Student thesis series INES nr 458

Assessing edge pixel classification and growing stock volume estimation in forest stands using a machine learning algorithm and Sentinel-2 data

Enzo Zerega

2018
Department of
Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Enzo Zerega (2018).
Assessing edge pixel classification and growing stock volume estimation in forest stands using a machine learning algorithm and Sentinel-2 data

Master degree thesis, 30 credits in *Geomatics*
Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *January* 2018 until *June* 2018

Disclaimer

Assessing edge pixel classification and growing stock volume estimation in forest stands using a machine learning algorithm and Sentinel-2 data

---

Enzo Zerega

Master thesis, 30 credits, in *Geomatics*

Supervisor: Lars Eklundh
Department of Physical Geography and Ecosystem Science
Lund University, Sweden

Co-Supervisor: Abdulhakim M. Abdi
Department of Physical Geography and Ecosystem Science
Lund University, Sweden

Exam committee:

Jonas Ardö
Department of Physical Geography and Ecosystem Science
Lund University, Sweden

Enass Said Al-Kharusi
Department of Physical Geography and Ecosystem Science
Lund University, Sweden

ABSTRACT

Sustainable forest management requires accurate and up-to-date baseline data regarding forest structural parameters and the definition of forest stand units. A crucial component in stand characterization is the identification of tree species and to correctly define the stand boundaries. The most common procedure used for this is land cover classification using remote sensing data. However, when classifying forests, misclassification of edge areas can be substantial, yet classification studies often exclude these areas from the accuracy assessment. With the launch of the Sentinel-2 (S2) mission, which provides medium-high spatial resolution satellite imagery, global coverage and high revisit times, new methodologies for defining and estimating stand parameters are being developed. In combination with machine learning algorithms, the use of S2 data to predict forest variables has demonstrated to generate highly accurate results. Random Forest (RF) is one of these algorithms, and has become increasingly popular in environmental studies during the last decade. The combination of the multi-temporal higher resolution S2 images with the ability of the RF algorithm to detect outliers may contribute to improve the classification accuracy of forest edge areas.

This study presents a methodology based on a combination of field data and S2 multi-temporal imagery that were analyzed using the RF algorithm in southern Sweden. The aim was to perform a land cover classification to identify forest patches of three tree species (Scots pine, Norway spruce and birch), and to test the inclusion of edge pixels as training data to improve the accuracy at edge areas. For this, a segmented accuracy assessment was proposed, where the accuracy was assessed for interior, intermediate and edge areas, as well as for entire forest patches. The RF algorithm was also used to estimate growing stock volume (GSV), another important forest stand parameter. The results indicate that higher accuracies at edge areas can be obtained when edge pixels are included in the training set. Moreover, the findings describe how different allocation schemes of the training and validation data affect the results. Bands covering the Red Edge, SWIR and a narrow segment of the NIR proved to be beneficial, together with the use of multi-temporal scenes. The GSV estimation yielded inferior results but was able to distinguish the S2 bands most correlated with tree volume. The present study contributes to a better characterization of forest stands and, consequently, to facilitate the generation of forest data required by environmental scientists and the forestry sector.

Keywords: Physical Geography, Ecosystem Analysis, Random Forest, remote sensing, accuracy assessment, land cover classification, training stage, multi-temporal, edge areas

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ANN	Artificial neural networks
BOA	Bottom-of-Atmosphere
ESA	European Space Agency
GSV	Growing stock volume
H_0	Null hypothesis
H_A	Alternative hypothesis
kNN	k-nearest neighbor
MSI	Multispectral instrument
NFI	National Forest Inventory
NIR	Near Infrared
OA	Overall accuracy
OOB	Out-of-bag
PA	Producer's accuracy
RF	Random Forest
RMSE	Root mean square error
S2	Sentinel-2
SPD	Sample plot data
SVM	Support vector machines
SWIR	Short-wave Infrared
TDS	Training data set
TOA	Top-of-Atmosphere
UA	User's accuracy
VDS	Validation data set
VIS	Visible

1. INTRODUCTION

Forest ecosystems cover nearly one third of the land surface of our planet and provide several social, economic and ecosystem services (Bond et al. 2009). As their importance has been acknowledged by the scientific community, sustainable forest management is being promoted worldwide (FAO 2015). This will increase the demand for information about forest resources, especially in terms of forest structural parameters such as tree species distribution, growing stock volume (GSV), above-ground biomass, tree height and age (Köhl 2003; Nink et al. 2015). Furthermore, many of these parameters can be used to plan forestry operations more efficiently and estimate yields more accurately (Barton et al. 2017).

In forestry, as well as in biodiversity studies on forest ecosystems, the characterization and measurement of forest structural parameters is crucial for the correct definition of forest stands and their boundaries (Lindenmayer et al. 1999; Wolter et al. 2015). A forest stand can be defined as a patch within a forest ecosystem presenting relatively uniform species composition, density and structure (Pretzsch 2009). Forest stands represent the basic unit on which forest ecosystem models can be applied and therefore their identification is of great value not only for scientific research, but also for the execution of sustainable forestry practices (Gao et al. 2014).

Remote sensing has been widely used to provide forest baseline data, particularly with the use of satellite-borne multispectral sensors. Satellites provide worldwide image coverage at a low cost, yet it is not common that they present both high spatial resolution and high revisit time (Immitzer et al. 2012). In an attempt to address this issue, the recently launched Sentinel-2 (S2) mission carries a Multispectral Instrument (MSI) sensor that can image the Earth in 13 spectral channels. The spatial resolution of the acquired data varies between 10, 20 and 60 meters. It presents global coverage with a high revisit time of 5 days, and at higher latitudes this number is reduced to 2-3 days. Moreover, S2 data products are freely available, making them a promising tool for environmental studies. Several applications of S2 data have been highlighted, such as measuring changes in land cover (Iurist et al. 2016), generating disaster maps (Demir et al. 2018), crop and forest monitoring (Immitzer et al. 2016), vegetation indices estimation (Addabbo et al. 2016), and water bodies monitoring (Pahlevan et al. 2017).

A key element when defining forest stands is to identify the dominant tree species. Previous studies performing tree species classifications using S2 data have shown positive results (Fassnacht et al. 2016; Vaglio et al. 2016). Nevertheless, as the S2 mission has been in operation only since 2015, research using data from the S2 mission is still incipient. For example, few studies focus solely on classifying forest

ecosystems, and even fewer in classifying tree species and estimating forest stand parameters (Mohajane et al. 2017; Chrysafis et al. 2017). Filling this knowledge gap requires assessing effective classification schemes using S2 data under different scenarios and forest ecosystems.

Tree species can be identified by performing an image classification procedure, where all pixels of an image are categorized into land cover classes. This process consists of training, classification and accuracy assessment stages (Lillesand et al. 2015). Conventionally, the training stage is performed with a collection of spatial data that is divided into training and validation data. From several reference fields within a study area, these data distinguish and designate the land cover classes present at those fields. The reference fields should represent the heterogeneity of each of the land cover classes being analyzed and are usually collected using some randomized sampling (Wu and Sun 2008). The training data is then used to train a classification algorithm and produce a land cover map of the area, whereas the validation data is used at a later stage to perform an accuracy assessment of the results. It is often the case that training and validation data are collected from areas with a certain distance from the reference fields boundaries, preventing the inclusion of edge areas that may not be representative (Canty 2010). When medium-high spatial resolution imagery such as those from the S2 mission are used for tree species classification, intra-class variability can be significant, as forests are seldom completely homogeneous areas (Wang and Weng 2010). This can lead to higher classification errors, particularly in the bordering areas of each class (Bruzzone and Demir 2014; Vaglio et al. 2016). Therefore, and due to the importance of accurately define the boundaries of forest stands, the training stage plays one of the most relevant roles during the whole image classification procedure (Foody and Mathur 2006). Together with the inclusion of multi-temporal or ancillary data, a refined training stage can significantly decrease classification errors at bordering areas (Lillesand et al. 2015). In fact, these improvements may be more significant than those related with the selection of the classification algorithm (Khatami et al. 2016).

The exclusion of forest bordering areas from the validation data results in that the accuracy assessment at edge areas is often overlooked (Sweeney and Evans 2012). Traditionally, accuracy assessments in land cover studies do not make any reference to how accurate a classification is at edge areas as the validation data comes solely from interior sampling (Hammond and Verbyla 1996). In studies where edge sampling has been considered, results show that accuracies at edge areas are lower than at interior areas (Muller et al. 1998; Zhu et al. 2000; Sweeney and Evans 2012; Meng et al. 2016). Hammond and Verbyla (1996) refers to accuracies obtained only

from interior-sampled data as optimistically biased, where the possible low accuracies of edge areas are ignored in benefit of a higher accuracy.

On the other hand, the continuous development of the so-called machine learning algorithms has led to an increase in their use as classification algorithms in land cover classification studies, showing highly accurate results for a wide range of conditions (Gislason et al. 2006; Novack et al. 2011; Qian et al. 2014), including the classification of edge areas of vegetation classes (García-Pedrero et al. 2017). Because they have the advantage to better handle outliers (Pal 2005; Belgiu and Drăguţ 2016), training data sampled from edge areas can be used to improve the accuracy of the classification at these areas (Shuai et al. 2015).

It has been only during the last ten years that machine learning algorithms have shown a significant increase in popularity among land cover classification studies, particularly the artificial neural networks (ANN), support vector machines (SVM), and random forest (RF) algorithms (Thanh Noi and Kappas 2017; Zhu et al. 2017). These are non-parametric methods, as they can produce statistically robust results even if the data are not normally distributed, which is often the case for remotely sensed data (Belgiu and Drăguţ 2016). The RF algorithm is especially interesting, as it is not sensitive to limited training data and it does not require complex parameter tuning (Immitzer et al. 2012). Additionally, RF gives estimates of the importance of each of the explanatory variables used during the classification (Pirotti et al. 2016). This can be used to quantify the level of importance of the different bands of a satellite image used in a classification. Another advantage is that it can be used both as a classification and regression method (Liaw and Wiener 2002), making possible the estimation of continuous data such as forest stand parameters (e.g., GSV, tree height, etc.) and environmental parameters in general.

1.1 Aim

The aim of this study is to classify tree species focusing on the accuracy at edge areas and to estimate GSV values, key aspects of forest stand definition. By focusing on edge areas, it is possible to compare the results of different allocation schemes of the training data and find the one that classifies the boundaries of forest stands more accurately. In combination with S2 data, the RF algorithm was used as the classification method during the tree species classification, as well as the regression method during the GSV estimation. The study area is located in the county of Skåne, Sweden, and only Scots pine (*Pinus silvestris* L.), Norway spruce (*Picea abies* (L.) Karst.), and birch (*Betula spp*.) were selected for the analysis.

1.2 Hypotheses and research questions

For the classification, satellite images covering the four seasons of the year were acquired between 2016 and 2018. A systematic approach during the training stage is proposed, where it is intended to assess the inclusion of edge pixels as part of the training and validation data and their effect on the resulting classification accuracy at forest edge areas. A traditional accuracy assessment approach, where edge pixels are excluded from the validation data, will be compared with the proposed approach presenting how the accuracy may vary depending on the training and validation data set used. In order to achieve this, three different training sets were created from forest patches of the analyzed tree species: one considering only interior areas, one combining interior and intermediate areas, and one combining interior, intermediate and edge areas. The following hypothesis was tested:

H_0: The accuracy at edge areas obtained by using the training sets excluding edge pixels is greater than the accuracies obtained by using the training set that included them.

H_A: The accuracy at edge areas obtained by using the training sets excluding edge pixels is less than the accuracies obtained by using the training set that included them.

However, there is another way of handling the training set containing edge pixels. Instead of training the RF algorithm using pixels from interior, intermediate and edge areas all combined, the pixels can be grouped according to those areas and used to classify subclasses of each target class. The idea behind this is to perform a more refined classification by dividing the heterogenous spectral response within forest patches into three more homogeneous areas. At a final stage, the subclasses can be aggregated into one single tree species class. The division of spectrally heterogenous classes into subclasses is a common practice used to improve classification accuracies of complex land cover classes (Lillesand et al. 2015). To compare this with the previous approach, a fourth training set was created containing pixels from interior, intermediate and edge areas grouped into subclasses. A hypothesis testing was performed to assess the results, where:

H_0: The statistical significance of the accuracy at edge areas obtained by using the training set where all pixels were combined into one single class is greater than when the training set divided into subclasses was used.

H_A: The statistical significance of the accuracy at edge areas obtained by using the training set where all pixels are combined into one single class is less than when the training set divided into subclasses was used.

It is important to notice that the training pixels included in each training set increase as more areas are covered. Therefore, there is a risk that the results obtained by using the different training sets are affected by the fact that they do not contain the same number of pixels. These data are referred as unbalanced data in the literature. Previous studies have shown that classification algorithms based on trees, such as the RF algorithm, can be sensitive to unbalanced training sets (Shan et al. 2006; Blagus and Lusa 2010; Janitza et al. 2013). To address this issue, the effect of balancing these data sets, so each training set contains the same number of pixels, was also assessed. Thus, four additional balanced training sets were created from the original unbalanced sets. The hypothesis testing was formulated as:

H_0: There is no statistically significant difference between the predicted land cover values obtained using unbalanced and balanced training sets.

H_A: There is statistically significant difference between the predicted land cover values obtained using unbalanced and balanced training sets.

For the regression, it was intended to assess the correlation and band importance between reflectance values of the S2 data with forest GSV. A S2 image from the summer season of 2016 was used to estimate GSV within the study area. Dissimilar results can be found in the literature. Chrysafis et al. (2017) showed negative but significant correlation between S2 bands and GSV values at Mediterranean forests. However, Puliti et al. (2018) found sub-optimal results when using S2 at boreal forests. Therefore, more studies using S2 images to estimate GSV should be carried out on different types of forest ecosystems.

In summary, this study intends to answer the following research questions:

- How accurate is the tree species classification at edge areas and the GSV estimation using the RF algorithm with S2 data?
- Is the RF algorithm affected by unbalanced and balanced training sets?
- Is it necessary to include edge pixels in the validation data when the interest is to accurately identify the borders of each class?
- Is it possible to increase the classification accuracy at edge areas by including edge pixels in the training data?
- Can the classification accuracy at edge areas be increased by grouping the training data according to the different areas that they represent, treating them as subclasses during the classification procedure?
- What is the band combination resulting in the highest accuracy at edge areas? Are multi-temporal images useful for tree species classification?
- What are the most important bands to estimate GSV?

This study contributes to the current state-of-the-art of tree species classification focused on edge areas and the estimation of forest GSV using medium-high spatial resolution satellite images in combination with a machine learning algorithm. This can be helpful for a better definition of forest stands and to contribute to more efficient and sustainable forestry practices.

2. REVIEW OF IMPORTANT CONCEPTS

There are a number of important specifications and concepts that can help the reader to understand the data and methodology proposed in this study. These are described in the following subsections.

2.1 Sentinel-2 data

According to the Sentinel-2 (S2) technical guide issued by the European Space Agency (ESA 2018), the S2 mission consists of two satellites that orbit around the Earth, passing above both poles (polar-orbit). The satellites received the names Sentinel-2A and Sentinel-2B and were launched on June 2015 and March 2017 respectively. Both satellites are placed in the same orbit, with a distance of 180° from each other. The Earth's coverage is from latitudes 84°N to 56°S. Overall, the mission presents a swath width of 290 km combined with high revisit times. The revisit time varies from 10 days at the equator to 5 days in mid-latitudes, and at higher latitudes this time can be reduced to 2-3 days. However, the quality of the images is subjected to the atmosphere's cloud conditions (Wang et al. 2017).

The aim of the S2 mission is to monitor land surface conditions, supporting a wide range of environmental applications. The high revisit time will help modeling of vegetation seasonality (Jönsson et al. 2018), and perform land cover change detection and time series analysis (Bruzzone et al. 2017). The satellites carry a Multispectral Instrument (MSI) producing optical imagery with medium-high resolution. The MSI samples 13 spectral bands of 10, 20 and 60 m spatial resolutions (Table 1). The radiometric resolution at which the instrument can sample is 12 bit, making possible to measure reflectance values within a range of 0 to 4095.

Table 1 Spectral bands sampled by the S2 satellites. Modified from ESA (2018).

Spatial resolution (m)	Band number	Central wavelength (nm)	Spectral region	Bandwidth (nm)
10	2	490	Blue	98
10	3	560	Green	45
10	4	665	Red	38
10	8	842	NIR	145
20	5	705	Red Edge	19
20	6	740	Red Edge	18
20	7	783	Red Edge	28
20	8A	865	NIR	33
20	11	1610	SWIR Clouds / Snow / Ice	143
20	12	2190	SWIR Clouds / Snow / Ice	242
60	1	443	Coastal aerosol	27
60	9	940	Water vapor	26
60	10	1375	SWIR Cirrus	75

NIR: Near Infrared. SWIR: Short-wave Infrared.

For vegetation studies, bands within the Red Edge region of the electromagnetic spectrum present especial interest. The Red Edge is located between the Red and Near

Infrared (NIR) regions, where vegetation shows a distinguishable increase in reflectance (Filella and Penuelas 1994), as it is shown in Figure 1. In forest ecosystems, Red Edge bands of S2 products have shown to be sensitive to different types of canopy structures (Immitzer et al. 2016; Forkuor et al. 2017), presenting high correlation with forest stand parameters such as GSV (Chrysafis et al. 2017).

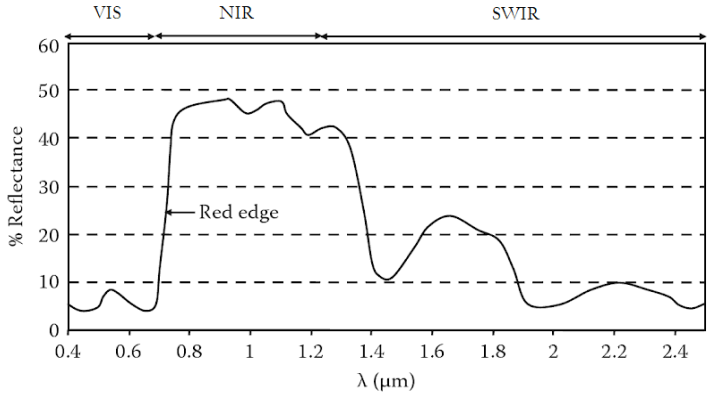

Figure 1 Green vegetation spectral response in the different regions of the electromagnetic spectrum. VIS: Visible. NIR: Near Infrared. SWIR: Short-wave Infrared. Modified from Chuvieco (2016). Copyright 2016. From 'Fundamentals of Satellite Remote Sensing' by E. Chuvieco. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc. License: 4373511175259.

The S2 data is available through the Copernicus Open Access Hub (https:// scihub.copernicus.eu/). Among the different data products that are generated within the S2 mission, only two are freely available to users: Level-1C and Level-2A. Level-1C product provides orthorectified images with Top-Of-Atmosphere (TOA) reflectance values. However, it is often necessary to work with Bottom-Of-Atmosphere (BOA) reflectance values, were TOA values are atmospherically corrected. Such data is provided in Level-2A products, which consist on orthorectified images with BOA reflectance values. Level-2A products were not available in the initial period of the S2 mission, however, Level-1C products can be atmospherically corrected using the Sen2Cor standalone package available at the ESA website (http://step.esa.int/main/third-party-plugins-2/sen2cor/).

2.2 Random Forest algorithm

Proposed by Breiman (2001), Random Forest (RF) is a non-parametric algorithm created to perform supervised classifications and regression analysis. Unlike parametric methods that assume the input data is normally distributed, non-parametric methods do not make any assumption regarding the data distribution. This has proven to be particularly beneficial when dealing with remote sensing data, which usually presents multi-modal distributions (Belgiu and Drăguț 2016). Other well-known examples are the Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms (Szuste et al. 2011). Several studies have highlighted the

advantages of the RF algorithm, including that it is user friendly as only two parameters need to be tuned (Immitzer et al. 2012), it can be used both as a classification method with categorical data and as a regression method with continuous data (Pang et al. 2006), it allows the identification of the most important variables (Cutler et al. 2007), overfitting is minimized when its parameters are correctly set (Probst and Boulesteix 2017), and has consistently shown highly accurate results (Pirotti et al. 2016; Belgiu and Drăguţ 2016).

The training data consists of a set of explanatory variables and their corresponding response variable (Antonisamy et al. 2017). In the present study, the explanatory variables are reflectance values of the different S2 bands, and the response variable is the tree species that they represent. In the case of the forest GSV estimation, the response variable is changed to continuous values of tree volume. A bootstrapping technique is applied to the input training set, where a subset of random observations is created. Typically, these observations are referred as in-bag samples and represent about two-thirds of the original training set. The in-bag samples are the data used to create the trees. The observations that are not included in the bootstrap samples represent the out-of-bag (OOB) samples and are used as validation data. OOB samples allow the computation of statistics about the accuracy of the model and to rank the importance of the explanatory variables.

