	0.88	0.42	0.29	
15	vnir	False	True	0.571	0.568	0.483	0.42	0.32	0.36	0.3	0.75	0.5	0.56	0.58	0.87	0.24	0.17	
16	vnir	False	False	0.561	0.556	0.499	0.44	0.27	0.51	0.34	0.76	0.5	0.61	0.56	0.84	0.21	0.08	
17	vnirfull	True	True	0.84	0.831	0.809	0.74	0.78	0.88	0.79	0.91	0.85	0.88	0.77	0.98	0.81	0.68	
18	vnirfull	True	False	0.841	0.834	0.802	0.71	0.76	0.87	0.77	0.89	0.9	0.88	0.75	0.97	0.67	0.74	
19	vnirfull	False	True	0.692	0.685	0.663	0.49	0.54	0.73	0.61	0.85	0.66	0.7	0.71	0.94	0.52	0.4	
20	vnirfull	False	False	0.699	0.688	0.645	0.49	0.54	0.65	0.63	0.84	0.68	0.71	0.7	0.9	0.67	0.3	
\mathcal{C}_V							0.245	0.408	0.337	0.306	0.073	0.172	0.193	0.105	0.051	0.517	0.741	

Table 25: Predicted results for Species (North) using Random Forest

Classifier Data							Percentage										
ID	Data	DTM	BS	OOB	CV	Kappa	AA	AP	AG	BP	FS	FE	LD	PA	PMu	PT	PM
1	all	True	True	0.833	0.829	0.824	14.46	3.4	1.7	1.03	31.13	8.48	3.58	33.91	0.88	0.25	1.18
2	all	True	False	0.83	0.822	0.817	13.79	3.57	1.62	1.08	30.43	8.68	3.47	35.22	0.86	0.21	1.06
3	all	False	True	0.719	0.717	0.653	13.18	5.34	1.2	1.39	23.71	10.19	7.98	35.09	1.16	0.17	0.6
4	all	False	False	0.718	0.713	0.661	12.1	4.97	1.03	1.41	23	10.91	7.3	37.36	1.26	0.18	0.48
5	specind	False	True	0.667	0.653	0.595	16.11	1.99	1.38	1.81	32.5	3.85	4.85	35.82	1.14	0.18	0.38
6	specind	False	False	0.661	0.648	0.59	14.19	1.55	1.07	1.93	32.25	4.09	4.46	38.78	1.25	0.17	0.27
7	spectral	False	True	0.641	0.624	0.551	21.54	3.92	2.18	1.27	35.27	2.38	4.22	28.05	0.47	0.25	0.44
8	spectral	False	False	0.637	0.626	0.54	18.99	2.06	1.18	1.23	33.93	2.97	3.72	34.34	1.07	0.24	0.27
9	swir	False	True	0.557	0.545	0.515	15.01	2.33	0.41	2.34	41.8	0.85	8.81	24.05	3.94	0.2	0.26
10	swir	False	False	0.566	0.562	0.533	13.48	2.12	0.38	2.04	41.85	1.18	8.86	26.32	3.55	0.08	0.13
11	swirfull	True	True	0.828	0.818	0.838	11.49	3.05	0.03	2.44	51.24	0.16	3.81	25.67	1.92	0.05	0.13
12	swirfull	True	False	0.812	0.803	0.829	10.87	2.46	0.03	2.27	51.73	0.16	3.86	26.26	2.21	0.05	0.11
13	swirfull	False	True	0.691	0.684	0.679	11.42	4.28	0.41	1.66	47.02	0.09	5.81	27.05	2.19	0.02	0.05
14	swirfull	False	False	0.674	0.674	0.671	10.54	3.46	0.47	1.51	47.52	0.15	6.23	27.87	2.21	0.01	0.03
15	vnir	False	True	0.593	0.584	0.461	10.6	3.43	1.9	4.85	48.86	3.97	9.35	14.1	2.45	0.37	0.13
16	vnir	False	False	0.606	0.591	0.465	10.57	2.67	1.6	4.68	48.6	2.95	9.74	13.88	4.57	0.68	0.08
17	vnirfull	True	True	0.849	0.841	0.791	14.22	4.24	0.03	4.99	54.47	0.2	2.48	17.75	1.49	0.06	0.08
18	vnirfull	True	False	0.835	0.827	0.774	14.2	4.06	0.02	5.17	53.5	0.18	2.42	18.38	1.91	0.06	0.11
19	vnirfull	False	True	0.719	0.713	0.626	13.54	7.85	0.24	5.37	48.69	0.17	2.65	17.7	3.67	0.09	0.02
20	vnirfull	False	False	0.713	0.701	0.612	13.16	8.29	0.2	5.67	46.9	0.16	2.98	18.45	4.06	0.1	0.03

Glossary

- bagging** Bootstrap aggregating (Breiman 1996a). 17
- BRDF** Bidirectional Reflectance Distribution Function. 24, 28
- BRF** Bi-hemispherical reflectance. 25
- BS** band selection. 34
- CAI** Cellulose Absorption Index. 10
- CATENA** Refers to the universal, operational infrastructure set up by the DLR, for the automatic processing of (optical) satellite and airborne image data. vii, 22
- CHM** Canopy Height Model. 15
- CIG** Chlorophyll Index Green. 10
- Coefficient of variation** (c_v) The ratio of the standard deviation σ to the mean μ . 43
- CV** Cross-validation. 16
- Digital Number** Pixel values not transformed into physically units. 22
- DLR** Deutsches Zentrum für Luft- und Raumfahrt (DLR). vii, 2, 3, 22
- DTM** Digital Terrain Model. 13, 33, 34
- ET** Extremely randomized trees. 32, 33, 62
- EV** Enhanced Vegetation Index. 10
- feature** Quantity describing an instance. 33
- GDAL** Geospatial Data Abstraction Library. 14
- HySpex** Line of hyperspectral cameras developed by Norsk Elektro Optikk. 8
- instance** Single data item. 31, 33
- Kappa** Cohen’s kappa coefficient. 19
- LAI** Leaf area index. 10–12
- LiDAR** Light Detection And Ranging. 13
- mNDVI** Modified NDVI. 12
- nadir** The nadir is the direction pointing directly below a particular location. 24
- NDLI** Normalized Difference Lignin Index. 11
- NDLma** Normalized Difference Leaf Mass (per area). 11
- NDVI** Normalized Difference Vegetation Index. 11, 15
- OOB** Out-of-bag error. 16
- PCA** Principal Component Analysis. 31
- REDNDVI** Red Edge NDVI. 12
- REIP** Red Edge Inflection Point. 12
- RF** Random Forest. 32, 33
- Seed** Refers to the seed used in generating randomness. 44
- Split** Refers to the split into training and test set, or the split of a node in a decision tree. 44
- SWIR** Short Wave Infrared. 3, 9, 34
- VI**s vegetation indices. 34, 45, 47, 49, 54, 59–61, 69, 70
- VNIR** Visible Near Infrared. 3, 9, 34

