Tree species classification using multi-temporal Sentinel-2 data

Trädslagsklassificering med multi-temporalt Sentinel-2 data

Magnus Persson

Work report 486 2018
Master thesis in Forest Sciences 30hp A2E
Forest Sciences – Masters Program

Supervisor:
Heather Reese
Eva Lindberg

Swedish University of Agricultural Sciences
Department of Forest Resource Management
901 83 UMEÅ
www.slu.se/srh
Tfn: 090-786 81 00

ISSN: 1401–1204
ISRN: SLU–SRG–AR–486–SE

Tree species classification using multi-temporal Sentinel-2 data

Trädslagsklassificering med multi-temporalt Sentinel-2 data

Magnus Persson

Keywords: Sentinel-2, tree species, Random Forest, Recursive Feature Elimination.

Master thesis in Forest Sciences at the Department of Forest Resource Management, 30 credits, EX0835, A2E
Forest Sciences – Masters Program
Supervisor: Heather Reese, SLU, Dept. of Forest Resource Management, Remote Sensing
Supervisor: Eva Lindberg, SLU, Dept. of Forest Resource Management, Remote Sensing
Examiner: Mats Nilsson, SLU, Dept. of Forest Resource Management, Remote Sensing

Swedish University of Agricultural Sciences
Department of Forest Resource Management
Publication: Umeå
Publication Year: 2018

ISSN 1401–1204
ISRN SLU–SRG–AR–486–SE

Preface

First and foremost, I would like to thank my main supervisors Heather Reese and assistant supervisor Eva Lindberg of the department of Forest Resource Management for their invaluable support. Their guidance has enabled me to canalise my focus and they have provided me with technical and scientific counselling which made this paper possible. The other staff members of the Forest Resource Department – whom I have hijacked for counselling when the software's stopped working – have also contributed. I would also like to pay tribute to the Ljungbergs' foundation which economical support to the Remote Sensing at SLU lab is indispensable. I also give my regards to the Swedish National Space Board (SNSB) which initiated this project. At last I give my thanks to my fellow colleagues: Carl Jansson, Lisa Wennerlund, and Anders Ellingsson, whom have made the long workdays in the Ljungbergs' remote sensing lab feel short and enjoyable.

Summary

The Sentinel-2 program provides the opportunity to monitor terrestrial ecosystems with a high temporal- and spectral resolution. In this study, the utilization of multi-temporal Sentinel-2 imagery and it's spectral variation due to phenology for classification of common tree species is evaluated at the forest estate Remningstorp in central Sweden.

The tree species classes to be classified were: Norway Spruce (*Picea abies*), Scots Pine (*Pinus silvestris*), Hybrid Larch *(Larix × marschlinsii)*, Silver Birch *(Betula pendula)* and Pedunculate Oak (*Quercus rubur*). The Random Forest classifier (RF) was fitted to four Sentinel-2 images taken during the vegetation period of 2017. The RF classifier was also coupled with the feature selection algorithm Recursive Feature Elimination to form a model with an optimal subset of bands. In addition to the classification, spectral profile plots were constructed for each species to visualize the possibility for identifying the less represented tree species.

The use of four satellite images from April 7th, May 27th, July 9th and October 19th resulted in a higher overall accuracy (86.4 %) compared to using single images (71.5 % – 79.4 %). The late spring image (May 27th) was found to be important since it always was included in the most accurate classifications, independently of the number of images.

The best combination of bands resulted in a model with 87.6 % in overall accuracy and included 37 of 40 bands. The highest ranked bands were all May bands except the red band, the SWIR 1-2 and red bands from April, July and October. The 5 tree species classes were classified with satisfying results and the Producer's Accuracy ranged from 73.7 % to 97.4 %.

Sammanfattning

Sentinel-2 satelliterna möjliggör övervakning av terrestra ekosystem med hög temporal- och spektral upplösning. I denna studie utvärderas möjligheten att nyttja den fenologiska variationen för att klassificera Sveriges vanligaste trädslag på skogsfastigheten Remningstorp i Västra Götaland.

I denna studie användes ett multi-temporalt Sentinel-2 dataset för att klassificera gran (*Picea abies*), tall (*Pinus silvestris*), hybridlärk *(Larix × marschlinsii)*, vårtbjörk *(Betula pendula)* och skogsek (*Quercus rubur*). Klassificeringsmetoden Random Forest (RF) användes för att utvärdera prestandan för olika kombinationer av fyra satellitbilder spridda över vegetationsperioden 2017. Recursive Feature Elimination (RFE) användes också tillsammans med RF för att hitta ett urval av band som bidrog mest till klassificeringens noggrannhet. Dessutom skapades spektrala kurvor för alla trädslag som komplement till klassificeringen och för att visualisera möjligheten att urskilja de mindre förkommande trädslagen på studieområdet.

Det multi-temporala datasetet innehållande alla satellitbilder (7 april, 27 maj, 9 juli och 19 oktober) resulterade i en noggrannhet på 86.4 %. Det är avsevärt bättre resultat än att endast använda enskilda satellitbilder (71.5 % – 79.4 %). Maj bilden var viktig då den alltid var med i den bästa modellen, oberoende av vilken av de andra satellitbilderna den kombinerades med.

Den bästa modellen från RFE resulterade i 87.6 % noggrannhet och innehöll 37 av 40 band. Enligt rankingen från RFE-modellen var de viktigaste banden alla band från maj-bilden utom det röda bandet samt SWIR 1–2 banden från april-, juli- och oktoberbilderna. Dessa resultat stärks av de spektrala kurvorna. Höga värden för ”Producer’s Accuracy” erhölls för gran, skogsek och vårtbjörk (90 %, 97.4 %, 95.6%), medan medelgoda värden erhölls för hybridlärk och tall (81.5 %, 73.7 %).

Table of contents

Introduction

Background

Information about tree species distribution in the forest landscape is valuable for many stakeholders concerned with forest management and forest conservation. Forest companies are interested in information regarding the standing stock and its distribution in age classes and species to plan for future cutting levels, use of machinery, forest management practices and assortment calculations. Knowledge about tree species' distribution in nature reserves and national parks is crucial to form effective treatments for conservation. This information is also indispensable for policy making. In addition, occurrence and spread of invasive species can damage natural ecosystems and are hard to detect and quantify without doing widespread field surveys.

There are many advantages of using remote sensing coupled with field-based forest inventories. Remote sensing usually captures data over larger areas, while field inventories may be sample-based according to a specified sampling design. By building a model – which uses the field inventory data to identify and predict target variables from the remote sensing data – a wall-to-wall prediction can be made for the whole area covered by the remote sensing data, allowing the user to detect areas likely to contain a certain class or value interest.

Sentinel-2

The Sentinel-2 satellite program is a part of the EU-led initiative Copernicus – formerly known as Global Monitoring for Environment and Security (GMES) – and the satellites are equipped with passive, optical sensors. Its purpose is to ensure the EU's capacity to provide and use geospatial information for environment and security monitoring. The European Space Agency (ESA) is responsible for the design, production and maintenance of the GMES Space missions and satellites. The Sentinel-2 mission will complement and bring continuity to other medium resolution satellite-programs, such as SPOT and Landsat (Drusch et al., 2012).

At this moment, Sentinel-2A and 2B are orbiting the Earth and were launched in 2015 and 2017, respectively. Copernicus plan to have the program operational for 15 years, and each satellite has a lifespan of 7.25 years but could stay operational for another 5 years if needed. To that end, two additional satellites will be launched in the coming years to preserve the twin-satellite concept (Drusch et al., 2012).

The system is travelling in a sun-synchronous polar orbit at an altitude of 786 km. The Multi-Spectral Instrument (MSI) is a push-broom system that enables a 290 km swath width, which is the largest to this day compared to other multi-spectral, medium spatial resolution optical missions such as Landsat and SPOT. A large swath width results in a shortened revisit time for Sentinel-2, which is 5 days at the equator and about 3 days closer to the poles, which will increase the possibility to get cloud- or haze free images (Drusch et al., 2012). The MSI is equipped with 13 bands of which 4 are in the red edge spectrum and two in the shortwave infrared spectrum (Table 1).

Table 1. Technical information regarding Sentinel 2 spectral bands and their spatial resolution

Tabell 1. *Teknisk information om Sentinel-2's spektralband och deras spatiala upplösning*

Band number	Name	Central wavelength (nm)	Band width (nm)	Spatial resolution (m)
1	Aerosol*	443	20	60
2	Blue	490	65	10
3	Green	560	35	10
4	Red	665	30	10
5	Red edge 1	705	15	20
6	Red edge 2	740	15	20
7	Red edge 3	783	20	20
8	Near-infra red (NIR)	842	115	10
8a	Red-edge 4	865	20	20
9	Water vapour*	945	20	60
10	Cirrus-cloud detection*	1375	30	60
11	SWIR 1	1610	90	20
12	SWIR 2	2190	180	20

** band not included in this study.*

Earlier studies

The spectral reflectance of vegetation across the wavelengths in the VIS-SWIR spectrum differs and this phenomenon is utilized by remote sensing analysts to separate forest types and tree species with passive sensors.

In the visible part of the spectrum (400 nm to 700 nm) leaves are mainly absorbing light due to the presence of foliar photosynthetic chlorophyll *a* and *b* and caratenoids (Clark and Roberts, 2012; Ustin et al., 2009). Leaf morphology effects how photons are scattering within air-cell wall and results in high reflectance in the NIR spectrum (700 – 1300 nm) (Clark et al., 2005). Water is driving the chemical absorption at 970 nm and 1200 nm and results in a drop in reflectance in these wavelengths (Asner, 1998). Leaves contain cellulose-, nitrogen- and lignin molecules which reflect high in the SWIR spectrum, but the spectral absorption of water overshadows this in vital leaves, but cellulose-, nitrogen- and lignin molecules in dry leaves reflect more since water is absent (Asner, 1998).