The basic units of the RF model are classification or regression trees. As explained by Kingsford and Salzberg (2008), classification or regression trees are types of decision trees. Decision trees are hierarchical tree-like models that establish relationships between the response and explanatory variables of a training data set. Each relationship is contained in a node, which is split repeatedly according to relationships between all explanatory variables and the response variable. At each split a new node is created and the tree grows as long as there are valuable relationships to make the prediction, hence no more splits can be performed.

The steps of the RF algorithm are presented in Figure 2. The algorithm works by making predictions using a set of classification/regression trees. The number of trees contained in the set is defined by the user and is one of the two parameters that can be tuned. This parameter is called *ntree*. When the desired *ntree* value is defined, different in-bag samples from the training data are created by bootstrapping. The excluded observations of each in-bag sample are set aside to form the OOB samples.

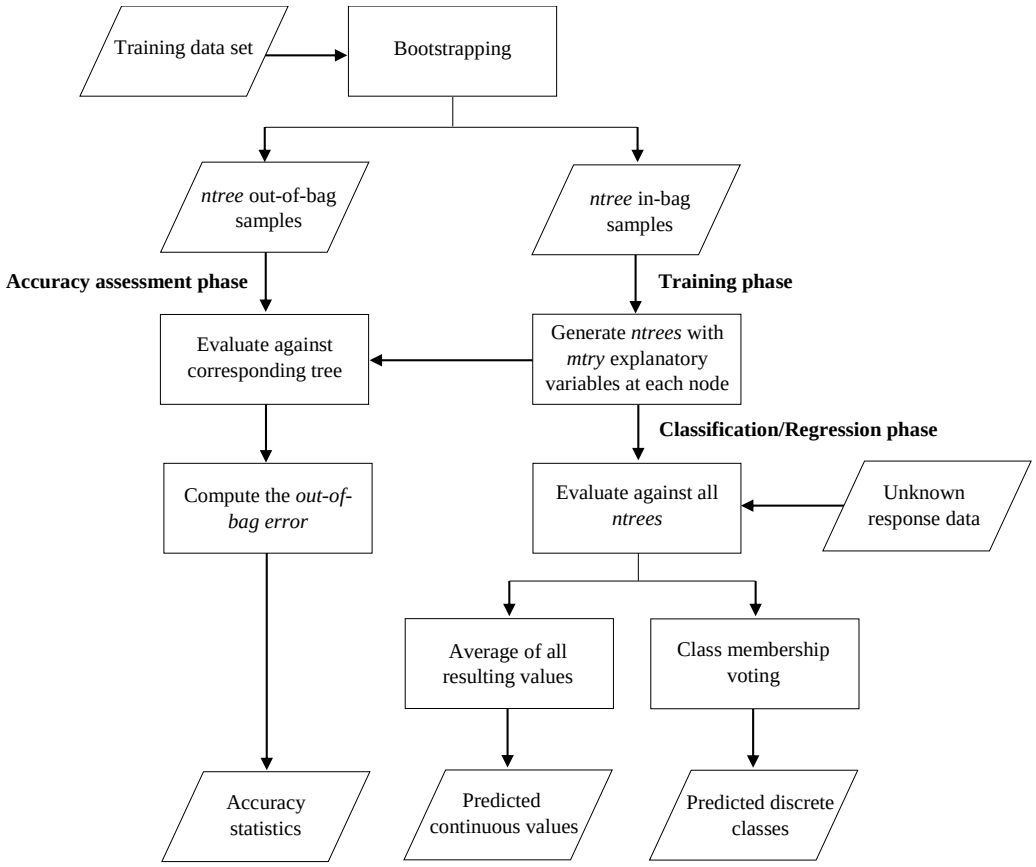

Figure 2 Flowchart of the Random Forest algorithm process.

The training stage starts when the in-bag samples are used to generate the trees. Opposed to traditional decision trees, each time the relationships between explanatory and response variables are being analyzed, only a random subset of the explanatory variables is considered. In other words, each time a node is split, not all explanatory variables are used. This procedure allows assessing the accuracy when particular variables were not considered in a tree, where an error estimate can be recorded and used to measure variable importance. The number of explanatory variables considered to make the split at each node is the remaining parameter that can be defined by the user. This parameter is referred as *mtry*. When no tuning is performed, these two parameters have proven to produce robust predictions using an *ntree* value of 500 and an *mtry* value equal to the square root of the number of explanatory variables in classification models. In regression models, the suggested number of *mtry* is the number of explanatory variables divided by 3 (Liaw and Weiner 2002).

When the algorithm has been trained and all trees of the RF model have been generated, the prediction can be made (Classification/Regression phase). A prediction is performed for each observation from a data set containing data with unknown response values. As described by Belgiu and Drăguţ (2016), each observation is evaluated against all the trees of the model and, when the process finishes, in the

classification case, each tree votes for a class membership. The class resulting with the highest number of votes is the one assigned to the observation. In the regression case, all trees predict continuous values that are averaged to obtain a single final value.

Finally, the accuracy statistics of the results are obtained by predicting response values of the OOB samples. Because the OOB samples come from the training data, the real response value is known. Thus, the accuracy can be estimated by assessing the proportion of correctly classified values comparing the predicted with the real response values. The proportion of misclassified values is commonly referred as OOB error. Similarly, when a regression is performed, the Root Mean Square Error (RMSE) can be computed as a measurement of the accuracy.

2.3 Accuracy assessment for land cover classification

When working with spatial data, a set of predictions obtained using the RF algorithm can be used to produce a map. Maps describing thematic or continuous spatial variables often present an accuracy assessment to state the quality of the map. This facilitates decision making when the information displayed needs to fulfill quality standards. Additionally, it allows to compare the results of different maps and the methodologies on which they were generated (Congalton and Green 2008). Foody (2002) reviewed how traditionally land cover classification maps are assessed in terms of their accuracy. He highlighted the risk of using the confusion matrix and the kappa coefficient to estimate statistical significance when the assumptions on which they rely are not satisfied. For example, the assumption of independence that states field observations must be independent is usually not fulfilled in these studies. Data dependency occurs when the same set of samples is used as the training data during the classification procedures being compared. Furthermore, many studies comparing land cover classification maps do not use statistical parameters to outline the level of significance of the results.

In the case that field samples are not independent, to test if the difference between land cover classifications is statistically significant, the McNemar's test can be used (Agresti 2006). Moreover, due to its non-parametric nature, the McNemar's test works well in remote sensing based classifications (Foody 2004; Momeni et al. 2016; Akar 2017). The test is performed after creating a 2 by 2 confusion matrix of the difference and agreement between two classifications (Table 2).

Table 2 Confusion matrix of 2 by 2 used to compute the McNemar's test of statistical significance. The variables in the matrix represent the difference and agreement between two classifications. Modified from Foody (2004).

	Classification 2		
Classification 1	Correct	Incorrect	∑
Correct	f_{11}	f_{12}	
Incorrect	f_{21}	f_{22}	
∑			

The McNemar's test is based on a chi-square distribution (X^2), and is calculated using the following equation (Foody, 2004):

$$X^2 = \frac{(|f_{12} - f_{21}| - 1)^2}{f_{12} + f_{21}} \quad (1)$$

where f_{12} represents the number of pixels that were correctly classified in Classification 1 but misclassified in Classification 2, and f_{21} is the number of pixels that were correctly classified in Classification 2 but misclassified in Classification 1.

The chi-square distribution is for two-sided tests, which means that only significant difference can be tested. In order to test the direction of the difference between two classifications, a one-sided test must be performed. For such cases, the McNemar's exact conditional test can be used, where the one-sided *p-value* (Fagerland et al. 2013) is obtained with equation 2:

$$p - value = \sum_{f_{12}=0}^{\min(f_{12},f_{21})} \binom{n}{f_{12}} \left(\frac{1}{2}\right)^n \quad (2)$$

where f_{12} represents the number of pixels that were correctly classified in Classification 1 but misclassified in Classification 2, and f_{21} is the number of pixels that were correctly classified in Classification 2 but misclassified in Classification 1, and $n = f_{12} + f_{21}$. Consequently, with the one-sided test it is possible to state with statistical significance if the accuracy of one classification is greater or less than the accuracy of another classification.

3. MATERIALS

3.1 Study area

This study is focused on higher latitude forest ecosystems in a 30x30 km study area located in the northeastern part of the county of Skåne, Sweden (55.56°N - 56.14°N, 13.31°E -14.25°E). Skåne is Sweden's southernmost county, covering a land surface area of 11,340 km^2. It has a warm temperate climate, with an average annual temperature of 7°C, oscillating between 0° and 16°C during winter and summer periods respectively (Blennow et al. 1999). The average annual precipitation is 661 mm $year^{-1}$, with minimum values close to 460 mm $year^{-1}$ and maximum values close to 960 mm $year^{-1}$ (Bärring et al., 2003).

Nearly 90% of the land surface comprises vegetation, making it an interesting area for vegetation studies. Agricultural fields represent 53% of the total land surface, dominating the southwestern part of the county, whereas forest ecosystems, representing 37%, dominate the northeastern part (Figure 3). Skåne is part of the temperate broadleaf and mixed forest biome, characterized by a mixture of frost-resistant deciduous and coniferous forests. The majority of the forest land is managed, and the most dominant tree species are Norway spruce (*Picea abies* (L.) Karst.), Scots pine (*Pinus sylvestris* L.), birch (*Betula spp.*), beech (*Fagus sylvatica*) and oak (*Quercus robur*) (SLU 2018). In Sweden, Norway spruce, Scots pine and birch are the species with most productive value (SLU 2017), and there is a large amount of data describing them, therefore, this study will be focused on these.

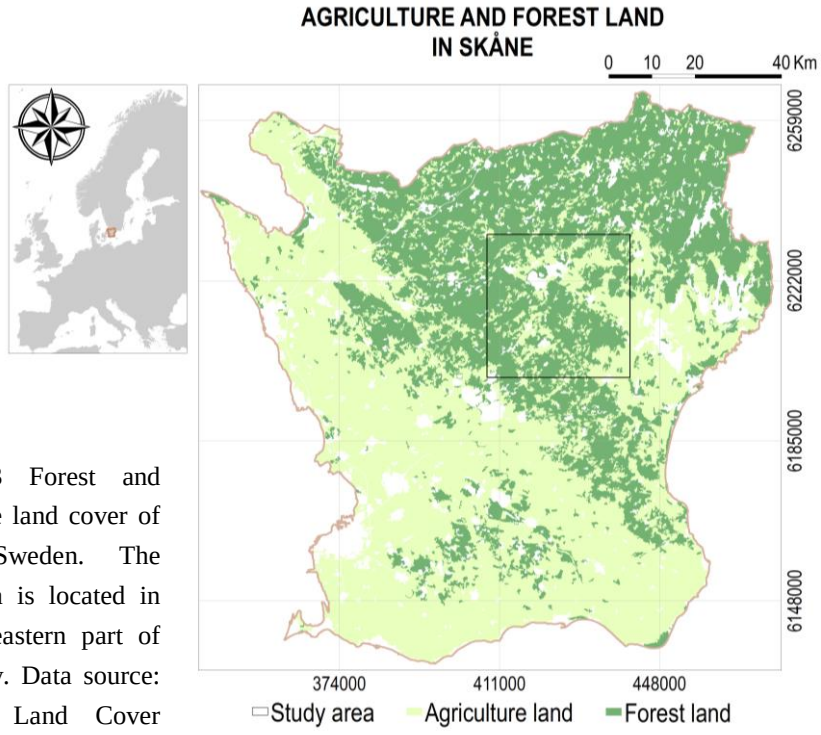

Figure 3 Forest and agriculture land cover of Skåne, Sweden. The study area is located in the northeastern part of the county. Data source: CORINE Land Cover (CLC) 2012.

3.2 Data

Sentinel-2 imagery

The Sentinel-2 (S2) images were downloaded from the Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/). According to the S2 mission tiling system Skåne is covered by four tiles, from which tile 33VVC covers all the study area extent. S2 tiles are 100x100 km^2 ortho-images in *SENTINEL-SAFE* format and projected in UTM/WGS84 coordinate system. The number of scenes selected for this study was restricted by the quality of the images, mainly affected by cloud and snow cover at this latitude. For each season of the year, and within the sensing period of 2016-2018, one cloud- and snow-free scene was selected for the classification procedure (Figure 4). Spring, fall and winter images were available at Level-2A, providing orthorectified Bottom-Of-Atmosphere (BOA) reflectance values. The summer image was only available at Level-1C, with Top-Of-Atmosphere (TOA) reflectance values. To work with BOA reflectance values, the image was atmospherically corrected using the Sen2Cor standalone package. The 60 m spatial resolution bands (bands 1, 9 and 10) were excluded from the analysis as their resolution is too coarse to capture differences in reflectance between forest interior and edge areas.

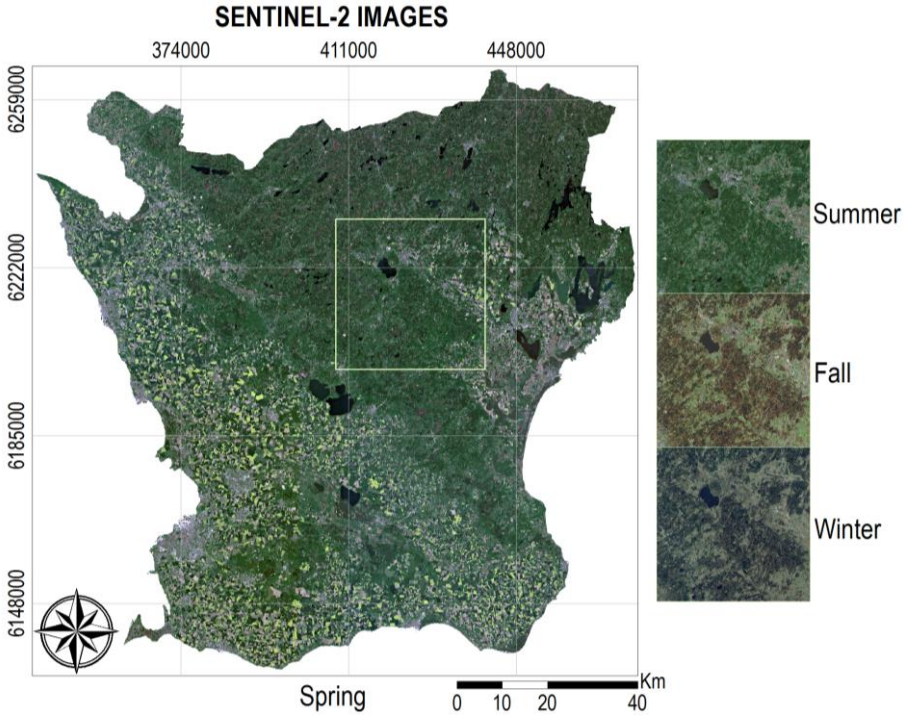

Figure 4 Sentinel-2 true color composites (432) of the different scenes used in the study. Images correspond to Level-2A products. Source: ESA Sentinel-2 image. Sensing dates: 2016-07-21 (summer); 2017-05-27 (spring); 2017-11-13 (fall); and 2018-01-07 (winter). Tiles: 33VUC, 33VVC (study area), 33UUB and 33UVB. Retrieved from https://scihub.copernicus.eu/.

Forest land cover mask

This study intended to assess and improve classification of edge areas of pine, spruce and birch patches, together with GSV estimation. Thus, non-target classes were excluded from the classification and regression analysis. This was achieved by using the Forest Map 2010 mask from *Sveriges lantbruksuniversitet* (SLU) that contains spatial data regarding forest land distribution. It is in the form of a 25 m spatial resolution raster image projected in RT 90 2,5 gon V. coordinate system. The map was created with the k-nearest neighbor (kNN) method using field data from the Swedish National Forest Inventory (NFI) and satellite images from 2008, 2009 and 2010. The first version of this map (kNN-Sweden 2000), obtained using the same methodology, had an estimated geometrical error of one pixel (Reese et al., 2003).

Training and validation data

The field data used both during the training stage and the accuracy assessment of the classification process was provided by the Swedish National Forest Inventory (NFI). The available Sample Plot Data (SPD) is a subset data set from field samples of forest variables. Each observation is collected at circular sample plots of about 154 m^2, where forest stand parameters such as tree species, volume and height are measured. The data set gathers observations collected between the years 2007 and 2016 and presents a positional accuracy of 5-6.5 m.

In order to corroborate the NFI field samples and to digitize training fields, several 0.25 m resolution orthophotos from 2016 covering the study area were acquired from *Lantmäteriet*. The 5x5 km images were projected in SWEREF 99 TM coordinate system and present an expected error of 1 m. A total of 36 orthophotos in *geoTIFF* format were needed to cover the study area.

4. METHODS

4.1 Sentinel-2 image pre-processing

All Level-2A images were processed using the Sentinel-2 Toolbox available with the Sentinel Application Platform (SNAP) 6.0.0, downloaded from ESA's website (http://step.esa.int/main/download/). Figure 5 details the pre-processing steps where the original S2 images were handled to generate a raster layer stack suitable to use for the classification/regression. The process started by applying an atmospheric correction to the Level-1C product (summer image) using the Sen2Cor standalone package. The correction generated a Level-2A product as the outcome. All Level-2A images were then resampled to the same spatial resolution so geoprocessing functions could be applied. To avoid data loss, the 20 m bands were resampled to the highest spatial resolution available, which was 10 m. The resampling was made using the nearest neighbor interpolation method and the results were exported to *geoTIFF* files. Using ArcMap 10.5.1, all resulting images were masked using the SLU forest map and the study area extent. Everything was reprojected from the original coordinate system UTM/WGS84 to SWEREF 99 TM. Finally, a 40 bands raster layer stack was created using the *composite bands* tool.

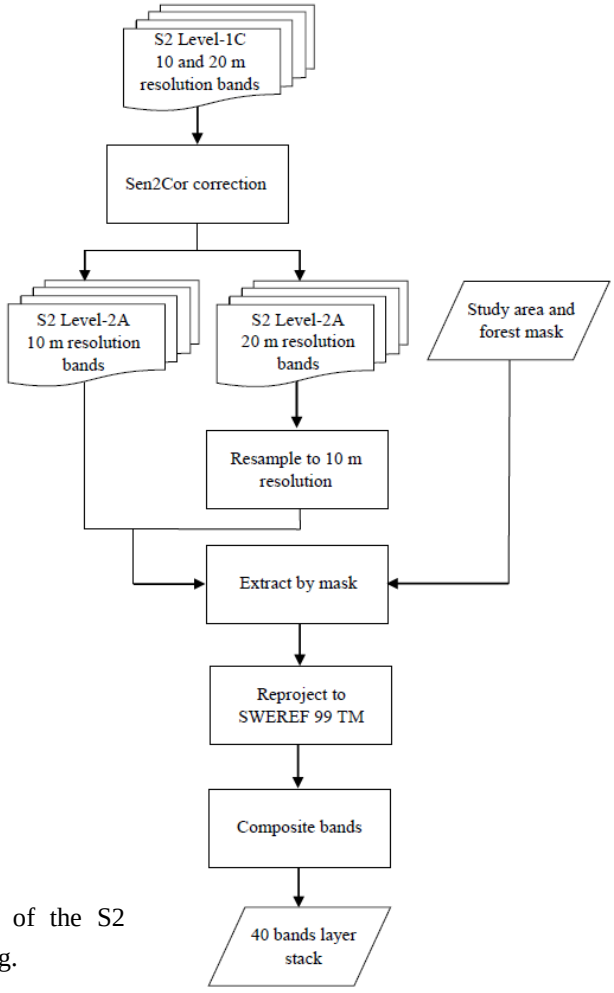

Figure 5 Flowchart of the S2 images pre-processing.

4.2 Forest edge areas classification

The classification of forest edge areas can be divided into training and classification stages. To assess the effect of including edge pixels as training data during the training stage, training sets containing edge pixels were created and compared against training sets that did not include them. Additionally, a comparison was made between using balanced and unbalanced training sets. For the classification, the *mtry* and *ntree* RF parameters were set using the default values, and the best band combination to use for the classification was assessed. Details of each stage are presented below.

4.2.1 Training stage

(i) Definition of the training areas

In ArcMap 10.5.1, using the NFI field data set, for each tree species several training locations were identified within the study area (Figure 6). Bands from different dates of the layer stack were evaluated at these locations to check if the forest patch had been harvested at the time the S2 image was acquired. Once a homogeneous forest patch was identified, using 0.25 m resolution orthophotos, the forest patch was manually digitized, including the patch edges, and labeled with its corresponding species name. When all training areas were digitized, the set of polygon features was saved to the ESRI shapefile format.