References

- Aggarwal, Charu C and Philip S Yu (2001). “Outlier detection for high dimensional data”. In: *ACM Sigmod Record*. Vol. 30. 2. ACM, pp. 37–46. DOI: 10.1145/376284.375668 (cit. on p. 30).
- Appice, Annalisa and Pietro Guccione (2016). “Exploiting Spatial Correlation of Spectral Signature for Training Data Selection in Hyperspectral Image Classification”. In: *Discovery Science: 19th International Conference, DS 2016, Bari, Italy, October 19–21, 2016, Proceedings*. Ed. by Toon Calders, Michelangelo Ceci, and Donato Malerba. Cham: Springer International Publishing, pp. 295–309. DOI: 10.1007/978-3-319-46307-0_19 (cit. on p. 67).
- Archer, Kellie J. and Ryan V. Kimes (2008). “Empirical characterization of random forest variable importance measures”. In: *Computational Statistics & Data Analysis* 52.4, pp. 2249–2260. DOI: 10.1016/j.csda.2007.08.015 (cit. on pp. 32, 36).
- Asner, Gregory P. (1998). “Biophysical and Biochemical Sources of Variability in Canopy Reflectance”. In: *Remote Sensing of Environment* 64.3, pp. 234–253. DOI: 10.1016/S0034-4257(98)00014-5 (cit. on p. 9).
- Asner, Gregory P., Jeffrey A. Hicke, and David B. Lobell (2003). “Per-Pixel Analysis of Forest Structure”. In: *Remote Sensing of Forest Environments: Concepts and Case Studies*. Ed. by Michael A. Wulder and Steven E. Franklin. Boston, MA: Springer US, pp. 209–254. DOI: 10.1007/978-1-4615-0306-4_8 (cit. on p. 9).
- Assmann, Ernst (1961). *Waldetragskunde: organische Produktion, Struktur, Zuwachs und Ertrag von Waldbeständen*. BLV Verlagsgesellschaft (cit. on p. 13).
- Aulinger, T, T Mette, KP Papathanassiou, et al. (2005). “Validation of Heights from Interferometric SAR and LIDAR over the Temperate Forest Site "Nationalpark Bayerischer Wald"”. In: *2nd International Workshop on Applications of Polarimetry and Polarimetric Interferometry*. (Frascati, Italy, Jan. 17–21, 2005) (cit. on p. 2).

- Bannari, A., D. Morin, F. Bonn, and A. R. Huete (1995). “A review of vegetation indices”. In: *Remote Sensing Reviews* 13.1-2, pp. 95–120. DOI: 10.1080/02757259509532298 (cit. on pp. 9, 35).
- Ben-Gal, Irad (2005). “Outlier Detection”. In: *Data Mining and Knowledge Discovery Handbook*. Ed. by Oded Maimon and Lior Rokach. Boston, MA: Springer US, pp. 131–146. DOI: 10.1007/0-387-25465-X_7 (cit. on p. 29).
- Bengio, Yoshua (2012). “Practical Recommendations for Gradient-Based Training of Deep Architectures”. In: *Lecture Notes in Computer Science*. Springer Nature, pp. 437–478. DOI: 10.1007/978-3-642-35289-8_26 (cit. on p. 38).
- Bergstra, James, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl (2011). “Algorithms for Hyper-parameter Optimization”. In: *Proceedings of the 24th International Conference on Neural Information Processing Systems*. NIPS’11. Granada, Spain, pp. 2546–2554 (cit. on p. 38).
- Bergstra, James and Yoshua Bengio (2012). “Random search for hyper-parameter optimization”. In: *The Journal of Machine Learning Research* 13.1, pp. 281–305 (cit. on p. 38).
- Bernard, Simon, Laurent Heutte, and Sébastien Adam (2009). “Influence of Hyperparameters on Random Forest Accuracy”. In: *Multiple Classifier Systems*. Springer Nature, pp. 171–180. DOI: 10.1007/978-3-642-02326-2_18 (cit. on p. 40).
- Biau, GÅřard, Luc Devroye, and GÅĄbor Lugosi (2008). “Consistency of random forests and other averaging classifiers”. In: *Journal of Machine Learning Research* 9.Sep, pp. 2015–2033 (cit. on p. 69).
- Bishop, Christopher M (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. New York: Springer (cit. on p. 38).
- Bolón-Canedo, Verónica, Noelia Sánchez-Marroño, and Amparo Alonso-Betanzos (2015). *Feature Selection for High-Dimensional Data*. Springer International Publishing. DOI: 10.1007/978-3-319-21858-8 (cit. on p. 35).
- Boschetti, M., L. Boschetti, S. Oliveri, L. Casati, and I. Canova (2007). “Tree species mapping with Airborne hyper-spectral MIVIS data: the Ticino Park study case”. In: *International Journal of Remote Sensing* 28.6, pp. 1251–1261. DOI: 10.1080/01431160600928542 (cit. on p. 2).

- Breiman, Leo (1996a). “Bagging predictors”. In: *Machine Learning* 24.2, pp. 123–140. DOI: 10.1007/bf00058655 (cit. on pp. 31 sq., 79).
- (1996b). *Out-of-bag estimation*. Tech. rep. Berkeley CA 94708: Statistics Department, University of California Berkeley, pp. 1–13. URL: <https://www.stat.berkeley.edu/~breiman/OOBestimation.pdf> (cit. on p. 16).
- (2001). “Random Forests”. In: *Machine Learning* 45.1, pp. 5–32. DOI: 10.1023/a:1010933404324 (cit. on pp. 31, 39).
- (1996c). “Technical Note: Some Properties of Splitting Criteria”. In: *Machine Learning* 24.1, pp. 41–47. DOI: 10.1023/a:1018094028462 (cit. on p. 36).
- Buddenbaum, H, M Schlerf, and J Hill (2005). “Classification of coniferous tree species and age classes using hyperspectral data and geostatistical methods”. In: *International Journal of Remote Sensing* 26.24, pp. 5453–5465. DOI: 10.1080/01431160500285076 (cit. on p. 2).
- Carlson, Toby N. and David A. Ripley (1997). “On the relation between NDVI, fractional vegetation cover, and leaf area index”. In: *Remote Sensing of Environment* 62.3, pp. 241–252. DOI: 10.1016/S0034-4257(97)00104-1 (cit. on p. 12).
- Caruana, Rich and Alexandru Niculescu-Mizil (2006). “An Empirical Comparison of Supervised Learning Algorithms”. In: *Proceedings of the 23rd International Conference on Machine Learning*. ICML ’06. Pittsburgh, Pennsylvania, USA: ACM, pp. 161–168. DOI: 10.1145/1143844.1143865 (cit. on p. 32).
- Cervone, Guido, Jessica Lin, and Nigel Waters, eds. (2014). *Data Mining for Geoinformatics*. Springer Nature. DOI: 10.1007/978-1-4614-7669-6 (cit. on p. 28).
- Chan, Jonathan Cheung-Wai and Desiré Paelinckx (2008). “Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery”. In: *Remote Sensing of Environment* 112.6, pp. 2999–3011. DOI: 10.1016/j.rse.2008.02.011 (cit. on p. 32).
- Chawla, Nitesh V., Nathalie Japkowicz, and Aleksander Kotcz (2004). “Editorial”. In: *ACM SIGKDD Explorations Newsletter* 6.1, p. 1. DOI: 10.1145/1007730.1007733 (cit. on p. 41).