Each species' reflectance at different wavelengths is also dependent on the current phenological stage. Flowering, leaf-onset and senescence changes the biophysical and structural properties, which is utilized to differentiate between tree species in multi-temporal satellite datasets (Boyd and Danson, 2005). These levels differ between tree specie's throughout the vegetation period due to leaf development and senescence.

In several studies conducted in the Great Lakes region, USA, improved results have been obtained in classifying forest types and tree species by using multi-temporal Landsat TM imagery (Mickelson et al., 1998; Reese et al., 2002; Wolter and Mladenoff, 1995). However, none of the previously mentioned studies have managed to obtain imagery from the same year since clouds/haze and the low temporal resolution of Landsat TM (16 days) has circumscribed them. Wolter and Mladenoff (1995) emphasize that classifying tree species with optical imagery is a hard task since the spectral variance is often greater within information classes then between them. Schriever & Congalton (2005) and Mickelson et al (1998) notes that satellite images from the start and the end of the growing season are important since spring and fall are when the phenological variation between tree species is the highest.

In recent years, research has shown that the spectral resolution of Sentinel-2 can be utilized for tree species classification. Two recent studies have shown that the red-edge bands and the SWIR-bands in the Sentinel-2 sensor are useful for discriminating between tree species using Sentinel-2 (Immitzer et al., 2016); (Nelson, 2017). The study by Nelson (2017) used multi-temporal Sentinel-2 imagery and a Random Forest (RF) classifier. The study area was situated in Ekerö, central Sweden. The information classes where; Norway Spruce (*Picea abies*), Scots Pine (*Pinus silvestris*), Mixed coniferous forest, Mixed coniferous/Deciduous forest, Deciduous forest, Deciduous hardwood forest, Deciduous forest with hardwoods. One image from each season of the year (May 2nd, July 21st and August 28th) was used, and the effect of combining all images on overall accuracy was evaluated. The results improved in general by using a multi-temporal approach, since it took species-specific phenological changes into account. The lowest overall accuracy was obtained by using only the autumn image from August 28th (~74.8 %). An interesting finding is that the best combination was the spring and the summer images (~85.2 %), hence excluding the early fall image. The best, uncorrelated band-combination of these three dates was red (B4), red edge 2 (B6), red edge 3 (B7) and SWIR-2 (B12). In Immitzer et al. (2016), Norway Spruce, Scots Pine, European Larch *(Larix decidua)*, Silver Fir (*Abies alba*), Common Beech (*Fagus sylvatica*), Oak (*Quercus sp.)* and a mixed class of broadleaves was classified in eastern Bavaria, Germany. RF was used with a single later summer/autumn image (August 13th, 2015). An overall accuracy of 66.2 % was achieved and the red-edge band 1 (B5), SWIR 1 (B11) and, surprisingly, the blue band (B2) were ranked as important variables. The low accuracy was explained with that the satellite image from August 13th failed to represent the spectral variation hone to senescence, few sample plots and heterogeneous species distribution of the field plots.

Spatial resolution and choice of classifier are also factors that influence the classification accuracy. When the spatial resolution is increased, the ability to detect single trees by species increases (Boyd and Danson, 2005). An object-based classification (OBC) with hyper-spectral satellite (HySpex-VNIR 1600 and HySpex-SWIR 320i) was carried out by (Dalponte et al., 2013). The study area was located in southeastern Norway, and the classes were Norway Spruce, Scots Pine and Silver Birch *(Betula pendula)*. It was concluded that finer

spatial resolution significantly increased the classification results of tree species in boreal forests, since the classification accuracy decreased by 20 % when the spatial resolution was resampled from 0.4 m to 1.5 m. Object-based classification and pixel-based classification (PBC) with Sentinel-2 imagery was also evaluated in Immitzer et al (2016), and it was concluded that similar overall accuracies were reached from these two approaches. However, the Kappa statistic increased from 0,357 to 0,588 with OBC. They deducted that the spatial resolution of the Sentinel-2 imagery was considered too low for identifying single tree crowns (objects) in the segmentation but could be applied to groups of trees. Object-based classification procedures are for that reason only applicable to sensors that have a finer spatial resolution than the single objects to be classified and the performance increases thereafter.

Sweden has produced countrywide estimates of forest variables, including stand age and tree species-specific stem volume/ha, within the project SLU Skogskartan, formerly known as kNN Sverige (Reese et al., 2003). Landsat TM and SPOT 5 data coupled with NFI data were used and overall stem volume estimations were reported to have an RMSE of 10 % on forest areas larger than 100 hectares. However, volume estimation for deciduous trees had poor accuracy, due to sparse reference data and use of single date imagery. In the near future, Sweden will produce a new version of SLU Skogskartan using Sentinel-2 data, which will include tree species (Mats Nilsson 2018, pers.comm., 16 Feb).

Background on the Random Forest classifier

The RF classifier is a supervised, non-parametric, ensemble method that has increased in popularity and usage in remote sensing applications in the last two decades for its high performance and ease of use. It is a powerful method that can be used for both regression and classification problems i.e., when the response variable is quantitative or qualitative. It is a robust method since no assumptions of normality are needed, can deal with highly correlated variables and is relatively insensitive to overfitting (Breiman, 2001). The RF algorithm has for that reason become very popular in remote sensing applications, since multi-spectral data are rarely normally distributed and uni-modal.

The RF classifier is built by training an ensemble of decision trees with samples, drawn with replacement, from the original dataset (i.e., bagging). Individual decision trees in the ensemble are formed by stratifying the feature space into regions by applying splitting-questions on the samples at each node. The questions are based on a random sample of predictors - in this case spectral bands - drawn from the original dataset, where the predictor that results in the purest split is chosen. Each split result in two daughter-nodes which are subjected to the same process with a new random sample of predictors. The class that gets the most votes at the terminal nodes is chosen. The measure of the purest split on the subset at each internal node is either the Gini criterion or Entropy. The Gini criterion is a measure of variance for the observation at each terminal node. The aim is to minimize the variance at the terminal nodes which increases the prediction performance. Entropy is a measure of uncertainty for a class or how pure a subset is at the nodes after a split and is supposed to be low. Only two-thirds of the sample from the original dataset are used for training each decision tree, and the last third (Out-Of-Bag) is used for validation in respective tree. It provides an instant estimate of the test error but independent validation is recommended since the OOB-error can be overestimated (Breiman, 2001; Friedman et al., 2001).

The RF classifier overcomes some of the major flaws of decision trees which suffer from high variance and have an inherently lower prediction accuracy. Bagging/bootstrap aggregation is a way to overcome the variance deficiency that individual decision trees suffer from, since the outcome for all decision trees in the ensemble is averaged (Friedman et al., 2001). Decision trees and individual trees in RF are built in the same way, but the latter with the difference that it decorrelates the trees and reduces overfitting of the model. Overfitting happens when the individual decision trees in an RF-model are trained with observations that are too similar. At each split in the tree, a sample *m* of total *p* predictors are candidates for the split (*mtry).* In bagged trees *mtry = p* but in RF *mtry* should be approximately = $p^{\wedge 0,5}$. Consequently, only a few of the predictors are candidates at each split. This procedure prevents the use of the best predictors at each split and decreases the possibility for overfitting (James et al., 2013). The drawback of RF is that the combination of unknown splitting rules at each node and large number of trees makes it hard to interpret (i.e., a black-box model). The variable ranking provides information about important variables but does not describe the form of relationship between the variables (Rodriguez-Galiano et al., 2012).

Non-parametric classifiers tend to outperform parametric classifiers when the complexity of the data increases. This has been demonstrated, for example, by Nitze et al (2012), who performed an agricultural land cover classification with multi-temporal imagery from four dates. Support Vector Machines (SVM), RF, Artificial Neural Network (ANN) and the Maximum Likelihood classifier (MLC) was used. They found that the machine learning algorithms outperformed MLC when all four images were used. RF performed worse than SVM and ANN when only 1 – 2 images were used.

Feature selection

Adding more variables to classification models causes it to increase in dimensionality, which is called the Hughes effect (Hughes, 1968). The high dimensional feature space requires a reference dataset that is sufficiently large to deal with all the predictors. Feature selection is about reducing the number of predictors in the dataset to a subset of predictors that best explains the response variable. Hence, these methods can be thought of as filters that clean the data from variables that do not add predictive power to the prediction. Additionally, feature selection is useful for gaining a better understanding of how the prediction accuracy is affected by excluding predictors (Guyon and Elisseeff, 2003).

Recursive (or Backward) Feature Elimination (RFE) is a feature selection procedure that includes fitting a model to a training dataset which contain all the variables, compute the model performance and then remove the variable with the least negative effect on the models' performance (lowest rank). This procedure is iterated until one variable is left. Additionally, a model with the best subset of variables according to overall accuracy is proposed (Guyon et al., 2002). Variable ranking is usually calculated in the modelling-process for *randomForest* in R and provides an understanding of which variables are important predictors for the response classes. The RFE can end up proposing a model that reaches a higher overall accuracy than done with RF alone. RFE is consequently a good tool for gaining better understanding of how different subsets of variables are affecting the performance of the model.