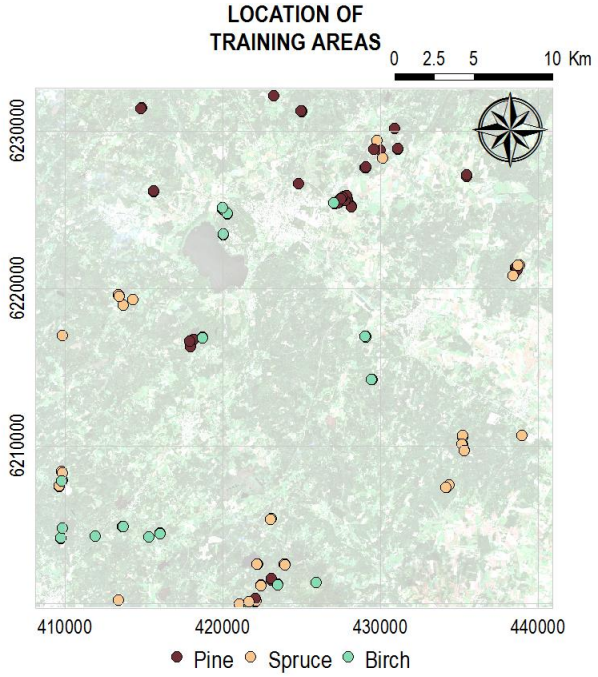

Figure 6 Location of training areas. The points represent the location, not the shape or area covered by each area. Background image: ESA Sentinel-2 image, true color composite (432). Tile: 33VVC. Acquisition date: 2016-07-21. Retrieved from https://scihub.copernicus.eu/

The training areas contained interior, intermediate and edge areas of the forest patches (Figure 7a). To compare and assess the outcome of the inclusion of edge pixels in the

classification of forest edge areas, new sets of training areas were generated where edge areas were not included. Two sets were created by using the geoprocessing *buffer* tool. A buffer of –20 and –40 m was applied, creating one set of training areas where all edge areas from the outer 20 m of the forest patch were removed (Figure 7b), and another where the outer 40 m were removed (Figure 7c).

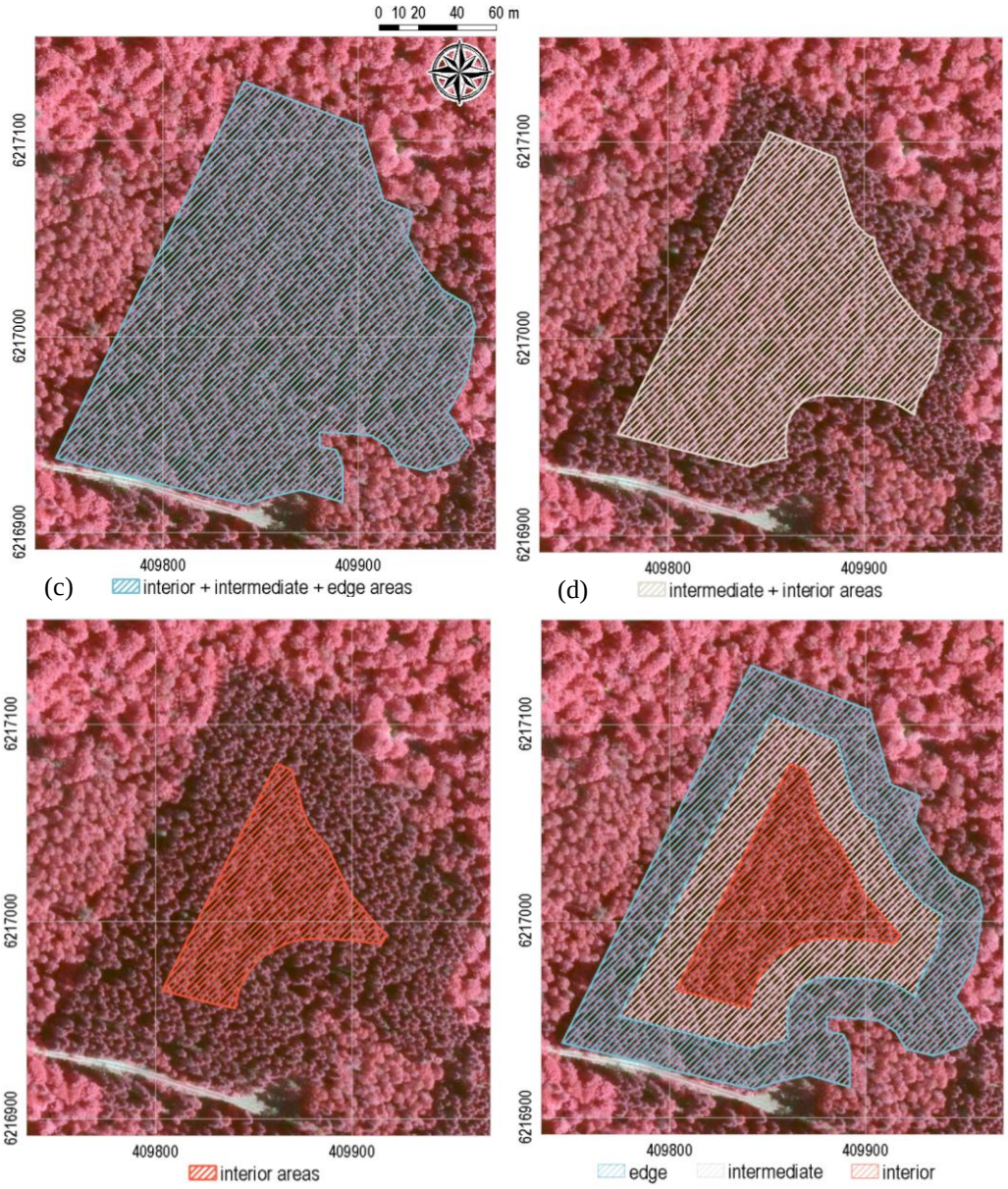

Figure 7 One example training area of a spruce patch showing the area covered according to four different training area definitions. (a) Training area covering the entire patch; (b) Training area covering intermediate and interior areas; (c) Training area covering interior areas; (d) Training area covering the entire patch divided by subclasses. Background image: Ortofoto-IR, 0.25 m. ©*Lantmäteriet* (2016).

With the described methodology, all areas inside a training field are used to train one single class in the classification algorithm. However, edge pixels may represent areas with spectral responses that deviate at some degree from the response at intermediate or interior areas of the forest patch. When land cover classes present such heterogenous response, the class can be divided into subclasses that represent more homogeneous areas. This may increase the classification accuracy at each area of the forest patches (Lillesand et al. 2015). The subclasses can be aggregated at a final stage into one single class. To compare this with the previous sets of training areas, a fourth set was created covering interior, intermediate and edge areas, but grouped into subclasses (Figure 7d). This was achieved merging all previous sets of training areas using the *merge* tool.

(ii) Training data sets creation

The sets of training areas were used to create training data sets containing the S2 bands reflectance values at each training patch, with their corresponding tree species label. Note the difference between the concepts set of training areas and training data set. As mention before, a set of training areas contains manually-digitized feature polygons stored in ESRI shapefile format. On the contrary, the training data set is a table with the values of each pixel from the 40 bands layer stack located at the training areas. The pixel values of each band represent the explanatory variables to be used in the RF algorithm. The table also includes the tree species occurring at each pixel, which corresponds to response variable. In other words, the training data set contains a series of observations (pixels) with the values of the explanatory and response variables, whereas the set of training areas represents the spatial location of each of these observations.

To generate the training data sets, the polygon feature sets were rasterized to 10 m resolution *geoTIFF* files with the *polygon to raster* conversion tool in ArcMap. At this point, a set of validation pixels were set aside. Stratified random sampling was performed using the *sampleStratified()* function from the *raster* package in R, to include the same proportion of validation data from all classes and patch areas. Thirty percent (30%) of pixels were sampled from the set of training areas containing interior, intermediate and edge areas grouped into subclasses. The validation pixels were removed from all sets of training areas using the *erase* tool in ArcMap. This tool requires vector data, so everything was converted again into ESRI shapefiles with the *raster to polygon* tool.

The training data sets were finally generated using several functions from the *raster* package in R (an example of the R script can be found in Appendix 2). For each of the training pixels, the corresponding values from all S2 bands in the raster layer stack

were extracted and stored into a table. For every pixel, the corresponding tree species value was also extracted. The data sets were named according to the allocation of the pixels within forest patches. Figure 8 summarizes the first steps of the training stage using a spruce training area as an example. TDS stands for training data set. The subscripts 1, 2 and 3 indicate interior, intermediate and edge pixels, respectively. The '+' symbol indicates that all pixels from different areas were combined, and the ',' symbol indicates that the pixels from the different areas were grouped into subclasses.

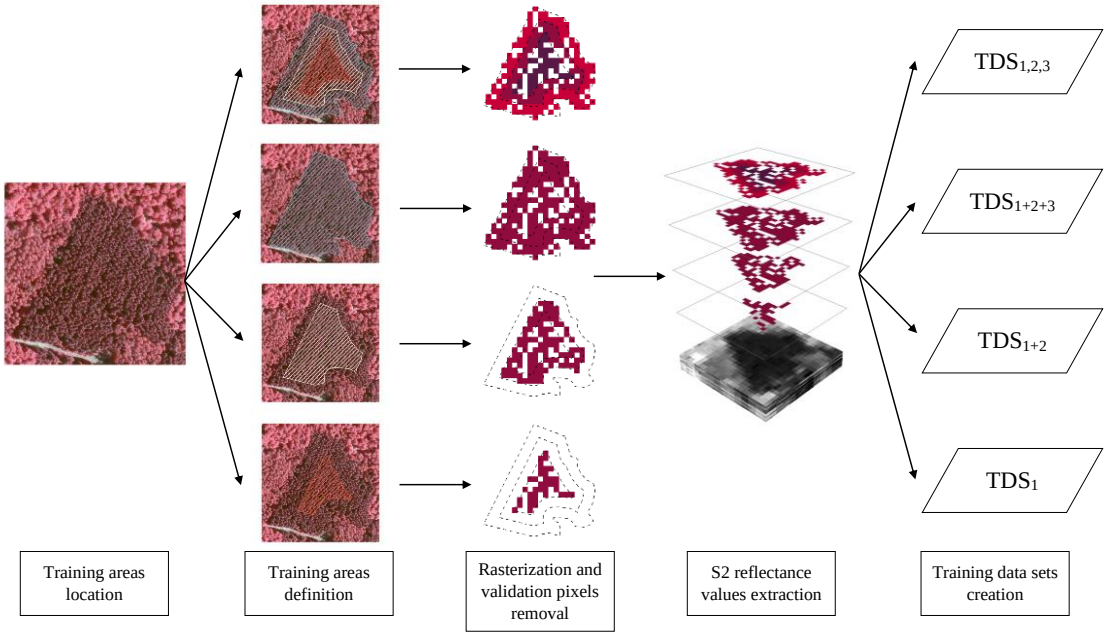

Figure 8 First steps of the training stage using a spruce training area as an example. The allocation of training pixels is presented for each training approach. TDS stands for training data set. The subscripts 1, 2 and 3 represent interior, intermediate and edge pixels respectively. The '+' symbol indicates that all pixels from different areas were combined. The ',' symbol indicates that the pixels of the different areas were grouped into subclasses. Background image: Ortofoto-IR, 0.25 m. ©*Lantmäteriet* (2016).

(iii) Unbalanced and balanced training sets

The classification was performed using the training sets described above. In each case, the number of pixels contained in each class or subclass varied depending on the number of training fields and their sizes. As classification results can be affected by the different sampling sizes of each class, and to assess if the RF algorithm was sensitive to unbalanced training data, new balanced training data sets were created. The data sets were balanced performing a stratified random sampling for each class extracting the same number of pixels every time. The chosen value was extracted from the class with the lowest pixel count from all training sets. This was performed using the *sampleStratified()* function in R.

(iv) Training the algorithm

The RF algorithm was trained using the *train()* function from the *caret* package in R. The function was set to perform a RF classification, for which the *randomForest* package was also needed (an example of the R script can be found in Appendix 2). The RF parameters *mtry* and *ntree* were set using the default values (*mtry* = square root of the number of bands and *ntree*= 500). Some studies have shown that tuning these parameters results in negligible improvements on the classification accuracy (Liaw 2002; Díaz-Uriarte and Alvarez de Andrés 2006; Immitzer et al. 2012; Belgiu and Drăguţ 2016). Moreover, when training the algorithm using the different training sets as the training data, tuning the parameters will likely produce different *mtry* and *ntree* values. As a consequence, classification accuracies may vary between each other due to differences in these parameters, affecting the measurement of the variation in accuracy produced when different training data sets are used.

4.2.2 Classification stage

To speed up the process, the classification procedure was done using the *clusterR()* function from the *raster* package in R, which allows parallel processing. The S2 layer stack was used as the unknown response input data for which the predictions were made. The predicted response values were saved to *geoTIFF* files to produce classification maps.

4.2.3 Best band combination

In R, the *varImp()* function from the *caret* package was used to obtain the importance of each explanatory variable (band). The variable importance is estimated by recording the predicted accuracy of the RF algorithm at each tree. The function analyzes the increase of the prediction error when OOB samples of each variable are permuted while the remaining variables are left unchanged (Liaw and Wiener 2002). Following Díaz-Uriarte and Alvarez de Andrés (2006), to obtain the best combination of explanatory variables using the RF algorithm, 20% of the variables with the lowest importance were discarded after running the algorithm using all the available bands. The process was repeated until all bands were discarded. The accuracy of the result was assessed at each run, and the band combination presenting the highest accuracy at edge areas was selected to generate the final tree species map. Because the process started with 40 bands, this resulted in a series of 13 classifications (Table 4).

Table 3 Number of bands used in each of the 13 classifications. 20% of the bands were discarded at each run (values were rounded up).

Classification number	Number of bands
1	40
2	32
3	25
4	20
5	16
6	12
7	9
8	7
9	5
10	4
11	3
12	2
13	1

4.3 Segmented accuracy assessment

4.3.1 Validation data sets creation

The validation data sets were created following the same process used for the creation of the training data sets (subsection 4.2.1). The first set of validation pixels included pixels from all areas of forest patches. However, to assess the classification accuracy at edge areas in particular, a validation set containing solely edge pixels was required. In the same way, the accuracies at intermediate and interior areas were needed for measuring how the accuracy changes depending on which area is considered in the validation set. Therefore, three additional sets of validation pixels were generated for each area of the forest patch. This methodology is proposed as a manner to perform a segmented accuracy assessment. The process is summarized in Figure 9. VDS stands for validation data set. The subscripts 1, 2 and 3 represent interior, intermediate and edge pixels respectively. The '+' symbol indicates that all pixels from different areas were combined.

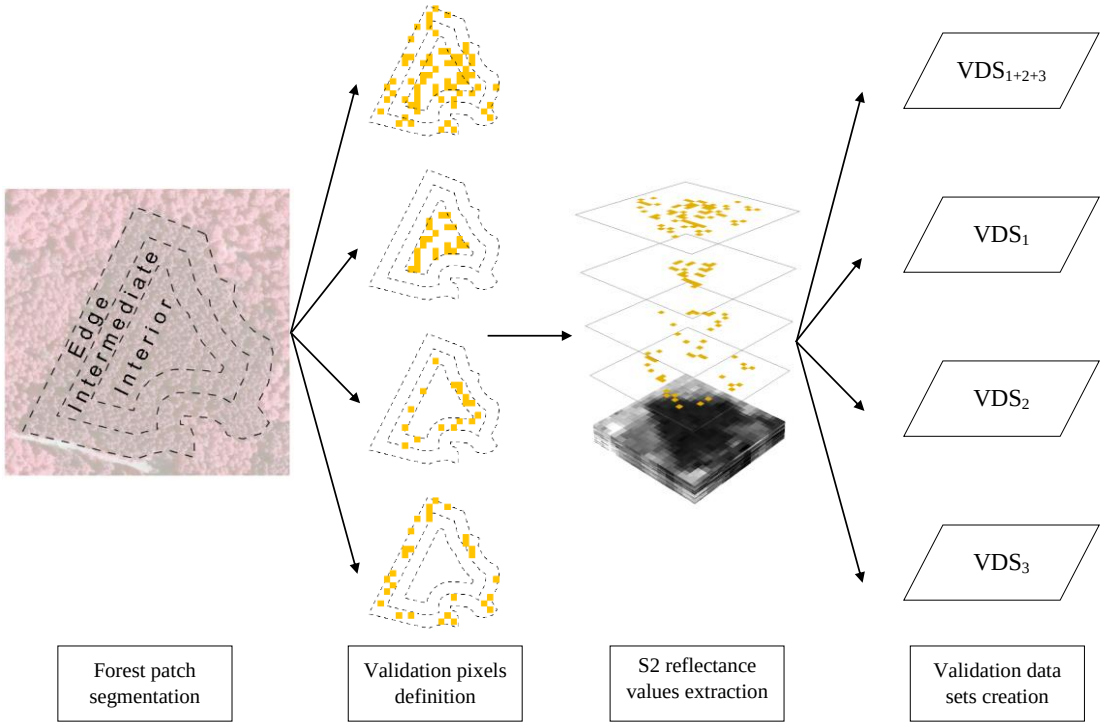

Figure 9 Validation data sets creation using a spruce patch as an example. VDS stands for validation data set. The subscripts 1, 2 and 3 represent interior, intermediate and edge pixels respectively. The '+' symbol indicates that all pixels from different areas were combined. Backround image: Ortofoto-IR, 0.25 m. ©*Lantmäteriet* (2016).

As described above, two different accuracy assessment approaches are presented. The first one was using the VDS_{1+2+3}, which contains a combination of training data from all areas of the forest patches. The advantage of using this approach is that it is not biased towards any specific area. However, with this procedure it is not possible to measure how accurate the classification was at edge or any particular area of the classified patches. Therefore, a second, more thorough approach, was used to perform a segmented accuracy assessment. This was done by using the VDS_1 (interior areas), VDS_2 (intermediate areas) and VDS_3 (edge areas).

For every validation data set, a confusion matrix was created, and the kappa coefficient of each classification result was extracted. From the confusion matrix, the Producer's and User's Accuracies (PA and UA respectively) at class level were estimated, together with the Overall Accuracy (OA) of the resulting land cover maps.

4.3.2 Statistical significance using the McNemar's test

During the first classification, the RF algorithm was run using the 40 bands as the explanatory variables. This was done for the four unbalanced and four balanced training sets. At this step, the McNemar's statistical test for categorical and non-parametric data was used to assess if the difference of the obtained results between using balanced and unbalanced training sets was statistically significant at edge areas. The test works for dependent data, which is the case of the data used in this study (all training sets come from the same set of pixels). As it is based on a chi-square distribution, the X^2 statistic was computed and a *p-value* was generated for each comparison. The confidence levels were set to 95%.

The *p-values* were used to perform a hypothesis testing, where H_0 assumed that there is no significant difference between the predicted land cover values obtained using unbalanced and balanced training sets. The H_A assumed that the difference between the results obtained using both training sets is indeed statistically significant (two-sided test). If the H_0 was rejected in favor of the H_A, the classification procedure would continue using only the balanced training sets, as the difference in the pixel count of each unbalanced training set would be altering the accuracies of the results. If H_0 was not rejected, and the class imbalance did not present any effect in the accuracy, the unbalanced training sets would be used for the subsequent steps of the study.

The second hypothesis testing was done after all 13 classifications were performed using each training set, and the classification with the band combination producing the highest accuracy at edge areas was found. For this classification, the H_0 assumed that the accuracy at edge areas obtained using the training set excluding edge and intermediate pixels (TDS_1) is greater than the accuracies obtained by using the training sets including them (TDS_{1+2}, TDS_{1+2+3} and $TDS_{1,2,3}$). The H_A assumed that the accuracy at edge areas obtained by using the TDS_1 is less than the accuracies obtained by using the TDS_{1+2}, TDS_{1+2+3} and $TDS_{1,2,3}$ (one-sided test). As the chi-square distribution is for two-sided tests, the McNemar's exact conditional test was used to compute the one-sided *p-values*.

It was also intended to know if there was a statistically significant improvement on the accuracy at edge areas when interior, intermediate and edge pixels were grouped into subclasses in the training set ($TDS_{1,2,3}$). This was compared against the results obtained by using the training set where pixels from all areas were combined into single classes (TDS_{1+2+3}). In this case, the H_0 argued that there is statistically significant evidence that the accuracy at edge areas obtained when using the TDS_{1+2+3} is greater than when the $TDS_{1,2,3}$ is used. The H_A argued that there is statistically

significant evidence that the accuracy obtained when using the TDS_{1+2+3} is less than the accuracy obtained when the $TDS_{1,2,3}$ is used. Once more, the McNemar's exact conditional test was used to calculate the one-sided *p-value*.

4.4 Growing stock volume estimation

The RF algorithm was tested as a regression model to estimate GSV. Forest growing stock volume (GSV) is defined as the volume of living trees with a certain diameter at breast height per m^3 ha^{-1} (FAO 2004). Using the NFI field samples, GSV was estimated using the S2 data. In order to identify useful bands, the Pearson's correlation coefficient (r) and coefficient of determination (R^2) were obtained between GSV and the bands used to perform the regression. However, in many cases, the acquisition date of the NFI samples did not match acquisition date of the S2 images. Tabulated mean annual volume increment by tree species for the Skåne county was used when a volume correction was needed (Table 5). Additionally, the *varImp()* function was called in R to obtain the ranking of the bands importance and to compare it with the r and R^2 values computed before the regression.

Table 4 Mean annual volume increment of Scots pine, Norway spruce and birch in Skåne, Sweden. Modified from SLU (2016).