- Chen, Chao, Andy Liaw, and Leo Breiman (2004). *Using random forest to learn imbalanced data*. Tech. rep. (cit. on p. 41).
- Clark, Matthew L, Dar A Roberts, and David B Clark (2005). “Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales”. In: *Remote sensing of environment* 96.3, pp. 375–398 (cit. on p. 2).
- Clutter, Jerome L, James C Fortson, Leon V Pienaar, Graham H Brister, Robert L Bailey, et al. (1983). *Timber management: a quantitative approach*. John Wiley & Sons, Inc. (cit. on p. 13).
- Cohen, J. (1960). “A coefficient of agreement for nominal scales”. In: *Educational and psychological measurement* 20.1, pp. 37–46. DOI: 10.1177/001316446002000104 (cit. on p. 19).
- Crawford, M.M., JiSoo Ham, Yangchi Chen, and J. Ghosh (2003). “Random forests of binary hierarchical classifiers for analysis of hyperspectral data”. In: *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003*. Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/warsd.2003.1295213 (cit. on p. 32).
- Cutler, Adele, D. Richard Cutler, and John R. Stevens (2012). “Random Forests”. In: *Ensemble Machine Learning*. Springer Nature, pp. 157–175. DOI: 10.1007/978-1-4419-9326-7_5 (cit. on p. 32).
- Czaja J. und Heurich, M. (2003). “GPS für Waldinventur im Nationalpark Bayerischer Wald”. In: *Oesterreichische Forstzeitung* 10/2003, pp. 30–32 (cit. on p. 7).
- Dalponte, M., Lorenzo Bruzzone, and Damiano Gianelle (2008). “Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas”. In: *Geoscience and Remote Sensing, IEEE Transactions on* 46.5, pp. 1416–1427. DOI: 10.1109/TGRS.2008.916480 (cit. on p. 2).
- (2012). “Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data”. In: *Remote Sensing of Environment* 123, pp. 258–270. DOI: 10.1016/j.rse.2012.03.013 (cit. on p. 2).
- Dalponte, M., H. O. Ørka, T. Gobakken, D. Gianelle, and E. Næsset (2013). “Tree Species Classification in Boreal Forests With Hyperspectral Data”. In: *IEEE*

- Transactions on Geoscience and Remote Sensing* 51.5, pp. 2632–2645. DOI: 10.1109/TGRS.2012.2216272 (cit. on p. 2).
- Daughtry, C.S.T., E.R. Hunt Jr., and J.E. McMurtrey (2004). “Assessing crop residue cover using shortwave infrared reflectance”. In: *Remote Sensing of Environment* 90.1, pp. 126–134. DOI: 10.1016/j.rse.2003.10.023 (cit. on p. 10).
- Dierschke, H. (1994). *Pflanzensoziologie: Grundlagen und Methoden*. UTB für Wissenschaft : Botanik, Ökologie, Agrar- und Forstwissenschaften. Ulmer (cit. on p. 15).
- DLR Remote Sensing Technology Institute (2016). “Airborne Imaging Spectrometer HySpex”. In: *Journal of large-scale research facilities*. DOI: 10.17815/jlsrf-2-151 (cit. on p. 8).
- E. Raymond Hunt, Jr., Paul C. Doraiswamy, James E. McMurtrey, et al. (2013). “A visible band index for remote sensing leaf chlorophyll content at the canopy scale”. In: *International Journal of Applied Earth Observation and Geoinformation* 21, pp. 103–112. DOI: 10.1016/j.jag.2012.07.020 (cit. on p. 10).
- Ebtehadj, Knorso (1973). “Application of ERTS-1 imagery in the fields of geology, agriculture, forestry, and hydrology to selected test sites in Iran”. In: *Goddard Space Flight Center Symp. on Significant Results obtained from the ERTS-1*. Ed. by S. C. Freden, E. P. Mercanti, and M. A. Becker. Vol. 1. NASA, pp. 1699–1714. URL: <https://ntrs.nasa.gov/search.jsp?R=19730019655> (cit. on p. 1).
- Fassnacht, Fabian E, Carsten Neumann, Michael Forster, et al. (2014). “Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites”. In: *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 7.6, pp. 2547–2561. DOI: 10.1109/JSTARS.2014.2329390 (cit. on p. 9).
- Fawcett, Tom (2006). “An introduction to ROC analysis”. In: *Pattern recognition letters* 27.8, pp. 861–874. DOI: 10.1016/j.patrec.2005.10.010 (cit. on p. 17).

