Classification of ground lichen using Sentinel-2 and airborne laser data

Helene Larsson

Work report 485 2018
Master thesis in Forest Sciences 30hp A2E
Forest Sciences – Masters Program

Supervisor:
Heather Reese

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–485–SE

Classification of ground lichen using Sentinel-2 and airborne laser data

Helene Larsson

Keywords:remote sensing, random forest, reindeer.

Master thesis in Forest Sciences at the Department of Forest Resource Management, 30 credits, EX0768, A2E
Forest Sciences – Masters Program
Supervisor: Heather Reese, 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–485–SE

Acknowledgements

This Master thesis of 30 hp was written as part of Jägmästarprogrammet, 300 hp, at the Swedish University of Agriculture (SLU) in Umeå, Sweden.

Several people have helped and supported me during my thesis work. I am deeply grateful for this, as without their support this thesis would not have been completed. I will not attempt to list everyone, since I will inevitably forget someone. I will, however, mention a few people who have been instrumental to my work. Firstly, I would like to extend my deepest gratitude to my thesis advisor Heather Reese, for sharing her knowledge on my subject and also for her infinite patience with my dilly-dallying. I would also like to thank my assistant thesis advisors Henrik Hedenås and Per Sandström.

I would also like to acknowledge the importance of the Ljungberg lab, which is financed by Ljungbergsfonden, both as a workplace and a community. I would like to thank my fellow master students as well as the employees of the remote sensing part of SRH, for making it such a creative and open place to work. I would also like to thank André Wästlund for help with processing some of the laser data. Lastly, I would like to mention Arvid Axelsson, who has been a great help with coding, thank you.

Helene Larsson
Umeå, 2017-09-02

Abstract

The northern part of Sweden has two overlapping land-use interests: forestry and reindeer husbandry. Forestry affects reindeer husbandry in several ways; most important of these is its impact on ground lichens. Lichens are the primary winter grazing resource for reindeer, therefore, mapping of lichens is of interest. The objective of this study is to evaluate the use of remote sensing data from the new Sentinel-2 satellite for the classification of ground lichen and to assess whether adding information derived from airborne laser scanning (ALS) will improve the result. The study area is situated in the reindeer husbandry area inland of Umeå, in the north of Sweden, and consists of two Sentinel-2 granules. Two Sentinel-2 images were used, one from 2015-08-19 and one from 2016-10-02. ALS-derived metrics was also used in the form of DEM, wetness index, a canopy density metric and forest height. Classification of lichen coverage was carried out with the Random Forest algorithm, and 90 field plots were used as training data. Due to the small field dataset, the evaluation method for this study was internal cross-validation. Fourteen different classification schemes were tried with the Random Forest algorithm. Classification scheme 6 (0-33 %, 34-66 % and 67-100 % lichen coverage) was the most interesting of the classification schemes with three classes, since it has the lowest out-of-bag error at 29 %. Classification scheme 4 (0-25 %, 26-50 % and 51-100 % lichen coverage), which is based on the Swedish National Forest Inventory's lichen class definition, also proved to be fairly accurate, with an out-of-bag error of 37 %. Overall, the analysis showed that bands 4 (red) and 8 (NIR) of the Sentinel-2 2015-08-19 image, along with ALS-derived canopy density were the most important variables. Wetness index was the least important variable. For the Sentinel-2 2016-10-02 image, bands 4 (red) and 5 (red-edge) were the most important. This study showed that a Sentinel-2 image from a one date during the summer season worked well for the classification of lichen into three classes, and that adding an ALS-derived canopy density metric could improve the results. The use of both Sentinel-2 images together did not give better classification results.

Contents

1 Introduction

The northern part of Sweden has two major, spatially overlapping land-use interests: forestry and reindeer husbandry (Berg 2010). The area affected by these conflicting interests is not insignificant, as 55 % of Sweden's land area belongs to the reindeer husbandry area (Sandström et al. 2016). Lichens are the primary grazing resource for reindeer, and therefore, the mapping of the lichen resource is of interest. The requirements of modern forestry and reindeer husbandry are not always compatible, which has led to conflict (Berg 2010; Sandström et al. 2003). At least part of this conflict stems from lacking knowledge about the distribution and quantity of reindeer lichens. Remote sensing is a data source which may be useful for creating spatially explicit maps with information about the lichen resource.

This Master thesis is based in part on a previous lichen remote sensing study by Gilichinsky et al. (2011). Since their study came out, new remote sensing data sources have become available which are of interest to test, namely Sentinel-2 and airborne laser scanning data.

1.1 Reindeer husbandry, forestry and lichen

Northern Sweden belongs to the boreal zone. Its forests are dominated by two coniferous tree species, Scots pine (*Pinus sylvestris* L.) and Norway spruce (*Picea abies* (L.) H. Karst.), interspersed with a few deciduous tree species, most commonly birch (*Betula* spp.) and aspen (*Populus tremula* L.) (Berg 2010). The climate is harsh and the vegetation period short, but plants and animals have adapted to these conditions. Reindeer (*Ragnifer tarandus* L.) have adapted by following a yearly migration pattern and by changing their food source with the seasons. Reindeer migrate from the mountainous inland of Sweden in the summer where they feed on shrubs, herbs, grass and fungi, to the forested areas closer to the coast during winter (Berg 2010; Gilichinsky et al. 2011). In the wintertime they forage for lichens which comprise as much as 50-80 % of their diet (Heggberget et al. 2002). For the most part they feed on ground-growing, mat-forming lichen of the genus *Cladonia*, primarily *Cladonia rangiferina*, *C. arbuscula*, *C. stellaris* as well as *Cetraria islandica*. Reindeer need more energy in the wintertime to compensate for colder temperatures and the difficulty in foraging for food. Wintertime reindeer pastures thus need to have a higher forage abundance (Nelson et al. 2013).

Reindeer husbandry in Sweden is the traditional livelihood of the indigenous Sami people. Today only Sami people can practice it (Sandström 2015; *Rennäringslag* 1971). The semi-domesticated reindeer still follow the migration patterns that their wild ancestors did, which makes this grazing system unique (Sandström 2015). It is this fact that makes it demand such large tracts of land. The reindeer husbandry area includes all seasonal grazing land, as well as migration routes. This region, as stated earlier, covers about 55 % of Sweden's land area, and more than 50 % of the productive forest land (Sandström et al. 2016). Within this area forestry and reindeer husbandry must coexist. Forestry affects reindeer husbandry in several ways; most important of these is its impact on ground lichens. Much of the conflict of interest between forestry and reindeer husbandry has to do with the management of lichen forests.

Modern forestry affects ground lichens in many ways (Sandström et al. 2003; Sandström 2015). Forestry as practiced in Sweden today is intensive and affects almost all forested land. It mainly follows this sequence: cleaning, thinning, clear-cutting, soil-scarification and planting of new trees. Soil-scarification is performed by turning over the topsoil to expose mineral soil beneath to improve the microclimate and nutrient availability for seedlings. This process destroys the slow-growing ground lichens, and can, depending on intensity, eradicate lichen from an area entirely. A decline in lichen cover has been noticed for the last few decades (Gilichinsky et al. 2011). Since the 1950s, the area of lichen-abundant forest (>50 % cover of ground lichens) has declined with 71 % (Sandström et al. 2016). This increases the grazing pressure on the remaining area.

1.2 Remote sensing of lichen

Earlier studies have attempted to map lichens using optical satellite data from Landsat, SPOT, and other available sensors (Nordberg and Allard 2002; Rees et al. 2003; Sandström et al. 2003; Tømmervik et al. 2003; Rautiainen et al. 2007; Gilichinsky et al. 2011; Nelson et al. 2013). Various types of lichen, both ground lichen and lichen growing on stone, has been studied. Some studies have focused on lichen growing on the tundra or heath, i.e. studies that were unimpeded by tree cover (Nordberg and Allard 2002; Falldorf et al. 2014). Other studies have attempted to map lichen under tree cover, as this study aims to do (Théau et al. 2005; Gilichinsky et al. 2011; Käyhkö and Pellikka 1994). As Théau et al. (2005) points out, the inclusion of a tree canopy complicates the detection of lichens due to the added pixel reflectance. Several studies have also incorporated forest data with the remote sensing data in their studies (Nordberg and Allard 2002; Colpaert et al. 2003; Johnson et al. 2003; Rees et al. 2003; Sandström et al. 2003; Tømmervik et al. 2003).

Several species of lichen, among them *Cladonia* spp., are light in colour and reflect more light in the blue to yellow spectrum as compared to green vegetation (Petzold and Goward 1988). This makes these lichens distinguishable from other vegetation. *Cladonia* spp. often also contain usnic acid, which is a pigment that is pale yellow in colour and also is distinct spectrally. Nelson et al. (2013) used Landsat 7 ETM+ images in their study and found that usnic lichens have a higher reflectance in the visible to NIR-range as compared to other vegetation and even to other kinds of lichen, especially in the the blue spectral band. This study was the first to concentrate solely on usnic lichens. Nelson et al. (2013) found these lichens to be the most detectable lichen cover group. They also found that usnic lichen cover is positively correlated with total lichen cover. This makes usnic lichen cover useful for mapping total lichen cover.

Käyhkö and Pellikka (1994) used SPOT XS bands 1-3 (green, red and NIR) for supervised classification of vegetation classes in Finland and Norway. This is an older study, which means an older sensor. Their study only determined lichen classes with sufficient accuracy in areas without tree cover. The study also found that certain combinations of tree density and ground-cover vegetation shared spectral characteristics which caused misclassifications. The study found that *Cladonia* had high reflectance in the visible to NIR spectra, but also that *Cladonia* could be hard to distinguish from green vegetation in the NIR band. In the NIR band of the SPOT XS satellite, chlorophyll absorption is weak which results in the vegetation appearing bright in that band. The high reflectance of *Cladonia* lichens in this spectrum can thus lead to confusion of the two. The best results of the study came from independent use of the red band for classification.

Nordberg (1998) used Landsat TM to develop a Normalized Difference Lichen Index (NDLI) which effectively assessed the spectral signatures of lichen. A few years later Nordberg and Allard (2002) proved Normalized Difference Vegetation Index (NDVI, where NDVI = (NIR−red)/(NIR+red)), to be better at predicting lichen cover. Nordberg and Allard used Landsat-5 TM images to detect lichen deterioration above the tree-line in the Swedish mountains. They also showed that lichen coverage increases the reflectance across all Landsat-5 TM bands. Greater lichen coverage gives higher reflectance, but grazing reduces the height of the lichen thallus, and thus its reflectance (Tømmervik et al. 2003). According to the Nordberg and Allard (2002) study, lichens have lower values than green vegetation in the NIR region. The study also indicated that the NIR band showed promise for distinguishing low coverage of lichen (<40 %). Their study also reinforced the known fact that lichen has a higher reflectance in the visible spectrum than green vegetation.

Théau et al. (2005) mapped lichen using Landsat-5 TM bands 3, 4 and 5 (red, NIR and SWIR bands), with a spectral mixture analysis (SMA) and enhancement-classification method (ECM). Their study showed that ECM was successful in separating lichen from non-lichen classes, but less accurate in separating between specific lichen classes. SMA also showed good results in separating lichen from non-lichen classes, but was better than the ECM at distinguishing between lichen classes.

Rautiainen et al. (2007) examined how understory vegetation influenced the reflectance of forest stands in Northern Finland. They used the SPOT HRVIR data in the green, red, NIR and SWIR bands. Their study showed that forest stands with a lichen understory could be distinguished from those with a dwarf shrub (heather (*Calluna vulgaris* (L.) Hull) and crowberry (*Empetrum nigrum* L.)) understory in the visible wavelength range. In the NIR spectrum however, this was only possible for sites with very sparse canopy cover.

Falldorf et al. (2014) used Landsat-5 TM images to develop a method for continuous estimation of lichen volume within lichen-dominated alpine heath in Norway. They used a 2D Gaussian regression model based on a Normalized Difference Lichen Index (NDLI = (SWIR-NIR)/(SWIR+NIR)) and Normalized Difference Moisture Index (NDMI = (NIR-SWIR)/(NIR+SWIR)). They found that there was a strong correlation between NDMI and lichen volume. NDMI contrasts the Landsat-5 TM NIR and SWIR bands (bands 4 and 5). *Cladonia* lichens are easily distinguishable in both of these bands.

In the study by Gilichinsky et al. (2011), images from the SPOT 5 and Landsat-7 satellites were used. The SPOT 5 data had a 10 meters pixel size and consisted of four spectral bands: green, red, NIR and SWIR. The Landsat-7 data had a pixel size of 25 meters and consisted of seven spectral bands: blue, green, red, NIR as well as two SWIR bands. The reference data for the study were taken from the Swedish National Forest Inventory (NFI) (Riksskogstaxeringen 2017) and the validation data came from an independent field inventory. Ground lichen cover was divided into three thematic classes based on percent lichen cover on the ground. These classes were: lichen poor (0-25 %), lichen moderate (25-50 %) and lichen abundant (50-100 %).