Parameter	Pine	Spruce	Birch
Annual increase in volume (m^3 ha^{-1})	6.1	12.4	4.0

4.5 Regression accuracy assessment

Due to the lack of enough NFI sample points available for the study area, the accuracy assessment was made by analyzing the OOB samples generated at each tree of the RF model. The Root Mean Square Error (RMSE) was extracted from the results. To reduce the error associated with the extrapolation in time of these parameters, only the summer image was used for the regression. This is the image with the closest acquisition date to the dates of when the NFI observations were collected.

The NFI samples also included tree height measurements. Estimating tree height was not part of the aim of this study, but since these data were available, a tree height estimation was performed using the same procedure as the one described for the GSV estimation. Tree height estimation can be done combining remotely sensed optical and altimetry data (Hansen et al. 2016), yet in this case only optical data from the S2 images were used for the regression. Thus, the results are not expected to be accurate. However, it can be interesting for the reader to know which S2 bands are correlated with tree height. These findings in combination with altimetry data could be used in future studies to estimate stand height more accurately. The results are presented in Appendix 3.

4.6 Non-target classes removal

The forest mask applied to all S2 bands of the raster layer stack was not highly accurate at pixel level. In some cases, it also included non-forest classes (e.g., artificial surfaces, water bodies and grasslands). Moreover, the forest mask does not include information regarding tree species distribution. Therefore, in all areas correctly designated as forest areas, tree species other than the three species of interest were included. It is suggested that all occurring classes in image land cover classifications should be considered, as all pixels end up being classified (Lillesand et al. 2015). Since this study is only valid for pine, spruce and birch classes, it was necessary to perform an additional land cover classification. The classification was carried out to identify all non-target land cover classes that were present in the masked S2 images. The output map was reclassified into target and non-target areas. The target areas were extracted and used as a mask to remove all non-target classes from the tree species classification and GSV regression results. This was achieved by using the *extract by mask* tool in ArcMap 10.5.1.

The classification was performed executing the RF algorithm with *mtry* and *ntree* default values. To train the algorithm, a training set was created including observations from all non-target classes present in the masked images. For non-forest classes, the observations were generated by using the 0.25 m resolution orthophotos to digitize and label training areas for each class. The locations of these areas were determined by visual interpretation of the orthophotos. For forest classes, both the NFI field data and the orthophotos were used to locate areas with non-target tree species. The non-target species were deciduous tree species and include oak, beech and aspen. Using the *sampleStratified()* function in R, 30% of pixels were sampled and removed from the training set as validation data to assess the quality of the results.

5. RESULTS

5.1 Unbalanced vs balanced training sets

There was a clear difference in the pixel count of unbalanced training sets (Table 5). The highest pixel count was 28,811 in the TDS_{1+2+3} and $TDS_{1,2,3}$. The lowest pixel count was 7,758 in the TDS_1, particularly in the birch class, with 2,343 pixels. This was the chosen value to balance each class of all training sets. Thus, the balanced training sets resulted in a pixel count of 7,029 pixels each.

Table 5 Unbalanced and balanced training sets pixel count by class.

Training set	Class	Unbalanced training set pixel count	Balanced training set pixel count
TDS_1	Pine	2809	2343
	Spruce	2606	2343
	Birch	2343	2343
	Total	7758	7029
TDS_{1+2}	Pine	6055	2343
	Spruce	5321	2343
	Birch	4474	2343
	Total	15850	7029
TDS_{1+2+3}	Pine	10963	2343
	Spruce	9435	2343
	Birch	8413	2343
	Total	28811	7029
$TDS_{1,2,3}$	Pine	10963	2343
	Spruce	9435	2343
	Birch	8413	2343
	Total	28811	7029

After the first run of the algorithm, using the 40 bands as the explanatory variables, the accuracy at edge areas was assessed (Table 6). The accuracies at edge areas obtained with unbalanced training sets show in three of the four training sets higher OA and kappa coefficient values than when the balanced training sets were used, particularly when the training sets included pixels from all areas of forest patches (TDS_{1+2+3} and $TDS_{1,2,3}$). When unbalanced and balanced training sets were tested using the McNemar's statistical test, the obtained difference between the unbalanced TDS_{1+2}, TDS_{1+2+3} and $TDS_{1,2,3}$, and their corresponding balanced data sets proved to be statistically significant with *p-value* < 0.001. The only exception was between the unbalanced and balanced TDS_1, probably because both contain a similar number of pixels (7,758 and 7,029 pixels respectively). This was the training set least affected by class imbalance. The result shows that the accuracies obtained when using the different training sets are affected by the number of pixels each training set contains. Therefore, the H_0 was rejected in favor of the H_A as there was no evidence showing that the mean accuracies between the results obtained using unbalanced and balanced

training sets was zero. The result led to the exclusion of the unbalanced training sets from subsequent steps of the classification process.

Table 6 Accuracy assessment at edge areas of the first classification (40 bands layer stack). The McNemar's significance test was performed between unbalanced and balanced training sets.

Training set		Edge OA (%)	Kappa coeff.		Edge OA (%)	Kappa coeff.	X^2	*p-value*
TDS_1	Unbalanced	68.5	0.53	Balanced	68.4	0.53	0.7	$p = 0.389$
TDS_{1+2}		73.2	0.60		74.3	0.61	84.3	$p < 0.001$
TDS_{1+2+3}		91.9	0.88		83.6	0.75	419.0	$p < 0.001$
$TDS_{1,2,3}$		94.0	0.91		84.2	0.76	537.9	$p < 0.001$

OA: Overall accuracy.

5.2 Classification accuracy assessment and band importance analysis

The RF algorithm was run 12 more times using the balanced training sets, each time removing 20% of the bands with the lowest importance. The segmented accuracy assessment resulted in different accuracy values depending on the area of the forest patch being evaluated and the training set used to train the algorithm. The details of these results are described in the following subsections. Overall, the highest accuracy at edge areas was obtained when 25 bands were used together with the $TDS_{1,2,3}$, with an OA of 85.4% and a kappa coefficient of 0.78 (Table 7). For intermediate areas, the highest accuracy resulted when using 16 bands and the TDS_{1+2}. Interior areas were most accurately classified when 4 bands and the TDS_1 were used. Finally, when the accuracy was assessed at entire forest patches, the highest accuracy was obtained when using 16 bands and the $TDS_{1,2,3}$. The complete accuracy assessment of all performed classifications is presented in Appendix 1.

Table 7 Highest accuracies obtained in the accuracy assessment. The table shows the combination between the number of bands and the training sets resulting in the highest accuracy for each validation area.

Forest patch area	Validation data set	Training set resulting in highest accuracy	Overall accuracy (%)	Kappa coefficient	Nº of bands used
Interior	VDS_1	TDS_1	98.2	0.97	4
Intermediate	VDS_2	TDS_{1+2}	92.9	0.89	16
Edge	VDS_3	$TDS_{1,2,3}$	85.4	0.78	25
Entire patch	VDS_{1+2+3}	$TDS_{1,2,3}$	90.4	0.86	16

5.2.1 Accuracy at interior areas

Figure 10a shows the bands used at each classification run together with the accuracies at interior areas. The accuracies are presented for the four training sets being compared. It is important to emphasize that interior areas represent the most homogeneous area of a forest patch. When the TDS_1 was used for the classification,

both training and validation sets represent what it is traditionally done in land cover classification studies, where edge pixels as training data are excluded and the accuracy assessment is not performed at edge areas. The obtained results support this approach. The accuracy at interior areas was higher when edge pixels were not used as training data (TDS_1). On the contrary, when training sets containing edge pixels (TDS_{1+2+3} and $TDS_{1,2,3}$) were used, the lowest accuracies were obtained.

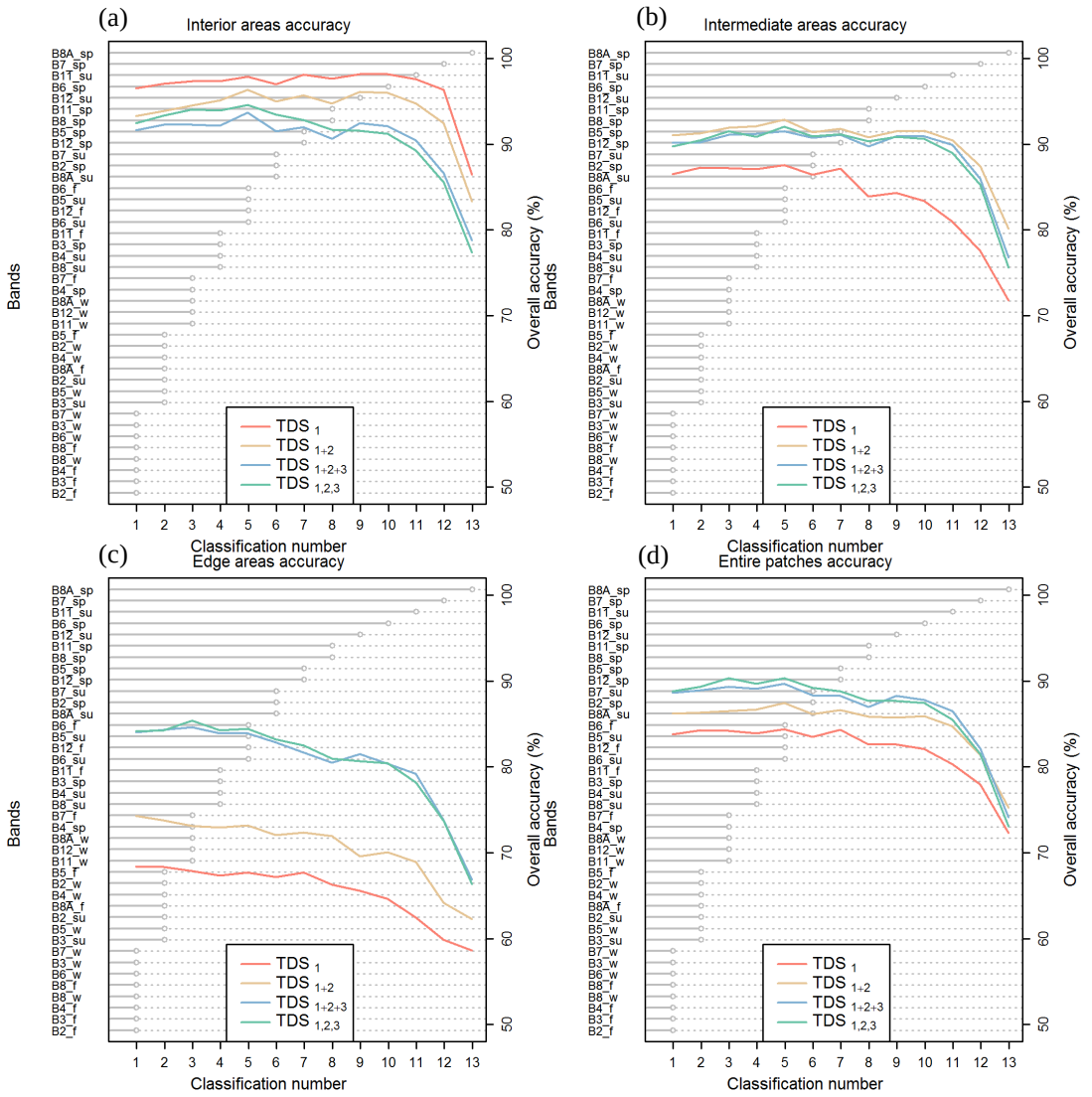

Figure 10 Bands used during the 13 RF classifications. They appear in order of importance, from the most important to the least important. The horizontal gray lines indicate if a band was used during a classification. Additionally, the accuracies obtained by using the four different training sets are presented when the accuracy assessment was performed at (a) interior, (b) intermediate, and (c) edge areas, as well as for (d) entire patches. su: summer. sp: spring. f: fall. w: winter.

5.2.2 Accuracy at intermediate areas

The highest accuracy at intermediate areas was obtained when the TDS_{1+2}, which includes interior and intermediate areas, was used during the classification (Figure 10b). Conversely, the lowest accuracy was obtained when pixels from intermediate areas were not included in the training set (TDS_1). The use of TDS_{1+2+3} and $TDS_{1,2,3}$ resulted in almost identical accuracies, presenting slightly lower values than when the TDS_{1+2} was used.

5.2.3 Accuracy at edge areas

When the accuracy was assessed at edge areas, the results contrasted completely with the results from the accuracy assessment at interior areas. Figure 10c shows the accuracies at edge areas, where there is a pronounced decrease in accuracy when the training sets excluding edge pixels were used (TDS_1 and TDS_{1+2}). The highest accuracy resulted when the $TDS_{1,2,3}$ was used, closely followed by when the TDS_{1+2+3} was used.

5.2.4 Accuracy at entire patches

In the case where pixels from interior, intermediate and edge areas were used as validation data, the results of the accuracy assessment show that the highest accuracies were obtained when edge pixels were included as training data (Figure 10d). This occurred when the $TDS_{1,2,3}$ and TDS_{1+2+3} were used. There was a clear drop in accuracy when using the TDS_1, which resulted in the training set with the lowest accuracy.

5.2.5 Band importance

The most important band proved to be band 8A, covering the NIR. It was followed by band 7 (Red Edge). From the top 10 bands with the highest importance, 7 corresponded to the spring image, whereas 3 corresponded to the summer image. The ranking included the remaining Red Edge bands (bands 5 and 6), bands covering the SWIR (bands 11 and 12) and band 8 (NIR). The bands from the winter image presented the lowest importance values, and all were discarded after the third classification. The bands from the fall image also showed low importance values, and all bands were discarded after the sixth classification.

As mention before, the training set used to obtain the highest accuracy at the different areas of forest patches varied depending on the area being analyzed. In the same way, the number of bands used to produce the highest accuracy was different for each training set. In fact, the highest accuracies were obtained when not all bands were used, where low importance bands had been removed. Table 8 presents the percentage of decrease in accuracy of the training sets that resulted in the highest accuracy at

each validation area for each of the 13 classifications. Some level of noise in the lowest importance bands (from winter and fall) might have affected the results. Nevertheless, the decrease in accuracy was very low in most cases. These results describe a trade-off between the number of bands used and the decrease in accuracy, which can be useful when data processing capabilities are limited.

Table 8 Percentage of decrease in OA of the training sets that resulted in the highest accuracy at each validation area where the accuracy assessment was performed. The values obtained for the 13 classifications are presented.

Classification number	Nº of bands used	Decrease in OA of the training sets that resulted in the highest accuracy by forest area (%)			
		Interior areas (TDS_1)	Intermediate areas (TDS_{1+2})	Edge areas ($TDS_{1,2,3}$)	Entire patches ($TDS_{1,2,3}$)
1	40	1.67	1.97	1.45	1.69
2	30	1.18	1.73	1.36	1.15
3	25	0.86	1.02	*	0.06
4	20	0.86	0.86	1.31	0.74
5	16	0.34	*	1.15	*
6	12	1.22	1.57	2.6	1.26
7	9	0.07	1.14	3.42	1.70
8	7	0.55	2.24	5.16	2.99
9	5	0.03	1.45	5.55	2.93
10	4	*	1.38	5.84	3.23
11	3	0.63	2.59	8.45	5.39
12	2	1.85	5.86	13.66	9.79
13	1	11.99	13.70	22.33	19.12

OA: Overall Accuracy. *: Classification that resulted in the highest accuracy.

5.2.6 Highest accuracy at edge areas

According to the described results, the best results for edge areas were obtained during the classification performed with 25 bands and using the $TDS_{1,2,3}$. To assess if there was statistically significant evidence to affirm that the results obtained by including intermediate and edge pixels as training data resulted in higher accuracies compared to when they were not included (TDS_1), the McNemar's exact conditional test was performed to extract the *p-values* (Table 9). The result showed that there was statistical evidence to reject the H_0, accepting the H_A in the training sets containing edge pixels, with *p-value* < 0.01. There was no statistical evidence showing that the accuracies between the results obtained when using the TDS_1 were greater than when the $TDS_{1,2,3}$ was used. This test proved that the accuracy is greater when edge pixels are included, rejecting the H_0 in favor of the H_A. The same conclusion was obtained when testing the TDS_1 against the TDS_{1+2} and TDS_{1+2+3}.

Table 9 McNemar's exact conditional test *p-values*. The comparison was made between including and excluding edge pixels as training data, and between using a training set with pixels divided into subclasses or with all pixels combined.

Training data set	Edge OA (%)	Training data set	Edge OA (%)	*p-value*
TDS_1	67.9	TDS_{1+2+3}	84.7	$p < 0.01$
		TDS_{1+2}	73.1	$p < 0.01$
		$TDS_{1,2,3}$	85.4	$p < 0.01$
$TDS_{1,2,3}$	85.4	TDS_{1+2+3}	84.7	$p = 0.07$

OA: Overall Accuracy

Table 9 also shows the statistical test between using the $TDS_{1,2,3}$ and the TDS_{1+2+3} to statistically assess if higher accuracies were obtained when the training pixels were grouped into subclasses than when all pixels were combined. In this case, the *p-value* shows that there is no evidence to reject the H_0. Furthermore, the increase in accuracy obtained with the $TDS_{1,2,3}$ compared to the accuracy of the TDS_{1+2+3} is less than 1%.

5.3 Non-target classes

The additional classes present in the S2 images masked using the SLU forest map are described in Table 10. The classification performed to remove these classes achieved OA values from 91 to 99% for all non-target classes. The highest accuracies were obtained for built-up areas, bare soil and water bodies classes. Figure 11 shows a true color composite image of the spring image where the non-target classes were removed. This was used to mask the final maps.

Table 10 Land cover classes used for the classification of non-target classes.

Land cover class	Description
Pine	Coniferous forest. Target class.
Spruce	Coniferous forest. Target class.
Birch	Deciduous forest. Target class.
Other deciduous	Mix of deciduous tree species (mainly oak, beech and aspen). Non-target class.
Bare soil	Recently harvested area. Non-target class.
Built-up areas	Includes roads, buildings and quarries. Non-target class.
Low vegetation	Grasslands, shrubs and bushes. Non-target class.
Water bodies	Lakes. Non-target class.

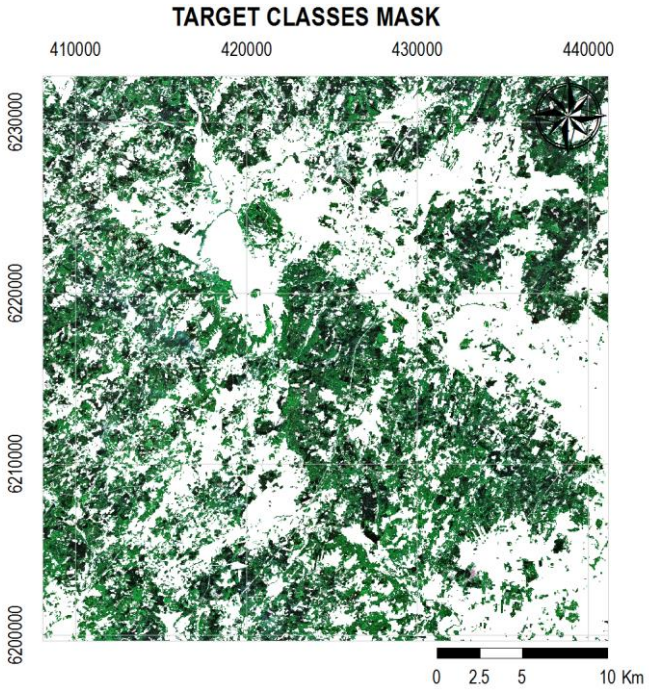

Figure 11 S2 true color composite (432) of the study area created using bands from the spring image and masked using the SLU forest map 2010. Additionally, non-target classes were removed by performing an additional RF classification. Image source: ESA Sentinel-2 image. Tile: 33VVC. Acquisition date: 2017-05-27. Retrieved from https://scihub.copernicus.eu/.

5.4 Final tree species classification

The final map showing the best tree species classification focusing on the accuracy at forest edge areas is presented in Figure 12. The classification was made using 25 bands and the $TDS_{1,2,3}$. The species with most area covered in the study area was birch, covering 241 km^2. Pine and spruce species covered 189 and 134 km^2 respectively. The segmented accuracy assessment performed for this classification is summarized in Table 11. Considering that forest ecosystems are complex classes, the resulting OA and kappa coefficients were high. The accuracies by class presented the highest UA and PA in the spruce and birch classes.