Clark and Roberts (2012) concluded that parametric models, such as Maximum Likelihood (ML), increased in overall accuracy from 68.3% to 81.6% by reducing the number of variables prior to the classification since noisy correlated variables are excluded from the dataset used for fitting the model. Dalponte et al (2013) noted that the accuracy of non-parametric models, such as RF, and Support Vector Machine (SVM) are not increased by a feature selection prior to the classification since they generally do not suffer from the Hughes phenomena. The opinions differ regarding if linearly correlated variables are a hazard or just redundant for RF. A recent study on the matter showed that a feature selection based on only using the most important, uncorrelated ($p< 0,90$) variables had a significantly higher classification accuracy compared to using all variables (Millard and Richardson, 2015). Guyon and Elisseeff (2003) state that even highly correlated variables can complement each other and that variables that are useless themselves can express important relationships when combined with others.

Research questions

The high temporal resolution and the inclusion of several red-edge bands in Sentinel-2 data makes it expedient for tree species classification since there is a higher likelihood of capturing images with phenological information and with more spectral information than Landsat 8 and SPOT 5. Conclusively, tree species classification in the boreal region could be made with a higher accuracy than done to date. The goal of this study is to perform a tree species classification using multi-temporal Sentinel-2 data with the RF classifier, and to address the following research questions:

1. What combinations of Sentinel-2 satellite image dates contribute to increased overall classification accuracy of tree species in the boreo-nemoral region of Sweden?
2. Which spectral bands are important for classifying tree species in the boreo-nemoral region of Sweden?

Material and methods

Study area

The study was carried out at the forest estate of Remningstorp and the neighbouring nature reserve Eahagen in the county of Västra Götaland in central Sweden (58°30’N, 13°40’E); (Figure 1). This part of Sweden is located within the boreo-nemoral region and the natural forest cover constitutes mainly of conifers along with a minor share of broadleaves. The Remningstorp forest estate is 1500 ha and the predominant silvicultural system used is clear-felling. The forest cover is constituted of Norway Spruce, Scots Pine, Hybrid Larch *(Larix* × *marschlinsii)* and Silver Birch, along with a small share of noble broadleaves.

Eahagen is a nature reserve characterized by a hilly landscape formed by the last ice age, and a large variety of nature types, ranging from wetlands to deciduous forests and meadows. There is a rich diversity of broadleaf tree species and a forest structure that has previously undergone silvo-pastoral practices. The tree species composition consists mainly of broadleaves native to Sweden, such as Pedunculate Oak (*Quercus rubur*), Wych Elm (*Ulmus glabra*), Norway Maple (*Acer platanoides*), Small-leaved lime (*Tilia cordata*), Ash (*Fraxinus excelsior*), Hornbeam (*Carpinus betulus*), and general deciduous tree species such as Wild Cherry (*Prunus avium*), Alder (*Alnus glutinosa*), Silver Birch and Aspen *(Populus tremula)*. The understory is mainly composed of bushy tree species such as Hazel (*Corylus avelana*), which is a common species in silvo-pastoral systems in Sweden.

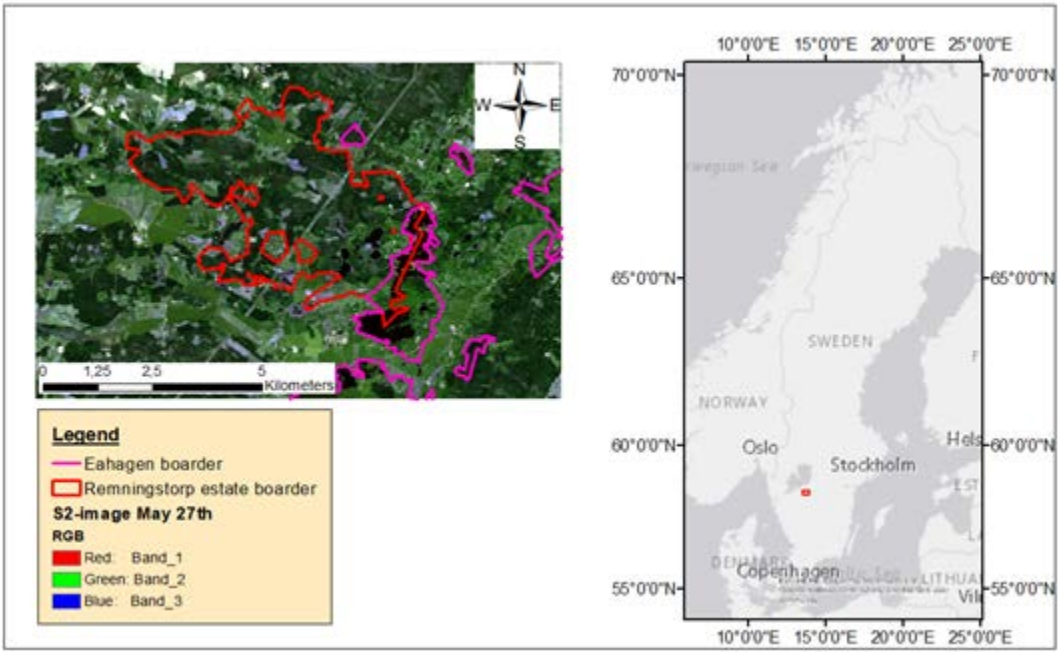

Figure 1. Map of the study area.

Figur 1. *Karta över studieområdet.*

Materials

Field data

The field inventory on Remningstorp was carried out 2016. The sampling design was systematic random sampling with field plots placed out in a grid to ensure a representative sample of the species composition on the estate. This dataset is mainly constituted of Norway Spruce, Scots Pine and Silver Birch. An additional inventory was carried out in the adjacent nature reserve Eahagen to complement the dataset with broadleaf tree species. Regarding the inventory in Eahagen, the location of the plot centre was flexible, since the aim of the inventory was to find plots that were dominated by a single tree species within homogenous stands (Lindberg, 2017).

The plots were reviewed individually, and the tree species composition was derived by calculating the basal area proportion of each tree species. Plots were assigned the information class of the species that made up 70 % or more of the total basal area. The plots which did not fulfil this criterion were not used. Other studies have used the same basal area threshold (Mickelson et al., 1998; Reese et al., 2002).

Remningstorp is actively managed and regeneration fellings could have been carried out after the inventory in 2016. To investigate this, a shapefile representing regeneration fellings carried out in the last decade was downloaded from the website of Skogsstyrelsen (The Swedish Forestry Board). It resulted in the removal of two plots in the Norway Spruce class. Plots located in young plantations - which were younger than 8 years - were omitted from the field data, since lower crown cover would introduce noise from the understory.

The field data in the present study – obtained from the systematic inventory of Remningstorp and Eahagen – was less extensive for some classes and two additional approaches (described below) were included to supplement the less represented classes.

1. Birch and Pedunculate Oak were complemented with field plots during a one-week inventory in 2017, by locating areas in the forest covered by the target species and recording coordinates with a handheld GPS.
2. The forest management plan of Remningstorp was queried for stands that constituted of at least 70 % of the target specie. Plots were placed out subjectively using aerial photointerpretation in parts of the stand which were dominated by a single tree species using an RGB aerial photo and a Colour-IR aerial photo, both with 0,25 × 0,25 m grid cell resolution.

Two shapefiles were created; the first contained all plots and the second only the field plots corresponding to the five information classes. The summary of the field data collection is shown in Table 2 and the location of the field plots divided by tree species is shown in Figure 2.

The information classes included in the study was Norway Spruce, Scots Pine, Hybrid Larch, Pedunculate Oak and Birch. The other four deciduous species (Wynch Elm, Alder, Aspen and Wild Cherry) were not included since they were unrepresented in the field data.

Table 2. Summary of field plots. The number of field plots is presented by inventory method and whether they were included in the classification or not

Tabell 2. *Sammanfattning av provytorna. Antal provytor uppdelade på inventeringsmetod och om de var inkluderade i klassificeringen eller inte*

Tree Species	**Field inventory 1[1]**	**Field inventory 2[2]**	**Aerial photo interpretation**	**Total**
Birch *(Betula ssp)**	24	3	18	45
Hybrid Larch *(Larix × marschlinsii)**	3		24	27
Pedunculate Oak (*Quercus rubur*)*	18	20		38
Scots Pine *(Pinus Silvestris)**	29		28	57
Norway Spruce *(Pices abies)**	100			100
Alder (*Alnus glutinosa*)	9	1		10
Aspen *(Populus tremula)*	3	1		4
Wild Cherry (*Prunus avium*)		5		5
Wych Elm (*Ulmus glabra*)		7		7
Total (Proportion) classification	174 (65,2%)	23 (8,6%)	70 (26,2%)	267
Total (Proportion) all species	186 (63,5%)	37 (12,6%)	70 ((23,9%)	293

[1] *The field inventory carried out in 2016 with systematic sampling design.*

[2] *The field inventory carried out in the fall of 2017 by placing out subjective plots.*

[*] *Tree species that were included in the classification.*

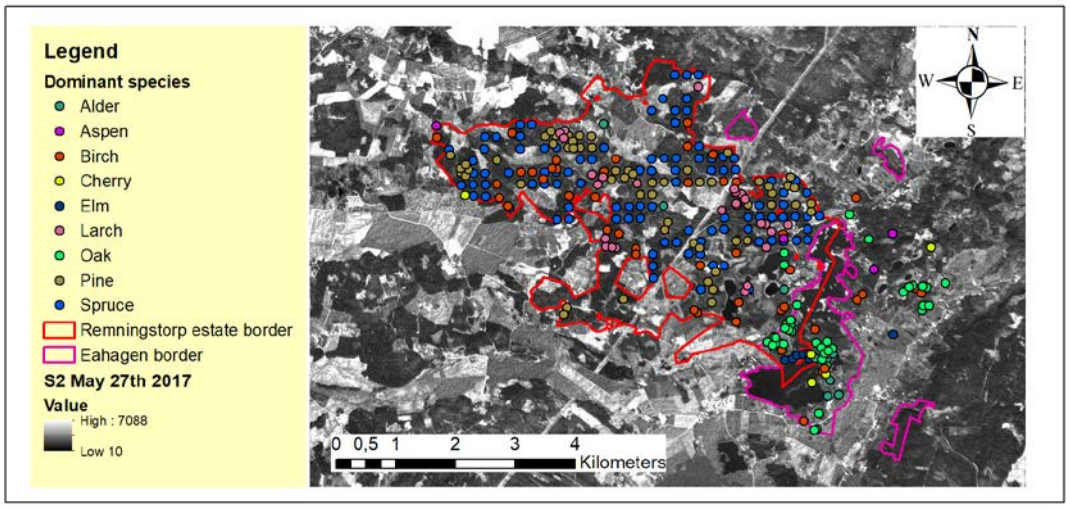

Figure 2. Map showing the geographical location of field plots and their tree species.