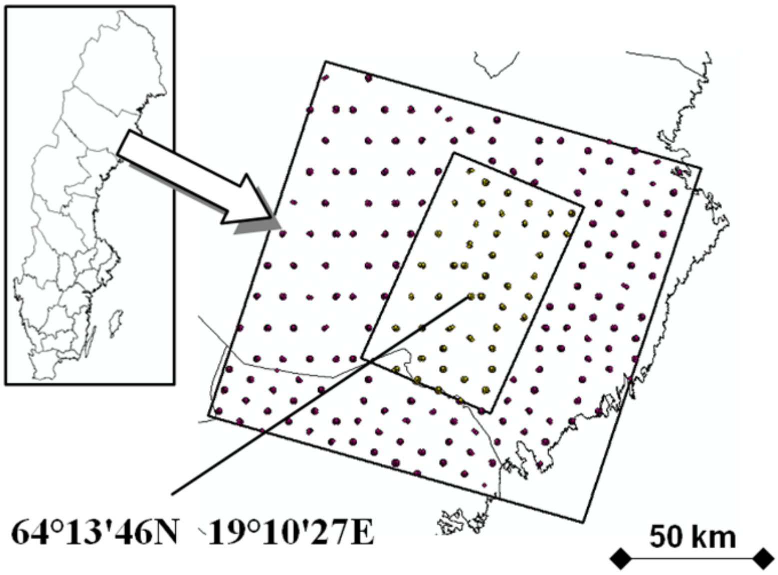


Figure 1.1. The original study area with outlines of the Landsat-7 (larger rectangle) and SPOT 5 (smaller rectangle) imagery in Västerbotten County in Northern Sweden (Gilichinsky et al. 2011). The dots represent the NFI field inventory plots used in the 2011 study.

These class definitions were based on the NFI's lichen class definitions. Three methods of supervised classification were used: Mahalanobis distance, SMA and maximum likelihood classification. The results indicated that it is possible to use data from the NFI as reference data for satellite data classification of ground lichen cover classes. The results showed higher accuracy for the classification using SPOT 5 images (10 meter pixels) as compared to using Landsat-7 images (25 meter pixels), which might have been due to the higher spatial resolution of the SPOT images. The accuracy was also higher when using a training dataset that only had two classes, lichen poor and lichen abundant, as compared to a dataset where the more diffuse lichen moderate class was included. As for the classification methods, Mahalanobis distance gave the best results.

The present study is tied to the earlier study by Gilichinsky et al. (2011) and uses part of the same study area (Figure 1.1), within the reindeer husbandry area in the county of Västerbotten. Gilichinsky et al. (2011) expressed a possibility to continue their research either by performing a more detailed classification, using other classification methods, or by adding airborne laser data or other forest parameters. This study will attempt to do that.

1.3 Objectives

The objective of this Master thesis is to evaluate the use of remote sensing data from the Sentinel-2 satellite for the classification of ground lichen and to assess whether the addition of information from airborne laser scanning data will improve the classification result.

2 Materials and Methods

2.1 Material

2.1.1 Study area

The study area is situated inland of Umeå, in Västerbotten County in the north of Sweden at a latitude of circa 64 degrees north and a longitude of about 18 degrees east (Figure 2.1). The area belongs to the boreal forest zone and as such the forest is primarily coniferous and dominated by managed Scots pine and Norway spruce. There are also a few species of deciduous trees, mostly birch and aspen (Gilichinsky et al. 2011).

Cladonia rangiferina, *C. arbuscula*, *C. stellaris* as well as *Cetraria islandica*, which are the four most significant species of reindeer lichen, are abundant within the area, making it important for reindeer husbandry (Gilichinsky et al. 2011). The area is used for winter grazing by several Sami villages, namely those of Vilhelmina North, Vapsten, Upmeje, Malå and Ran. The area has a mixed land ownership, where land is owned by the state, forestry companies as well as small private owners.

2.1.2 Satellite data

Sentinel-2 satellite images were used in the study. The new study area falls within the Sentinel-2 granules with the reference numbers 33WXM and 33VXL. The old study area also included part of the 34WDS granule. This granule was removed from the study due to pixelwise spectral differences in the pixel values within the area of overlap. The values differed between 10 and 50 Digital Numbers (DNs) for different bands of the Sentinel-2 2015-08-19 image.

The Sentinel-2A satellite was launched on 2015-06-23 and has a 290 km wide-swath that captures high-resolution, multi-spectral images. Sentinel-2 collects 13 spectral bands with three different spatial resolutions (i.e., pixel sizes): blue, green, red and NIR with a pixel size of 10 meters, six bands within the red-edge (680-750 nm), NIR and SWIR range with a pixel size of 20 meters, and coastal aerosol, water vapour and cirrus clouds with a pixel size of 60 meters (European Space Agency 2016) (Table 2.1). In this study bands 2 through 8, 11 and 12 were used; that is, all the bands with 10 meter and 20 meter resolution aside from band 8a. The 20 meter bands were resampled to 10 m. No radiometric corrections, such as atmospheric correction, was performed on the satellite images. Sweref99 TM was the coordinate system used for the entire study.

The Sentinel-2 satellite has four different bands within the red-edge spectrum (European Space Agency 2017d). The importance of these for mapping lichen are as of yet unknown. Furthermore, the Sentinel-2 satellite has a finer spatial resolutions

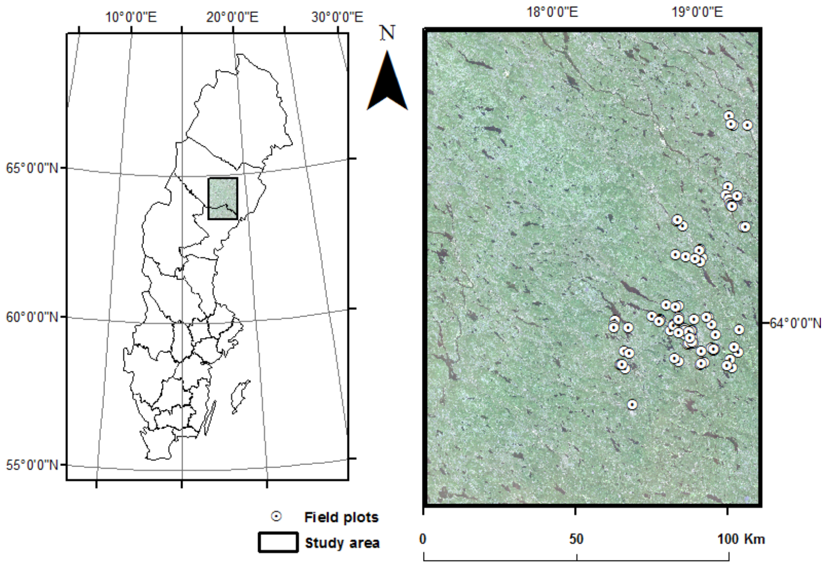


Figure 2.1. Left map: the new study area is situated in Västerbotten County, northern Sweden. Right map: the study area is made up of two Sentinel-2 granules (33VXL and 33WXM). The dots represent the field plots. The Sentinel-2 image is from 2015-08-19, and displayed in true colour.

compared to the Landsat satellite. Gilichinsky et al. (2011) indicated that the spatial resolution might be important for detecting smaller patches of lichen, with 10 meters pixels giving more accurate results than the 25 meter pixels, but also that it might give a more accurate classification over all.

The Sentinel-2 satellite also has a high radiometric resolution (European Space Agency 2017b). Radiometric resolution measures the ability of the instrument to distinguish differences in light intensity or reflectance. Radiometric resolution is often defined as a bit number, typically in the range of 8 to 16 bits. The radiometric resolution of the Sentinel-2 satellite data is 12 bit. This means that the DNs in the image can be acquired over a range of 0 to 4095 light intensity values.

The temporal resolution of Sentinel-2A is ten days. Despite this frequent imaging, only two images from 2015 and 2016 were sufficiently cloud-free to be used in the study. They are the images from 2015-08-19 and 2016-10-02. These two images were downloaded from the Copernicus Open Access Hub (European Space Agency 2017a). However, 2017-03-07 the Sentinel-2B satellite, the twin of the Sentinel-2A, was sent up (European Space Agency 2017c). This will increase the frequency of the coverage to once every five days at the equator, and roughly once every three days for Scandinavia (European Space Agency 2017e).

2.1.3 Airborne laser data

The National Mapping Agency (Lantmäteriet 2017) performed a national airborne laser scanning (ALS) campaign over Sweden between the years 2009 and 2016, where the point density is approximately 0.5 laser points per square meter over most of the country (Lantmäteriet 2016). The aim was to produce a digital elevation model (DEM), which has been done at a 2 meter grid cell resolution. The ALS point data over the

Table 2.1. The different spectral bands captured by the Sentinel-2 satellite (European Space Agency 2016)

Band	Spatial resolution (m)	Central wavelength (nm)	Bandwidth (nm)
1 - Coastal aerosol	60	443	20
2 - Blue	10	490	65
3 - Green	10	560	35
4 - Red	10	665	30
5 - Red edge	20	705	15
6 - Red edge	20	740	15
7 - Red edge	20	783	20
8a - Red edge	20	865	20
8 - NIR	10	842	115
9 - Water vapour	60	945	20
10 - SWIR, cirrus	60	1375	30
11 - SWIR	20	1610	90
12 - SWIR	20	2190	180

ground level can be used to estimate height and density metrics about the vegetation and forest. This study used four ALS-derived metrics: a DEM, the Saga wetness index (SWI), a canopy density metric in the form of a vegetation ratio (vegkvot) and the 95^{th} height percentile (p95) which gives a good indication of forest height.

The 2 meter grid cell size DEM was downloaded from the National Mapping Agency and resampled to a grid cell size of 10 meters using bilinear interpolation. The wetness index was calculated from the DEM in R Studio using Saga GIS (Brenning 2008). The forest height and canopy density metrics were calculated from the height normalized National Mapping Agency point cloud using FUSION's command of 'grid metrics' and 'csv-to-grid' (McGaughey 2016). The canopy density was calculated according to the vegetation ratio, which is derived by dividing the number of first returns above a threshold of 1.5 meters by the total number of first returns. All ALS data used in the study were calculated to a grid cell size of 10 meters that spatially matched the pixels of the Sentinel-2 images.

The ALS-derived metrics, DEM, wetness index, canopy density, and forest height, were the other datasets used together with the Sentinel-2 data. The field plot coordinates were used to extract the values from these datasets and used in the Random Forest classification.

2.1.4 Field data

The reference data from this project came from two sources: the earlier field inventory by Gilichinsky et al. (2011) and a new field inventory. The Gilichinsky et al. (2011) field inventory was carried out 2006-2008 and was comprised of 229 plots . It was hypothesized that the lichen classes would not have changed significantly, and that the plots could be re-used for lichen classification. To check this, ten of the old plots were re-inventoried during early fall of 2016.

The earlier inventory plots were placed according to a stratified random sample within the study area, where the strata were based on an unsupervised classification of a Landsat-7 ETM+ image (Gilichinsky et al. 2011). The plots were divided evenly between three strata: forest with low probability of lichen, forest with medium probability of lichen and forest with a high probability of lichen. To minimize time spent locating the plots they were placed within 300 meters of roads. The plots were squares of 50 meters by 50 meters. Data on lichen and moss abundance, lichen height, the field layer, the tree layer and forest management actions were gathered. Lichen and moss abundance were estimated based on presence/absence in 100 points distributed 2 meters apart from North to South, in 4 rows which were 10 meters apart.

In addition to the 2006-2008 field data, 20 new plots were inventoried in Fall 2016. To reduce time spent doing inventory, the new plots were squares of 20 meters by 20 meters, which fit the size of the Sentinel-2 pixels. The plots were placed using a stratified random sample within the study area, where the strata were based on an unsupervised classification of the 2015-08-19 Sentinel-2 image. Six different strata were created, water, agricultural land, forest with low probability of lichen, forest with some probability of lichen, forest with medium probability of lichen, and forest with high probability of lichen. To minimize time spent locating the plots they were placed within 300 meters of roads. In the field, data on lichen and moss abundance, field layer, bush layer, tree layer and forest management actions were gathered. Lichen and moss abundance were estimated based on presence/absence in 100 points distributed 1 meter apart from North to South, in 5 rows which were 5 meters apart. The methods of inventory were based on those used by Gilichinsky et al. (2011).

2.2 Method

Classification was carried out in the following steps: preparation of satellite images and training data, change analysis and removal of unsuitable plots from the old dataset, classification and accuracy assessment. The process is described in further detail in the sections below.

2.2.1 Data preparation

The Sentinel-2 20 meter pixel bands were resampled to 10 meter pixels, using bilinear interpolation. Following that, all the bands used in the study, bands 2-8 and 11-12, were merged together into one file per date and granule. The two Sentinel-2 granules were mosaiked together. The southernmost part of the mosaics included the coastal area close to Örnsköldsvik. As it was likely that the training data would be too dissimilar to be applicable there, the southernmost parts were removed from the study area.

Change analysis was performed on the 2006-2008 field plots using image differencing in Arc Map. The images used were the 2015-08-19 Sentinel-2 image and a mosaic of SPOT images from 2008 and 2009, reflecting the forest state during of the previous field data inventory. SPOT images had to be chosen from two different years due to the inability to find cloud-free images that covered the entire study area within one year. The change analysis was performed using the NIR, red and green wavelengths, in that order. The analysis was done to find field plots that in the old inventory had been mature forest, but that since then had been clear-cut. 63 plots affected by clear-cutting since the earlier lichen inventory were removed in this step.