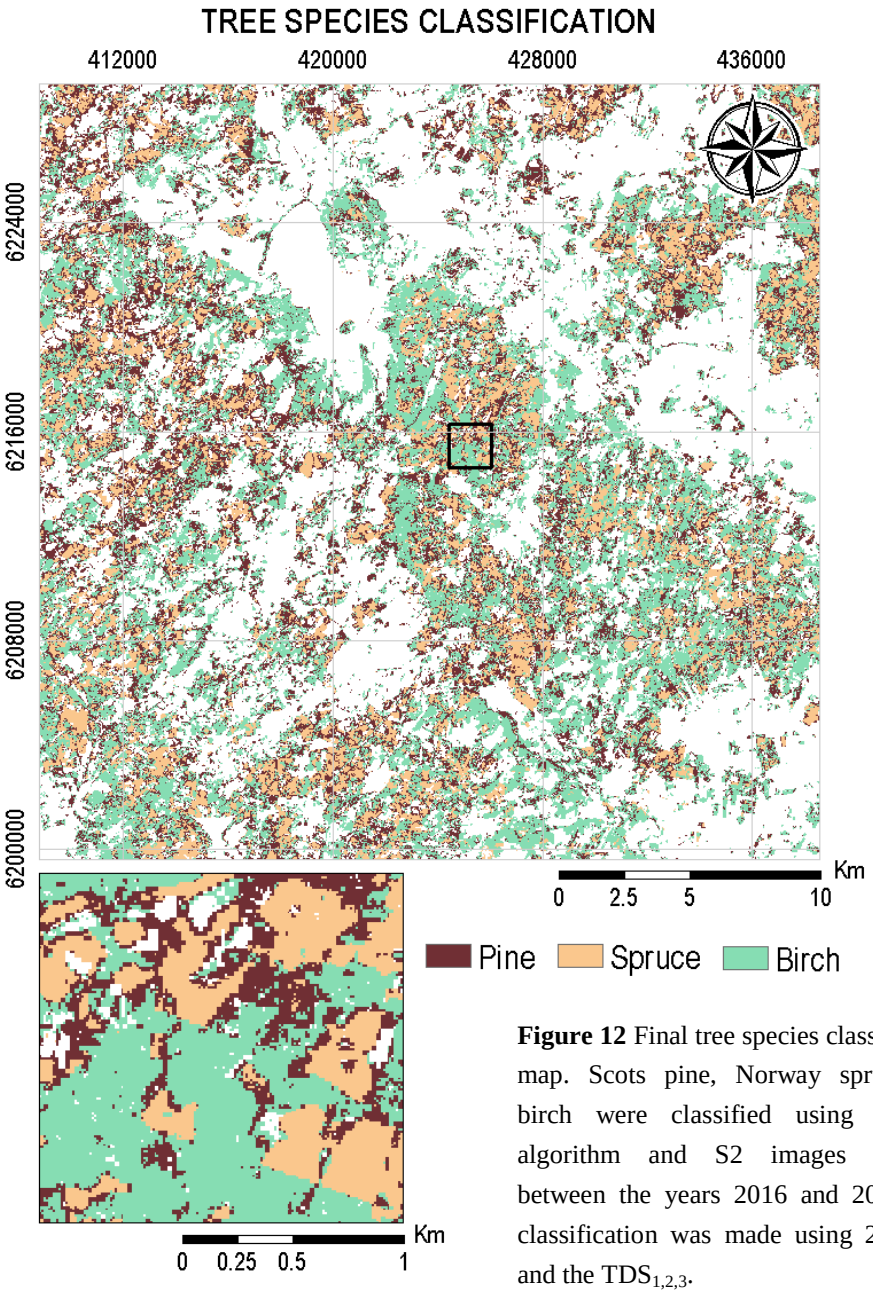

Figure 12 Final tree species classification map. Scots pine, Norway spruce and birch were classified using the RF algorithm and S2 images obtained between the years 2016 and 2018. The classification was made using 25 bands and the $TDS_{1,2,3}$.

Table 11 Segmented accuracy assessment of final classification.

Forest patch area	Overall accuracy (%)	Kappa coefficient	Class	User's accuracy (%)	Producer's accuracy (%)
Interior	94.1	0.91	Pine	87.8	94.2
			Spruce	95.2	94.5
			Birch	99.2	93.5
Intermediate	91.5	0.87	Pine	83.9	92.2
			Spruce	92.2	91.7
			Birch	98.5	90.8
Edge	85.4	0.78	Pine	80.4	79.9
			Spruce	86.1	86.8
			Birch	89.7	89.6
Entire patch	90.3	0.85	Pine	84.0	88.5
			Spruce	91.2	91.0
			Birch	95.8	91.3

To visually compare the final classification map with the maps obtained when the rest of the training sets were used, Figure 13 shows a close look to an example area within the study area. The maps produced using the four training approaches described in this study are presented. There is a clear difference between the maps in terms of how the edge of the forest patches were classified. Classifications made by using edge pixels as training data produced more homogeneous and larger forest patches (TDS_{1+2+3} and $TDS_{1,2,3}$).

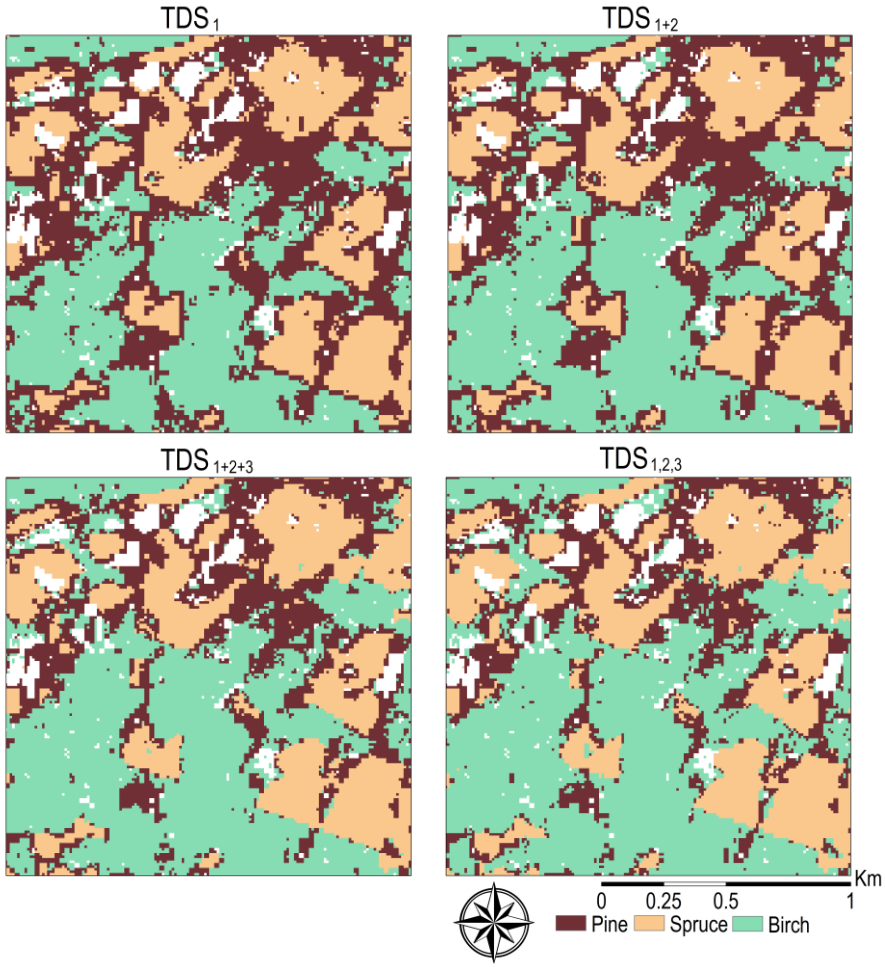

Figure 13 Tree species classification maps created by using the RF algorithm and S2 data. Each map corresponds to the result obtained when using one of the four training sets described in this study. The classification was made using 25 bands.

The differences between the results obtained with the $TDS_{1,2,3}$ and the rest of the training sets are highlighted in Figure 14. Almost all differences are located at edge areas. The maps of the training sets that excluded edge pixels (TDS_1 and TDS_{1+2}) are the ones presenting the largest differences. For interior, more homogeneous areas all classifications produced similar results. This arises the need for considering edge pixels as part of the training data in forest land cover classification. It is within edge areas where the misclassifications are concentrated.

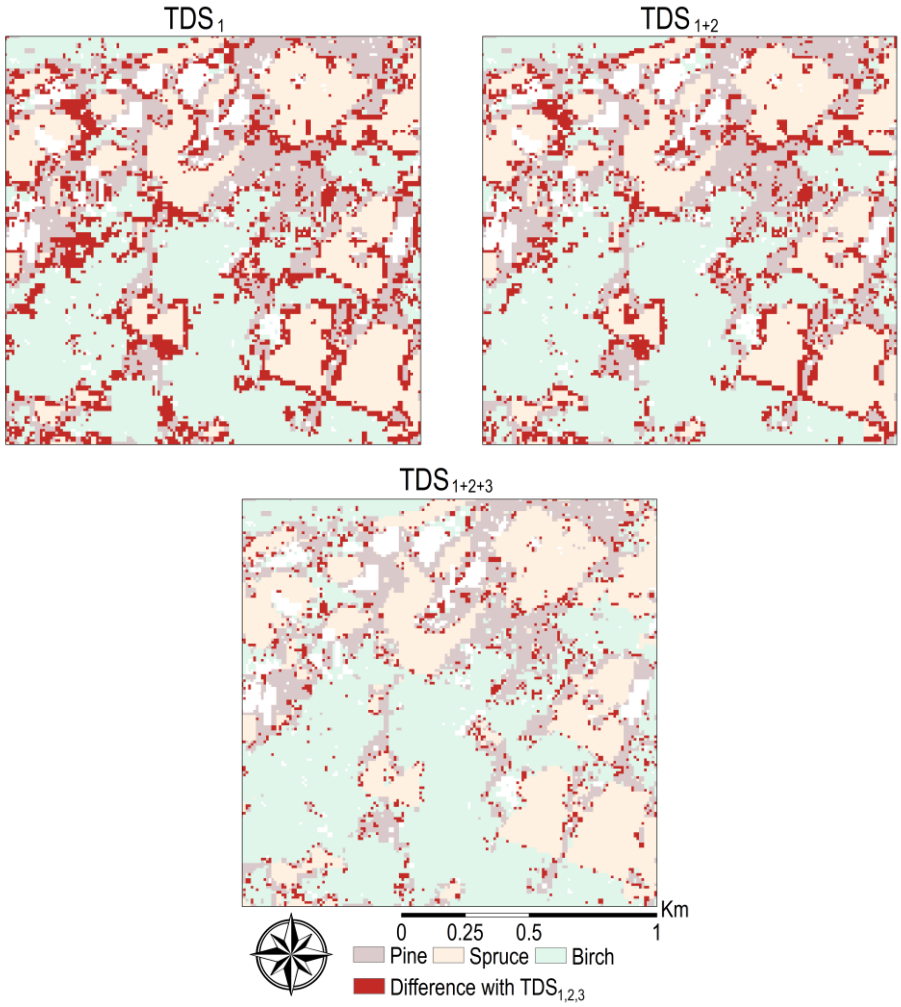

Figure 14 Visualization of the differences between the final classification map (using TDS_{1+2+3}) and the rest of the classification maps (using the remaining training data sets).

5.5 GSV estimation

The correlation analysis performed on S2 bands of the summer image produced weak correlation values. The correlation coefficient (r) and the coefficient of determination (R^2) between the reflectance values of each band and tree volume are presented in Table 12. The highest correlations were found in bands 6, 7 8 and 8A. They presented a decreasing relationship with tree volume values. However, the highest R^2 values were not larger than 0.27. This means that only a 27% of the variation in volume was due to the decreasing trend. The r and R^2 parameters assess linear relationships, and it can be possible that the relationship between the S2 data and volume values was not linear. In such case, r and R^2 were not adequately assessing the nature of data variation.

Table 12 Correlation coefficient (*r*) and coefficient of determination (R^2) obtained between S2 bands from the summer image and tree volume values of the NFI field collected samples**.**

S2 band	*r*	R^2
2	0.02	0.01
3	-0.20	0.03
4	0.07	0.01
5	-0.27	0.08
6	-0.52	0.27
7	-0.52	0.27
8	-0.52	0.27
8A	-0.51	0.26
11	-0.26	0.07
12	-0.13	0.02

This was compared with the results of the band importance analysis performed using the RF algorithm. The same bands showed the highest importance values (Figure 15). However, band 5 also scored a high importance value, even higher than bands 7,8 and 8A. The accuracy assessment using the statistics generated by the RF model resulted in a RMSE of 124 m^3 ha^{-1}. This error value is relatively high considering that the GSV values range between 46 and 513 m^3 ha^{-1}.

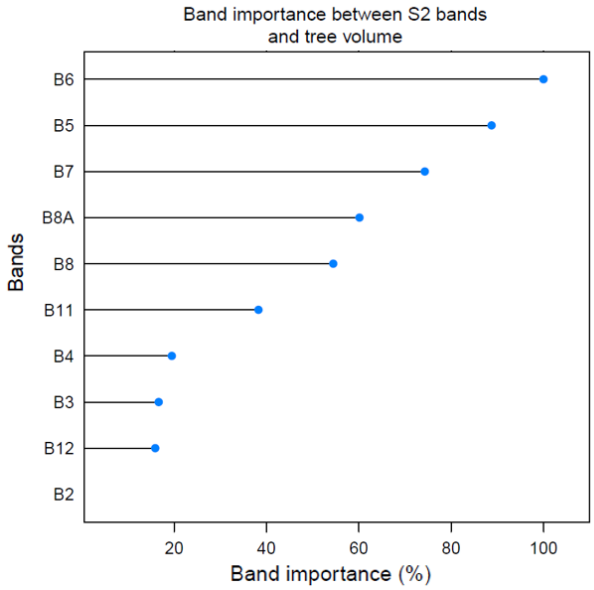

Figure 15 Band importance between S2 bands and tree volume obtained from the RF regression. Only bands from the summer image were used for the regression.

The forest GSV map generated after performing the RF regression is shown in Figure 16. The tree species showing the highest volume values was spruce with an average of 314 m^3 ha^{-1}. It was followed by pine with 210 m^3 ha^{-1} and birch with 143 m^3 ha^{-1}. Finally, the mean GSV for the study area resulted in 208 m^3 ha^{-1}.

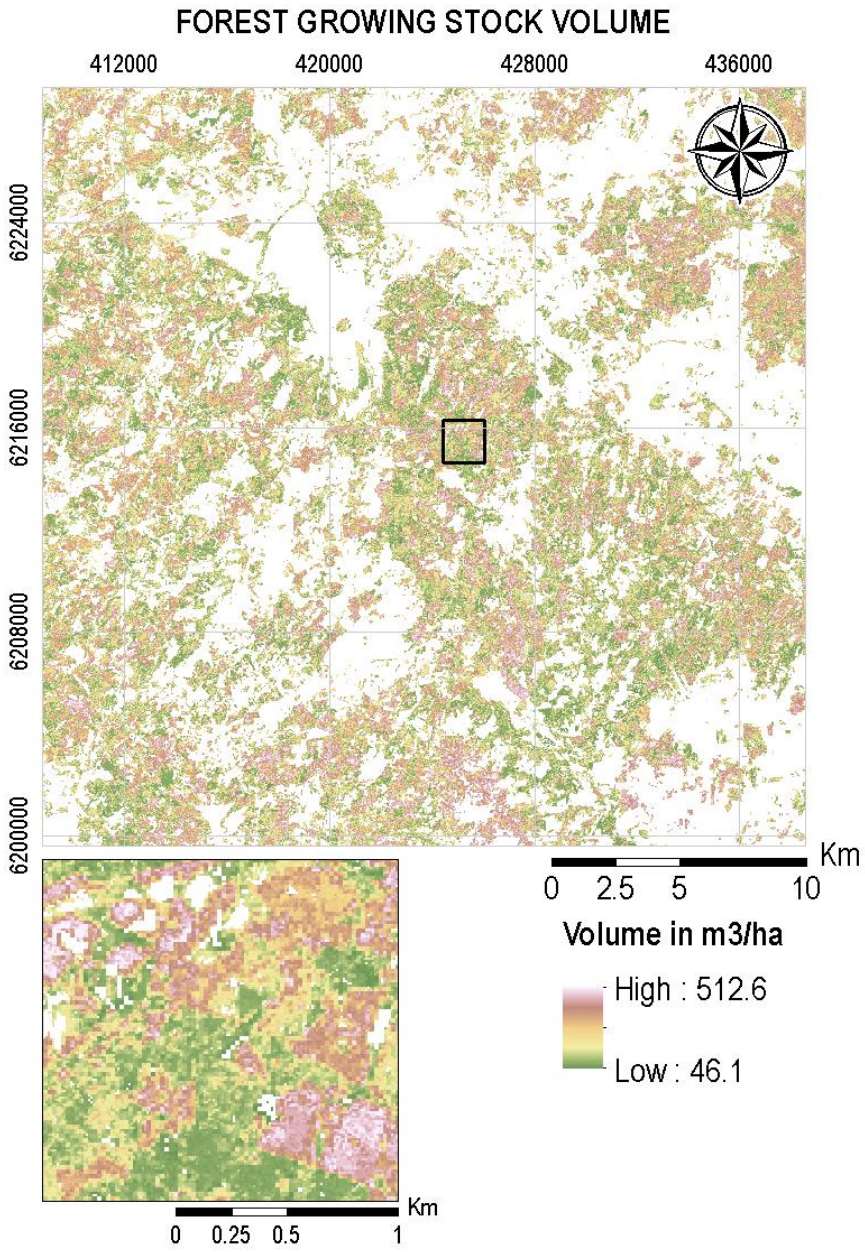

Figure 16 Forest growing stock volume estimated by a RF regression using the NFI data and S2 bands from the summer image.

6. DISCUSSION

6.1 Inclusion of edge pixels as training data

One of the aims of this study was to test a tree species classification procedure including edge pixels as training and validation data to improve accuracies of forest edge areas. The results proved that higher accuracies at edge areas can be obtained when edge pixels are included in the training data set. In fact, the lowest accuracy at edge areas was obtained when only interior areas were sampled as training data.

The performed accuracy assessment at edge areas, as well as the performed at entire patches, presented higher accuracies when edge and intermediate pixels were included in the training set. It seems that the multimodal distribution of the spectral responses in forest patches was captured more accurately when pixels representing all areas of the patch were considered. The results are consistent with what was concluded by Colditz (2015), who stated that for land cover classes presenting heterogeneous spectral responses higher accuracies are obtained when the sampling of training pixels includes the complete spectral range occurring in the target class. Similarly, the study conducted by Shuai et al. (2015) determined that including pixels located at edge segments of the spectral distributions of vegetation classes improves the OA when focusing the classification in these particular classes. Both studies sustain that an adequate training stage is critical to obtain high accuracies in land cover classification. Additionally, it should be highlighted that when the sample of training data includes observations only from homogeneous areas, there is a higher risk of exposing the data to a subjective treatment. The present study proposes a methodology based on a systematic and non-biased sampling where random pixels representing the diversity of spectral values in forest patches are sampled for the training and validation sets.

6.2 Accuracy assessment of edge areas

Opposed to the traditional approach where the classification accuracy is measured only with validation data from interior sampling, the proposed segmented accuracy analysis accomplished to estimate accuracies for edge, intermediate and interior areas. The result of this method presents evidence on how accuracy values can be biased when the validation set only includes interior, or homogeneous areas, where the accuracy shows its highest values. This was demonstrated when the training set containing only interior pixels was used to train the classification algorithm. In that case, the accuracy showed notably high values reaching an OA of 98.2%. This contrasts the lower values obtained at edge areas when the same training set was used, where the highest OA was 68.6%. Previous studies assessing accuracy of edge areas show the same results. Sweeney and Evans (2012) proposed a methodology to include

the error of edge areas into the land cover map accuracy assessment. Their results show that traditional error estimation is not representative of all areas of a classified map, emphasizing on the disparity of the error between interior and edge areas. Furthermore, most misclassifications occur at edges of land cover classes (Zhu et al. 2000; Sweeney and Evans 2012). Yet usually little detail regarding the validation data sampling technique is reported in land cover classification studies (Hammond and Verbyla 1996). Studies such as the carried out by Muller et al. (1998) work as an example, as they stated in the results of their land cover classification that the validation data came solely from homogeneous areas, making clear that the classification results were mainly valid for homogeneous areas.

Based on the obtained results, it was possible to improve the accuracy of edge areas. This was achieved when the RF algorithm was trained with the training set containing pixels from edge, intermediate and interior areas grouped into subclasses ($TDS_{1,2,3}$). The accuracy at edge areas increased significantly, resulting in a value of 85.4%. Additionally, in the same classification a similar value was obtained when the accuracy was measured in entire patches, where the OA was 87.5%. These findings contribute to a better recognition of forest edges, crucial aspect in the definition of forest stands. Furthermore, other ecological areas such as priority conservation zones for endangered species can be better identified in terms of their boundaries (Lu et al. 2012). There was no statistical significance in the improvement in accuracy when comparing the $TDS_{1,2,3}$ and TDS_{1+2+3}. The RF algorithm generated equally good results in both cases, suggesting that it was not necessary to divide the target classes into subclasses.

6.3 Random Forest algorithm

Land cover classification studies using RF classifier have consistently shown good results when complex classes are considered (Pal 2005; Gislason et al. 2006; Rodriguez-Galiano et al. 2012). The same is true for regression studies estimating environmental variables (Prasad et al. 2006; Abdel-Rahman et al. 2013), where it can even outperform traditional multiple linear regression (Mutanga et al. 2012; Oliveira 2012). This explains the good results obtained during the tree species classification. RF also works well with the presence of outliers in the training data (Pal 2005; Colditz 2015; Belgiu and Drăguţ 2016), which may be the reason of the high accuracy obtained when edge pixels were included in the training data. The estimation of GSV did not performed as good, resulting in a RMSE of 124 m^3 ha^{-1}. In that case, the small size of the training set could have affected the results. The extrapolation in time of volume values to match them with the acquisition date of the S2 image represents an additional source of error.