- Feilhauer, Hannes, Gregory P. Asner, Roberta E. Martin, and Sebastian Schmidtlein (2010). “Brightness-normalized partial least squares regression for hyperspectral data”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 111.12, pp. 1947–1957. DOI: 10.1016/j.jqsrt.2010.03.007 (cit. on p. 29).
- Fernández-Delgado, Manuel, Eva Cernadas, Senén Barro, and Dinani Amorim (2014). “Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?” In: *Journal of Machine Learning Research* 15, pp. 3133–3181 (cit. on p. 32).
- Foody, Giles M. (2002). “Status of land cover classification accuracy assessment”. In: *Remote sensing of environment* 80.1, pp. 185–201. DOI: 10.1016/S0034-4257(01)00295-4 (cit. on pp. 16, 19).
- Foody, Giles M., I. M. J. Sargent, P. M. Atkinson, and J. W. Williams (2004). “Thematic labelling from hyperspectral remotely sensed imagery: trade-offs in image properties”. In: *International Journal of Remote Sensing* 25.12, pp. 2337–2363. DOI: 10.1080/01431160310001654969 (cit. on p. 35).
- Fuan, Tsai and W.D. Philpot (2002). “A derivative-aided hyperspectral image analysis system for land-cover classification”. In: *IEEE Transactions on Geoscience and Remote Sensing* 40.2, pp. 416–425. DOI: 10.1109/36.992805 (cit. on p. 31).
- Gadow, Klaus von (2003). *Waldstruktur und Wachstum*. Universitaetsverlag Goettingen, p. 241. URL: <http://resolver.sub.uni-goettingen.de/purl?isbn-3-930457-32-6> (cit. on p. 13).
- Gamon, John A., Christopher B. Field, Michael L. Goulden, et al. (1995). “Relationships Between NDVI, Canopy Structure, and Photosynthesis in Three Californian Vegetation Types”. In: *Ecological Applications* 5.1, pp. 28–41. DOI: 10.2307/1942049 (cit. on p. 12).
- Geurts, Pierre (2002). “Contributions to decision tree induction: bias/variance tradeoff and time series classification”. PhD thesis. University of Liège, Belgium. URL: <http://www.montefiore.ulg.ac.be/services/stochastic/pubs/2002/Geu02> (cit. on p. 43).

- Geurts, Pierre, Damien Ernst, and Louis Wehenkel (2006). “Extremely randomized trees”. In: *Machine Learning* 63.1, pp. 3–42. DOI: 10.1007/s10994-006-6226-1 (cit. on pp. 32 sq., 40).
- Gislason, Pall Oskar, Jon Atli Benediktsson, and Johannes R. Sveinsson (2006). “Random Forests for land cover classification”. In: *Pattern Recognition Letters* 27.4, pp. 294–300. DOI: 10.1016/j.patrec.2005.08.011 (cit. on p. 32).
- Gitelson, Anatoly and Mark Merzlyak (1994). “Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation”. In: *Journal of Plant Physiology* 143.3, pp. 286–292. DOI: 10.1016/S0176-1617(11)81633-0 (cit. on p. 12).
- Gitelson, Anatoly, Andrés Viña, Timothy J. Arkebauer, et al. (2003). “Remote estimation of leaf area index and green leaf biomass in maize canopies”. In: *Geophysical Research Letters* 30.5. 1248. DOI: 10.1029/2002GL016450 (cit. on p. 10).
- Gong, Peng, Ruiliang Pu, and Bin Yu (1997). “Conifer species recognition: An exploratory analysis of in situ hyperspectral data”. In: *Remote Sensing of Environment* 62.2, pp. 189–200. DOI: 10.1016/S0034-4257(97)00094-1 (cit. on p. 2).
- Gregorutti, Baptiste, Bertrand Michel, and Philippe Saint-Pierre (2016). “Correlation and variable importance in random forests”. In: *Statistics and Computing*. DOI: 10.1007/s11222-016-9646-1 (cit. on pp. 36 sq.).
- Grubbs, Frank E. (1969). “Procedures for Detecting Outlying Observations in Samples”. In: *Technometrics* 11.1, pp. 1–21. DOI: 10.1080/00401706.1969.10490657 (cit. on p. 30).
- Guyon, I. and A. Elisseeff (2003). “An Introduction to Variable and Feature Selection.” In: *Journal of Machine Learning* 3, pp. 1157–1182 (cit. on p. 68).
- Ham, J., Yangchi Chen, M.M. Crawford, and J. Ghosh (2005). “Investigation of the random forest framework for classification of hyperspectral data”. In: *IEEE Transactions on Geoscience and Remote Sensing* 43.3, pp. 492–501. DOI: 10.1109/tgrs.2004.842481 (cit. on p. 32).

- Hänsch, R and O Hellwich (2016). “Task-dependent Band-selection of Hyperspectral images by projection-based Random Forests.” In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* III-7, pp. 263–270. DOI: 10.5194/isprs-annals-III-7-263-2016 (cit. on pp. 35, 68).
- Hastie, Trevor, Robert Tibshirani, and Jerome H. Friedman (2009). *The elements of statistical learning : data mining, inference, and prediction*. Springer series in statistics. New York: Springer. DOI: 10.1007/978-0-387-84858-7 (cit. on pp. 16, 31 sq., 38).
- He, Haibo and E.A. Garcia (2009). “Learning from Imbalanced Data”. In: *IEEE Transactions on Knowledge and Data Engineering* 21.9, pp. 1263–1284. DOI: 10.1109/tkde.2008.239 (cit. on p. 41).
- Heurich, Marco, Peter Krzystek, Fabian Polakowsky, et al. (2015). “Erste Waldinventur auf Basis von Lidardaten und digitalen Luftbildern im Nationalpark Bayerischer Wald”. In: *Forstliche Forschungsberichte Muenchen* 214.Part 3, pp. 101–113 (cit. on p. 2).
- Heurich, Marco and Markus Neufanger (2005). *Die Wälder des Nationalparks Bayerischer Wald: Ergebnisse der Waldinventur 2002/2003 im Geschichtlichen und Waldökologischen Kontext*. Nationalparkverwaltung Bayerischer Wald (cit. on pp. 5 sq.).
- Heurich, Marco, S Schadeck, H Weinacker, and Peter Krzystek (2004). “Forest parameter derivation from DTM/DSM generated from lidar and digital modular camera (DMC)”. In: *XX ISPRS Congress. Istanbul, Turcja*. Vol. XXXV. URL: <http://www.isprs.org/proceedings/XXXV/congress/comm2/papers/103.pdf> (cit. on p. 2).
- Heurich, Marco, Thomas Schneider, and Eckhard Kennel (2003). “Laser scanning for identification of forest structures in the Bavarian forest national park”. In: *Proceedings of the ScandLaser scientific workshop on airborne laser scanning of forests*. Swedish University of Agricultural Sciences, Department of Forest Resource Management and Geomatics, pp. 98–107. URL: http://pub.epsilon.slu.se/9060/1/hyyppa_et_al_120814.pdf (cit. on p. 2).