Figur 2. *Karta som visar den geografiska positionen för provytorna uppdelat på trädslag.*

Satellite data

To evaluate the seasonal effect of different combinations of satellite images, a subset of imagery from four dates spread over the vegetation period of 2017 was included (Table 3). The quality criteria for the images was no or minimal cloud/haze covering the study area and taken at different phenological stages (i.e., between leaf-out to end of senescence).

Naturkalendern (Bolmgren, 2017) is a phenological network directed by the Swedish University for Agricultural Sciences (SLU). Volunteers observe and report geographical, species-specific phenological events such as leafing, flowering, budding, senescence and leaf-fall, and it is presented in a web-GIS called Naturkalendern (“Naturkalendern,” 2017). This data source was useful for two reasons: finding satellite imagery that resembles a succession in phenology for the species and explaining why certain images are more useful in a classification than others. No records for phenological events are kept for the Remningstorp estate or Eahagen but observation from the surrounding area in the county was reviewed and assumed to be representative of the field plots at hand.

The satellite imagery was obtained from the Swea-portal of the Swedish National Space Board (SNSB) website on 2017-10-05. Images with processing level 1C were readily available, which entails that the data had been corrected for radiometric and geometric discrepancies but not for atmospheric (Drusch et al., 2012). To that end, atmospheric correction was carried out on Top-Of Atmosphere (ToA) Level 1C images using Sentinel Application Platform (SNAP) software provided by ESA (*Sentinel Application Plattform (SNAP)*, 2018). The library Sentinel-2 toolbox includes the algorithm Sen2cor, which transforms Level 1C Sentinel-2 imagery to Level 2A Bottom of Atmosphere (BoA) reflectance (Mueller-Wilm, 2017).

After the corrections, all 10 remaining bands (Table 1) in each of the four scenes were imported to ArcMap 10.5. The 20 m bands were resampled to 10 m spatial resolution with

nearest neighbour and merged into raster datasets corresponding to their original date. The haze- and cloud reduction was assessed by visually comparing the processed image with the non-processed one. Some haze was found in the October 19th image, but not in the three others. A shapefile containing the field plot coordinates was also used to see how the cells aligned with the plot boundary. Some haze covered four Norway Spruce plots and they were excluded since the haze could have disturbing effect.

The geometric correctness was assessed for each image by comparing distinct features in the landscape in each image, but no tendencies for geographic offset were found.

Data processing was further carried out in the statistical software R (RStudio Team, 2016). The rasters were imported as RasterStacks and the geographic coordinate system was set to SWEREF 99 TM. The satellite images were later clipped to the extent of the study area.

Table 3. The Sentinel-2 imagery included in the study

Tabell 3. *Sentinel-2 bilder som användes i studien.*

Image acquisition date	Tile	Granule name
2017-04-07	33VVE	S2A_OPER_MSI_L1C_TL_MPS__20170413T142452_A009357_T33VVE_N02.04
2017-05-27	33VVE	S2A_OPER_MSI_L1C_TL_SGS__20170527T154136_A010072_T33VVE_N02.05
2017-07-09	33VVE	S2A_OPER_MSI_L1C_TL_SGS__20170709T141958_A010687_T33VVE_N02.05
2017-10-19	33VVE	S2B_OPER_MSI_L1C_TL_SGS__20171019T140029_A003237_T33VVE_N02.05

Method

Multi-temporal imagery

Multi-temporal Sentinel-2 imagery introduces a lot of data which benefits from using a classification method that can deal with high dimensional data sets. For this reason, the RF classifier was used to classify the satellite data. To evaluate the multi-temporal approach and the significance of the different image dates, a series of models were fitted according to the subsets in Table 4.

The data processing and modelling was carried out in RStudio along with the latest version of R (3.4.2;(RStudio Team, 2016). The shapefile containing the plots of the 5 information classes (Norway Spruce, Scots Pine, Hybrid Larch, Pedunculate Oak and Birch) was imported. A weighted average of the spectral values – which corresponded to an area-weighted fraction of each cell that was covered by the 10 m radius plot – was extracted for each individual plot and satellite image and stored in separate data sets. The plot radius differed between the inventories which could introduce inconsistencies if the spectral values were extracted from the pixels. To that end, 10 m radius was used for all plots throughout the study, assuming that tree species composition, assigned to the plot, did not change by the reduction of the plot size from 12 m radius plots.

The subsets in each group are every possible combination of satellite images from different dates. The data sets containing the spectral information from all bands extracted from the

Table 4. The different combinations of satellite imagery for which RF-models were fitted

Tabell 4. *Olika kombinationer av satellitbilder för vilka modeller skapades*

Group	Subset	Abbreviation	Number of bands
Single	April 7th	A	10
	October 19th	O	10
	July 9th	J	10
	May 27th	M	10
Double	July 9th/October 19th	JO	20
	April 7th/July 9th	AJ	20
	May 27th/July 9th	MJ	20
	April 7th/October 19th	AO	20
	May 27th/October 19th	MO	20
	April 7th/May 27th	AM	20
Triple	May 27th/July 9th/October 19th	MJO	30
	April 7th/July 9th/October 19th	AJO	30
	April 7th/May 27th/July 9th	AMJ	30
	April 7th/May 27th/October 19th	AMO	30
All	April 7th/May 27th/July 9th/October 19th	AMJO	40

satellite images along with a data set with the class for each plot were merged corresponding to the subset in Table 4. RF models were fitted to each data set using the *randomForest*-package (Liaw and Wiener, 2017). Each model was built with the default settings for the parameters with *ntree* = 500 and *mtry* = $\sqrt{number\ of\ bands}$. The cross-validated overall accuracies for each model were used to evaluate their performance. The RF classifier was implemented with the *randomForest*-package (Liaw and Wiener, 2017) and the models was trained and evaluated with the *caret*-package (Williams et al., 2017).

Data exploration and band importance

Spectral profile plots were constructed for each satellite images as an initial evaluation of the spectral fingerprint of the 5 information classes in the classification and the four additional deciduous tree species: Wynch Elm, Alder, Aspen and Wild Cherry.

The shapefile containing the plots for all nine species was used to extract the spectral information from each Sentinel-2 image and was stored in data sets corresponding to each image. The spectral reflectance was averaged for all plots divided on bands and tree species and plotted. The spectral profile gives an indication of how species' reflectance differs between the wavelength bands and furthermore discloses which bands and from which season that could aid in the classification. Plots for their standard deviation as also formed by plotting the standard deviation

The RFE was used evaluate if a better model can be formed by a subset of bands from each satellite image. The band ranking and the confusion matrix of the best model was later utilized to relate the results to individual tree species. The classification procedure followed the one described for RF, but the lowest ranked band was eliminated after each iteration. New models were formed after each iteration based on sequentially fewer bands and the bands were ranked according to the model with the optimal subset of bands.

The data set containing the subset with all 40 bands and a data set with each plot corresponding class served as input to the RFE model. RFE was implemented with the *caret*-package (Williams et al., 2017).

Accuracy assessment

The models were trained and evaluated with the *caret*-package (Williams et al., 2017), which allows the user to apply an independent validation of the classification instead of relying on the OOB error. Both the classifications models and the RFE were evaluated by a *K*-fold cross-validation approach. The complete dataset is randomly split up into *K*-samples that are about the same size and the classifier is trained with the *K*-1 samples and the model is validated with the *k*th sample. This process is iterated for each *K* and the prediction error for each session is combined and averaged into one estimate of the prediction error (Friedman et al., 2001). In this study, *K*=10 was applied.

Results

The first subchapter states the results for the models created with RF for different subsets of satellite imagery. The second subchapter contains the results of the model created with RF coupled with the feature selection algorithm RFE.

Multi-temporal imagery

Adding images to the model resulted in higher overall accuracy (Figure 3) given that the best combination of satellite images in each group was used (M=79.4 %; MO = 83.4 %; AMO = 86.4 %; AMJO = 86.4 %), but the effect on accuracy appeared to decrease with each additional image.

Single satellite images as input to the RF classifier turned out to perform relatively well in separating the tree species at hand, however there was a notable difference in accuracy between May and the other images. Higher accuracies were obtained if the May image was combined with any image from another date. Regarding the combinations of three images, only the AMO subset outperformed the best double subset. The July image was never in the best combination in any of the double and triple groups, but the April and May image was always in the best combinations. The addition of the July image to AMO combination did not make any difference. Confusion matrices for the single images and the best models in the double and triple group are shown in the Appendix.