A disadvantage of machine learning algorithms is that they can require longer processing times both during the classification and training stages (Pirotti et al. 2016), particularly when large amounts of multispectral and multi-temporal images are used (Schneider 2012). However, with the combination of these algorithms with parallel processing the execution time can be drastically reduced (Plaza et al. 2009). Thanks to the *train()* function from the *caret* package of the R software that allows parallel processing, the RF algorithm was trained using parallel processing. For the classification/regression stage, the same was possible by using the *clusterR()* function from the *raster* package. This resulted in overall short processing times considering that a 10 m resolution layer stack of up to 40 bands in a 30x30 km window was used.

In terms of the different sizes of the training data for each class, it was concluded that the difference between the results obtained when using the unbalanced and balanced training sets was statistically significant. It is possible to find in the literature studies showing that the use of unbalanced data produces results favoring the class with the largest size (Shan et al. 2006; Blagus and Lusa 2010; Janitza et al. 2013). Nonetheless, there are also studies showing that RF algorithm is not sensitive to unbalanced data sets (Pal 2005; Freeman et al. 2012). Blagus and Lusa (2010) assigns the dissimilar conclusions to differences in the classification methodologies, the variability between classes and the level of class imbalance. They suggest that balanced data sets should be preferred unless there is evidence for particular cases that the RF algorithm is not affected by unbalanced data. In the present study, training sets with larger number of pixels were indeed favored by an improvement in the OA. This led to the selection of the balanced training sets to perform the classification analysis.

6.4 Sentinel-2 important bands

The variable importance analysis of the tree species classification successfully identified the best band combination to use in order to generate the most accurate outcome. The best result was obtained when 25 bands, with bands from all seasons, were used. This result was the expected, as the use of multi-temporal images in land cover classification is usually beneficial (Eisavi et al. 2015; Khatami et al. 2016; Puletti et al. 2017). The most important bands corresponded to the spring and summer images, probably because they cover the growing season period. Of the spectral bands sampled by the S2 MSI sensor, band 8A (NIR) proved to be the most important of them all. Unlike the band 8, which also covers the NIR, band 8A is much narrower, presenting a bandwidth of 33 nm compared to the 145 nm that band 8 has. The combination of band 8A with other narrow bands, such as the Red Edge bands (bands 5, 6 and 7), proved to be of great value to obtain higher classification accuracies. This conclusion has already been outlined by previous studies using S2 data in classification of vegetation (Sibanda et al. 2015; Forkuor et al 2017). Moreover, this

agrees with the characteristic spectral responses of pine, spruce and birch, where the main differences occur in the range from 700 to 850 nm (Jaaskelainen et al. 1994). Pine and spruce present more similar spectral responses between each other, which can explain that the highest accuracy at entire patches was obtained for birch (UA = 95.8% and PA = 91.3%), whereas pine and spruce resulted in slightly lower accuracies (UA = 84.0% and PA = 88.5% for pine and UA = 91.2% and PA = 91.0% for spruce).

Similarly, bands 11 and 12 (SWIR) scored high band importance values and proved to be relevant for the tree species classification. This complies with the wide evidence showing the importance of the SWIR in forest classifications (Immitzer et al. 2016; Karasiak et al. 2017; Sothe et al. 2017). The SWIR covers the water absorption wavelengths, capturing differences in vegetation water content (Serrano et al. 2000). Thus, due to the different water content in pine, spruce and birch foliage, they present clear differences within the SWIR region (Lukeš et al. 2013).

The correlation analysis of summer bands and GSV showed values between -0.51 and -0.52 for most Red Edge bands (bands 6 and 7) and NIR bands (bands 8 and 8A). Similarly, when the band importance was assessed using the RF algorithm, bands 5,6,7,8 and 8A presented the highest importance values. Other studies estimating GSV using S2 data reported similar results for these bands. Chrysafis et al. (2017) for example, obtained a RMSE of 63.11 m^3 ha^{-1} when estimating GSV of Mediterranean forests using RF and S2 data. They also found that the S2 band most correlated with volume values is band 11 (SWIR), with r = -0.68. However, in the present study band 11 produced a much lower correlation (r = -0.26). Another example, presented by Puliti et al. (2018), concluded that Red Edge and SWIR bands produced the highest correlation values when being compared with GVS in boreal forests.

6.5 Limitations

The limitations of the present study involve the error associated with the absence of more field data apart from the NFI data set. Particularly for the GSV estimation, where only a small number of samples was available. In the case of the tree species classification, errors may have occurred during the digitization of the training patches. This process was not made with field visits, but solely with 0.25 m resolution orthophotos and NFI observations. However, regardless the study limitations, the prediction of tree species distribution focused on edge areas, and the estimation of GSV represent an important step to identify forest stands. If more reference field data is available it is possible to achieve better results, and other forest structural parameters could be estimated with S2 data and the RF algorithm. In this context,

related research presents enormous opportunities for improving the characterization of forest ecosystems and, consequently, of forestry practices.

7. CONCLUSION

As the demand for information about forest ecosystems increases, a constant development and improvement of methodologies for predicting forest parameters is required. The combined use of the recently and freely available S2 satellite images and machine learning algorithms can contribute to this matter.

The present study addressed the problematic of classifying forest edge areas by assessing the inclusion of edge pixels in the training set when using the RF algorithm. It also highlighted how the accuracy assessment in land cover studies can be optimistically biased towards spectrally homogeneous areas. Thus, a segmented accuracy assessment was proposed to quantify the accuracy at different areas of forest patches.

The results showed that for Scots pine, Norway spruce and birch the accuracy of forest edge areas can be increased by including edge pixels as training data using the RF algorithm. This suggests that the classifier was able to effectively handle outliers occurring at edge areas. Dividing the training set of one tree species into subclasses grouping pixels according to their location within the forest patch resulted only in a small increase in accuracy at edge areas, which was not statistically significant. On the other hand, balancing the training sets proved to be necessary, as data imbalance affected the accuracy values by favoring the accuracy of training sets with larger sample sizes.

The RF algorithm has the advantage that it can predict continuous data by working as a regression model. This was assessed for GSV estimation, but the outcome presented a high RMSE value compared to previous studies. The band importance analysis showed more consistent results, where bands covering the Red Edge, SWIR and a narrow area of the NIR demonstrated to deliver most of the information to predict forest variables. These spectral regions sampled by the S2 mission together with the use of multi-temporal images are useful for vegetation studies.

The findings of this study can help to better understand how the accuracy varies in spectrally heterogenous areas and how different allocation schemes of the training and validation data affect the results. Moreover, improving accuracies at edge areas and estimating GSV contributes to more precise definitions of forest stands. Further research assessing the use S2 data with machine learning algorithms is needed, especially regarding the estimation of other forest structural parameters.

8. REFERENCES

Abdel-Rahman, E. M., F. B. Ahmed, and R. Ismail. 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing*, 34**:** 712-728. DOI: 10.1080/01431161.2012.713142

Addabbo, P., M. Focareta, S. Marcuccio, C. Votto, and S. L. Ullo. 2016. Contribution of Sentinel- 2 data for applications in vegetation monitoring. *ACTA IMEKO*, 5**:** 44-54.

Agresti, A. 2006. *An introduction to categorical data analysis*. John Wiley & Sons, Inc.

Akar, Ö. 2017. Mapping land use with using Rotation Forest algorithm from UAV images. *European Journal of Remote Sensing*, 50**:** 269-279. DOI: 10.1080/22797254.2017.1319252

Antonisamy, B., Prasanna Premkumar, and Solomon Christopher. 2017. *Principles and Practice of Biostatistics*. Elsevier.

Bärring, L., P. Jönsson, J. O. Mattsson, and R. Åhman. 2003. Wind erosion on arable land in Scania, Sweden and the relation to the wind climate—a review. *Catena*, 52**:** 173-190. DOI: 10.1016/s0341-8162(03)00013-4

Barton, I., G. Király, K. Czimber, M. Hollaus, and N. Pfeifer. 2017. Treefall Gap Mapping Using Sentinel-2 Images. *Forests*, 8. DOI: 10.3390/f8110426

Belgiu, M., and L. Drăguţ. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114**:** 24-31. DOI: 10.1016/j.isprsjprs.2016.01.011

Blagus, R., and L. Lusa. 2010. Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 523.

Blennow, K., L. Bärring, P. Jönsson, M.-L. Lindersson, J. O. Mattsson, and P. Schlyter. 1999. Klimat, sjöar och vattendrag. In *Atlas över Skåne*, 184. Lantmäteriverket och Svenska sällskapet för antropologi och geografi.

Bond, I., M. Grieg-Gran, S. Wertz-Kanounnikoff, P. Hazlewood, S. Wunder, and A. Angelsen, 2009. Incentives to sustain forest ecosystem services: A review and lessons for REDD. International Institute for Environment and Development, London, UK, with CIFOR, Bogor, Indonesia, and World Resources Institute, Washington D.C., USA., Report. [in Swedish, English summary]

Bruzzone, L., F. Bovolo, C. Paris, Y. T. Solano-Correa, M. Zanetti, and D. Fernández-Prieto. 2017. Analysis of Multitemporal Sentinel-2 Images in the Framework of the ESA Scientific Exploitation of Operational Missions. *IEEE*.

Bruzzone, L., and B. Demir. 2014. A Review of Modern Approaches to Classification of Remote Sensing Data. In *Land Use and Land Cover Mapping in Europe*, 127-143.

Canty, M. J. 2010. *Image analysis, classification, and change detection in remote sensing*. CRC Press.

Chrysafis, I., G. Mallinis, S. Siachalou, and P. Patias. 2017. Assessing the relationships between growing stock volume and Sentinel-2 imagery in a Mediterranean forest ecosystem. *Remote Sensing Letters*, 8**:** 508-517. DOI: 10.1080/2150704x.2017.1295479

Chuvieco, E. 2016. *Fundamentals of satellite remote sensing: an environmental approach*. CRC Press.

Colditz, R. 2015. An Evaluation of Different Training Sample Allocation Schemes for Discrete and Continuous Land Cover Classification Using Decision Tree-Based Algorithms. *Remote Sensing*, 7**:** 9655-9681. DOI: 10.3390/rs70809655

Congalton, R. G., and K. Green. 2008. *Assessing the Accuracy of Remotely Sensed Data*. CRC Press.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. C. Gibson, and J. J. Lawler. 2007. Random Forests for Classification in Ecology. *Ecology*, 88**:** 2783-2792.

Demir, N., Y. E. Eryilmaz, and S. Oy. 2018. Post-hurricane damage assessment on greenhouse fields with use of SAR data. *GeoInformation For Disaster Management*, XLII-3/W4**:** 191-195. DOI: 10.5194/isprs-archives-XLII-3-W4-191-2018

Díaz-Uriarte, R., and S. Alvarez de Andres. 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7: 3. DOI: 10.1186/1471-2105-7-3

Eisavi, V., S. Homayouni, A. M. Yazdi, and A. Alimohammadi. 2015. Land cover mapping based on random forest classification of multitemporal spectral and thermal images. *Environmental Monitoring and Assessment,* 187**:** 291. DOI: 10.1007/s10661-015-4489-3

ESA. 2018. Sentinel-2 MSI Technical Guide. Retrieved Apr-2018 2018, from https://sentinels. copernicus.eu/ web/sentinel/technical-guides/sentinel-2-msi

Fagerland1, M. W., S. Lydersen, and P. Laake. 2013. The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. *BMC Medical Research Methodology,* 91.

FAO, 2004. Global Forest Resources Assessment Update 2005. Terms and Definitions., Forestry Department FAO, Report, Rome

FAO, 2015. Global Forest Resources Assessment 2015: How are the world's forests changing?, Food and Agriculture Organization of the United Nations, Report, Rome.

Fassnacht, F. E., H. Latifi, K. Stereńczak, A. Modzelewska, M. Lefsky, L. T. Waser, C. Straub, and A. Ghosh. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment,* 186**:** 64-87. DOI: 10.1016/j.rse.2016.08.013

Filella, I., and J. Penuelas. 1994. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing,* 15**:** 1459-1470. DOI: 10.1080/01431169408954177

Foody, G. M. 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment,* 80**:** 185-201.

Foody, G. M. 2004. Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering & Remote Sensing,* 70**:** 627-633.

Foody, G. M., and A. Mathur. 2006. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment,* 103**:** 179-189. DOI: 10.1016/j.rse.2006.04.001

Forkuor, G., K. Dimobe, I. Serme, and J. E. Tondoh. 2017. Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GIScience & Remote Sensing,* 55**:** 331-354. DOI: 10.1080/15481603.2017.1370169

Freeman, E. A., G. G. Moisen, and T. S. Frescino. 2012. Evaluating effectiveness of down-sampling for stratified designs and unbalanced prevalence in Random Forest models of tree species distributions in Nevada. *Ecological Modelling,* 233**:** 1-10. DOI: 10.1016/j.ecolmodel.2012.03.007

Gao, T., M. Hedblom, T. Emilsson, and A. B. Nielsen. 2014. The role of forest stand structure as biodiversity indicator. *Forest Ecology and Management,* 330**:** 82-93. DOI: 10.1016/j.foreco.2014.07.007

García-Pedrero, A., C. Gonzalo-Martín, and M. Lillo-Saavedra. 2017. A machine learning approach for agricultural parcel delineation through agglomerative segmentation. *International Journal of Remote Sensing,* 38**:** 1809-1819. DOI: 10.1080/01431161.2016.1278312

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. 2006. Random Forests for land cover classification. *Pattern Recognition Letters,* 27**:** 294-300. DOI: 10.1016/j.patrec.2005.08.011

Hammond, T. O., and D. L. Verbyla. 1996. Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing,* 17**:** 1261-1266.

Hansen, M. C., P. V. Potapov, S. J. Goetz, S. Turubanova, A. Tyukavina, A. Krylov, A. Kommareddy, and A. Egorov. 2016. Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data. *Remote Sensing of Environment,* 185**:** 221-232. DOI: 10.1016/j.rse.2016.02.023

Immitzer, M., C. Atzberger, and T. Koukal. 2012. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing,* 4**:** 2661-2693. DOI: 10.3390/rs4092661

Immitzer, M., F. Vuolo, and C. Atzberger. 2016. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing,* 8. DOI: 10.3390/rs8030166

Iurist, N., F. Stătescu, and I. Lateş. 2016. Analysis of Land Cover and Land Use Changes Using Sentinel-2 Images. *Present Environment and Sustainable Development,* 10. DOI: 10.1515/pesd-2016-0034

Jaaskelainen, T., R. Silvennoinen, J. Hiltunen, and J. P. S. Parkkinen. 1994. Classification of the reflectance spectra of pine, spruce, and birch. *Applied Optics,* 33**:** 2356-2362.

Janitza, S., C. Strobl, and A.-L. Boulesteix. 2013. An AUC-based permutation variable importance measure for random forests. *BMC Bioinformatics,* 119.

Jönsson, P., Z. Cai, E. Melaas, M. Friedl, and L. Eklundh. 2018. A Method for Robust Estimation of Vegetation Seasonality from Landsat and Sentinel-2 Time Series Data. *Remote Sensing,* 10. DOI: 10.3390/rs10040635

Karasiak, N., D. Sheeren, M. Fauvel, J. Willm, J.-F. Dejoux, and C. Monteil, 2017. Mapping Tree Species of Forests in Southwest France using Sentinel-2 Image Time Series. European Union, Report. [in Swedish, English summary]

Khatami, R., G. Mountrakis, and S. V. Stehman. 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment,* 177**:** 89-100. DOI: 10.1016/j.rse.2016.02.028

Kingsford, C., and S. L. Salzberg. 2008. What are decision trees? *Nature Biotechnology,* 26**:** 1011-1013.

Köhl, M. 2003. New approaches for multi resource forest inventories. In *Advances in Forest Inventory for Sustainable Forest Management and Biodiversity Monitoring,* eds. P. Corona, M. Köhl, and M. Marchetti. Kluwer Academic Publishers.

Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. *R News,* 2**:** 18-22.

Lillesand, T., R. W. Kiefer, and J. Chipman. 2015. *Remote Sensing and Image Interpretation*. New York: Wiley.

Lindenmayer, A. B., C. R. Margules, and D. B. Botkin. 1999. Indicators of Biodiversity for Ecologically Sustainable Forest Management. *Conservation Biology,* 14**:** 941–950.

Liu, M., X. Cao, Y. Li, J. Chen, and X. Chen. 2016. Method for land cover classification accuracy assessment considering edges. *Science China Earth Sciences,* 59**:** 2318-2327. DOI: 10.1007/s11430-016-5333-5

Lu, N., C.-X. Jia, H. Lloyd, and Y.-H. Sun. 2012. Species-specific habitat fragmentation assessment, considering the ecological niche requirements and dispersal capability. *Biological Conservation,* 152**:** 102-109. DOI: 10.1016/j.biocon.2012.04.004

Lukeš, P., P. Stenberg, M. Rautiainen, M. Mõttus, and K. M. Vanhatalo. 2013. Optical properties of leaves and needles for boreal tree species in Europe. *Remote Sensing Letters,* 4**:** 667-676. DOI: 10.1080/2150704x.2013.782112

Mohajane, M., A. Essahlaoui, F. Oudija, M. El Hafyani, and A. Cláudia Teodoro. 2017. Mapping Forest Species in the Central Middle Atlas of Morocco (Azrou Forest) through Remote Sensing Techniques. *ISPRS International Journal of Geo-Information,* 6. DOI: 10.3390/ijgi6090275

Momeni, R., P. Aplin, and D. Boyd. 2016. Mapping Complex Urban Land Cover from Spaceborne Imagery: The Influence of Spatial Resolution, Spectral Band Set and Classification Approach. *Remote Sensing,* 8. DOI: 10.3390/rs8020088

Muller, S. V., D. A. Walker, F. E. Nelson, N. A. Auerbach, J. G. Bockheim, S. Guyer, and D. Sherba. 1998. Accuracy Assessment of a Land-Cover Map of the Kuparu k River Basin, Alaska: Considerations for Remote Regions. *Photogrammetric Engineering & Remote Sensing,* 64**:** 619-628.

Mutanga, O., E. Adam, and M. A. Cho. 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation,* 18**:** 399-406. DOI: 10.1016/j.jag.2012.03.012

Nink, S., J. Hill, H. Buddenbaum, J. Stoffels, T. Sachtleber, and J. Langshausen. 2015. Assessing the Suitability of Future Multi- and Hyperspectral Satellite Systems for Mapping the Spatial

Distribution of Norway Spruce Timber Volume. *Remote Sensing,* 7**:** 12009-12040. DOI: 10.3390/rs70912009

Novack, T., T. Esch, H. Kux, and U. Stilla. 2011. Machine Learning Comparison between WorldView-2 and QuickBird-2-Simulated Imagery Regarding Object-Based Urban Land Cover Classification&#xD. *Remote Sensing,* 3**:** 2263-2282. DOI: 10.3390/rs3102263

Oliveira, S., F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J. M. C. Pereira. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management,* 275**:** 117-129. DOI: 10.1016/j.foreco.2012.03.003

Pahlevan, N., S. Sarkar, B. A. Franz, S. V. Balasubramanian, and J. He. 2017. Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. *Remote Sensing of Environment,* 201**:** 47-56. DOI: 10.1016/j.rse.2017.08.033

Pal, M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing,* 26**:** 217-222. DOI: 10.1080/01431160412331269698

Pang, H., A. Lin, M. Holford, B. E. Enerson, B. Lu, M. P. Lawton, E. Floyd, and H. Zhao. 2006. Pathway analysis using random forests classification and regression. *Bioinformatics,* 22**:** 2028-2036. DOI: 10.1093/bioinformatics/btl344

Pirotti, F., F. Sunar, and M. Piragnolo. 2016. Benchmark of Machine Learning Methods for Classification of a Sentinel-2 Image. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B7**:** 335-340. DOI: 10.5194/isprsarchives-XLI-B7-335-2016

Plaza, A., J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, et al. 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment,* 113**:** S110-S122. DOI: 10.1016/j.rse.2007.07.028

Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems,* 9**:** 181-199. DOI: 10.1007/s10021-005-0054-1

Pretzsch, H. 2009. *Forest Dynamics, Growth and Yield*. Springer-Verlag Berlin Heidelberg.

Probst, P., and A.-L. Boulesteix. 2017. To tune or not to tune the number of trees in random forest? *arXiv:1705.05654 [stat.ML]*.