- Ho, Tin Kam (1998). “The random subspace method for constructing decision forests”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.8, pp. 832–844. DOI: 10.1109/34.709601 (cit. on p. 31).
- Holmgren, J., Å. Persson, and U. Söderman (2008). “Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images”. In: *International Journal of Remote Sensing* 29.5, pp. 1537–1552. DOI: 10.1080/01431160701736471 (cit. on p. 2).
- Horler, D. N. H., M. Dockray, and J. Barber (1983). “The red edge of plant leaf reflectance”. In: *International Journal of Remote Sensing* 4.2, pp. 273–288. DOI: 10.1080/01431168308948546 (cit. on p. 12).
- Huete, A, K Didan, T Miura, et al. (2002). “Overview of the radiometric and biophysical performance of the MODIS vegetation indices”. In: *Remote Sensing of Environment* 83.1-2, pp. 195–213. DOI: 10.1016/s0034-4257(02)00096-2 (cit. on p. 11).
- Hui, K. Y. and M. Gratzl (1996). “Anomalies of Convolutional Smoothing and Differentiation”. In: *Analytical Chemistry* 68.6, pp. 1054–1057. DOI: 10.1021/ac950697k (cit. on p. 27).
- Hyypä, Juha, H. Hyypä, D. Leckie, et al. (2008). “Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests”. In: *International Journal of Remote Sensing* 29.5, pp. 1339–1366. DOI: 10.1080/01431160701736489 (cit. on p. 13).
- Hyypä, Juha and Mikko Inkinen (1999). “Detecting and estimating attributes for single trees using laser scanner”. In: *The photogrammetric journal of Finland* 16.2, pp. 27–42 (cit. on p. 2).
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani (2013). “Tree-Based Methods”. In: *An Introduction to Statistical Learning: with Applications in R*. New York, NY: Springer New York, pp. 303–335. DOI: 10.1007/978-1-4614-7138-7_8 (cit. on p. 16).
- Janecek, Andreas, Wilfried N. Gansterer, Michael Demel, and Gerhard Ecker (2008). “On the Relationship Between Feature Selection and Classification Accuracy”. In: *Third Workshop on New Challenges for Feature Selection in Data Mining and Knowledge Discovery, FSDM 2008, held at ECML-PKDD*

- 2008, Antwerp, Belgium, September 15, 2008, pp. 90–105. URL: <http://www.jmlr.org/proceedings/papers/v4/janecek08a.html> (cit. on p. 35).
- Jurgens, C. (1997). “The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data”. In: *International Journal of Remote Sensing* 18.17, pp. 3583–3594. DOI: 10.1080/014311697216810 (cit. on p. 12).
- Keenan, Rodney J., Gregory A. Reams, Frédéric Achard, et al. (2015). “Dynamics of global forest area: Results from the {FAO} Global Forest Resources Assessment 2015”. In: *Forest Ecology and Management* 352. Changes in Global Forest Resources from 1990 to 2015, pp. 9–20. DOI: 10.1016/j.foreco.2015.06.014 (cit. on p. 1).
- King, Roger L., Chris Ruffin, F.E. LaMastus, and D.R. Shaw (1999). “The analysis of hyperspectral data using Savitzky-Golay filtering-practical issues. Part II”. In: *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS’99*. Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/igarss.1999.773512. URL: <http://dx.doi.org/10.1109/IGARSS.1999.773512> (cit. on p. 27).
- Kirvida, L and Greg R. Johnson (1973). “Automatic interpretation of ERTS data for forest management”. In: URL: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19730019588.pdf> (cit. on p. 1).
- Koch, Barbara (2010). “Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.6, pp. 581–590. DOI: 10.1016/j.isprsjprs.2010.09.001 (cit. on p. 13).
- Köhler, Claas Henning (2016). “Airborne Imaging Spectrometer HySpex”. In: *Journal of large-scale research facilities JLSRF* 2. DOI: 10.17815/jlsrf-2-151 (cit. on p. 22).
- Koukal, T. and W. Schneider (2010). “Analysis of BRDF characteristics of forest stands with a digital aerial frame camera”. In: *ISPRS TC VII Symposium – 100 Years ISPRS*. (Vienna, Austria). Vol. XXXVIII. IAPRS (cit. on p. 24).
- Koukal, Tatjana, Clement Atzberger, and Werner Schneider (2014). “Evaluation of semi-empirical BRDF models inverted against multi-angle data from a digital

- airborne frame camera for enhancing forest type classification”. In: *Remote Sensing of Environment* 151, pp. 27–43. DOI: 10.1016/j.rse.2013.12.014 (cit. on p. 67).
- Krauß, T., P. d’Angelo, M. Schneider, and V. Gstaiger (2013). “The fully automatic optical processing system CATENA at DLR”. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-1/W1, pp. 177–183. DOI: 10.5194/isprsarchives-xl-1-w1-177-2013 (cit. on p. 22).
- Kriegel, Hans-Peter, Peer Kröger, Erich Schubert, and Arthur Zimek (2009). “Outlier Detection in Axis-Parallel Subspaces of High Dimensional Data”. In: *Advances in Knowledge Discovery and Data Mining: 13th Pacific-Asia Conference, PAKDD 2009 Bangkok, Thailand, April 27-30, 2009 Proceedings*. Ed. by Thanaruk Theeramunkong, Boonserm Kijsirikul, Nick Cercone, and Tu-Bao Ho. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 831–838. DOI: 10.1007/978-3-642-01307-2_86 (cit. on p. 30).
- Kubat, Miroslav, Robert C. Holte, and Stan Matwin (1998). In: *Machine Learning* 30.2/3, pp. 195–215. DOI: 10.1023/a:1007452223027 (cit. on p. 41).
- Landis, J. Richard and Gary G. Koch (1977). “The Measurement of Observer Agreement for Categorical Data”. In: *Biometrics* 33.1, p. 159. DOI: 10.2307/2529310 (cit. on p. 19).
- Latifi, Hooman, Fabian E. Fassnacht, Jörg Müller, et al. (2015). “Forest inventories by LiDAR data: A comparison of single tree segmentation and metric-based methods for inventories of a heterogeneous temperate forest”. In: *International Journal of Applied Earth Observation and Geoinformation* 42, pp. 162–174. DOI: 10.1016/j.jag.2015.06.008 (cit. on p. 13).
- Laurikkala, Jorma (2001). “Improving Identification of Difficult Small Classes by Balancing Class Distribution”. In: *Artificial Intelligence in Medicine*. Springer Nature, pp. 63–66. DOI: 10.1007/3-540-48229-6_9 (cit. on p. 41).
- Lawrence, Robert D and James H Herzog (1975). “Geology and forestry classification from ERTS-1 digital data”. In: *Photogrammetric Engineering and Remote Sensing* 41.10. URL: <https://ntrs.nasa.gov/search.jsp?R=19750063550> (cit. on p. 1).