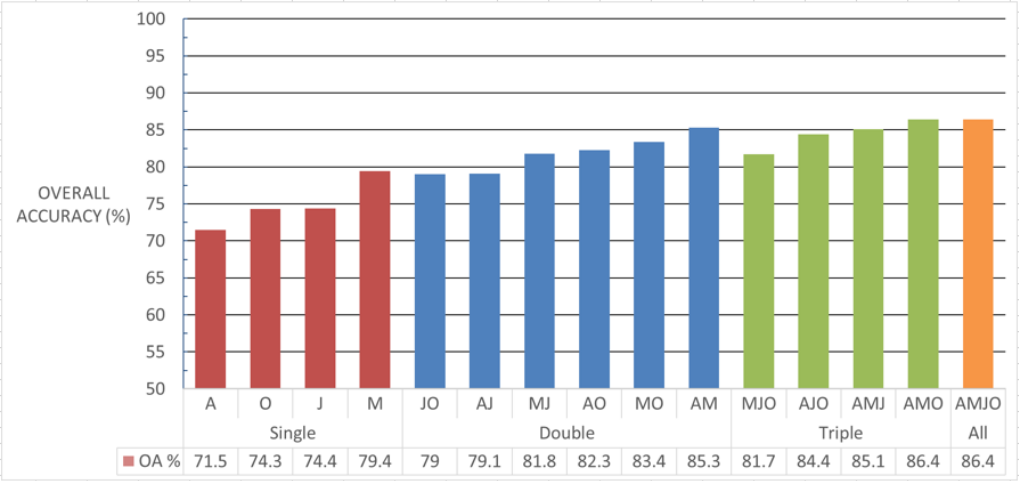

Figure 3. The overall accuracy for all individual RF models produced with all bands ranging from a single image to all images. The colour division is based on the amount of satellite images used in each group and the single letters determine from which month: *A = April 7; M = May 27; J= July 9; O = October 19.*

Figur 3. *Overall accuracy för alla individuella RF-modeller producerade med alla band löpande från en enskild till samtliga satellitbilder. Färgindelningen anger antalet satellitbilder som ingår och de enskilda bokstäverna från vilken månad: A = 7 april; M = 27 maj; J= 7 juli; O = 19 oktober.*

Band importance and tree species separation

The spectral signatures for each tree species gives an indication of how species' reflectance differs between the wavelength bands and furthermore discloses which bands and what season could aid in the classification. The species-specific change in mean reflectance from one image date to another is illustrated in Figure 4. The standard deviations for the spectral curves are shows in Figure 5.

The May 27th and July 9th images show higher values of spectral reflectance for each tree species in each band compared to the April 4th and October 19th image and most obvious separation in reflectance is provided by the bands in the infrared spectrum. Notably, the four red edge bands and NIR in the May and July images have high separation between the tree species, while the SWIR bands in the April and October images also show good separation.

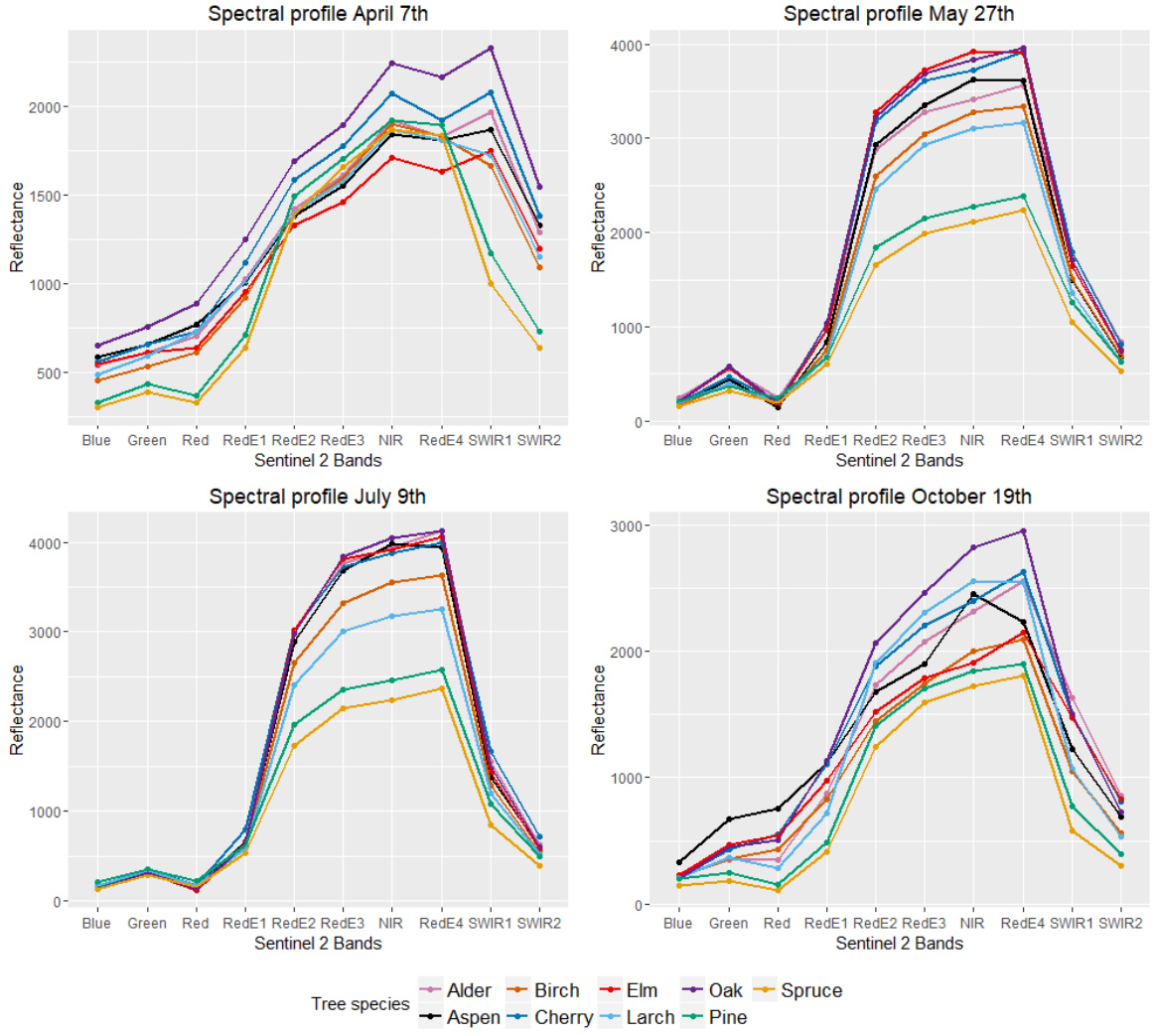

Figure 4. Spectral profile specified per tree species and band for each satellite image.

Figur 4. *Spektrala profiler uppdelat på trädslag, band och satellitbild.*

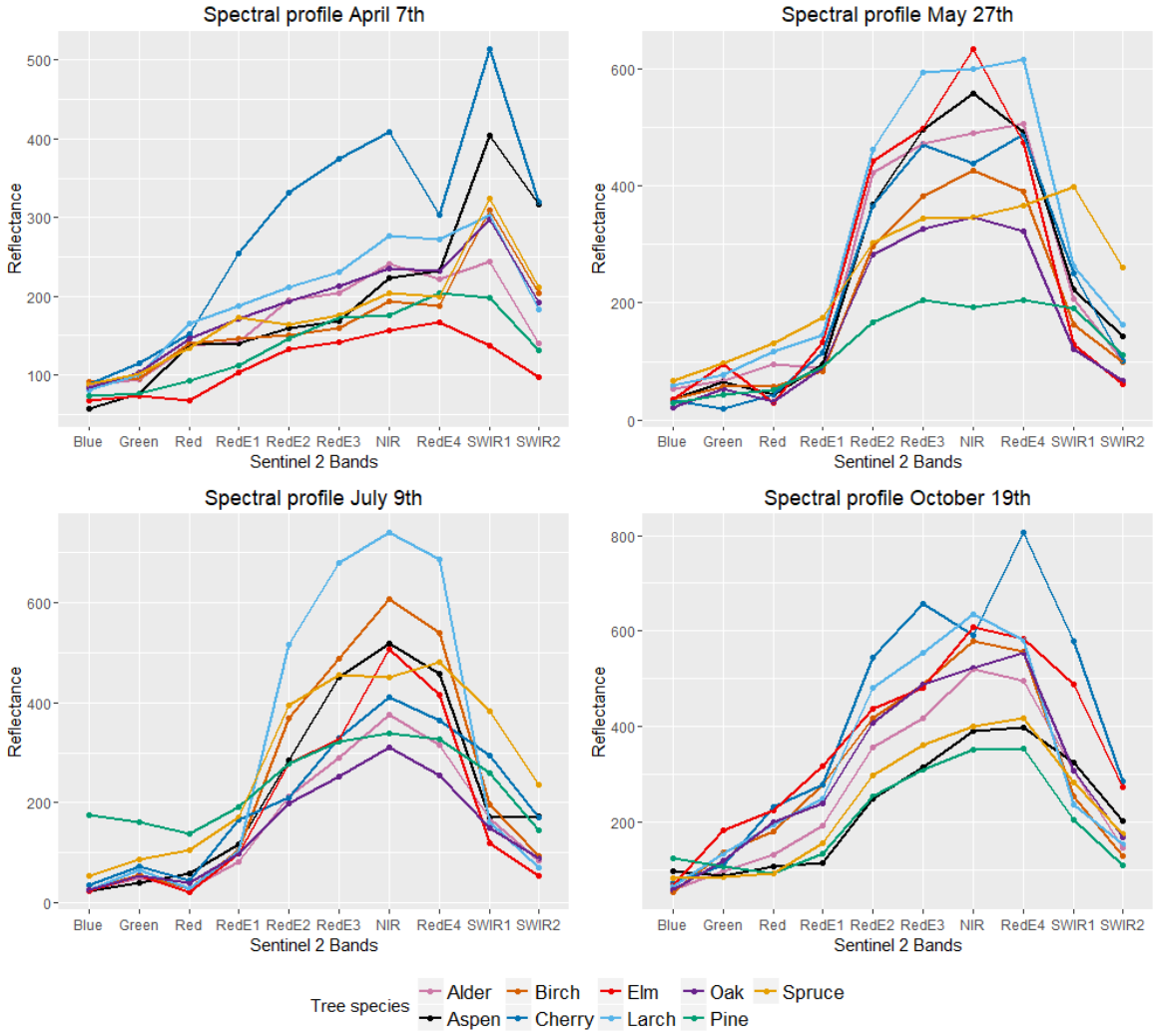

Figure 5. The standard deviation of the spectral reflectance. The plots are specified per tree species and band for each satellite image.