Polygons were created from the field plot boundaries and used to cut out the pixels within those polygons. The mean values for the plots were extracted and added to the training data.

According to Gilichinsky et al. (2011), 229 plots were inventoried between 2006 and 2008. Only 214 of those plots were made available for this study. Of those, 151 plots remained after the change analysis. Thirty-nine of the remaining plots fell within the Sentinel-2 granule 34WDS. This granule, and thus the field plots within it, were removed from the study. When the ALS-derived metrics were added, seven plots were removed both from both the old and the new plots due to N/A-values. Lastly, to achieve a more balanced training dataset, which is more suitable for Random Forest (Reese et al. 2014), a further 33 plots with 0 % lichen were removed using random selection. In the end 71 old plots and 19 new plots, totaling 90 plots, were used in the Random Forest classification.

The final field dataset is presented in Table 2.2. Number of plots dominated by the three different tree species and by the different dwarf shrub species are presented in Table 2.3

Table 2.2. Mean, minimum and maximum values of the final field variables for the 90 plots in the training dataset, as well as standard deviation of the mean. Basal area is in square meters per hectare, coverages are in percentages.

Variables	Mean	StDev	Min	Max
Basal area	13.3	5.4	1	27
Number of tree stories	1.2	0.4	1	2
Lichen coverage	26.2	24.4	0	88
Field layer coverage	61.2	20.2	0	95
Canopy coverage	32.5	16.2	5	70

Table 2.3. Number of field plots dominated by the different tree and dwarf shrub species.

Dominant tree species	Number of plots
Broadleaves	1
Pine	86
Spruce	3
Dominant dwarf shrub species	**Number of plots**
Bilberry	27
Heather	38
Lingon	25

2.2.2 Random Forest

Random Forest is an algorithm used for classification and regression (Liaw and Wiener 2002; Breiman 2001; Gislason et al. 2006; Rodriguez-Galiano et al. 2012). It is an ensemble learning technique that grows many decision trees and then lets them vote for the most common class. Random Forest increases the diversity of decision trees by growing them from different subsets of the training data. It does not overfit to its training set like single decision trees do, either. It is a non-parametric classifier, and does not require that the input data are normally distributed, and allows the input of multiple data sources. Random Forest is useful for classification when there are a lot of different variables, but not such a great need for variable selection for optimizing the model. It also assesses the relative importance of the different variables of input features during the classification, which is useful for multi-source studies (Rodriguez-Galiano et al. 2012). The difficulty with the Random Forest algorithm is that it is a "black-box method" that may be hard to understand.

The Sentinel-2 bands and the ALS-derived metrics were used as variables for classification with the Random Forest algorithm in seven combinations of different variable sets (Table 2.4).

Table 2.4. The three different sets of variables used to run the Random Forest, either for one data source or in combination. s215 is the 2015-08-19 Sentinel-2 image, s216 is the 2016-10-02 image. Laser refers to the ALS-derived variables: DEM, canopy density (vegkvot), tree height (p95) and wetness index (SWI).

s215	s216	laser
Band 2	Band 2	DEM
Band 3	Band 3	vegkvot
Band 4	Band 4	p95
Band 5	Band 5	SWI
Band 6	Band 6	
Band 7	Band 7	
Band 8	Band 8	
Band 11	Band 11	
Band 12	Band 12	

Due to the small field dataset, the evaluation method for this study was the internal cross-validation which assesses model performance. This method is not an independent accuracy assessment, but is nonetheless relevant when comparing models relative to each other.

As accuracy assessment, Tables over out-of-bag errors were generated in R using the Random Forest package. Out-of-bag error measures the prediction error of the Random Forest. Out-of-bag error is the mean prediction error of training sample X, that only uses trees that did not have X in their bootstrap sample. Breiman (1996) showed that out-of-bag error provides the same accuracy as using a test set of the same size as the training set.

Figures of the mean decrease in accuracy for the three most interesting classification schemes were generated in R using the VarImpPlot routine in the Random Forest package in for the three classification schemes that had the best result. The mean decrease in accuracy is determined when the out-of-bag errors are calculated. The more the elimination of a particular variable decreases the accuracy, the more important that variable is considered. This means that variables that have a large mean decrease in accuracy are deemed to be more important for classification. To further analyse important variables, scatter plots of the two most important variables from the three different sets of variables were generated for the three most interesting classification schemes.

Mean and standard deviation of the Digital Number for each band were calculated for the different classes and the variables used to run Random Forest of the three best classification schemes.

2.2.3 Classification schemes

Fourteen different classification schemes were tried with the Random Forest algorithm (Table 2.5). Eleven of the classification schemes were based on lichen coverage. Three of those incorporated either dominant dwarf shrub type or basal area as grounds for the class division. Three additional classes were calculated solely from dwarf shrub type or from basal area. This was done as it was of interest to see if results were relating to vegetation types co-existing with lichen.

There were six different schemes with four classes (schemes 1-3, 5, 9 and 11), four schemes had three classes (schemes 4, 6-7 and 13), three schemes had two different classes (schemes 8, 12 and 14) and lastly one scheme had six classes (scheme 10). Classification scheme 4 (Table 2.5) was based on the one Gilichinsky et al. (2011) used, which in turn is based on the NFI division of lichen coverage classes. The classification schemes divided lichen coverage into different classes, and different numbers of classes, to test what class division would prove to be more accurate in the Random Forest classification. This was done to test whether different definitions of lichen classes would affect model performance and eventually classification accuracy. Additional data, such as dominant dwarf shrub species and basal area, was added to test whether they had any bearing on the classification.

Table 2.5. Classification schemes 1 through 14. Numbers in classification schemes 1 - 11 represent percent lichen coverage. For classification schemes 9 - 11, footnotes 1, 2, 3 and 4 are co-existing dwarf shrub type where 1 is bilberry type, 2 is mixed heather and lingon type, 3 is heather type and 4 is lingon type. Footnote 5 denotes basal area$\leq$13 and footnote 6, basal area>13. For classification schemes 12 - 14 the numbers have the same meaning as the footnotes.

Classification	Class number					
scheme	1	2	3	4	5	6
1	0.00	1-33	34-66	67-100		
2	0-25	26-50	51-75	76-100		
3	0-10	11-50	50-70	71-100		
4	0-25	26-50	51-100			
5	0-10	11-30	31-49	50-100		
6	0-33	34-66	67-100			
7	0-10	11-55	56-100			
8	0-49	50-100				
9	$0-50^1$	$0-50^2$	$51-100^1$	$51-100^2$		
10	$0-33^1$	$0-33^2$	$34-66^1$	$34-66^2$	$67-100^1$	$67-100^2$
11	$0-39^5$	$0-39^6$	$40-100^5$	$40-100^6$		
12	5	6				
13	1	3	4			
14	1	2				

3 Results

3.1 Comparison of out-of-bag errors

The most accurate classification scheme was number 8 (Table 3.1), with only 2 classes (0-50 % and 51-100 %) and an out-of-bag error of 19 %. Four of the classification schemes with four classes, schemes 1-3 and 5 (Table 2.5), were not very accurate. Classification scheme 6 (0-33 %, 34-66 % and 67-100 %) is probably the most interesting of the classification schemes with three classes, since it has the lowest out-of-bag error at 29 % out of those. Classification scheme 4 (0-25 %, 26-50 % and 51-100 %), which is the same as the one used in the old study, also proved to be fairly accurate with an out-of-bag error of 37 %. Classification scheme 2 (0-25 %, 26-50 %, 51-75 % and 76-100 %) was the most accurate of the classification schemes with four classes. Classification schemes 9 through 11 used additional data other than lichen cover percentage for class division, but this did not increase accuracy.

Classification schemes 12-14 were not based on lichen coverage (Table 3.1). Classification scheme 12, which was based on basal area, was fairly accurate, as was classification scheme 14, with two classes based on dominant dwarf shrub species. Scheme 13, with its three classes based on dominant dwarf shrub species, was less accurate.

Of the single variable sets used, the Sentinel-2 image from August 2015 gave lower out-of-bag errors as compared to the Sentinel-2 image from October 2016. Using the ALS-derived variables generally gave a higher out-of-bag error than using the Sentinel-2 2015-08-19 image, with four exceptions. Those were classification schemes 3, 7 and 8, where the out-of-bag errors either were the same as that for the Sentinel-2 2015-08-19 image and thus lower than that for the Sentinel-2 2016-10-02 image, or equal to that of the Sentinel-2 2015-08-19 image.

Combining several variable sets gave mixed results. Using both Sentinel-2 images gave a lower out-of-bag error than for just the 2015-08-19 image for nine of the classification schemes, the same out-of-bag error for classification scheme 6, and higher out-of-bag error for classification schemes 1, 5, 10 and 11.

The Sentinel-2 2015-08-19 image in combination with the ALS-derived variables gave better results. This combination gave a lower out-of-bag error than just the 2015-08-19 image for ten of the classification schemes, the same out-of-bag error for classification scheme 8, and higher out-of-bag error for classification schemes 5, 10 and 11.

In contrast, the combination of the 2016-10-02 Sentinel-2 image and the ALS-derived variables was less accurate. This combination gave a lower out-of-bag error than the 2015-08-19 image alone for just four of the classification schemes, schemes 2, 3, 8 and 14. It generated the same out-of-bag error for classification schemes 1, 4 and 7, and a higher out-of-bag error for the remaining seven classification schemes. Using

Table 3.1. Comparison of the out-of-bag errors for the different classification schemes and variable sets. Numbers are in percentages. s215 is the 2015-08-19 Sentinel 2 image, s216 is the 2016-10-02 image. Laser refers to the ALS-derived variables: DEM, canopy density (vegkvot), tree height (p95) and wetness index (SWI). The numbers in **bold** represent the 2-class, 3-class and 4-class classifications with the lowest out-of-bag errors.

Classification	Out-of-bag errors for the variable sets						
scheme	s215	s216	laser	s215+ s216	s215+ laser	s216+ laser	s215+ s216+ laser
1	57	61	58	60	52	57	54
2	40	46	53	38	**37**	38	**37**
3	50	54	50	46	46	49	48
4	38	46	51	37	37	38	37
5	49	62	51	57	51	50	49
6	32	39	38	32	**29**	34	32
7	51	53	51	47	44	51	46
8	20	29	20	**19**	20	**19**	20
9	53	62	60	47	52	56	47
10	54	68	67	61	58	60	62
11	49	60	67	52	53	61	57
12	37	38	50	34	36	43	39
13	50	56	60	44	48	54	44
14	34	37	37	29	33	32	29

all the variables gave lower out-of-bag errors than only the 2015-08-19 Sentinel-2 image for eight classifications schemes equal out-of-bag errors for classifications schemes 5, 6 and 8, and higher out-of-bag errors for classifications schemes 10, 11 and 12.

3.2 Maps

Several maps of the classification were produced. Overviews of classification schemes 2, 4 and 6 using the Sentinel-2 2015-08-19 image and ALS-derived variables, as well as using only Sentinel-2 variables are presented in Figure 3.1 and 3.2. Larger, individual maps are presented in the appendix, Appendix figures 13-20.

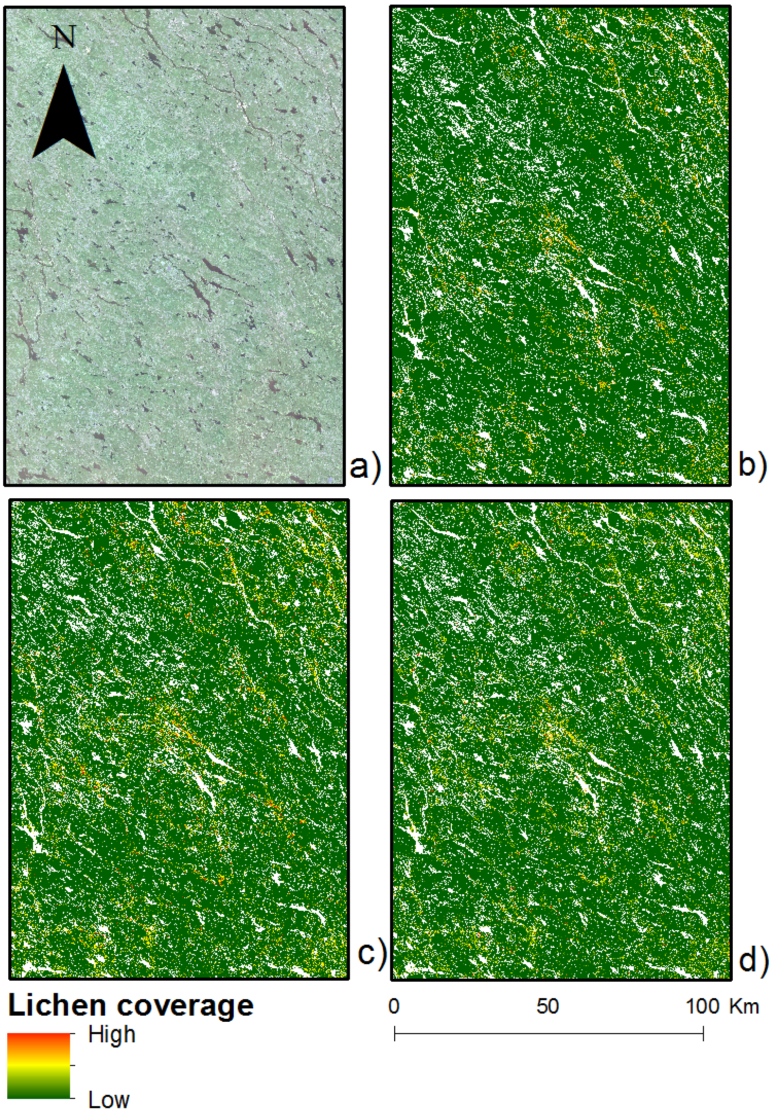


Figure 3.1. The study area, a) the Sentinel-2 2015-08-19 image shown in true colour, b) classification scheme 2 with 4 classes, c) classification scheme 4 with 3 classes, d) classification scheme 6 with 3 classes. b-d) were classified with the Random Forest algorithm using the Sentinel-2 image from 2015-08-19 and ALS-derived variables (DEM, p95, canopy density (vegkvot) and wetness index).