Puletti, N., F. Chianucci, and C. Castaldi. 2017. Use of Sentinel-2 for forest classification in Mediterranean environments. *Annals Of Silvicultural Research*. DOI: 10.12899/ASR-1463

Puliti, S., S. Saarela, T. Gobakken, G. Ståhl, and E. Næsset. 2018. Combining UAV and Sentinel-2 auxiliary data for forest growing stock volume estimation through hierarchical model-based inference. *Remote Sensing of Environment,* 204**:** 485-497. DOI: 10.1016/j.rse.2017.10.007

Qian, Y., W. Zhou, J. Yan, W. Li, and L. Han. 2014. Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery. *Remote Sensing,* 7**:** 153-168. DOI: 10.3390/rs70100153

Reese, H., M. Nilsson, T. G. Pahlén, O. Hagner, S. Joyce, U. Tingelöf, M. Egberth, and H. Olsson. 2003. Countrywide Estimates of Forest Variables Using Satellite Data and Field Data from the National Forest Inventory. *AMBIO: A Journal of the Human Environment,* 32**:** 542-548. DOI: 10.1579/0044-7447-32.8.542

Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing,* 67**:** 93-104. DOI: 10.1016/j.isprsjprs.2011.11.002

Schneider, A. 2012. Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sensing of Environment,* 124**:** 689-704. DOI: 10.1016/j.rse.2012.06.006

Serrano, L., S. L. Ustin, D. A. Roberts, J. A. Gamon, and J. Peñuelas. 2000. Deriving Water Content of Chaparral Vegetation from AVIRIS Data. *Remote Sensing of Environment,* 74**:** 570-581.

Shan, Y., D. Paull, and R. I. McKay. 2006. Machine learning of poorly predictable ecological data. *Ecological Modelling*, 195**:** 129-138. DOI: 10.1016/j.ecolmodel.2005.11.015

Shuai, G., J. Zhang, L. Deng, and X. Zhu. 2015. Edge-pixels-based support vector data description for specific land-cover distribution mapping. *Journal of Applied Remote Sensing*, 9. DOI: 10.1117/1.Jrs.9.096034

Sibanda, M., O. Mutanga, and M. Rouget. 2015. Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 110**:** 55-65. DOI: 10.1016/j.isprsjprs.2015.10.005

SLU, 2016. Skogsdata 2016. Institutionen för skoglig resurshushållning, Report, Umeå

SLU, 2017. Skogsdata 2017. Institutionen för skoglig resurshushållning, Report, Umeå.

SLU. 2018. Riksskogstaxeringen. Officiell statistik om de svenska skogarna. Retrieved Apr 2018, from http://skogsstatistik.slu.se/pxweb/en/OffStat/OffStat__ProduktivSkogsmark__Virkesf%C3%B6rr%C3%A5d/PS_Virkesf_tr%C3%A4dslag_diameter_tab.px/table/tableViewLayout2/?rxid=0e29b5c6-ad64-4c2e-87e1-22a729055c2a

Sothe, C., C. Almeida, V. Liesenberg, and M. Schimalski. 2017. Evaluating Sentinel-2 and Landsat-8 Data to Map Sucessional Forest Stages in a Subtropical Forest in Southern Brazil. *Remote Sensing*, 9. DOI: 10.3390/rs9080838

Sweeney, S. P., and T. P. Evans. 2012. An edge-oriented approach to thematic map error assessment. *Geocarto International*, 27**:** 31-56. DOI: 10.1080/10106049.2011.622052

Szuster, B. W., Q. Chen, and M. Borger. 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography*, 31**:** 525-532. DOI: 10.1016/j.apgeog.2010.11.007

Thanh Noi, P., and M. Kappas. 2017. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel)*, 18. DOI: 10.3390/s18010018

Vaglio Laurin, G., N. Puletti, W. Hawthorne, V. Liesenberg, P. Corona, D. Papale, Q. Chen, and R. Valentini. 2016. Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data. *Remote Sensing of Environment*, 176**:** 163-176. DOI: 10.1016/j.rse.2016.01.017

Wang, G., and Q. Weng. 2013. *Remote Sensing of Natural Resources*. Taylor & Francis.

Wang, Q., G. A. Blackburn, A. O. Onojeghuo, J. Dash, L. Zhou, Y. Zhang, and P. M. Atkinson. 2017. Fusion of Landsat 8 OLI and Sentinel-2 MSI Data. *IEEE Transactions on Geoscience and Remote Sensing*, 55**:** 3885-3899. DOI: 10.1109/tgrs.2017.2683444

Wolter, P. T., P. A. Townsend, and B. R. Sturtevant. 2009. Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data. *Remote Sensing of Environment*, 113**:** 2019-2036. DOI: 10.1016/j.rse.2009.05.009

Wu, Q., and L. Sun. 2008. Sampling methods using remote sensing and global positioning system for crop acreage estimation at national scale in China. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, XXXVII**:** 1337-1342.

Zhu, X. X., D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer. 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5**:** 8-36. DOI: 10.1109/mgrs.2017.2762307

Zhu, Z., U. Yang, S. V. Stehman, and R. L. Czaplewski. 2000. Accuracy Assessment for the U.S. Geological Survey Regional Land-Cover Mapping Program: New York and New Jersey Region. *Photogrammetric Engineering & Remote Sensing*, 66**:** 1425-1435.

Appendix 1: Complete classification accuracy assessment

Classification 1 – 40 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	91.7	0.87	Pine	87.8	87.6	93.3	0.90	Pine	91.4	89.0
			Spruce	87.8	95.5			Spruce	89.7	96.2
			Birch	99.5	92.2			Birch	98.8	94.9
Intermediate	90.2	0.85	Pine	83.9	88.6	91.1	0.87	Pine	88.1	87.6
			Spruce	89.4	93.2			Spruce	89.9	95.0
			Birch	97.2	88.9			Birch	95.2	90.8
Edge	84.0	0.76	Pine	82.4	75.5	74.3	0.61	Pine	81.3	59.4
			Spruce	83.0	88.2			Spruce	69.8	84.8
			Birch	86.7	89.6			Birch	71.9	89.0
Entire patch	88.6	0.83	Pine	84.7	83.6	86.3	0.79	Pine	86.9	76.7
			Spruce	86.8	92.3			Spruce	83.2	92.4
			Birch	94.4	90.2			Birch	88.7	91.8

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	96.6	0.95	Pine	94.7	94.9	92.5	0.89	Pine	85.3	92.0
			Spruce	96.0	97.6			Spruce	93.2	92.8
			Birch	98.9	97.3			Birch	98.9	92.6
Intermediate	86.5	0.80	Pine	83.6	79.3	89.8	0.85	Pine	80.5	90.1
			Spruce	89.0	90.3			Spruce	90.1	90.7
			Birch	87.0	90.4			Birch	97.9	88.6
Edge	68.4	0.53	Pine	79.5	52.5	84.2	0.76	Pine	77.9	79.0
			Spruce	64.7	81.2			Spruce	85.1	85.4
			Birch	61.1	88.9			Birch	89.5	88.0
Entire patch	83.8	0.76	Pine	86.0	72.3	88.8	0.83	Pine	81.4	86.8
			Spruce	83.2	90.3			Spruce	89.7	89.7
			Birch	82.3	92.7			Birch	95.4	89.8

Classification 2 – 32 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	92.3	0.88	Pine	89.0	88.4	94.5	0.92	Pine	92.6	91.3
			Spruce	88.8	95.6			Spruce	91.9	96.7
			Birch	99.1	93.1			Birch	99.1	95.8
Intermediate	91.1	0.87	Pine	84.8	90.5	91.9	0.88	Pine	89.0	89.1
			Spruce	90.4	93.4			Spruce	91.5	95.2
			Birch	98.3	89.6			Birch	95.4	91.7
Edge	84.7	0.77	Pine	82.8	76.6	73.1	0.60	Pine	80.6	57.9
			Spruce	82.9	88.6			Spruce	65.8	83.8
			Birch	88.2	89.8			Birch	73.0	88.6
Entire patch	89.4	0.84	Pine	85.5	84.9	86.5	0.80	Pine	87.4	77.0
			Spruce	87.4	92.6			Spruce	83.1	92.4
			Birch	95.2	90.8			Birch	89.2	92.3

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	97.1	0.96	Pine	95.8	95.3	93.4	0.90	Pine	86.5	93.4
			Spruce	96.3	98.2			Spruce	94.5	93.6
			Birch	99.0	97.6			Birch	99.0	93.1
Intermediate	87.3	0.81	Pine	84.0	80.9	90.5	0.86	Pine	81.5	91.4
			Spruce	89.1	90.4			Spruce	91.6	90.6
			Birch	88.7	90.8			Birch	98.4	89.6
Edge	68.4	0.53	Pine	79.3	52.5	84.3	0.74	Pine	79.5	78.2
			Spruce	64.6	81.2			Spruce	84.4	86.1
			Birch	61.3	88.5			Birch	88.8	88.7
Entire patch	84.3	0.76	Pine	86.4	72.9	89.3	0.84	Pine	82.5	87.2
			Spruce	83.3	90.5			Spruce	90.1	90.1
			Birch	83.0	92.8			Birch	95.4	90.5

Classification 3 – 25 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	92.3	0.88	Pine	89.0	88.4	94.5	0.92	Pine	92.6	91.3
			Spruce	88.8	95.6			Spruce	91.9	96.7
			Birch	99.1	93.1			Birch	99.1	95.8
Intermediate	91.1	0.87	Pine	84.8	90.5	91.9	0.88	Pine	89.0	89.1
			Spruce	90.4	93.4			Spruce	91.5	95.2
			Birch	98.3	89.6			Birch	95.4	91.7
Edge	84.7	0.77	Pine	82.8	76.6	73.1	0.60	Pine	80.6	57.9
			Spruce	82.9	88.6			Spruce	65.8	83.8
			Birch	88.2	89.8			Birch	73.0	88.6
Entire patch	89.4	0.84	Pine	85.5	84.9	86.5	0.80	Pine	87.4	77.0
			Spruce	87.4	92.6			Spruce	83.1	92.4
			Birch	95.2	90.8			Birch	89.2	92.3

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	97.4	0.96	Pine	96.2	96.0	94.1	0.91	Pine	87.8	94.2
			Spruce	96.6	98.2			Spruce	95.2	94.5
			Birch	99.3	98.0			Birch	99.2	93.5
Intermediate	87.3	0.81	Pine	84.9	80.3	91.5	0.87	Pine	83.9	92.2
			Spruce	89.2	91.0			Spruce	92.2	91.7
			Birch	87.6	91.0			Birch	98.5	90.8
Edge	67.9	0.52	Pine	78.9	51.9	85.4	0.78	Pine	80.4	79.9
			Spruce	63.7	81.3			Spruce	86.1	86.8
			Birch	61.1	87.6			Birch	89.7	89.6
Entire patch	84.2	0.76	Pine	86.7	72.7	90.3	0.85	Pine	84.0	88.5
			Spruce	83.2	90.8			Spruce	91.2	91.0
			Birch	82.7	92.8			Birch	95.8	91.3

Classification 4 – 20 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	92.2	0.88	Pine	88.5	88.2	95.1	0.93	Pine	93.5	91.9
			Spruce	89.0	95.4			Spruce	92.7	96.6
			Birch	99.0	93.0			Birch	99.1	96.9
Intermediate	91.3	0.87	Pine	84.9	90.8	92.1	0.88	Pine	89.6	88.8
			Spruce	90.7	93.8			Spruce	90.9	95.7
			Birch	98.1	89.4			Birch	95.7	91.9
Edge	83.9	0.76	Pine	82.0	75.7	73.0	0.59	Pine	80.1	57.8
			Spruce	82.0	87.2			Spruce	65.2	83.2
			Birch	87.9	89.8			Birch	73.6	88.6
Entire patch	89.1	0.84	Pine	85.1	84.6	86.7	0.80	Pine	87.8	77.1
			Spruce	87.2	92.1			Spruce	82.7	92.4
			Birch	95.0	90.8			Birch	89.5	92.7

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	97.4	0.96	Pine	96.4	95.8	94.0	0.91	Pine	88.0	93.7
			Spruce	96.8	98.0			Spruce	94.8	93.9
			Birch	98.9	98.4			Birch	99.1	94.2
Intermediate	87.1	0.81	Pine	83.7	80.8	90.9	0.86	Pine	82.2	91.5
			Spruce	88.5	89.7			Spruce	92.3	90.7
			Birch	89.1	91.2			Birch	98.1	90.6
Edge	67.4	0.51	Pine	77.8	51.2	84.3	0.76	Pine	79.5	77.9
			Spruce	62.8	79.2			Spruce	83.9	85.4
			Birch	61.5	88.2			Birch	89.4	89.7
Entire patch	83.9	0.76	Pine	86.0	72.6	89.71	0.85	Pine	83.2	87.4
			Spruce	82.6	89.7			Spruce	90.4	90.0
			Birch	83.2	93.1			Birch	95.6	91.5

Classification 5 – 16 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	93.7	0.91	Pine	90.7	90.6	96.4	0.95	Pine	94.5	94.6
			Spruce	90.8	95.8			Spruce	95.1	97.3
			Birch	99.6	94.7			Birch	99.6	97.2
Intermediate	91.5	0.87	Pine	85.2	91.0	92.9	0.89	Pine	89.8	90.5
			Spruce	91.1	93.2			Spruce	92.6	95.4
			Birch	98.3	90.5			Birch	96.3	92.8
Edge	84.0	0.76	Pine	82.2	75.3	73.2	0.60	Pine	81.0	58.0
			Spruce	81.1	87.2			Spruce	66.0	83.9
			Birch	88.6	90.6			Birch	72.5	88.8
Entire patch	89.7	0.85	Pine	86.0	85.2	87.5	0.81	Pine	88.4	78.4
			Spruce	87.6	92.1			Spruce	84.5	92.7
			Birch	95.5	92.0			Birch	89.5	93.3

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	97.9	0.97	Pine	96.3	97.4	94.6	0.92	Pine	90.0	93.7
			Spruce	97.9	97.9			Spruce	94.8	94.8
			Birch	99.5	98.4			Birch	99.0	95.2
Intermediate	87.6	0.81	Pine	84.0	81.6	92.1	0.88	Pine	84.9	92.4
			Spruce	89.0	90.2			Spruce	93.2	91.7
			Birch	89.8	91.2			Birch	98.0	92.1
Edge	67.7	0.52	Pine	78.6	51.8	84.4	0.77	Pine	79.4	78.3
			Spruce	63.1	80.3			Spruce	84.0	85.6
			Birch	61.4	88.2			Birch	89.8	89.4
Entire patch	84.4	0.77	Pine	86.3	73.2	90.4	0.86	Pine	84.8	87.9
			Spruce	83.4	90.2			Spruce	90.7	90.7
			Birch	83.6	93.1			Birch	95.6	92.3

Classification 6 – 12 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	91.5	0.87	Pine	87.1	87.6	95.0	0.93	Pine	94.4	91.1
			Spruce	89.3	94.2			Spruce	92.5	97.6
			Birch	98.1	92.8			Birch	98.1	96.8
Intermediate	90.8	0.86	Pine	84.8	89.3	91.4	0.87	Pine	87.9	88.1
			Spruce	90.2	93.5			Spruce	90.9	94.6
			Birch	97.4	89.6			Birch	95.5	91.7
Edge	82.9	0.74	Pine	80.9	74.1	72.1	0.58	Pine	81.0	56.8
			Spruce	80.5	86.2			Spruce	64.0	82.7
			Birch	87.2	89.6			Birch	71.2	89.1
Entire patch	88.4	0.83	Pine	84.2	83.2	86.2	0.79	Pine	87.8	76.0
			Spruce	86.7	91.3			Spruce	82.5	92.2
			Birch	94.2	90.7			Birch	88.3	92.8

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	97.0	0.96	Pine	96.1	95.0	93.5	0.90	Pine	86.7	93.6
			Spruce	96.2	97.8			Spruce	94.6	92.9
			Birch	98.8	98.3			Birch	99.1	93.9
Intermediate	86.4	0.80	Pine	82.9	79.4	91.0	0.86	Pine	83.4	90.7
			Spruce	87.5	89.6			Spruce	91.6	91.6
			Birch	88.8	90.8			Birch	97.9	90.6
Edge	67.2	0.51	Pine	78.3	51.4	83.2	0.75	Pine	78.1	76.5
			Spruce	61.1	79.2			Spruce	82.7	84.2
			Birch	62.1	88.2			Birch	88.7	89.1
Entire patch	83.5	0.75	Pine	85.8	71.8	89.2	0.84	Pine	82.8	86.6
			Spruce	81.5	89.6			Spruce	89.7	89.7
			Birch	83.3	92.9			Birch	95.3	91.2

Classification 7 – 9 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	92.0	0.88	Pine	87.5	88.4	95.7	0.94	Pine	93.8	93.4
			Spruce	90.4	93.8			Spruce	94.2	97.1
			Birch	98.0	93.7			Birch	99.1	96.7
Intermediate	91.1	0.87	Pine	85.6	89.6	91.8	0.88	Pine	87.6	89.0
			Spruce	90.6	94.4			Spruce	91.0	94.6
			Birch	97.3	89.6			Birch	96.8	92.0
Edge	81.7	0.73	Pine	80.3	72.2	72.4	0.59	Pine	79.4	57.5
			Spruce	79.9	85.2			Spruce	65.7	82.3
			Birch	84.9	89.3			Birch	72.0	87.7
Entire patch	88.3	0.82	Pine	84.5	82.9	86.6	0.80	Pine	87.0	77.4
			Spruce	86.9	91.2			Spruce	83.6	91.9
			Birch	93.4	90.9			Birch	89.3	92.5

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	98.1	0.97	Pine	96.8	97.6	92.9	0.89	Pine	86.4	91.8
			Spruce	98.7	98.6			Spruce	93.7	91.9
			Birch	98.9	98.3			Birch	98.5	94.8
Intermediate	87.2	0.81	Pine	84.0	80.2	91.2	0.87	Pine	83.8	90.9
			Spruce	88.2	90.7			Spruce	91.7	92.2
			Birch	89.3	91.2			Birch	98.0	90.5
Edge	67.7	0.52	Pine	78.6	51.9	82.5	0.74	Pine	78.0	75.6
			Spruce	60.9	79.6			Spruce	82.1	83.0
			Birch	63.7	88.3			Birch	87.4	89.3
Entire patch	84.3	0.77	Pine	86.5	73.0	88.8	0.83	Pine	82.8	85.7
			Spruce	82.5	90.5			Spruce	89.1	89.1
			Birch	84.0	93.1			Birch	94.7	91.6

Classification 8 – 7 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	90.7	0.86	Pine	84.9	86.9	94.8	0.92	Pine	92.5	91.6
			Spruce	88.5	93.3			Spruce	93.2	96.5
			Birch	98.6	91.8			Birch	98.6	96.0
Intermediate	89.8	0.85	Pine	84.1	87.1	90.8	0.86	Pine	88.0	86.4
			Spruce	89.9	93.6			Spruce	90.7	94.6
			Birch	95.2	88.7			Birch	93.8	91.7
Edge	80.5	0.71	Pine	78.8	70.7	72.0	0.58	Pine	77.7	57.6
			Spruce	80.7	83.9			Spruce	68.5	80.8
			Birch	82.1	88.9			Birch	69.7	86.8
Entire patch	87.0	0.80	Pine	82.6	81.1	85.8	0.79	Pine	86.0	76.5
			Spruce	86.4	90.3			Spruce	84.1	91.0
			Birch	91.9	89.8			Birch	87.4	91.8

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	97.7	0.97	Pine	96.1	96.9	91.7	0.88	Pine	84.5	90.1
			Spruce	97.7	98.0			Spruce	91.9	92.1
			Birch	99.2	98.1			Birch	98.7	92.7
Intermediate	83.9	0.76	Pine	81.8	74.5	90.3	0.86	Pine	82.8	89.5
			Spruce	85.5	89.5			Spruce	91.2	92.2
			Birch	84.5	89.3			Birch	96.6	89.4
Edge	66.3	0.49	Pine	76.3	50.7	81.0	0.72	Pine	76.9	73.3
			Spruce	60.7	76.8			Spruce	82.3	82.4
			Birch	61.8	87.7			Birch	83.8	88.1
Entire patch	82.6	0.74	Pine	84.7	70.7	87.7	0.82	Pine	81.4	83.8
			Spruce	81.3	88.9			Spruce	88.6	88.9
			Birch	81.8	92.2			Birch	93.0	90.1

Classification 9 – 5 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	92.5	0.89	Pine	87.1	90.2	96.1	0.94	Pine	94.2	94.2
			Spruce	91.9	94.3			Spruce	94.9	97.6
			Birch	98.5	93.0			Birch	99.3	96.6
Intermediate	91.0	0.86	Pine	84.7	89.5	91.5	0.87	Pine	88.1	87.9
			Spruce	92.0	92.6			Spruce	92.7	94.7
			Birch	96.2	90.7			Birch	93.9	92.1
Edge	81.5	0.72	Pine	81.5	72.1	69.6	0.54	Pine	75.4	54.9
			Spruce	80.3	84.9			Spruce	67.0	78.7
			Birch	82.7	89.6			Birch	66.4	85.6
Entire patch	88.3	0.82	Pine	84.4	83.2	85.8	0.79	Pine	85.9	76.3
			Spruce	88.1	90.7			Spruce	84.8	90.9
			Birch	92.4	91.2			Birch	86.5	91.9