Figur 5. *Standardavvikelsen för den spektrala reflektansen uppdelat på band, trädslag och satellitbild.*

The standard deviation follows the average spectral reflectance in Figure 4, but the magnitude varies between tree species which could affect the discrimination.

The model formed with the optimal band-combination (Table 5) was constituted of 37 of total 40 variables and has an overall accuracy of 87.6 % and a Kappa of 82.9 % (Table 6). The confusion matrix (Table 6) shows the Producer's accuracy and the User's accuracy resulting from this model. Birch, Norway Spruce and Pedunculate Oak have higher Producer's accuracies than Hybrid Larch and Pine. Confusion occurs between Scots Pine with Norway Spruce and between Hybrid Larch and Birch, hence their lower Producer's accuracies.

The highest ranked bands are all bands from May 27th except the red band, the SWIR 1-2 bands from April 7th, July 9th and October 19th. The red band from April 7th, the blue and red edge 4 band from July 9th and the red band from October 19th were also ranked high.

Figure 6 shows the overall accuracy for the models built with a subset of bands. The overall accuracy decreased gradually but a tipping point is reached when one third of the bands are left. Using roughly a third of the total number of bands (19) resulted in a 2.4 % decrease in accuracy compared to the best model of 37 bands. The 19 highest ranked bands in Table 5 are however not the same as this subset but does most likely constitute of a large part of them.

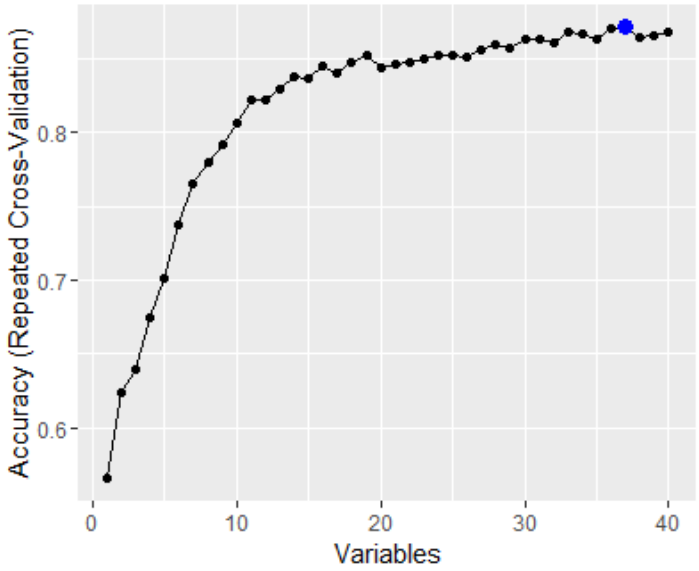

Figure 6. The resulting overall accuracy for every model derived from the RFE. The blue dot marks the model with the optimal number and combination of spectral bands.

Figur 6. *Overall accuracy för alla modeller producerade i den stegvisa variabel elimineringen (RFE). Den blå pricken anger modellen med det optimala antalet och kombinationen* ***spektralband.***

Table 5. The band-ranking of the best model suggested by the RFE

Tabell 5. *Rankningen av band från den bästa RFE-modellen*

Rank	Variable	Date	Rank	Variable	Date
1	RedE2	May 27th	**19**	Red	October 19th
2	SWIR2	April 7th	**20**	NIR	July 9th
3	NIR	May 27th	**21**	RedE3	July 9th
4	SWIR1	October 19th	**22**	RedE1	October 19th
5	RedE3	May 27th	**23**	RedE3	April 7th
6	RedE4	May 27th	**24**	Blue	April 7th
7	Blue	May 27th	**25**	Green	April 7th
8	Green	May 27th	**26**	RedE2	July 9th
9	SWIR1	April 7th	**27**	Red	July 9th
10	RedE1	May 27th	**28**	RedE3	October 19th
11	Blue	July 9th	**29**	RedE2	April 7th
12	SWIR1	May 27th	**30**	Red	May 27th
13	Red	April 7th	**31**	RedE2	October 19th
14	SWIR1	July 9th	**32**	RedE1	April 7th
15	SWIR2	May 27th	**33**	NIR	April 7th
16	SWIR2	October 19th	**34**	RedE4	October 19th
17	RedE4	July 9th	**35**	NIR	October 19th
18	SWIR2	July 9th	**36**	RedE4	April 7th
			37	Green	October 19th

Table 6. The confusion matrix for the RFE-model with the optimal number and combination of spectral bands

Tabell 6. *Confusion matrix för RFE-modellen med det optimala antalet och kombinationen av spektralband*

	Reference data						
	Birch	**Hybrid Larch**	**Pedunculate Oak**	**Scots Pine**	**Norway Spruce**	**Total**	**Producer's accuracy (%)**
Birch	**43**	0	1	0	1	45	95.6
Hybrid Larch	3	**22**	2	0	0	27	81.5
Pedunculate Oak	1	0	**37**	0	0	38	97.4
Scots Pine	2	0	0	**42**	13	57	73.7
Norway Spruce	3	1	1	5	**90**	100	90.0
Total	52	23	41	47	104	**267**	
User's accuracy (%)	82.7	95.7	90.2	89.4	86.5		
Overall accuracy (%)	87.6						
Kappa (%)	82.9						

Discussion

Multi-temporal imagery

The best model of the subsets was constituted of satellite imagery from all dates and had and overall accuracy of 86.4 %. The successive addition of a satellite image increased the overall accuracy if the highest performing model in each group was chosen (M → AM → AMO → AMJO). This means that additional satellite images complement the lack of spectral information contained in single images and that a multi-temporal dataset increases the predictive power of the model. These results align with other studies regarding classifying coniferous and broadleaf forest types conducted in the boreal zone (Nelson, 2017; Reese et al., 2002; Wolter and Mladenoff, 1995).

The same accuracy was reached for the AMO and the AMJO combination. Bands from the July image were redundant in the highest performing models, probably since the tree species reflected similarly to the May image and did not contribute with any additional information. Moreover, the July image only resulted in slightly higher overall accuracy in the triple-subset if it was combined with AO. Mid-summer images have proven to be less accurate for tree species classification compared to spring and fall images (Mickelson et al., 1998; Schriever and Congalton, 2005), which explains the lack of contribution of the July image.

By viewing the spectral profiles, it is obvious that the April, May and October images vary most spectrally. The May image performed well on its own and was always in the best combination; this is likely due to the phenological variation between species being highest in late spring. The April image performed poorly on its own, but together with the May image it made up the best model in the group. The difference in the confusion matrices for May and April/May (appendix, Table 11) is that the Producers accuracies for Birch and Hybrid Larch are greatly improved (+10 % and + 20 %). The reason could be that the two species are captured in leaf-out condition in April. The April/May/October combination obtained slightly higher accuracies (appendix, Table 12). Hybrid Larch was less confused with Birch and Norway Spruce, probably due to difference in leaf-senescence. Conclusively, satellite imagery that coincide with phenological events for tree species are more likely to increase the overall accuracy of the model. Images from the same part of the year can still add information that increases the predictive accuracy but only to some extent. The overall accuracy of the best models in each group chosen (M → AM → AMO → AMJO) did not increase linearly, which also support this claim.

The April image is from a time of year when leaf development has not started. Conifers are probably the only species in the April image that reflect from leaves, which provides an argument for including this image. The nature of the reflectance of the deciduous tree species in Figure 4 is however unknown for early spring and could be attributed to the understory and stem rather than the target tree species. Using satellite images from early spring could for that reason introduce noise which makes the model unreliable and hard to generalize and draw conclusions from. The October image probably has the same characteristics.

The results of this study reinforce the claim that timing of satellite image acquisition is crucial to obtain satisfying accuracy, but that the timing is governed by the occurrence of clouds and haze. The timing and the number of suitable satellite images can be conceived as nearly random since it will vary between years and study areas. The exclusion of the July image would not have made any difference in overall accuracy if one of the four satellite images would have randomly been omitted. Largest drop in accuracy would occur if the May image would fall out. Sentinel-2 A and B has since the spring of 2017 provided new satellite imagery every 2th to 3th day for Sweden, but the occurrence of cloud and haze over the study area resulted in only four suitable images for this study.