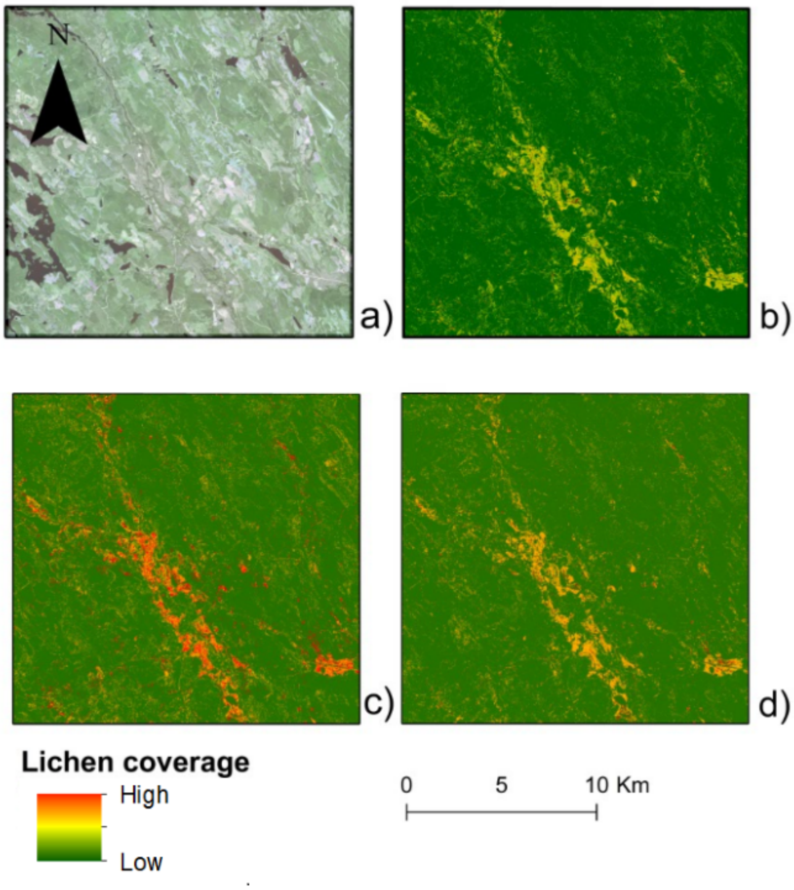


Figure 3.2. Part of the study area, a) the Sentinel-2 2015-08-19 image shown in true colour, b) classification scheme 2 with 4 classes, c) classification scheme 4 with 3 classes, d) classification scheme 6 with 3 classes. b-d) were classified with the Random Forest algorithm using the Sentinel-2 image from 2015-08-19.

3.3 Variable importance

Classification schemes 2, 4 and 6 were deemed the most interesting to present more results for. This is due to the fact that scheme 2 was the most accurate of the classification schemes that had four classes, scheme 4 was the same as that in the previous study (Gilichinsky et al. 2011) and scheme 6 was the most accurate of the schemes with three classes. Figure 3.3 shows the mean decrease in accuracy and thus the variable importance in the model for classification schemes 2, 4 and 6 when all variables were used. Appendix Figures 1-6 show the mean decrease in accuracy for the other variables combinations for the same classification schemes.

The analysis showed that when using all the variables, Sentinel-2 2015-08-19 bands 4 (red) and 8 (NIR) were the most important. For classification schemes 2 and 4 Sentinel-2 2015-08-19 bands 5 (red-edge) and 6 (red-edge) were fourth and fifth most important (Figure 3.3). For classification scheme 6 (Figure 3.3), however, 6 (red-edge) and 7 (red-edge) were fourth and fifth. Band 2 (blue) was the most important Sentinel-

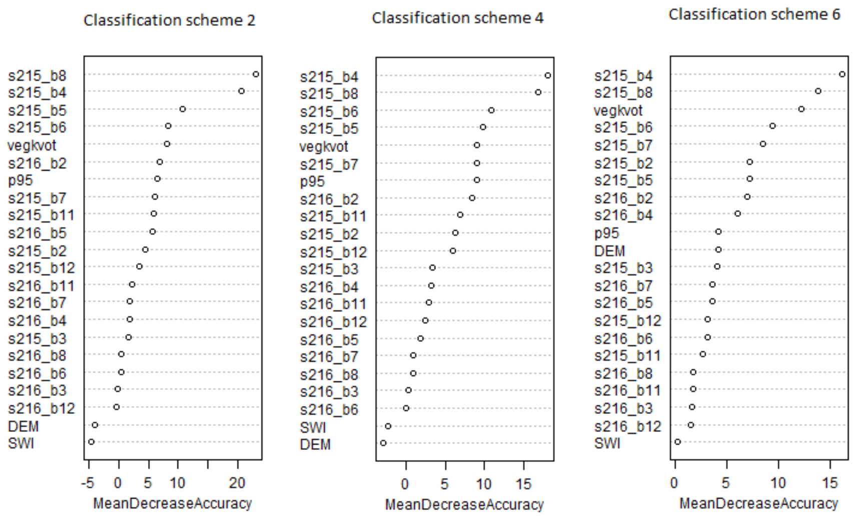


Figure 3.3. The mean decrease in accuracy for classification scheme 2, 4 and 6 and all Sentinel-2 2015-08-19 and 2016-10-02 bands as well as ALS-derived variables (DEM, forest height (p95), canopy density (vegkvot) and wetness index (SWI)). The variables are ordered top-to-bottom as most to least important.

2 2016-10-02 band for all three classification schemes, and 2016-10-02 bands were generally less important than 2015-08-19 bands. For the ALS-derived variables, canopy density remained the most important variable, as fifth and third most important for the different schemes. Forest height was also fairly important over all, but wetness index was the least important.

Overall, the analysis showed that the Sentinel-2 2015-08-19 image bands 4 (red) and 8 (NIR) along with ALS-derived canopy density metric were the most important variables for all three schemes and all different combinations of variables (Figure 3.3 and Appendix figures 1-6). Forest height was of middling importance, while wetness index remained the least important variable. For the Sentinel-2 2016-10-02 image, bands 4 (red) and 5 (red-edge) were generally the most important.

3.4 Bandwise mean and variation

Tables 3.2-3.4 show standard deviation of the mean Digital Number for each band. The standard deviation was greater for the majority of the classes with less lichen cover and smaller for the classes with greater lichen cover. The standard deviation for the Sentinel-2 bands within the visible spectrum were generally smaller than those within the red edge and NIR spectra.

Table 3.2. Mean and standard deviation of the Digital Number for each band for the four classes and all possible variables for classification scheme 2

Variables	Mean and standard deviation for the lichen cover classes							
	0-25 %		26-50 %		51-75 %		76-100 %	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
s215 b2	807	70.61	815.39	38.42	830.99	28.61	849.6	32.1
s215 b3	656.3	93.2	652.84	43.23	669.1	35.5	695.4	28.9
s215 b4	421.8	122.6	440.4	71.8	461.9	43.9	504.4	54.7
s215 b5	713.8	201.5	690.1	70.1	738.3	73.7	773.7	83
s215 b6	1630	420.7	1471.6	165.4	1471.1	167.4	1414.3	53.7
s215 b7	1951.1	521	1722.8	207.5	1733.2	201.5	1633.7	39.8
s215 b8	1931.6	572.3	1709.2	196.7	1674.6	182.2	1600.1	114.4
s215 b11	1012.8	294.9	1039.2	175.5	1161.1	142.4	1239	183
s215 b12	454.8	155.7	480.9	103.5	544.6	70.7	612.3	109.1
s216 b2	1057.3	43.9	1065.1	30.4	1069.1	23.4	1065.8	27.4
s216 b3	729.79	60.2	742.59	50.77	759.7	45.9	759.59	16.61
s216 b4	496.4	112.4	492.1	60.3	513.3	51.2	522.2	29.2
s216 b5	700.7	188.1	702.9	111.8	771.3	97.6	732.3	52
s216 b6	1264.3	308.1	1295.2	232.8	1449	243.3	1322.3	72.1
s216 b7	1444.8	377.6	1462.9	266.7	1650.2	281.9	1474.3	82.9
s216 b8	1475.4	460.2	1498	287.4	1627.3	312.1	1467.6	122.9
s216 b11	685.2	309.3	663	191.5	798.6	181.9	787.7	104.7
s216 b12	324.1	169.4	302.1	94.9	364.8	82.2	373	48.6
DEM	265.35	49.87	256.59	47.01	260.27	29.02	258.04	16.35
SWI	4.81	1.279	4.576	1.178	4.305	1.067	4.4	2.05
vegkvot	50.58	25.44	44.88	16.11	31.51	16.66	47.9	24.5
p95	13.852	5.155	13.088	3.782	14.21	4.94	15.735	0.494

Table 3.3. Mean and standard deviation of the Digital Number for each band for the three classes and all possible variables for classification scheme 4

Variables	Mean and standard deviation for the lichen cover classes					
	0-25 %		26-50 %		51-100 %	
	Mean	SD	Mean	SD	Mean	SD
s215 b2	807	70.61	815.39	38.42	835.29	29.19
s215 b3	656.3	93.2	652.84	43.23	675.2	34.92
s215 b4	421.8	122.6	440.4	71.8	471.7	47.9
s215 b5	713.8	201.5	690.1	70.1	746.5	73.9
s215 b6	1630	420.7	1471.6	165.4	1458	148.7
s215 b7	1951.1	521	1722.8	207.5	1710.2	180.6
s215 b8	1931.6	572.3	1709.2	196.7	1657.4	167.7
s215 b11	1012.8	294.9	1039.2	175.5	1179	148.2
s215 b12	454.8	155.7	480.9	103.5	560.2	81.3
s216 b2	1057.3	43.9	1065.1	30.4	1068.4	23.2
s216 b3	729.79	60.2	742.59	50.77	759.7	40.3
s216 b4	496.4	112.4	492.1	60.3	515.4	46.1
s216 b5	700.7	188.1	702.9	111.8	762.3	88.8
s216 b6	1264.3	308.1	1295.2	232.8	1419.8	219.8
s216 b7	1444.8	377.6	1462.9	266.7	1609.6	258.3
s216 b8	1475.4	460.2	1498	287.4	1590.5	283.7
s216 b11	685.2	309.3	663	191.5	796.1	163.3
s216 b12	324.1	169.4	302.1	94.9	366.7	74
DEM	265.35	49.87	256.59	47.01	259.76	26.02
SWI	4.81	1.279	4.576	1.178	4.327	1.249
vegkvot	50.58	25.44	44.88	16.11	35.29	18.96
p95	13.852	5.155	13.088	3.782	14.56	4.34

Table 3.4. Mean and standard deviation of the Digital Number for each band for the three classes and all possible variables for classification scheme 6

Variables	Mean and standard deviation for the lichen cover classes					
	0-33 %		34-66 %		67-100 %	
	Mean	SD	Mean	SD	Mean	SD
s215 b2	807.82	67.69	818.98	34.42	840.7	34.9
s215 b3	655.5	88.2	654.93	42.58	689.4	27.4
s215 b4	423.5	117.9	445.3	63.9	484.7	54
s215 b5	710.2	190	699	72.2	767	74.9
s215 b6	1608.9	400.7	1458.5	165.4	1508.3	153.6
s215 b7	1920.7	498.3	1710.9	204.7	1763.1	192.8
s215 b8	1896.8	547	1704.2	199.3	1692.5	155.1
s215 b11	1013.2	281.7	1071.7	175.9	1201.3	158
s215 b12	457.7	149.9	495.8	100.2	575.4	95.1
s216 b2	1056.1	41.5	1069.6	29.6	1072.2	25.6
s216 b3	728.87	56.92	749.05	50.59	770.5	47.1
s216 b4	493.1	106.2	501.5	59.7	523.4	53.5
s216 b5	697.5	178	719.9	111.4	778.4	105.1
s216 b6	1259.5	291.5	1334.3	242.8	1451.1	250.6
s216 b7	1441.1	357.4	1497.5	276	1656	305
s216 b8	1462.1	436.3	1549.5	294.5	1609	308
s216 b11	677.7	293.2	698	196.6	819	191.4
s216 b12	318.8	160.6	318.9	95.6	381.1	85.9
DEM	267.59	49.31	247.27	41.09	270.06	20.77
SWI	4.776	1.273	4.563	1.147	4.202	1.386
vegkvot	50.22	24.77	41.48	15.99	36.98	20.37
p95	13.827	4.925	13.264	4.29	14.62	3.88

3.5 Scatter plots

Scatter plots are shown in Figures 3.4-3.6 for the two most important spectral bands for classification schemes 2, 4 and 6. No clear patterns for class division based on the spectral bands emerge. However, it is possible to see that the class with the lowest lichen cover percentages is the largest class, and that it stretches over the entire scatter plot, i.e., it has a large standard deviation.