- Lenhard, Karim, Andreas Baumgartner, and Thomas Schwarzmaier (2015). “Independent Laboratory Characterization of NEO HySpex Imaging Spectrometers VNIR-1600 and SWIR-320m-e”. In: *IEEE Transactions on Geoscience and Remote Sensing* 53.4, pp. 1828–1841. DOI: 10.1109/tgrs.2014.2349737 (cit. on p. 23).
- Leutenegger, Stefan, Margarita Chli, and Roland Y. Siegwart (2011). “BRISK: Binary Robust invariant scalable keypoints”. In: *2011 International Conference on Computer Vision*. Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/iccv.2011.6126542 (cit. on p. 23).
- Li, Ying, Haokui Zhang, and Qiang Shen (2017). “Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network”. In: *Remote Sensing* 9.1, p. 67. DOI: 10.3390/rs9010067 (cit. on p. 69).
- Lim, Kevin, Paul Treitz, Michael Wulder, Benoit St-Onge, and Martin Flood (2003). “LiDAR remote sensing of forest structure”. In: *Progress in Physical Geography* 27.1, pp. 88–106. DOI: 10.1191/0309133303pp360ra (cit. on p. 2).
- Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou (2008). “Isolation forest”. In: *2008 Eighth IEEE International Conference on Data Mining*. IEEE, pp. 413–422. DOI: 10.1109/ICDM.2008.17 (cit. on p. 30).
- (2012). “Isolation-based anomaly detection”. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6.1, p. 3. DOI: 10.1145/2133360.2133363 (cit. on p. 30).
- Liu, Huan and Hiroshi Motoda, eds. (1998). *Feature Extraction, Construction and Selection*. Springer Nature. DOI: 10.1007/978-1-4615-5725-8 (cit. on p. 35).
- Louppe, Gilles (2014). “Understanding Random Forests: From Theory to Practice”. PhD thesis. University of Liège, Belgium. URL: <https://arxiv.org/abs/1407.7502v3> (cit. on p. 38).
- Louppe, Gilles, Louis Wehenkel, Antonio Sutera, and Pierre Geurts (2013). “Understanding variable importances in forests of randomized trees”. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems*. Lake Tahoe, Nevada, pp. 431–439 (cit. on p. 36).

- Maignan, F., F.-M Bréon, and R. Lacaze (2004). “Bidirectional reflectance of Earth targets: evaluation of analytical models using a large set of spaceborne measurements with emphasis on the Hot Spot”. In: *Remote Sensing of Environment* 90.2, pp. 210–220. DOI: 10.1016/j.rse.2003.12.006 (cit. on p. 25).
- Maire, Gueric le, Christophe François, Kamel Soudani, et al. (2008). “Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass”. In: *Remote Sensing of Environment* 112.10, pp. 3846–3864. DOI: 10.1016/j.rse.2008.06.005 (cit. on p. 11).
- Martens, Sven (2012). “Erste Erfahrungen bei der Herleitung ertragskundlicher Parameter fuer Einzelbaeume aus LiDAR Daten”. In: DOI: 10.13140/2.1.4661.7607 (cit. on p. 2).
- Means, Joseph E., Steven A. Acker, Brandon J. Fitt, et al. (2000). “Predicting forest stand characteristics with airborne scanning lidar”. In: *Photogrammetric Engineering and Remote Sensing* 66.11, pp. 1367–1372. DOI: 10.1016/S0034-4257(01)00290-5 (cit. on p. 13).
- Melillo, Jerry M., John D. Aber, and John F. Muratore (1982). “Nitrogen and Lignin Control of Hardwood Leaf Litter Decomposition Dynamics”. In: *Ecology* 63.3, pp. 621–626. DOI: 10.2307/1936780 (cit. on p. 11).
- Mellor, Andrew, Samia Boukir, Andrew Haywood, and Simon Jones (2015). “Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 105, pp. 155–168. DOI: 10.1016/j.isprsjprs.2015.03.014 (cit. on p. 41).
- Menze, Bjoern H., B. Michael Kelm, Ralf Masuch, et al. (2009). “A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data”. In: *BMC Bioinformatics* 10.1, p. 213. DOI: 10.1186/1471-2105-10-213 (cit. on p. 36).
- Miglani, Anshu, Shibendu S. Ray, D. P. Vashishta, and Jai Singh Parihar (2011). “Comparison of Two Data Smoothing Techniques for Vegetation Spectra Derived From EO-1 Hyperion”. In: *Journal of the Indian Society of Remote Sensing* 39.4, pp. 443–453. DOI: 10.1007/s12524-011-0103-5 (cit. on p. 27).

- Moss, D.M. and B.N. Rock (1991). “Analysis of Red Edge Spectral Characteristics and Total Chlorophyll Values for Red Spruce (*Picea Rubens*) Branch Segments from Mt. Moosilauke, Nh, Usa”. In: *[Proceedings] IGARSS'91 Remote Sensing: Global Monitoring for Earth Management*. (June 3–6, 1991). Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/igarss.1991.579470 (cit. on p. 12).
- Müller, Rupert, Manfred Lehner, Peter Reinartz, and Manfred Schroeder (2005). “Evaluation of spaceborne and airborne line scanner images using a generic ortho image processor”. In: *Proc. of High Resolution Earth Imaging for Geospatial Information, ISPRS Hannover Workshop, Commission I WG*. Vol. 5, p. 2005 (cit. on p. 23).
- Næsset, Erik (1997). “Determination of mean tree height of forest stands using airborne laser scanner data”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 52.2, pp. 49–56. DOI: 10.1016/S0924-2716(97)83000-6 (cit. on p. 6).
- Nagler, Pamela L, Y Inoue, E.P Glenn, A.L Russ, and C.S.T Daughtry (2003). “Cellulose absorption index (CAI) to quantify mixed soil–plant litter scenes”. In: *Remote Sensing of Environment* 87.2–3, pp. 310–325. DOI: 10.1016/j.rse.2003.06.001 (cit. on p. 10).
- Naidoo, L, MA Cho, R Mathieu, and G Asner (2012). “Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 69, pp. 167–179. DOI: 10.1016/j.isprsjprs.2012.03.005 (cit. on p. 2).
- Nicodemus, F.E., J.C. Richmond, and J.J. Hsia (1977). *Geometrical considerations and nomenclature for reflectance*. Vol. 160. US Department of Commerce, National Bureau of Standards (cit. on p. 24).
- Nicodemus, Kirstin (2011). “Letter to the Editor: On the stability and ranking of predictors from random forest variable importance measures”. In: *Briefings in Bioinformatics* 12.4, pp. 369–373. DOI: 10.1093/bib/bbr016 (cit. on p. 36).
- Ollinger, S. V. (2010). “Sources of variability in canopy reflectance and the convergent properties of plants”. In: *New Phytologist* 189.2, pp. 375–394. DOI: 10.1111/j.1469-8137.2010.03536.x (cit. on p. 9).