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	98.2	0.97	Pine	96.3	98.2	91.6	0.87	Pine	84.5	88.9
			Spruce	98.7	98.3			Spruce	92.1	91.8
			Birch	99.6	98.1			Birch	98.3	93.0
Intermediate	84.3	0.76	Pine	81.6	75.2	90.9	0.86	Pine	83.7	90.2
			Spruce	86.3	89.6			Spruce	92.2	92.0
			Birch	85.0	89.4			Birch	96.6	90.3
Edge	65.6	0.48	Pine	73.0	50.0	80.7	0.71	Pine	77.9	71.9
			Spruce	61.9	74.6			Spruce	80.0	83.1
			Birch	61.8	87.1			Birch	84.1	88.3
Entire patch	82.7	0.74	Pine	83.5	71.1	87.7	0.82	Pine	82.0	83.4
			Spruce	82.3	88.1			Spruce	88.1	89.1
			Birch	82.1	92.1			Birch	93.0	90.6

Classification 10 – 4 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	92.1	0.88	Pine	87.1	89.3	96.0	0.94	Pine	94.2	93.8
			Spruce	90.1	93.5			Spruce	94.2	97.2
			Birch	98.4	93.5			Birch	99.6	97.0
Intermediate	91.0	0.86	Pine	84.9	89.4	91.6	0.87	Pine	88.5	87.9
			Spruce	91.7	92.8			Spruce	92.5	94.4
			Birch	96.3	90.6			Birch	93.9	92.6
Edge	80.4	0.71	Pine	80.2	70.7	70.1	0.55	Pine	75.0	54.9
			Spruce	78.2	83.7			Spruce	67.6	79.1
			Birch	82.7	88.8			Birch	67.6	86.6
Entire patch	87.8	0.82	Pine	84.1	82.5	85.9	0.79	Pine	85.9	76.4
			Spruce	87.0	90.1			Spruce	84.8	90.7
			Birch	92.4	91.1			Birch	87.0	92.6

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	98.2	0.97	Pine	96.5	98.1	91.3	0.87	Pine	83.6	90.0
			Spruce	98.8	98.5			Spruce	91.1	91.2
			Birch	99.3	98.1			Birch	99.0	92.4
Intermediate	83.4	0.75	Pine	78.9	74.9	90.7	0.86	Pine	84.0	90.1
			Spruce	86.7	86.8			Spruce	91.8	92.6
			Birch	84.6	89.2			Birch	96.2	89.3
Edge	64.6	0.47	Pine	70.2	48.8	80.4	0.71	Pine	78.6	71.7
			Spruce	61.4	72.3			Spruce	79.2	82.7
			Birch	62.3	87.7			Birch	83.5	88.2
Entire patch	82.1	0.73	Pine	81.9	70.6	87.5	0.81	Pine	82.0	83.3
			Spruce	82.3	86.6			Spruce	87.5	88.9
			Birch	82.1	92.1			Birch	92.9	90.0

Classification 11 – 3 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	90.4	0.86	Pine	82.6	88.2	94.8	0.92	Pine	91.6	92.7
			Spruce	91.4	91.5			Spruce	94.4	95.2
			Birch	97.4	91.5			Birch	98.4	96.5
Intermediate	89.9	0.85	Pine	82.7	88.1	90.5	0.86	Pine	86.1	86.4
			Spruce	92.3	92.5			Spruce	91.8	94.1
			Birch	94.8	89.1			Birch	93.5	91.1
Edge	79.2	0.69	Pine	80.1	68.3	69.0	0.53	Pine	74.0	54.1
			Spruce	77.2	84.4			Spruce	66.5	79.1
			Birch	80.3	88.1			Birch	66.4	84.0
Entire patch	86.5	0.80	Pine	81.8	80.4	84.8	0.77	Pine	84.0	75.1
			Spruce	87.0	89.6			Spruce	84.2	90.0
			Birch	90.7	89.6			Birch	86.1	91.0

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	97.6	0.96	Pine	95.6	97.1	89.3	0.84	Pine	80.3	87.0
			Spruce	98.8	98.3			Spruce	91.3	90.4
			Birch	98.4	97.4			Birch	96.4	90.4
Intermediate	81.0	0.71	Pine	76.9	70.6	89.0	0.83	Pine	81.1	86.9
			Spruce	86.1	87.2			Spruce	92.3	91.1
			Birch	79.9	86.6			Birch	93.5	88.9
Edge	62.5	0.44	Pine	66.4	46.7	77.2	0.67	Pine	76.8	67.9
			Spruce	61.7	71.8			Spruce	77.5	83.3
			Birch	59.3	82.8			Birch	80.3	85.6
Entire patch	80.3	0.71	Pine	79.7	68.3	85.5	0.78	Pine	79.4	79.7
			Spruce	82.2	86.4			Spruce	87.0	88.4
			Birch	79.1	89.7			Birch	90.1	88.4

Classification 12 – 2 bands

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	86.6	0.80	Pine	73.7	84.6	92.5	0.89	Pine	86.7	90.6
			Spruce	89.0	84.1			Spruce	93.2	91.3
			Birch	97.2	90.8			Birch	97.5	95.3
Intermediate	86.0	0.79	Pine	75.6	82.9	87.5	0.81	Pine	81.3	82.5
			Spruce	87.9	85.9			Spruce	89.7	88.6
			Birch	94.5	88.8			Birch	91.4	91.3
Edge	73.8	0.61	Pine	73.1	60.1	64.2	0.46	Pine	71.1	48.3
			Spruce	69.0	79.6			Spruce	55.3	73.1
			Birch	79.2	85.8			Birch	66.1	85.6
Entire patch	82.1	0.73	Pine	74.1	74.3	81.4	0.72	Pine	79.7	70.0
			Spruce	82.0	83.4			Spruce	79.4	85.4
			Birch	90.3	88.6			Birch	85.0	91.2

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	96.4	0.95	Pine	92.3	96.7	85.6	0.78	Pine	72.1	83.1
			Spruce	98.1	95.3			Spruce	88.2	83.4
			Birch	98.7	97.2			Birch	96.4	89.6
Intermediate	77.6	0.66	Pine	71.7	66.1	85.3	0.78	Pine	74.4	81.9
			Spruce	81.3	82.0			Spruce	86.8	85.3
			Birch	79.8	86.3			Birch	94.6	88.1
Edge	59.9	0.40	Pine	64.8	44.3	73.7	0.61	Pine	71.2	61.2
			Spruce	54.8	66.8			Spruce	70.1	77.9
			Birch	60.1	83.9			Birch	79.9	85.4
Entire patch	78.0	0.67	Pine	76.3	65.3	81.5	0.72	Pine	72.6	74.1
			Spruce	78.1	82.4			Spruce	81.7	82.4
			Birch	79.5	89.9			Birch	90.3	87.8

Classification 13 – 1 band

	TDS_{1+2+3}					TDS_{1+2}				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	78.8	0.68	Pine	62.8	70.9	83.3	0.75	Pine	73.3	76.3
			Spruce	80.9	76.4			Spruce	82.6	81.5
			Birch	92.7	87.8			Birch	94.1	91.7
Intermediate	76.8	0.65	Pine	62.7	69.4	80.2	0.70	Pine	73.0	70.7
			Spruce	81.7	77.4			Spruce	82.1	82.6
			Birch	85.9	82.5			Birch	85.5	87.8
Edge	66.9	0.50	Pine	60.2	52.9	62.3	0.43	Pine	66.2	47.4
			Spruce	66.6	68.9			Spruce	57.2	67.6
			Birch	73.9	82.4			Birch	63.4	83.7
Entire patch	74.1	0.61	Pine	61.9	63.5	75.3	0.63	Pine	70.8	62.7
			Spruce	76.4	74.4			Spruce	74.0	77.8
			Birch	84.1	84.3			Birch	81.0	88.1

	TDS_{1}					$TDS_{1,2,3}$				
Forest patch area	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**	**OA (%)**	**Kappa coeff.**	**Class**	**UA (%)**	**PA (%)**
Interior	86.4	0.80	Pine	78.5	81.3	77.4	0.66	Pine	61.5	68.6
			Spruce	85.7	83.7			Spruce	78.3	74.8
			Birch	95.2	94.1			Birch	92.3	87.4
Intermediate	71.8	0.58	Pine	61.5	59.7	75.6	0.63	Pine	60.5	67.4
			Spruce	79.1	73.3			Spruce	80.4	76.5
			Birch	74.7	84.0			Birch	85.9	81.7
Edge	58.6	0.38	Pine	60.4	43.7	66.3	0.50	Pine	59.4	52.3
			Spruce	58.2	61.6			Spruce	66.1	69.0
			Birch	57.3	85.3			Birch	73.5	81.2
Entire patch	72.3	0.58	Pine	66.8	59.3	73.1	0.60	Pine	60.4	61.9
			Spruce	74.3	73.2			Spruce	74.9	73.5
			Birch	75.8	88.3			Birch	83.9	83.5

Appendix 2: R Script

```
### Uncomment if packages are not installed
# install.packages("raster")
# install.packages("caret")
# install.packages("randomForest")
# install.packages("parallel")
# install.packages("foreach")
# install.packages("doParallel")


### Load necessary packages
library(raster)
library(caret)
library(randomForest)
library(parallel)
library(foreach)
library(doParallel)


### Define variables
layerstack_path <- "filepath" # Path to the layer stack containng the
layers to be used as explanatory variables.
training_shape_path <- "filepath" # Path to the ESRI shapefile
(polygon) with the training areas.
validation_shape_path <- "filepath" # Path to the ESRI shapefile
(polygon) with the validation areas.
responseCol <- "column name" # Name of the response column in the
training_shape and validation_shape.
output_raster <- "file path" # Define the file path of the output
raster generated during the classification (include the extension).
seed <- numeric value # Choose a numeric value as the input in
set.seed().
RF_mtry <- numeric value # Enter mtry value to be used in the RF
algorithm.
RF_ntree <- numeric value # Enter ntree value to be used in the RF
algorithm.
```

```
### Loading files
layerstack <-brick(layerstack_path)
names(layerstack) <- paste0("B", c(1:length(names(layerstack))))
training_shape <- shapefile(training_shape_path)
validation_shape <- shapefile(validation_shape_path)

### Creating training and validation data sets
cl <- makeCluster(detectCores())
registerDoParallel(cl)
training_data_frame <-
foreach(i=1:length(unique(training_shape[[responseCol]])),
.packages=c("raster"), .combine = rbind) %dopar% {
  training_data_frame = data.frame(matrix(vector(), nrow = 0, ncol =
length(names(layerstack)) + 1))
  category <- unique(training_shape[[responseCol]])[i]
  categorymap <- training_shape[training_shape[[responseCol]] ==
category,]
  dataSet <- extract(layerstack, categorymap)
  dataSet <- dataSet[!unlist(lapply(dataSet, is.null))]
  if(is(training_shape, "SpatialPointsDataFrame")){
    dataSet <- cbind(dataSet, class = as.numeric(category))
    training_data_frame <- rbind(training_data_frame, dataSet)
  }
  if(is(training_shape, "SpatialPolygonsDataFrame")){
    dataSet <- lapply(dataSet, function(x){cbind(x, class =
as.numeric(rep(category, nrow(x))))})
    df <- do.call("rbind", dataSet)
    training_data_frame <- rbind(training_data_frame, df)
  }
  return(training_data_frame)
}


validation_data_frame <-
foreach(i=1:length(unique(validation_shape[[responseCol]])),
.packages=c("raster"), .combine = rbind) %dopar% {
  validation_data_frame = data.frame(matrix(vector(), nrow = 0, ncol
= length(names(layerstack)) + 1))
  category <- unique(validation_shape[[responseCol]])[i]
  categorymap <- validation_shape[validation_shape[[responseCol]] ==
category,]
  dataSet <- extract(layerstack, categorymap)
  dataSet <- dataSet[!unlist(lapply(dataSet, is.null))]
  if(is(validation_shape, "SpatialPointsDataFrame")){
```

```
    dataSet <- cbind(dataSet, class = as.numeric(category))
    validation_data_frame <- rbind(validation_data_frame, dataSet)
  }
  if(is(validation_shape, "SpatialPolygonsDataFrame")){
    dataSet <- lapply(dataSet, function(x){cbind(x, class =
as.numeric(rep(category, nrow(x))))})
    df <- do.call("rbind", dataSet)
    validation_data_frame <- rbind(validation_data_frame, df)
  }
  return(validation_data_frame)
}


stopCluster(cl)
registerDoSEQ()

### Training the RF algorithm
set.seed(seed)
tunegrid <- expand.grid(.mtry=RF_mtry)
control <- trainControl(method="repeatedcv", number=10, repeats=3,
allowParallel = TRUE)


cl <- makeCluster(detectCores())
registerDoParallel(cl)


RF_model <- train(as.factor(class) ~., method = "rf", data =
training_data_frame, metric="Accuracy", tuneGrid=tunegrid,
ntree=RF_ntree, trControl=control)
# For regression use:
# RF_model <- train(class ~., method = "rf", data =
training_data_frame, metric="RMSE", tuneGrid=tunegrid,
ntree=RF_ntree, trControl=control)


stopCluster(cl)
registerDoSEQ()


print(RF_model)


### Classification stage
beginCluster()
RF_classification <- clusterR(layerstack, raster::predict, args =
list(model = RF_model))
endCluster()
writeRaster(RF_classification,output_raster)
```

```
### Accuracy assessment
predicted_values <- predict(RF_model, validation_data_frame)
confusionMatrix(predicted_values,validation_data_frame$class)
```

Appendix 3: Tree height estimation

In many cases, the acquisition date of the NFI samples did not match acquisition date of the S2 images. Height development curves were applied to height values when a correction was necessary. The equations of these curves are the following (Johansson et al. 2013):

$$H2 = \frac{(H1 + d + r)}{[2 + (4 \times \beta \times A2^{b2})/(H1 - d + r)]}$$

$$d = \beta \times asi^{b2}$$

$$r = [(H1 - d)^2 + 4 \times \beta \times H1 \times A1^{b2}]^{0.5}$$

where H1 is the measured upper tree height, A1 is the tree age, A2 is the tree age on which the tree height is estimated, H2 is the estimated tree height and β, *asi* and *b2* are tabulated values that depend on the tree species (Table A3.1).

Table A3.1 β, *asi* and *b2* values for Scots pine, Norway spruce and birch (Johansson et al. 2013).

Parameter	Pine	Spruce	Birch
asi	25	10	7
β	7395.6	1495.3	394
b2	-1.7829	-1.5978	-1.387

The correlation analysis performed on S2 bands from the summer image produced weak correlation values. The correlation coefficient (r) and the coefficient of determination (R^2) between the reflectance values of each band and tree height are presented in Table A3.2. The results are similar to what was obtained when the GSV was analyzed. The highest correlations were found in bands 6, 7 8 and 8A. They presented a decreasing relationship with tree height values. However, the highest R^2 values were not larger than 0.27. This means that only a 27% of the variation in height values was due to the decreasing trend. It seems that the relationship between the S2 data and height values was not linear, and r and R^2 were not adequately assessing the nature of the data variation.

Table A3.2 Correlation coefficient (r) and coefficient of determination (R^2) obtained between S2 bands from the summer image and the NFI tree height values of field collected samples**.**

S2 band	r	R^2
2	-0.03	0.00
3	-0.22	0.05
4	0.01	0.00
5	-0.31	0.10
6	-0.51	0.26
7	-0.50	0.25
8	-0.46	0.21
8A	-0.49	0.24
11	-0.20	0.04
12	-0.10	0.01

On the other hand, when comparing these results with the results of the band importance analysis performed using the RF algorithm, other bands proved to be more important (Figure A3.1). Bands 5, 4, 6 and 12 were the four most important bands. The accuracy assessment using the statistics generated by the RF algorithm resulted in a RMSE of 5.24 m. The error value is relatively high considering that the height values range between 6 and 27 m.

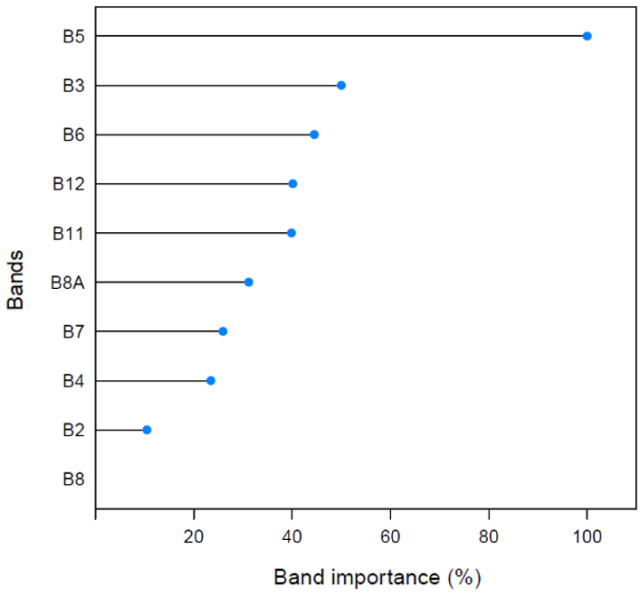

Figure A3.1 Band importance between S2 bands and tree height obtained from the RF regression. Only the bands from the summer image were used for the regression.

The tree height map generated after performing the RF regression is shown in Figure A3.2.

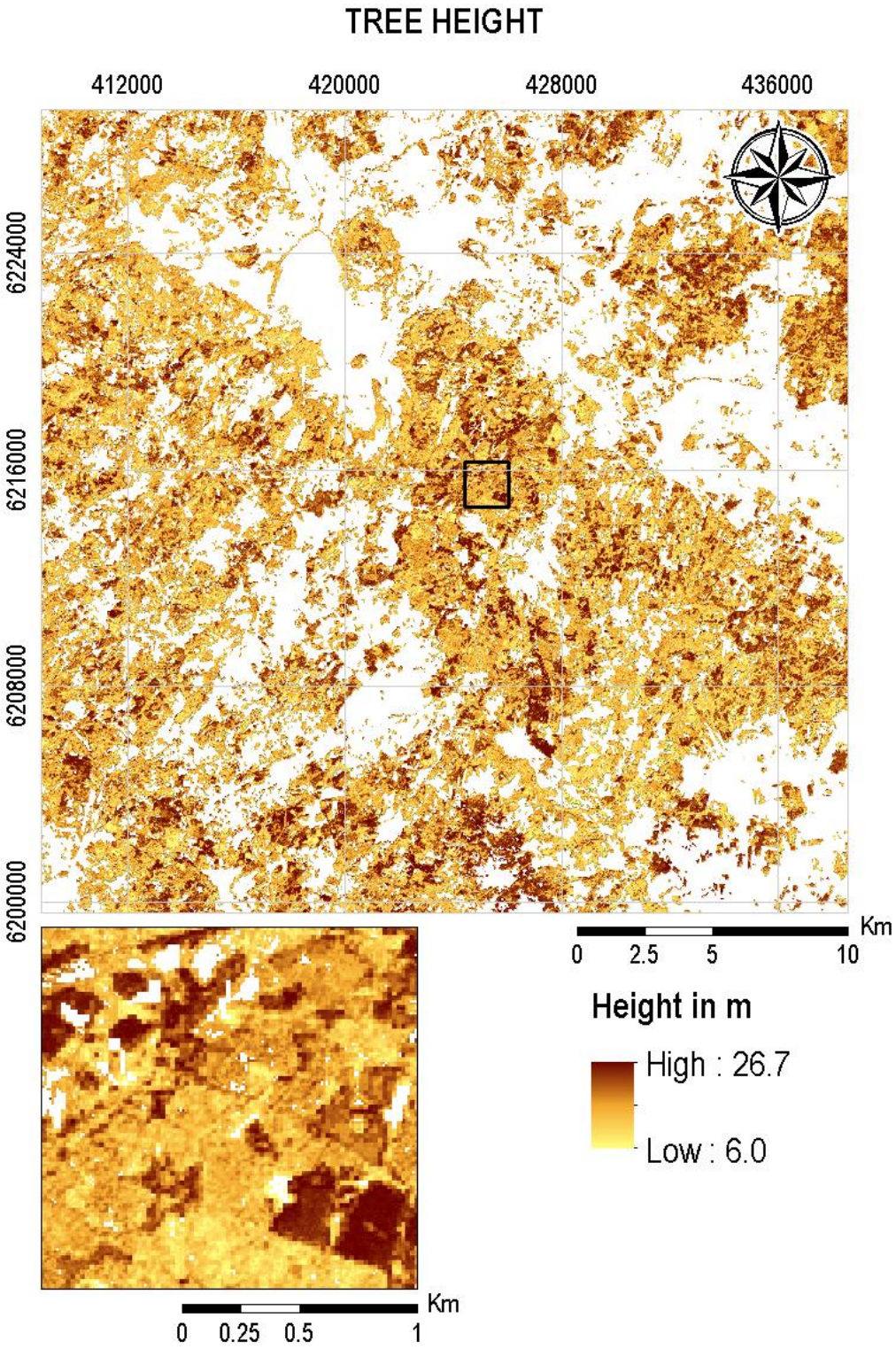

Figure A3.2 Tree height estimated by a RF regression using NFI data and S2 bands from the summer image.

References of Appendix 3

Johansson, U., P. M. Ekö, B. Elfving, T. Johansson, and U. Nilsson, 2013. Nya höjdutvecklingskurvor för bonitering. SLU, Report. [in Swedish, English summary]