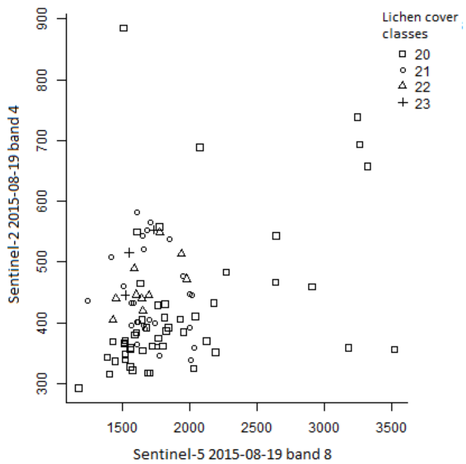


Figure 3.4. Scatter plot of the two most important Sentinel-2 2015-08-19 bands, band 8 and band 4, for classification scheme 2. Each point represents a value for each of the two spectral bands, as well as a class. The classes are 20: 0-25 %, 21: 26-50 %, 22: 51-75 % and 23 76-100 % lichen coverage.

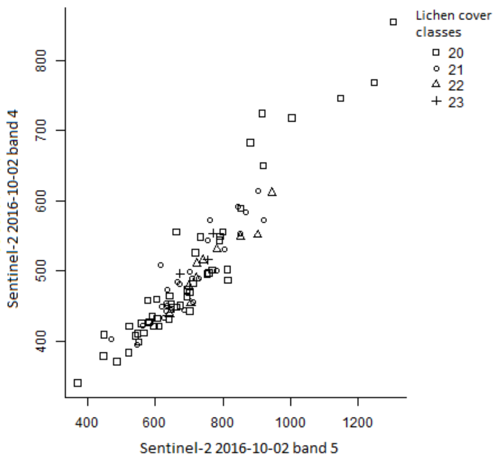


Figure 3.5. Scatter plot of the two most important Sentinel-2 2016-10-02 bands, band 5 and band 4, for classification scheme 2. Each point represents a value for each of the two spectral bands, as well as a class. The classes are 20: 0-25 %, 21: 26-50 %, 22: 51-75 % and 23 76-100 % lichen coverage.

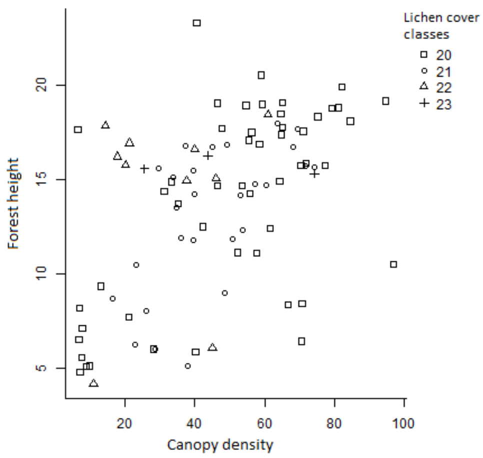


Figure 3.6. Scatter plot of the two most important ALS-derived variables, canopy density and forest height (p95), for classification scheme 2. Each point represents a value for each of the two spectral bands, as well as a class. The classes are 20: 0-25 %, 21: 26-50 %, 22: 51-75 % and 23 76-100 % lichen coverage.

4 Discussion

4.1 Utility of Sentinel-2 data for lichen cover classification

The objective of this study was to evaluate the use of data from the Sentinel-2 satellite for the classification of ground lichen and to assess whether the addition of information from airborne laser scanning data would improve the results. The Sentinel-2 satellite is new and particularly interesting for classification with its increased number of red-edge spectral data bands, frequent imaging, and bands with 10-20 meter spatial resolution. Using Sentinel-2 images is also interesting as a means to classify or map lichens over even larger areas than the study area used in this thesis.

As for the 2015-08-19 Sentinel-2 image versus the 2016-10-02 image, the 2015-08-19 image generally gave higher accuracy, both by itself and in combination with other data. This might be due to the fact that the August image is more useful for ground vegetation classification. While the lichen stays the same spectrally over the year, it is very rare that a single pixel would consist entirely of lichen. This is especially true for the classes with lower lichen coverage. Other vegetation, such as trees, dwarf shrubs, herbs, and grasses, contribute to the spectral information. Coniferous trees might not be very affected spectrally by the changing season, but other species groups, such as deciduous trees and dwarf shrubs, are. There is probably a difference in senescence between August and October, which will affect the result of the classification. Among the dwarf shrubs, the lingon and heather are not as different spectrally between the seasons as bilberry is. This is the reason for the creation of the classification schemes that are based on the co-existing dominant dwarf shrub species. The hope was that the additional information would increase the accuracy, which it failed to do.

The study area being situated so far north also means that the change in time of year between the August of 2015 and October of 2016 Sentinel-2 images impacts the day length and the position of the sun. This will in turn affect the quality of the light and the length of shadows. Especially shadows might interfere with classification results.

For the Sentinel-2 2015-08-19 image, band 4 (red) and 8 (NIR), were the most important for the classification. For the classifications using the 2016-10-02 image, bands 4 (red) and 5 (red-edge), were the most important. For the classification solely using the ALS-derived metrics, canopy density and tree height were most important for classification schemes 2 and 4, while canopy density and DEM were the most important for scheme 6.

The fact that band 8 had such a great impact on classification was not surprising, as NIR is known to be of great use for classification of vegetation. Band 4 (red) was also important for classification. Käyhkö and Pellikka (1994) also found that the red band was useful for identifying lichen, thus the importance of the red band was not entirely unexpected. However, it was expected that band 2 (blue) would be more important for classifying lichen. Previous studies had indicated that the blue band would be useful for identifying lichen (Nelson et al. 2013; Petzold and Goward 1988).

One challenge in this study was the newness of the Sentinel-2 satellite. There were some problems with the images. Initially, the plan for the study had been to include the 34-WDS granule. However, there were significant discrepancies between the WDS and the WXM granules both spatially and spectrally, for pixels that should have matched. The discrepancies were said to be due to geographic corrections to two different UTM zones, which are performed on a granule-to-granule basis rather than a scene-to-scene basis (Rosengren 2016). This, together with the problem of finding cloud-free images for the WDS granule, led to the decision to exclude it from the study.

The newness of the Sentinel-2 satellite also meant that, at the time of the start of the study, there was only little more than a year's worth of images to choose from, namely the period between the launch of the Sentinel-2A satellite in June 2015 and the beginning phase of this study, October 2016. During this period only two cloud-free images could be found. It would have been interesting to use at least three images in the study, but this was not possible. The fact that the images chosen fall during different times of the year also affects the results. The 2015-08-19 image is from August, whereas the 2016-10-02 image is from October, when senescence has already started. The greatest spectral differences between vegetation types occur during spring and autumn, right before the leaves fall.

4.2 Utility of combining Sentinel-2 and airborne laser data for lichen cover classification

As stated, the objective of this study was to evaluate the use of data from the Sentinel-2 satellite for the classification of ground lichen cover and to assess whether the addition of information from airborne laser scanning data would improve the results. For the most part adding ALS-derived metrics did improve classification results. Adding ALS-derived metrics to the Sentinel-2 2015-08-19 image gave lower out-of-bag error for ten of the 14 different classification schemes, and ALS-derived metrics improved the result of the classification using both satellite images as it gave a lower out-of-bag error for eight of the classification schemes. The addition of ALS-derived metrics to the 2016-10-02 image however only improved the out-of-bag errors for four of the classification schemes.

As for the most influential ALS-derived variables, canopy density in the form of a vegetation ratio was unquestionably the most important, while forest height and elevation from the DEM were also quite important for the results. As this study attempts to classify lichen coverage on forest land, both canopy density and tree height have an impact on spectral composition, as well as the amount of lichen in an area. Lichen grows more frequently and in greater abundance in open forests, i.e. forests with a lower canopy density. Canopy density would also be expected to be greater for a spruce dominated forest, in which there would be little or no lichen, than for a pine forest, where lichens are more likely to grow. Furthermore, a greater canopy density would shade out the ground, making it harder to image using remote sensing (Théau et al. 2005; Käyhkö and Pellikka 1994).

It was initially surprising that wetness index had the least impact of all variables, both ALS-derived metrics and spectral bands, seeing how lichen abundance tends to correlate well with dryness. However, the wetness index does not take soil texture into account, nor is this information available at a suitable spatial resolution. Coarse sand in river flood plains that are very dry could thus accidentally be labelled as 'wet' due to their position in the landscape.

4.3 Class definitions of lichen cover

Over all, the classification schemes with fewer classes showed greater accuracy, which was to be expected. A classification scheme with three lichen cover classes, similar to the one used by Gilichinsky et al. (2011), may be the most reasonable to use since it has an accuracy that would be acceptable in a mapping of lichen. Classification scheme 2 (0-25 %, 26-50 %, 51-75 % and 76-100 % lichen coverage) gave more detail by dividing Gilichinsky et al. (2011)'s highest percentage class of lichen coverage into two classes. This division, however, might not be worthwhile, given the lower accuracy of the model.

No clear pattern for class division could be discerned based on the scatter-plots of the two most important variables. Ideally, the individual classes would have been grouped together, or showed some other pattern. This was not the case.

The standard deviations of the mean Digital Numbers for each band were generally greater for the classes with less lichen cover and smaller for the classes with greater lichen cover. A larger standard deviation shows that the data points are spread out from the mean, i.e. that there is a large variation within the different classes in this case. This is not unexpected, as there were fewer plots in the dataset with higher percentages of lichen cover but also due to the fact that areas with greater lichen cover are more homogeneous than those with less or no lichen. The former class consists largely of dry pine heath with dwarf shrubs, whereas the latter class can be characterized as 'all other forest land', meaning various tree species on various types of land and so forth. This makes the possible spectral signatures of this class very varied, as compared to the classes with the greatest lichen coverage. Perhaps a more detailed division of this class would increase the accuracy of the classification.

The hope was that adding information regarding the co-existing species in the field layer would help explain spectral variation and improve the model performance. While this did not result in higher accuracies in this study, perhaps due to the limited field data, it may still be of value for future studies to consider. Another reason why the field-layer data did not increase accuracy might be the fact that such data is of more use for sparse forests types than for more dense forest, where the canopy shades out the field-layer. This study included all forest land within the study area.

4.4 Field data

Within remote sensing, it is important to have accurate and representative field data. Without reliable field reference data, it is hard to use the satellite image classify the actual landscape with any accuracy.

This study uses field data from a previous study (Gilichinsky et al. 2011) along with 20 newly inventoried plots. Starting with the old plots, 63 were removed in the change analysis due to clear-cutting, which was expected. However, plots were also missing from the expected dataset from the earlier study. The missing plots were mostly those with a higher percentage lichen coverage, which skewed the dataset towards no-lichen plots. Furthermore, both among the old and new plots, plots were removed due to N/A values in the ALS-derived metrics, which the random forest algorithm is unable to process. An additional 39 plots were removed when the WDS Sentinel-2 granule was removed from the study area. This further unexpectedly reduced the training data. To somewhat balance the training data between no-lichen and lichen plots, 33 no-lichen plots were eliminated before running the random forest algorithm. In the end, only 90 of an initially possible 249 plots remained.

The number of field plots might be the greatest weakness of the study. In hindsight, inventorying an additional 20 to 30 new plots would have greatly benefited the study.

4.5 Comparison to previous study and recommendations for future studies

Gilichinsky et al. (2011) utilized lichen information collected by the Swedish National Forest Inventory (NFI) as training data to classify lichen cover using one Landsat-7 TM and one SPOT 5 satellite image. For the SPOT 5 image they used a maximum of 943 NFI plots as training data, and for the larger Landsat-7 scene they used a maximum of 3670 NFI plots as training data. The lichen cover classes were: 0-25 %, 26-50 % and 51-100 %.Gilichinsky et al. tested three different classification methods: Mahalanobis distance, maximum likelihood and spectral mixture analysis (SMA). They got high accuracy for the classification of the SPOT 5 image when using the Mahalanobis distance classifier, with an overall accuracy 84.3 %. The classification accuracy for the Landsat-7 image was slightly lower at 76.8 %, using the maximum likelihood classification.