- Pal, M. (2005). “Random forest classifier for remote sensing classification”. In: *International Journal of Remote Sensing* 26.1, pp. 217–222. DOI: 10.1080/01431160412331269698 (cit. on p. 32).
- Peddle, Derek R., H. Peter White, Raymond J. Soffer, John R. Miller, and Ellsworth F. LeDrew (2001). “Reflectance processing of remote sensing spectroradiometer data”. In: *Computers & Geosciences* 27.2, pp. 203–213. DOI: 10.1016/S0098-3004(00)00096-0 (cit. on p. 23).
- Pedregosa, F, G. Varoquaux, A. Gramfort, et al. (2011). “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12, pp. 2825–2830 (cit. on pp. 16, 32).
- Pettorelli, Nathalie (2013). *The normalized difference vegetation index*. Oxford University Press. DOI: 10.1093/acprof:osobl/9780199693160.001.0001 (cit. on p. 12).
- Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press (cit. on p. 27).
- Pretzsch, Hans (2009). *Forest dynamics, growth, and yield. From Measurement to Model*. Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-88307-4 (cit. on p. 3).
- (1995). “Zum Einfluss des Baumverteilungsmusters auf den Bestandeszuwachs”. In: *Allgemeine Forst und Jagdzeitung* 166.9-10, pp. 190–201 (cit. on p. 3).
- Pu, Ruiliang (2009). “Broadleaf species recognition within situhyperspectral data”. In: *International Journal of Remote Sensing* 30.11, pp. 2759–2779. DOI: 10.1080/01431160802555820 (cit. on p. 2).
- Qi, Yanjun (2012). “Random Forest for Bioinformatics”. In: *Ensemble Machine Learning*. Springer Nature, pp. 307–323. DOI: 10.1007/978-1-4419-9326-7_11 (cit. on p. 32).
- Reitberger, Josef, Peter Krzystek, and Uwe Stilla (2008). “Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees”. In: *International Journal of Remote Sensing* 29.5, pp. 1407–1431. DOI: 10.1080/01431160701736448 (cit. on p. 13).

- Richards, John A. (2012a). “Feature Reduction”. In: *Remote Sensing Digital Image Analysis*. Springer Nature, pp. 343–380. DOI: 10.1007/978-3-642-30062-2_10 (cit. on p. 35).
- (2012b). “Supervised Classification Techniques”. In: *Remote Sensing Digital Image Analysis*. Springer Nature, pp. 247–318. DOI: 10.1007/978-3-642-30062-2_8 (cit. on p. 31).
- Rouse J. W., Jr., R. H. Haas, J. A. Schell, and D.W. Deering (1974). “Monitoring vegetation systems in the Great Plains with ERTS”. In: *NASA special publication* 351, p. 309. URL: <https://ntrs.nasa.gov/search.jsp?R=19740022614> (cit. on p. 11).
- Ruffin, Chris, Roger L. King, and Nicolas H. Younan (2008). “A Combined Derivative Spectroscopy and Savitzky-Golay Filtering Method for the Analysis of Hyperspectral Data”. In: *GIScience & Remote Sensing* 45.1, pp. 1–15. DOI: 10.2747/1548-1603.45.1.1 (cit. on p. 27).
- Sammut, Claude and Geoffrey I Webb (2011). *Encyclopedia of machine learning*. Springer Science & Business Media. DOI: 10.1007/978-0-387-30164-8 (cit. on pp. 17, 31).
- Savitzky, Abraham and Marcel JE Golay (1964). “Smoothing and differentiation of data by simplified least squares procedures.” In: *Analytical chemistry* 36.8, pp. 1627–1639. DOI: 10.1021/ac60214a047 (cit. on p. 27).
- Schaepman-Strub, G., M.E. Schaepman, T.H. Painter, S. Dangel, and J.V. Martonchik (2006). “Reflectance quantities in optical remote sensing—definitions and case studies”. In: *Remote Sensing of Environment* 103.1, pp. 27–42. DOI: 10.1016/j.rse.2006.03.002 (cit. on p. 24).
- Schlaepfer, Rodolphe and Chris Elliott (2000). “Ecological and Landscape Considerations in Forest Management: The End of Forestry?” In: *Sustainable Forest Management*. Ed. by Klaus von Gadow, Timo Pukkala, and Margarida Tomé. Dordrecht: Springer Netherlands, pp. 1–67. DOI: 10.1007/978-94-010-9819-9_1 (cit. on p. 1).
- Schläpfer, Daniel and Rudolf Richter (2014). “Evaluation of BRECOR BRDF effects correction for Hypspec, CASI, APEX imaging spectroscopy data”. In: *Proc. 6th IEEE WHISPERS*, p. 4 (cit. on p. 25).

- Schläpfer, Daniel and Rudolf Richter (2011). “Spectral polishing of high resolution imaging spectroscopy data”. In: *Proceedings of the 7th SIG-IS Workshop on Imaging Spectroscopy, Edinburgh, UK*, pp. 11–13 (cit. on p. 27).
- Schläpfer, Daniel, Rudolf Richter, and Tal Feingersh (2015). “Operational BRDF effects correction for wide-field-of-view optical scanners (BREFCOR)”. In: *Geoscience and Remote Sensing, IEEE Transactions on* 53.4, pp. 1855–1864. DOI: 10.1109/TGRS.2014.2349946 (cit. on pp. 25, 28).
- Schlerf, Martin, Clement Atzberger, and Joachim Hill (2005). “Remote sensing of forest biophysical variables using HyMap imaging spectrometer data”. In: *Remote Sensing of Environment* 95.2, pp. 177–194. DOI: 10.1016/j.rse.2004.12.016 (cit. on p. 9).
- Serrano, Lydia, Josep Peñuelas, and Susan L Ustin (2002). “Remote sensing of nitrogen and lignin in Mediterranean vegetation from {AVIRIS} data: Decomposing biochemical from structural signals”. In: *Remote Sensing of Environment* 81.2–3, pp. 355–364. DOI: 10.1016/S0034-4257(02)00011-1 (cit. on p. 11).
- Shannon, C. E. (1948). “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27.3, pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x (cit. on p. 3).
- Shi, Di and Xiaojun Yang (2016). “An Assessment of Algorithmic Parameters Affecting Image Classification Accuracy by Random Forests”. In: *Photogrammetric Engineering & Remote Sensing* 82.6, pp. 407–417. DOI: 10.14358/PERS.82.6.407 (cit. on p. 43).
- Simpson, E. H. (1949). In: 163.4148, pp. 688–688. DOI: 10.1038/163688a0 (cit. on p. 3).
- Sims, Daniel A and John A Gamon (2002). “Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages”. In: *Remote Sensing of Environment* 81.2–3, pp. 337–354. DOI: 10.1016/S0034-4257(02)00010-X (cit. on p. 9).
- Sommer, Carolin (2015). “Feature based tree species classification using airborne hyperspectral and LiDAR data for the Bavarian Forest National Park”. MA the-