The image acquisition time in this study was not optimal since the April image was taken before leaf-onset and the October image is at the end of senescence for all deciduous tree species. Future studies in this region should focus on obtaining additional imagery from the first part of May (if possible) when the reflectance is higher and mid-fall during senescence when there is a gradient in phenological activity. The timing for a species-specific phenological event is dependent on the local climate and the best date for capturing these will therefor differ depending on geographical location.

Band importance

The best model obtained from the RFE-procedure resulted in a model constituted of 37 bands with an overall accuracy of 87.6 % and a Kappa of 82.9 %. The increased overall accuracy of the RFE model compared to the models built with different subsets containing all bands suggests that improved results can be attained by selecting bands.

The highest ranked bands were all May bands except the red band, the SWIR 1-2 bands from April, July and October, along with the blue and red edge 4 band from July and the red bands from April and October. The spectral profile plots support their importance, but the standard deviation plots suggest overlap in spectral reflectance for some bands. The other variables included in the best model with 37 bands are not unnecessary, but their contribution is only marginal. Presumably, they are highly correlated, but as Guyon and Elisseeff (2003) state, even highly correlated variables are not redundant.

There are some key differences between previous studies and the current one regarding the tree species that were classified, the timing and processing of the satellite imagery, the amount of training data and the geographic location which influences the band ranking and complicates comparison. The bands ranking in this study conform with the results obtained in other studies to some degree. Immitzer et al., (2016) and Nelson (2017) reports that some of the red edge bands (6, 7, and 8a) and SWIR 2 are among the most important. However, dissimilarities occur regarding the importance of individual bands in the visible spectrum. The results from Immitzer et al. (2016) propose the blue band, but Nelson (2017) reports that the red band was more important, probably since an image from May 2nd was included. The blue band is usually discarded as an undesirable band for tree species classification, since it is most affected by atmospheric haze caused by Rayleigh scattering from the atmosphere. None of the previously mentioned studies performed an atmospheric correction on the satellite imagery to obtain Bottom of Atmosphere (BoA) reflectance, which can have introduced inconsistencies. Nelson (2017) omitted some of the

most correlated variables, which also affected subset of bands used, hence the variable importance result.

The results from this study states that the SWIR bands 1-2 from all dates were important. Clark and Roberts (2012) suggests that the species-specific variation of nitrogen, cellulose and lignin can be detected in the SWIR spectrum in dry leaves. Leaf-senescence has gone a long way by October 19th and the withdrawal of water from the leaves could explain the importance of the SWIR 1-2 bands. The difference between tree species is pronounced in the red edge bands in the May image. Clark and Roberts (2012) suggests that the red edge bands are highly correlated to the content of chlorophyll *a* and *b* in green leaves. The May image is from a time of the year when the tree species in this study are in different states of leaf development, which could explain the high rank of the red edge bands. Solid conclusions of what biochemical substances that reflect cannot be made since no field-measurements of leaf water content was carried out and not in the scope of the study.

The red edge bands and the SWIR bands have lower spatial resolution (20 m) than the bands in the visible spectrum and NIR band (10 m). The spectral reflectance in these bands probably contained spectral information of neighbouring trees outside of the field plot - possibly of another tree species. Noise could also have been introduced in cases when a tree species barely surpassed 70 % in BA on a field plot. The spectral signal from these bands is possibly a bit more unreliable and should be taken in account in the sampling design. The occurrence of this in the study was marginal since most plots where located in homogenous stands.

Performance on species level

In this study, Birch, Pedunculate Oak and Norway Spruce were classified with high Producer's accuracies in the RFE-model with 37 bands. Hybrid Larch and Scots Pine reached lower accuracies and were confused with Birch and Pedunculate Oak, and the latter with Norway Spruce. Scots Pine was only confused with Norway Spruce which probably is due to that Norway Spruce occurred in the understory in some of the field plots. Additionally, Norway Spruce was the most common class in the RF model, which is the class with the highest probability in uncertain cases. The reason for confusion in the case of Hybrid Larch is probably due to a small training dataset which – to some degree – failed to provide a coherent spectral reflectance since the standard deviation is large in all dates. Interestingly, Hybrid Larch had 95.7 % in User's accuracy and a similar pattern is seen for Scots Pine, since only Norway Spruce was misclassified as Scots Pine. Of the broadleaves, only one Birch-field plot was misclassified as any of the conifers, which indicates that conifers and broadleaves can be sufficiently separated in future studies.

The information classes used in this study were few compared to the previously mentioned studies and did not include any mixture of tree species. Plots with pure information classes can easily be confused with plot that contain mixtures of that class. The results obtained in this study is therefore rather optimistic, however, the scope of this study was to evaluate how well tree species could be separated and not mixtures.

The broadleaves species not included in the classification have interesting spectral characteristics which provide a notion of how well they could be separated in future studies. The standard deviation plots suggest that there would be a high confusion between these broadleaves species and the other five. Wild Cherry and Wynch Elm are slightly separated from the other tree species in the April image (red edge 2-4) but otherwise not. However, the flowers of Wild Cherry are all white in mid-May, which could aid separation. Aspen and Alder are reflecting similarly throughout the year but separates slightly in May (NIR). The use of Soil Topographic Index (STI, (Buchanan et al., 2014) as a variable in the classification could improve the classification of species that naturally grow adjacent to stream and in wetlands. Future studies that strive to classify additional deciduous tree species will need several well-timed satellite images, larger training datasets and could utilize additional geographical variables, such as STI.

The quality of the field data

The training data must have a sufficient number of samples (training data set size) to handle the increasing number of dimensions attained by adding predictors and the inter- and intra-species spectral variation. The accuracy of the model will decrease if this is not addressed (Hughes, 1968). Several studies have evaluated the effect of training data set size on land cover classification (Colditz, 2015) and on tree species classification (Nelson, 2017), using the RF classifier. The results from both studies imply that the classification accuracy increases with the size of the training data set in each class. Nelson (2017) reports the lowest accuracy with 10 to 25 plots in each spectral class (77 % in overall accuracy) but it increased linearly to 86 % if 150 plots were included. Reese (2011) reported that 30 to 50 plots per class were enough to represent the spectral variation of individual alpine vegetation classes. The satisfying results in overall accuracy suggests that the training data had a sufficient size to account for the spectral variation. Hybrid Larch was misclassified to a greater extent than the others which is probably due to that it had the least number of plots.

The reference dataset used in this study is imbalanced, favouring Norway Spruce over the other classes, even after addition of plots to the underrepresented classes. Unbalanced datasets tend to overclassify the most abundant and misclassify the scarce (Chen et al., 2004). Although the problem in modelling is evident, measures to even out the proportion of the classes do not increase the classification accuracy results, in contrary it has been shown to reduce it (Zhu et al., 2016). Recent studies that evaluated the sampling design and its effect on the classification (Colditz, 2015; Dalponte et al., 2013), have deduced that an area-proportional allocation of training samples per class yields the best results in the classification compared to using equally large training samples for each class. Colditz (2015) claims that classes which are more widespread in the landscape need more training samples to take account for the spectral variability.

Species-specific basal area > 70 % was used as proxy for crown cover on the field plots, which might have introduced some inconsistencies since the relationship between basal area and crown width differs between the species present in this study.

The field data did not represent all the development stages for the species used in the classification. The Pedunculate Oak field plots only originate from stands with mature trees with fully developed crowns. The other information classes had representation of different age classes but field plots from young stands that had not reached full crown closure were omitted from the field data since a pure spectral response was preferred. Future studies that focus on a larger study area should aim on representing all development classes in the dataset.

The Oak class originates from roughly two stand types: abandoned pasture land with mature Pedunculate Oak in the overstory and Hazel in the understory, and managed Pedunculate Oak stands with grass subspecies in the field vegetation. The crown cover ratio is high for all Oak plots, which reinforces the fact that the spectral signal is pure, but if an abundant understory starts leafing out earlier it could introduce noise. The Hazel understory could be a source of error since it was very abundant, but according to Naturkalendern leaf-out started at roughly the same time for both species in the second half of May. It is therefore not likely that it should have intervened. The April and October image are for that reason more likely to contribute to the discrimination between coniferous species and deciduous species, since the former group does not undergo such drastic change in leaf occurrence during the year.

Field plots that were located in plantations of Norway spruce younger than 8 years were excluded from the Norway spruce information class. The age could have been set higher in order to ensure closed canopies, hence a purer spectral reflectance. Ground vegetation or single stems of Birch could have distorted the spectral signature. Since the site index of the Norway spruce was high, it was assumed that the canopy could have been fairly closed at the time of the image acquisition. However, the potential effect of these plots is limited since they are very few related to the total amount of field plots in this class. Their implications on the classification result is considered negligible.

Conclusions

- The best model of the subsets was constituted of satellite imagery from all dates and had and overall accuracy of 86.4 %. The successive addition of a satellite image increased the overall accuracy if the highest performing model in each group was chosen (M → AM → AMO → AMJO).
- The May image was crucial, and the overall accuracies increased with 7 % by adding three other images. The April image was important since it captured the tree species in leaf-out condition. The October image was important since it represented the tree species during senescence. The July image was redundant but images from the same part of the year can still add information but the increase in overall accuracy is marginal.
- The RFE procedure coupled with RF obtained an overall accuracy of 87.6 % and a Kappa of 82.9 %. All bands from May 27th except the red band, the SWIR 1-2 bands from April 7th, July 9th and October 19th. The red band from April 7th, the blue and red edge 4 band from July 9th and the red band from October 19th were also ranked high.
- Little misclassification occurred between evergreen coniferous species and deciduous species, which indicates that these species groups can be separated well in future studies.
- The broadleaf species which were not included in the classification (Wynch Elm, Alder, Aspen and Wild Cherry) reflected similarly which could aggravate spectral separation. Future studies that strive to classify these should consider using several well-timed satellite images, larger training datasets and additional geographical variables, such as STI, to achieve acceptable accuracies.