This study is based on the Gilichinsky et al. (2011) study. It uses largely the same study area, and utilizes the earlier study's validation data as training data. This study uses other data for classification, however, and a different method, namely the Random Forest algorithm to try several different classification schemes, compared to the earlier study's one classification scheme.

The classification results of this study are not as accurate as those of the earlier study (Gilichinsky et al. 2011). The best out-of-bag error was 19 % for the present study's classification scheme 8, which has two classes. Generally, though, the out-of-bag errors fall in the span between 20 % and 60 %. One of the lichen coverage classification schemes in this study was the same as the one Gilichinsky et al. (2011) tested, namely classification scheme 4. The most accurate result for that classification scheme was an out-of-bag error of 37 %.

However the purpose of this study was not to get higher classification accuracies than Gilichinsky et al., but to test new data sources: the Sentinel-2 satellite and ALS-derived metrics. In that, this study succeeded. Sentinel-2 is indeed useful for classification, as are the ALS-derived metrics. A single summer date of Sentinel-2 imagery was sufficient, as compared to combining Summer and Fall images, for good results. That said, it would have been interesting to use three or four Sentinel-2 images, and to include data about soil texture and fertility in addition to the ALS-derived metrics to perhaps increase accuracy.

In the future for the field of lichen mapping as a whole, quantifying lichen biomass would be interesting and of potential use for the forestry and reindeer husbandry land-use. With the debate about the effect of global warming on the environment, and the influence of forestry on lichen cover, it would also be of interest to study change in lichen cover over time.

Bibliography

Berg, A. (2010). "Reindeer herding and modern forestry the historical impacts on forests of two main land users in northern Sweden". PhD thesis. Umeå: Dept. of Forest Ecology and Management, Swedish University of Agricultural Sciences.

Breiman, L. (1996). "Out-of-bag estimation". In:

Breiman, L. (2001). "Random Forests". In: *Machine Learning* 45.1, pp. 5–32. ISSN: 0885-6125.

Brenning, A. (2008). "Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models". In: *SAGA – Seconds Out (= Hamburger Beitraege zur Physischen Geographie und Landschaftsoekologie, vol. 19)*. J. Boehner, T. Blaschke, L. Montanarella, pp. 23–32.

Colpaert, A., Kumpula, J., and Nieminen, M. (2003). "Reindeer pasture biomass assessment using satellite remote sensing". In: *Arctic* 56.2. ISSN: 0004-0843.

European Space Agency (2017a). *Open Access Hub*. URL: `https://scihub.copernicus.eu/` (visited on 2017-09-06).

European Space Agency (2017b). *Radiometric - Resolutions - Sentinel-2 MSI - User Guides - Sentinel Online*. URL: `https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric` (visited on 2017-09-06).

European Space Agency (2017c). *Sentinel-2 - ESA Operational EO Missions*. URL: `https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/sentinel-2` (visited on 2017-09-06).

European Space Agency (2017d). *Sentinel-2 - Heritage - Sentinel Online*. URL: `https://sentinel.esa.int/web/sentinel/missions/sentinel-2/heritage` (visited on 2017-09-06).

European Space Agency (2017e). *Sentinel-2 - Missions - Sentinel Online*. URL: `https://sentinel.esa.int/web/sentinel/missions/sentinel-2` (visited on 2017-09-06).

European Space Agency (2016). *SENTINEL-2 MSI Introduction*. URL: `https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi` (visited on 2016-10-03).

Falldorf, T., Strand, O., Panzacchi, M., and Tømmervik, H. (2014). "Estimating lichen volume and reindeer winter pasture quality from Landsat imagery". In: *Remote Sensing of Environment* 140, pp. 573–579. ISSN: 0034-4257.

Gilichinsky, M., Sandström, P., Reese, H., Kivinen, S., Moen, J., and Nilsson, M. (2011). "Mapping ground lichens using forest inventory and optical satellite data". In: *International Journal of Remote Sensing* 32.2, pp. 455–472. ISSN: 0143-1161.

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R. (2006). "Random Forests for land cover classification". In: *Pattern Recognition Letters* 27.4, pp. 294–300. ISSN: 0167-8655.

Heggberget, T. M., Gaare, E., and Ball, J. P. (2002). "Reindeer (Rangifer tarandus) and climate change: Importance of winter forage". In: *Rangifer* 22.1, pp. 13–31. ISSN: 1890-6729.

Johnson, C. J., Alexander, N. D., Wheate, R. D., and Parker, K. L. (2003). "Characterizing woodland caribou habitat in sub-boreal and boreal forests". In: *Forest Ecology and Management* 180.1, pp. 241–248. ISSN: 0378-1127. DOI: 10.1016/S0378-1127(02)00563-7. URL: http://www.sciencedirect.com/science/article/pii/S0378112702005637.

Käyhkö, J. and Pellikka, P. (1994). "Remote sensing of the impact of reindeer grazing on vegetation in northern Fennoscandia using SPOT XS data". In: *Polar Research* 13.1, pp. 115–124. ISSN: 0800-0395.

Lantmäteriet (2017). *Lantmäteriet*. URL: https://www.lantmateriet.se/ (visited on 2017-09-09).

Lantmäteriet (2016). *Product description: Laser data*. URL: https://www.lantmateriet.se/globalassets/kartor-och-geografisk-information/hojddata/produktbeskrivningar/eng/lidar_data.pdf.

Liaw, A. and Wiener, M. (2002). "Classification and Regression by randomForest". In: *R News* 2.3, pp. 18–22. URL: http://CRAN.R-project.org/doc/Rnews/.

McGaughey, R. (2016). *FUSION/LDV: Software for LIDAR Data Analysis and Visualization*. URL: http://forsys.cfr.washington.edu/fusion/%7BFUSION%7D%5C_manual.pdf.

Nelson, P. R., Roland, C., Macander, M. J., and Mccune, B. (2013). "Detecting continuous lichen abundance for mapping winter caribou forage at landscape spatial scales". In: *Remote Sensing of Environment* 137, pp. 43–54. ISSN: 0034-4257.

Nordberg, M.-L. (1998). "Vegetation and biomass changes in mountainous areas in Sweden using satellite and airborne imaging scanner data". In: *Information for sustainability*, pp. 431–435.

Nordberg, M.-L. and Allard, A. (2002). "A remote sensing methodology for monitoring lichen cover". In: *Canadian Journal of Remote Sensing* 28.2, pp. 262–274. ISSN: 0703-8992.

Petzold, D. E. and Goward, S. N. (1988). "Reflectance spectra of subarctic lichens". In: *Remote Sensing of Environment* 24.3, pp. 481–492. ISSN: 0034-4257. DOI: 10.1016/0034-4257(88)90020-X. URL: http://www.sciencedirect.com/science/article/pii/003442578890020X.

Rautiainen, M., Suomalainen, J., Mõttus, M., Stenberg, P., Voipio, P., Peltoniemi, J., and Manninen, T. (2007). "Coupling forest canopy and understory reflectance in the Arctic latitudes of Finland". In: *Remote Sensing of Environment* 110.3, pp. 332–343. ISSN: 0034-4257.

Rees, W., Williams, M., and Vitebsky, P. (2003). "Mapping land cover change in a reindeer herding area of the Russian Arctic using Landsat TM and ETM+ imagery and indigenous knowledge". In: *Remote Sensing of Environment* 85.4, pp. 441–452. ISSN: 0034-4257.

Reese, H., Nyström, M., Nordkvist, K., and Olsson, H. (2014). "Combining airborne laser scanning data and optical satellite data for classification of alpine vegetation". In: *International Journal of Applied Earth Observation and Geoinformation* 27, pp. 81–90.

Rennäringslag (1971). SFS 1971:437. Stockholm.

Riksskogstaxeringen (2017). *Riksskogstaxeringen - Externwebben.* URL: http://www.slu.se/centrumbildningar-och-projekt/riksskogstaxeringen/ (visited on 2017-09-09).

Rodriguez-Galiano, V., Chica-Olmo, B., Ghimire, J., Rogan, M., and Rigol-Sanchez, J. (2012). "An assessment of the effectiveness of a random forest classifier for land-cover classification". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 67.1, pp. 93–104. ISSN: 09242716.

Rosengren, M. (2016). Personal communication.

Sandström, P. (2015). "A toolbox for co-production of knowledge and improved land use dialogues: the perspective of reindeer husbandry". PhD thesis. Umeå: Department of Forest Resource Management, Swedish University of Agricultural Sciences. ISBN: 978-91-576-8238-3.

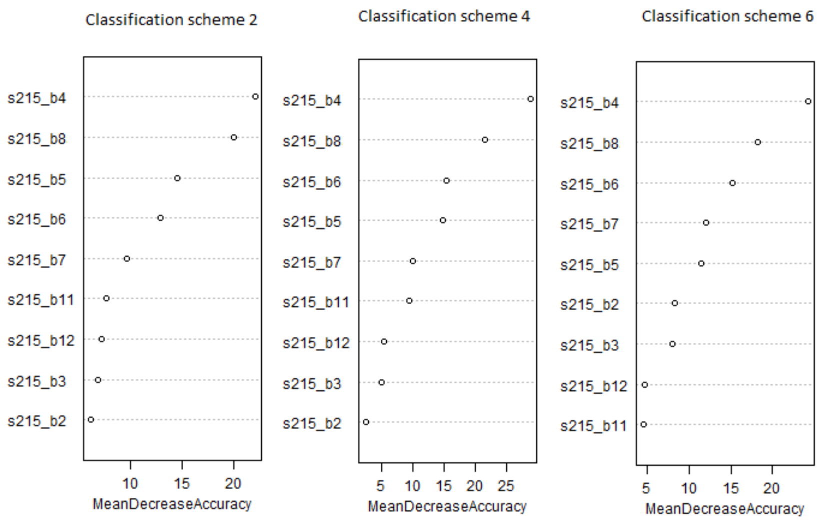
Sandström, P., Cory, N., Svensson, J., Hedenås, H., Jougda, L., and Borchert, N. (2016). "On the decline of ground lichen forests in the Swedish boreal landscape: Implications for reindeer husbandry and sustainable forest management". In: *Ambio* 45.4, pp. 415–429.

Sandström, P., Pahlén, T. G., Hagner, O., Olsson, H., Egberth, M., Edenius, L., Tømmervik, H., Hemberg, L., Baer, K., Stenlund, T., and Brandt, L. G. (2003). "Conflict Resolution by Participatory Management: Remote Sensing and GIS as Tools for Communicating Land-use Needs for Reindeer Herding in Northern Sweden". In: *Ambio* 32.8, pp. 557–567. ISSN: 00447447.

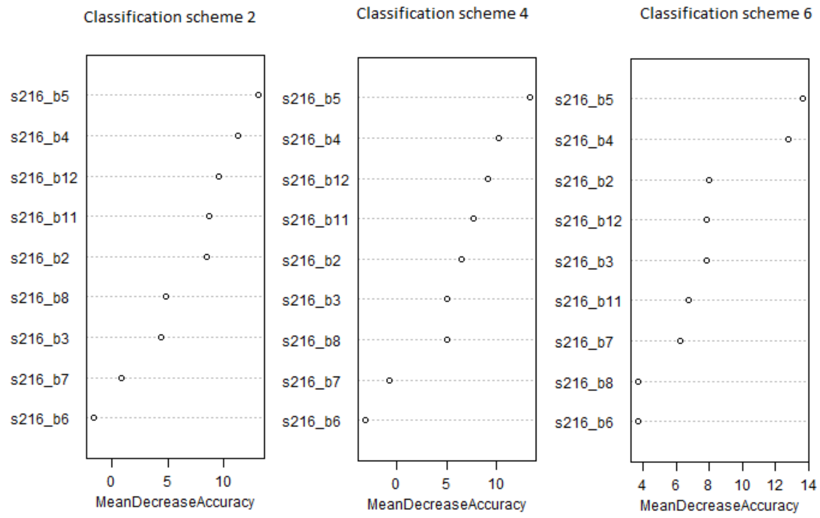
Théau, J., Peddle, D., and Duguay, C. (2005). "Mapping lichen in a caribou habitat of Northern Quebec, Canada, using an enhancement_classification method and spectral mixture analysis". In: *Remote Sensing of Environment* 94.2, pp. 232–243. ISSN: 0034-4257.

Tømmervik, H., Høgda, K. A., and Solheim, I. (2003). "Monitoring vegetation changes in Pasvik (Norway) and Pechenga in Kola Peninsula (Russia) using multitemporal Landsat MSS/TM data". In: *Remote Sensing of Environment* 85.3, pp. 370–388. ISSN: 0034-4257.