- sis. Ludwig-Maximilians-Universität München Fakultät für Geowissenschaften (cit. on pp. 3, 6, 21).
- Sommer, Carolin, Stefanie Holzwarth, Uta Heiden, et al. (2016). “Feature-based tree species classification using hyperspectral and Lidar data in the Bavarian Forest National Park”. In: *9th EARSeL Imaging Spectroscopy Workshop*. Vol. 14. EARSeL eProceedings, pp. 49–70 (cit. on p. 9).
- Straub, Christoph and Barbara Koch (2011). “Estimating Single Tree Stem Volume of *Pinus sylvestris* Using Airborne Laser Scanner and Multispectral Line Scanner Data”. In: *Remote Sensing* 3.12, pp. 929–944. DOI: 10.3390/rs3050929 (cit. on p. 2).
- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis (2008). “Conditional Variable Importance for Random Forests”. In: *BMC Bioinformatics* 9.1, p. 307. DOI: 10.1186/1471-2105-9-307 (cit. on p. 36).
- Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn (2007). “Bias in random forest variable importance measures: Illustrations, sources and a solution”. In: *BMC Bioinformatics* 8.1, p. 25. DOI: 10.1186/1471-2105-8-25 (cit. on p. 36).
- Tiede, Dirk, Thomas Blaschke, and Marco Heurich (2004). “Object-based semi automatic mapping of forest stands with Laser scanner and Multi-spectral data”. In: *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36, pp. 328–333. URL: <http://www.isprs.org/proceedings/XXXVI/8-W2/TIEDE.pdf> (cit. on p. 2).
- Tolosi, L. and T. Lengauer (2011). “Classification with correlated features: unreliability of feature ranking and solutions”. In: *Bioinformatics* 27.14, pp. 1986–1994. DOI: 10.1093/bioinformatics/btr300 (cit. on p. 36).
- Tsai, Fuan and William Philpot (1998). “Derivative Analysis of Hyperspectral Data”. In: *Remote Sensing of Environment* 66.1, pp. 41–51. DOI: 10.1016/S0034-4257(98)00032-7. URL: [http://dx.doi.org/10.1016/S0034-4257\(98\)00032-7](http://dx.doi.org/10.1016/S0034-4257(98)00032-7) (cit. on p. 27).
- Vaiphasa, Chaichoke (2006). “Consideration of smoothing techniques for hyperspectral remote sensing”. In: *ISPRS Journal of Photogrammetry and Remote*

- Sensing* 60.2, pp. 91–99. DOI: 10.1016/j.isprsjprs.2005.11.002 (cit. on p. 27).
- Vauhkonen, Jari, Hans Ole Ørka, Johan Holmgren, et al. (2013). “Tree Species Recognition Based on Airborne Laser Scanning and Complementary Data Sources”. In: *Forestry Applications of Airborne Laser Scanning*. Springer Nature, pp. 135–156. DOI: 10.1007/978-94-017-8663-8_7 (cit. on p. 13).
- Vogelmann, J. E., B. N. Rock, and D. M. Moss (1993). “Red edge spectral measurements from sugar maple leaves”. In: *International Journal of Remote Sensing* 14.8, pp. 1563–1575. DOI: 10.1080/01431169308953986 (cit. on p. 12).
- Walentowski, Helge, Jörg Ewald, Anton Fischer, Christian Kölling, and Winfried Türk (2006). *Handbuch der natürlichen Waldgesellschaften Bayerns: ein auf geobotanischer Grundlage entwickelter Leitfaden für die Praxis in Forstwirtschaft und Naturschutz*. Second. Geobotanica-Verlag. URL: <http://d-nb.info/970516398> (cit. on pp. 4 sq.).
- Wang, Quan, Samuel Adiku, John Tenhunen, and André Granier (2005). “On the relationship of NDVI with leaf area index in a deciduous forest site”. In: *Remote Sensing of Environment* 94.2, pp. 244–255. DOI: 10.1016/j.rse.2004.10.006 (cit. on p. 12).
- Wang, Shuo and Xin Yao (2012). “Multiclass Imbalance Problems: Analysis and Potential Solutions”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42.4, pp. 1119–1130. DOI: 10.1109/tsmcb.2012.2187280 (cit. on p. 41).
- Wei, Yao, Peter Krzystek, and Marco Heurich (2012). “Identifying standing dead trees in forest areas based on 3D single tree detection from full waveform lidar data”. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, p. 7. DOI: 10.5194/isprsannals-I-7-359-2012 (cit. on p. 2).
- Wu, Chaoyang, Zheng Niu, and Shuai Gao (2012). “The potential of the satellite derived green chlorophyll index for estimating midday light use efficiency in maize, coniferous forest and grassland”. In: *Ecological Indicators* 14.1, pp. 66–73. DOI: 10.1016/j.ecolind.2011.08.018 (cit. on p. 10).

Yu, Shiqi, Sen Jia, and Chunyan Xu (2017). “Convolutional neural networks for hyperspectral image classification”. In: *Neurocomputing* 219, pp. 88–98. DOI: 10.1016/j.neucom.2016.09.010 (cit. on p. 69).

Zimek, Arthur, Erich Schubert, and Hans-Peter Kriegel (2012). “A survey on unsupervised outlier detection in high-dimensional numerical data”. In: *Statistical Analysis and Data Mining* 5.5, pp. 363–387. DOI: 10.1002/sam.11161 (cit. on p. 30).

Online sources

GDAL Development Team (2016). *GDAL - Geospatial Data Abstraction Library, Version 2.1.0*. Open Source Geospatial Foundation. URL: <http://www.gdal.org> (cit. on p. 14).

Köhler, Claas Henning and M. Schneider (2015). *HySpec Product Guide*. Deutsches Zentrum für Luft- und Raumfahrt. URL: <http://www.dlr.de/opairs> (cit. on pp. 8, 22).

Richter, Rudolf and Daniel Schläpfer (2016). *ATCOR-4 User Guide, Version 7.1.0*. ReSe Applications. URL: https://www.rese-apps.com/pdf/atcor4_manual.pdf (cit. on pp. 23, 25).

B Eidesstattliche Erklärung

Ich erkläre hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch bei keiner anderen Prüferin/ keinem anderen Prüfer als Prüfungsleistung eingereicht.

English translation

I hereby confirm that this thesis is my own work and that if any text passages or diagrams from books, papers, the web or other sources have been copied or in any other way used, all references including those found in electronic media have been acknowledged and fully cited.

The thesis at hand has not yet been submitted as thesis in this or a similar form.

Place, Date

Yannic Timothy Fetik