References

Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sens. Environ. 64, 234–253.

Boyd, D.S., Danson, F.M., 2005. Satellite remote sensing of forest resources: three decades of research development. Prog. Phys. Geogr. 29, 1–26. https://doi.org/10.1191/0309133305pp432ra

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Buchanan, B.P., Fleming, M., Schneider, R.L., Richards, B.K., Archibald, J., Qiu, Z., Walter, M.T., 2014. Evaluating topographic wetness indices across central New York agricultural landscapes. Hydrol. Earth Syst. Sci. 18, 3279–3299. https://doi.org/10.5194/hess-18-3279-2014

Chen, C., Liaw, A., Breiman, L., 2004. Using Random Forest to Learn Imbalanced Data. Stat. Tech. Rep.

Clark, M., Roberts, D., Clark, D., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sens. Environ. 96, 375–398. https://doi.org/10.1016/j.rse.2005.03.009

Clark, M.L., Roberts, D.A., 2012. Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. Remote Sens. 4, 1820–1855. https://doi.org/10.3390/rs4061820

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

Dalponte, M., Orka, H.O., Gobakken, T., Gianelle, D., Naesset, E., 2013. Tree Species Classification in Boreal Forests With Hyperspectral Data. IEEE Trans. Geosci. Remote Sens. 51, 2632–2645. https://doi.org/10.1109/TGRS.2012.2216272

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sens. Environ. 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Friedman, J., Hastie, T., Tibshirani, R., 2001. The elements of statistical learning. Springer series in statistics New York.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422.

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theory 14, 55–63. https://doi.org/10.1109/TIT.1968.1054102

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, Springer Texts in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-7138-7

Liaw, A., Wiener, M., 2017. Package 'RRF.'

Lindberg, E., 2017. Remningstorp inventering 2016.

Mickelson, J.G., Civco, D.L., Silander, J.A., 1998. Delineating forest canopy species in the northeastern United States using multi-temporal TM imagery. Photogramm. Eng. Remote Sens. 64, 891–904.

Millard, K., Richardson, M., 2015. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. Remote Sens. 7, 8489–8515. https://doi.org/10.3390/rs70708489

Mueller-Wilm, U., 2017. Sen2Cor Configuration and User Manual.

Naturkalendern [WWW Document], 2017. . Naturkalendern. URL http://www.naturenskalender.se/index.php (accessed 10.1.17).

Nelson, M., 2017. Evaluating Multitemporal Sentinel-2 data for Forest Mapping using Random Forest.

Reese, H., Nilsson, M., Pahlén, T.G., Hagner, O., Joyce, S., Tingelöf, U., Egberth, M., Olsson, H. akan, 2003. Countrywide estimates of forest variables using satellite data and field data from the national forest inventory. AMBIO J. Hum. Environ. 32, 542–548.

Reese, H.M., Lillesand, T.M., Nagel, D.E., Stewart, J.S., Goldmann, R.A., Simmons, T.E., Chipman, J.W., Tessar, P.A., 2002. Statewide land cover derived from multiseasonal Landsat TM data: a retrospective of the WISCLAND project. Remote Sens. Environ. 82, 224–237.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

RStudio Team, 2016. RStudio: Integrated Development for R. RStudio Inc, Boston, MA.

Schriever, J.., Congalton, R.., 2005. Evaluating seasonal variability as an aid to cover-type mapping from Landsat Thematic Mapper data in the northeast. Photogramm. Eng. Remote Sens. 3, 321–327.

Sentinel Application Plattform (SNAP), 2018. . Brockmann Consult, Array Systems Computing and C-S.

Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., Gamon, J.A., Zarco-Tejada, P., 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sens. Environ. 113, S67–S77. https://doi.org/10.1016/j.rse.2008.10.019

Williams, C.K., Engelhardt, A., Cooper, T., Mayer, Z., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., Kuhn, M.M., 2017. Package ‘caret.’

Wolter, P., Mladenoff, D.J., 1995. Improved Forest Classification in the Northern Lake States Using Multi-Temporal Landsat Imagery. Photogramm. Eng. Remote Sens. 61, 1129–1143.

Zhu, Z., Gallant, A.L., Woodcock, C.E., Pengra, B., Olofsson, P., Loveland, T.R., Jin, S., Dahal, D., Yang, L., Auch, R.F., 2016. Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. ISPRS J. Photogramm. Remote Sens. 122, 206–221. https://doi.org/10.1016/j.isprsjprs.2016.11.004

Appendix 1.

Table 7. Confusion matrix for single image April

Tabell 7. *Confusion matrix för singelbild April*

	Reference data						
	Birch	**Hybrid Larch**	**Pedunculate Oak**	**Scots Pine**	**Norway Spruce**	**Total**	**Producer's accuracy (%)**
Birch	30	6	4	3	2	45	66.7
Hybrid Larch	15	9	3	0	0	27	33.3
Pedunculate Oak	6	2	30	0	0	38	78.9
Scots Pine	3	0	0	41	13	57	71.9
Norway Spruce	0	0	3	14	83	100	83.0
Total	54	17	40	58	98	267	
User's accuracy (%)	55.6	52.9	75.0	70.7	84.7		
Overall accuracy (%)	71.5						
Kappa (%)	62.2						

Table 8. Confusion matrix for single image May

Tabell 8. *Confusion matrix för singelbild Maj*

	Reference data						
	Birch	**Hybrid Larch**	**Pedunculate Oak**	**Scots Pine**	**Norway Spruce**	**Total**	**Producer's accuracy (%)**
Birch	**38**	1	2	2	2	45	84.4
Hybrid Larch	4	**14**	1	2	6	27	51.9
Pedunculate Oak	2	0	**36**	0	0	38	94.7
Scots Pine	4	3	0	**40**	10	57	70.2
Norway Spruce	4	2	1	6	**87**	100	87.0
Total	52	20	40	50	105	**267**	
User's accuracy (%)	73.1	70.0	90.0	80.0	82.9		
Overall accuracy (%)	79.4						
Kappa (%)	72.6						

Table 9. Confusion matrix for single image July

Tabell 9. *Confusion matrix för singelbild Juli*

	Reference data						
	Birch	**Hybrid Larch**	**Pedunculate Oak**	**Scots Pine**	**Norway Spruce**	**Total**	**Producer's accuracy (%)**
Birch	**33**	3	5	0	4	45	73.3
Hybrid Larch	8	**9**	2	5	3	27	33.3
Pedunculate Oak	4	1	**33**	0	0	38	86.8
Scots Pine	2	1	0	**41**	13	57	71.9
Norway Spruce	5	3	1	9	**82**	100	82.0
Total	52	17	41	55	102	**267**	
User's accuracy (%)	63.5	52.9	80.5	74.5	80.4		
Overall accuracy (%)	74.4						
Kappa (%)	66						

Table 10. Confusion matrix for single image October

Tabell 10. *Confusion matrix för singelbild oktober*

	Reference data						
	Birch	**Hybrid Larch**	**Pedunculate Oak**	**Scots Pine**	**Norway Spruce**	**Total**	**Producer's accuracy (%)**
Birch	**33**	0	8	1	3	45	73.3
Hybrid Larch	5	**11**	1	4	6	27	40.7
Pedunculate Oak	6	1	**31**	0	0	38	81.6
Scots Pine	4	2	2	**39**	10	57	68.4
Norway Spruce	3	5	0	10	**82**	100	82.0
Total	51	19	42	54	101	**267**	
User's accuracy (%)	64.7	57.9	73.8	72.2	81.2		
Overall accuracy (%)	74.3						
Kappa (%)	65.9						

Table 11. Confusion matrix for the combination of April and May

Tabell 11. *Confusion matrix för kombinationen april och maj*

	Reference data						
	Birch	**Hybrid Larch**	**Pedunculate Oak**	**Scots Pine**	**Norway Spruce**	**Total**	**Producer's accuracy (%)**
Birch	**42**	1	0	0	2	45	93.3
Hybrid Larch	5	**19**	1	0	2	27	70.4
Pedunculate Oak	1	0	**37**	0	0	38	97.4
Scots Pine	3	0	0	**43**	11	57	75.4
Norway Spruce	4	1	1	7	**87**	100	87.0
Total	55	21	39	50	102	**267**	
User's accuracy (%)	76.4	90.5	94.9	86.0	85.3		
Overall accuracy (%)	85.3						
Kappa (%)	80.6						

Table 12. Confusion matrix for the combination of April, May and October

Tabell 12. *Confusion matrix för kombinationen april, maj och oktober*

	Reference data						
	Birch	**Hybrid Larch**	**Pedunculate Oak**	**Scots Pine**	**Norway Spruce**	**Total**	**Producer's accuracy (%)**
Birch	**42**	0	1	1	1	45	93.3
Hybrid Larch	3	**22**	1	0	1	27	81.5
Pedunculate Oak	1	0	**37**	0	0	38	97.4
Scots Pine	3	0	0	**41**	13	57	71.9
Norway Spruce	3	1	1	6	**89**	100	89.0
Total	52	23	40	48	104	**267**	
User's accuracy (%)	80.8	95.7	92.5	85.4	85.6		
Overall accuracy (%)	86.4						
Kappa (%)	81.9						