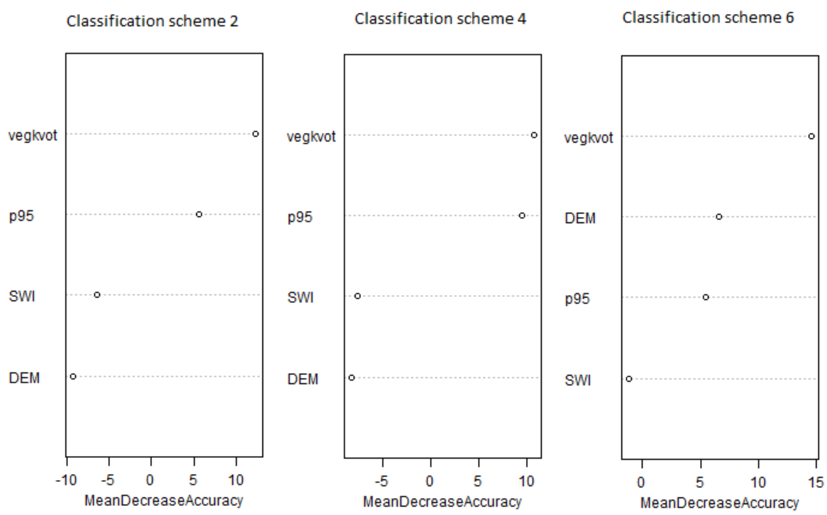
Appendix



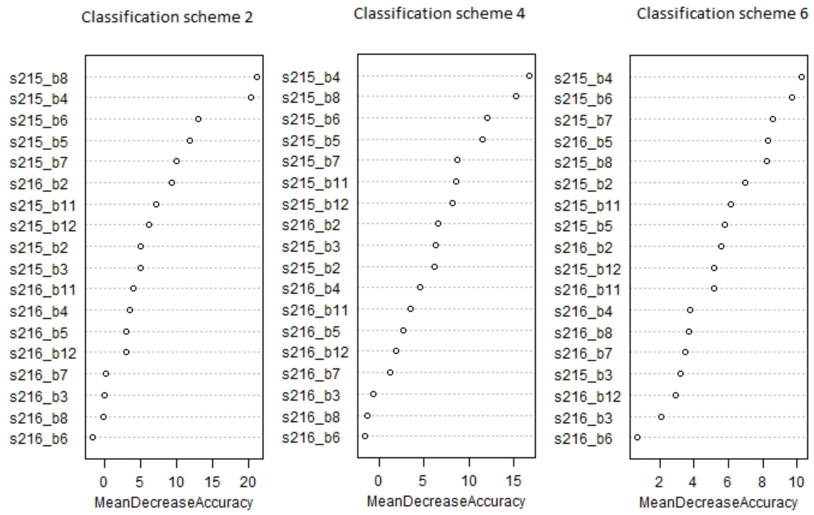
Appendix figure 1. The mean decrease in accuracy for classification scheme 2, 4 and 6 and all Sentinel-2 2015-08-19 bands. The variables are ordered top-to-bottom as most to least important.



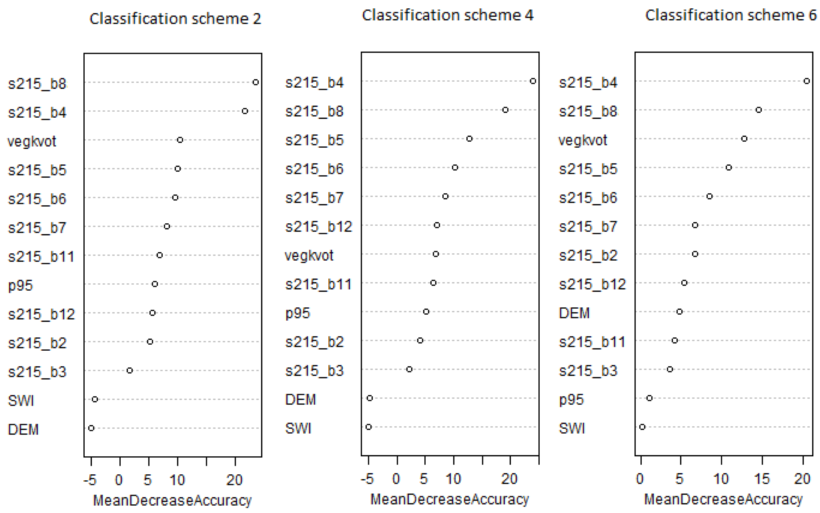
Appendix figure 2. The mean decrease in accuracy for classification scheme 2, 4 and 6 and all Sentinel-2 2016-10-02 bands. The variables are ordered top-to-bottom as most to least important.



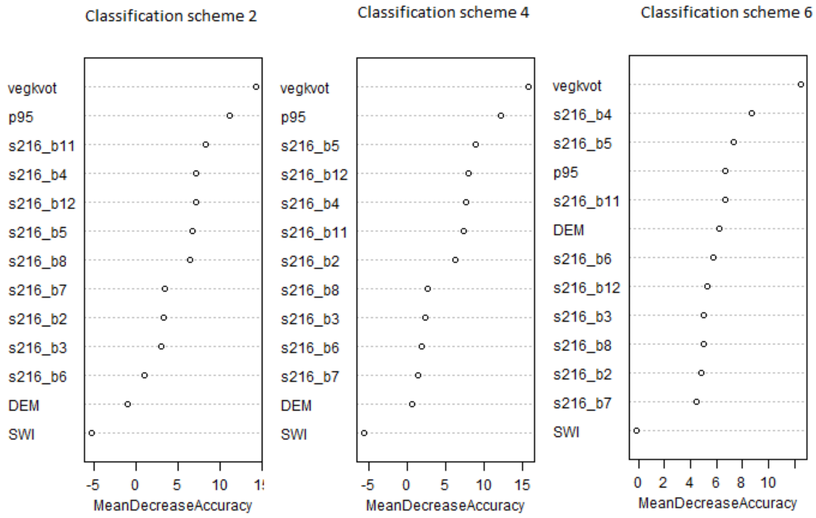
Appendix figure 3. The mean decrease in accuracy classification scheme 2, 4 and 6 and all ALS-derived variables (DEM, forest height (p95), canopy density (vegkvot) and wetness index (SWI)). The variables are ordered top-to-bottom as most to least important.



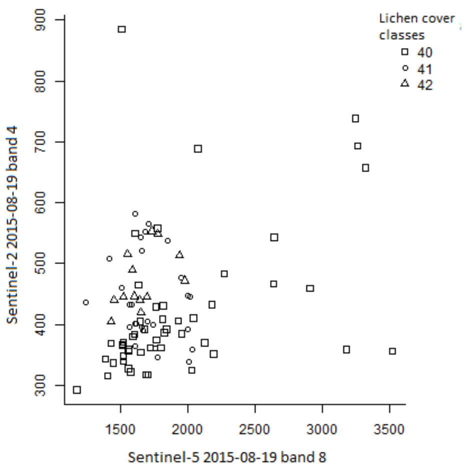
Appendix figure 4. The mean decrease in accuracy for classification scheme 2, 4 and 6 and all Sentinel-2 2015-08-19 and 2016-10-02 bands. The variables are ordered top-to-bottom as most to least important.



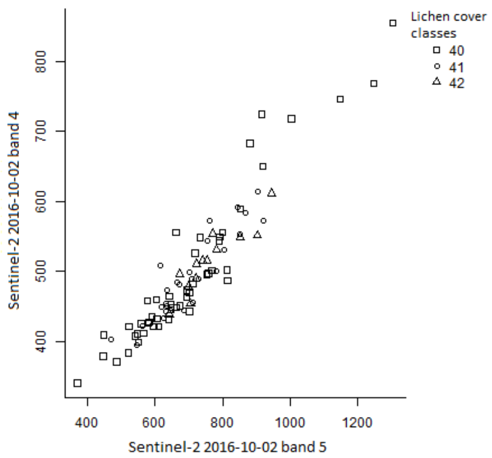
Appendix figure 5. The mean decrease in accuracy for classification scheme 2, 4 and 6 and all Sentinel-2 2015-08-19 bands and ALS-derived variables (DEM, forest height (p95), canopy density (vegkvot) and wetness index (SWI)). The variables are ordered top-to-bottom as most to least important.



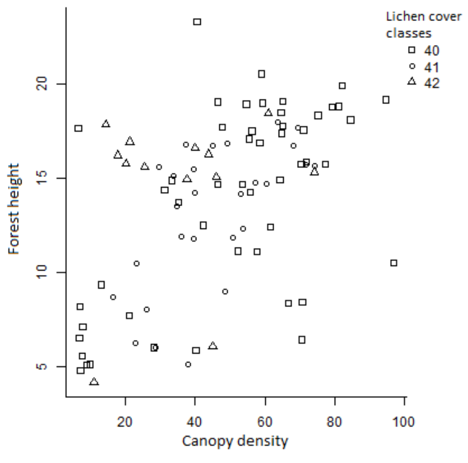
Appendix figure 6. The mean decrease in accuracy for classification scheme 2, 4 and 6 and all Sentinel-2 2016-10-02 bands and ALS-derived variables (DEM, forest height (p95), canopy density (vegkvot) and wetness index (SWI)). The variables are ordered top-to-bottom as most to least important.



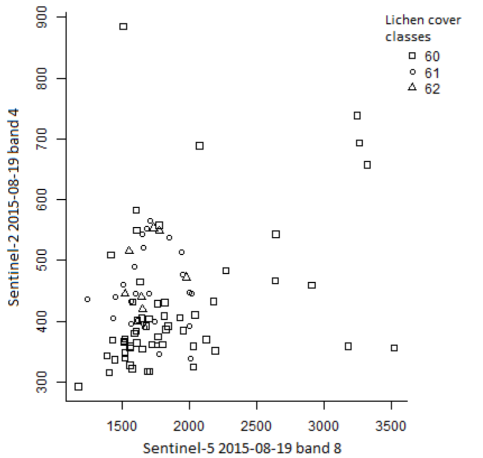
Appendix figure 7. Scatter plot of the two most important Sentinel-2 2015-08-19 bands, band 8 and band 4, for classification scheme 4. The classes are 40: 0-25 %, 41: 26-50 % and 42: 51-100 % lichen coverage.



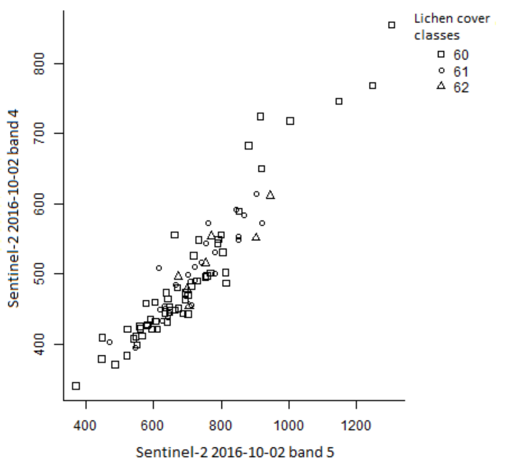
Appendix figure 8. Scatter plot of the two most important Sentinel-2 2016-10-02 bands, band 5 and band 4, for classification scheme 4. The classes are 40: 0-25 %, 41: 26-50 % and 42: 51-100 % lichen coverage.



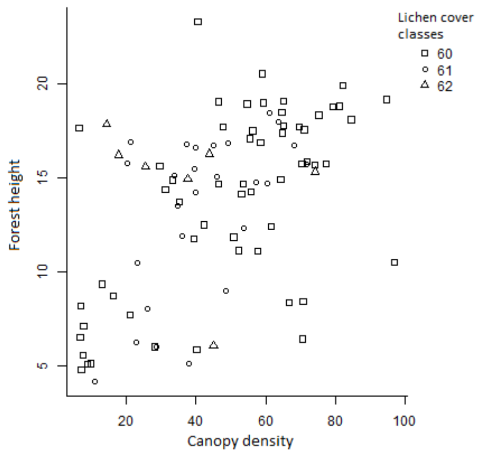
Appendix figure 9. Scatter plot of the two most important ALS-derived variables, vegetation ratio and forest height, for classification scheme 4. The classes are 40: 0-25 %, 41: 26-50 % and 42: 51-100 % lichen coverage.



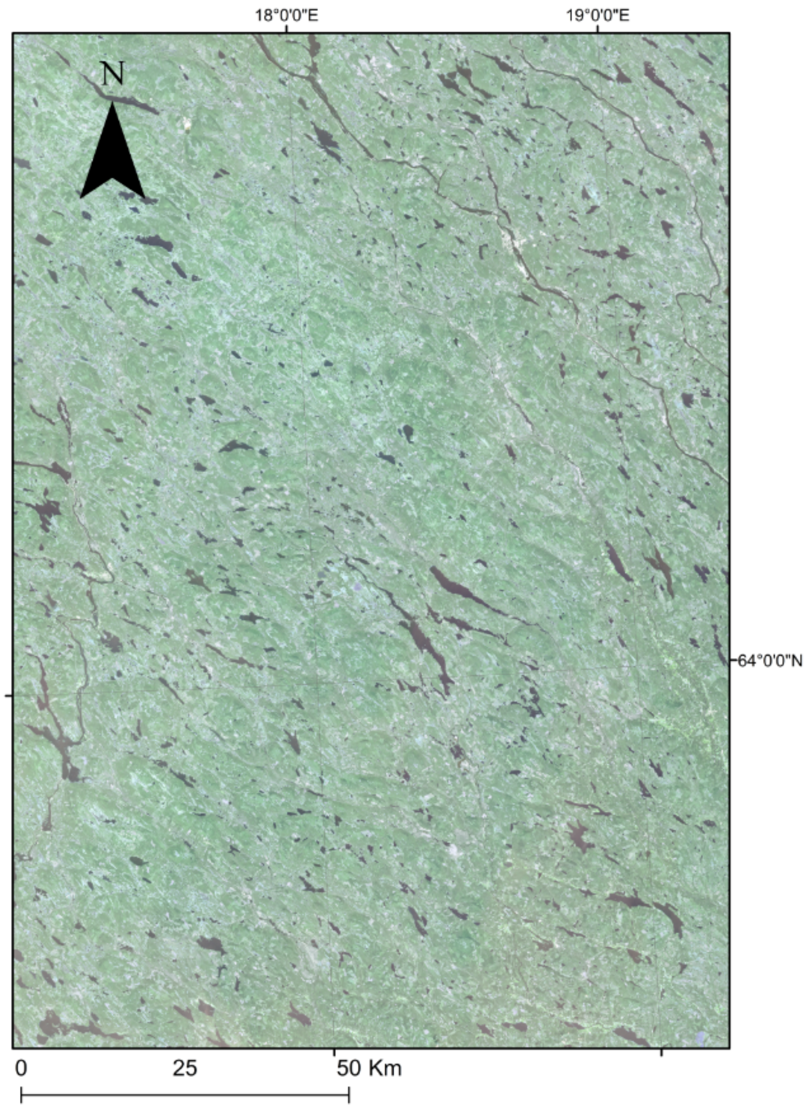
Appendix figure 10. Scatter plot of the two most important Sentinel-2 2015-08-19 bands, band 8 and band 4, for classification scheme 6. The classes are 60: 0-33 %, 61: 34-66 % and 62: 67-100 % lichen coverage.



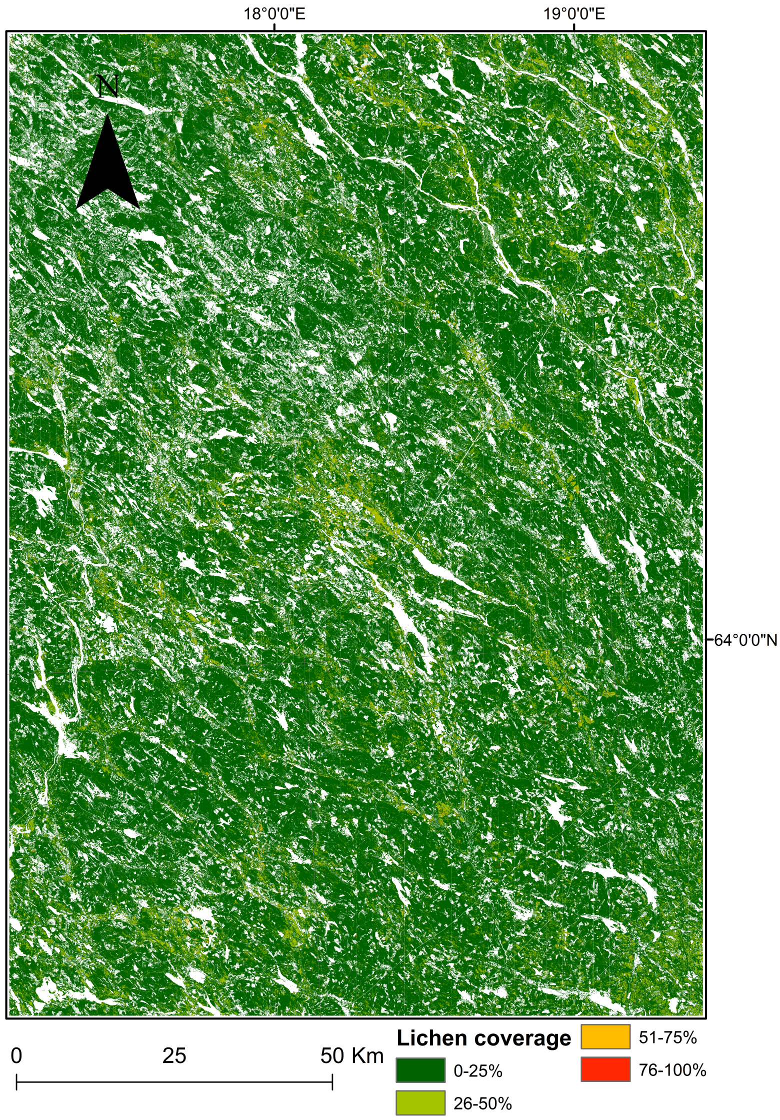
Appendix figure 11. Scatter plot of the two most important Sentinel-2 2016-10-02 bands, band 5 and band 4, for classification scheme 6. The classes are 60: 0-33 %, 61: 34-66 % and 62: 67-100 % lichen coverage.



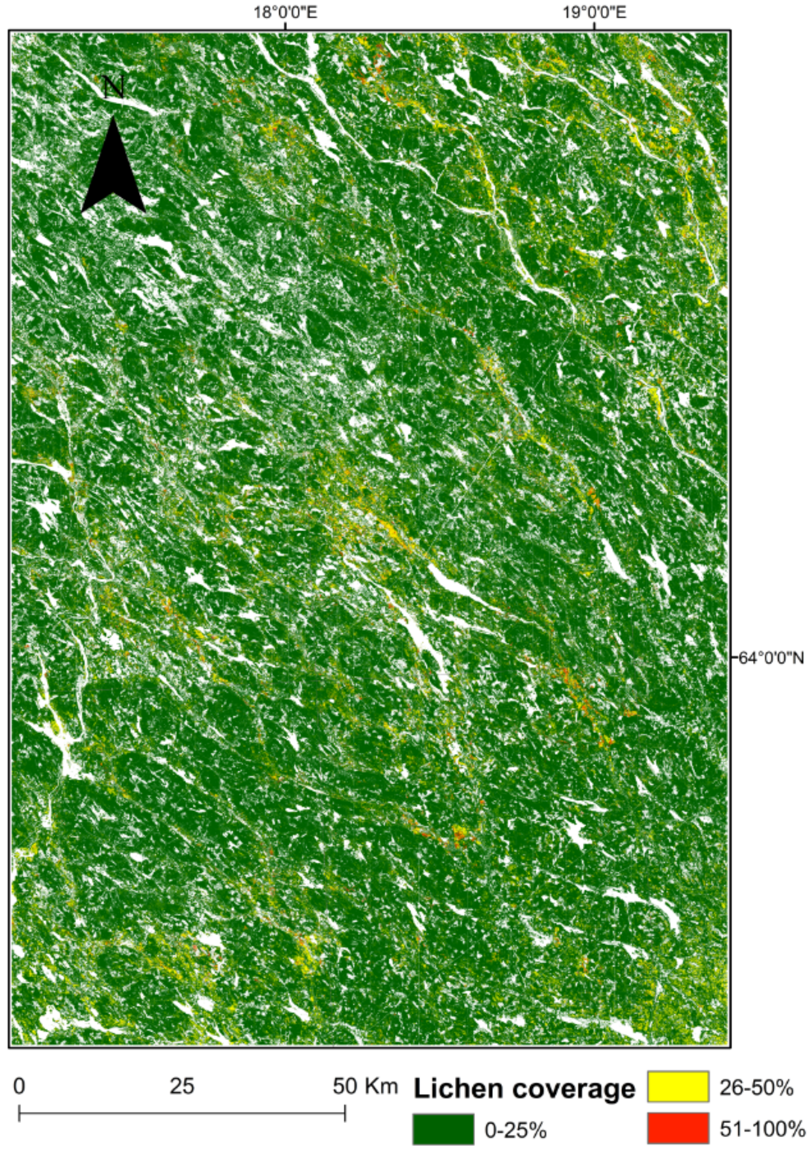
Appendix figure 12. Scatter plot of the two most important ALS-derived variables, vegetation ratio and forest height, for classification scheme 6. The classes are 60: 0-33 %, 61: 34-66 % and 62: 67-100 % lichen coverage.



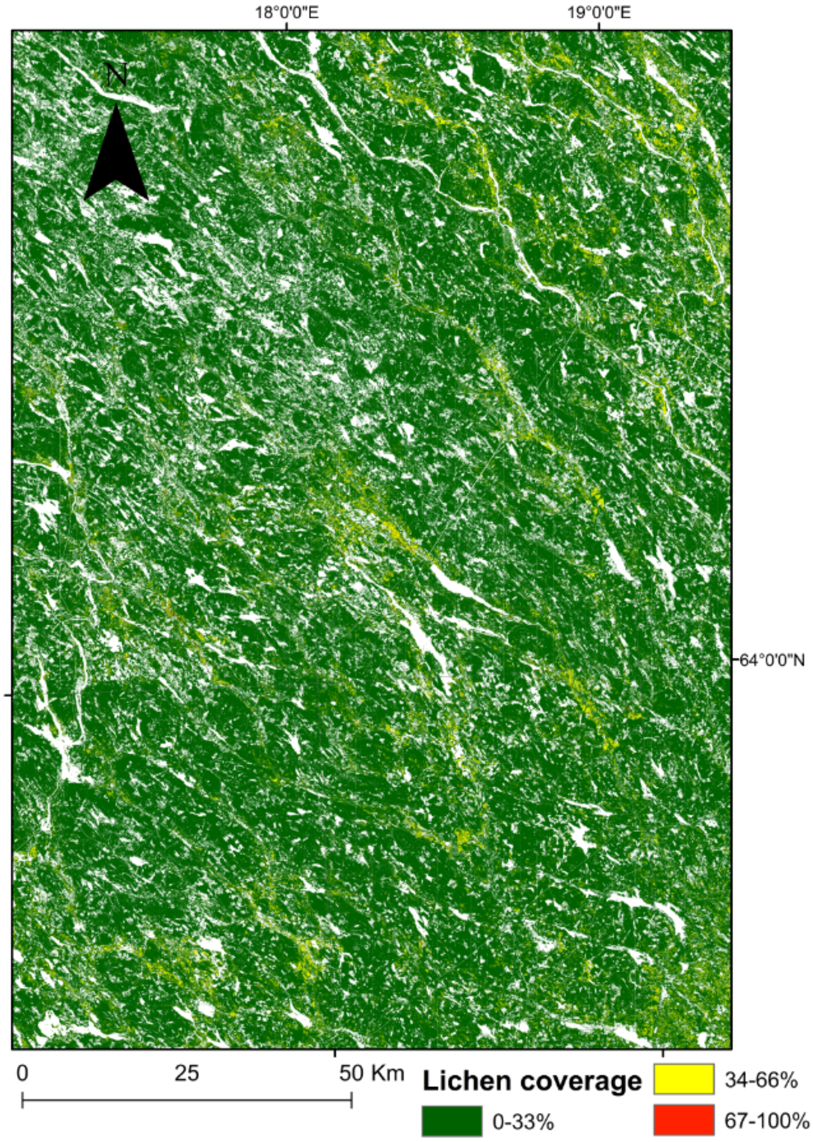
Appendix figure 13. Map of the study area. The Sentinel-2 image from 2015-08-19 is shown in true colour.



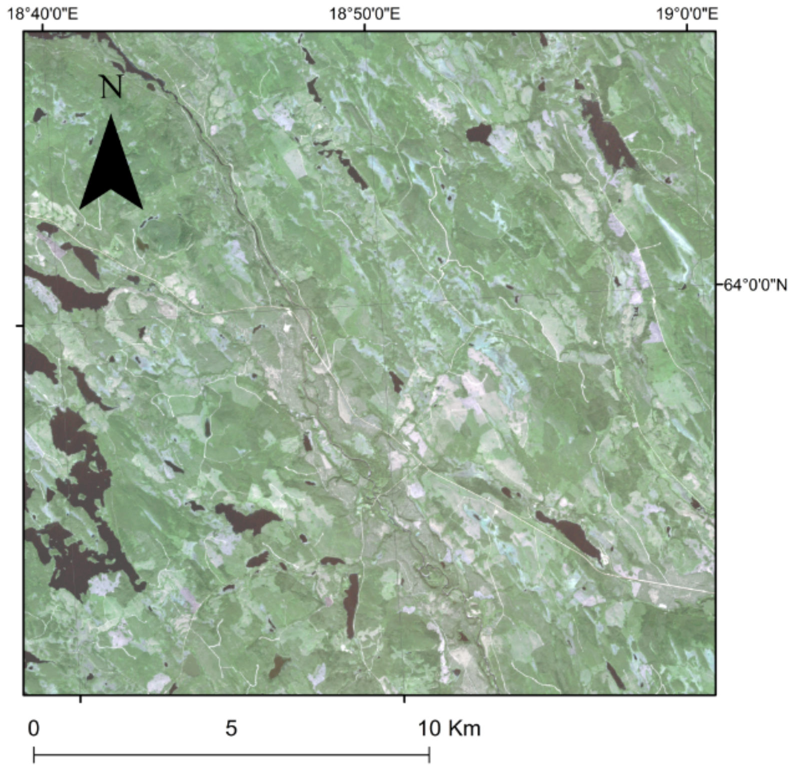
Appendix figure 14. Lichen coverage map of the study area, classified with the Random Forest algorithm according to classification scheme 2, using the Sentinel-2 image from 2015-08-19 and ALS-derived variables (DEM, forest height, canopy density and wetness index).



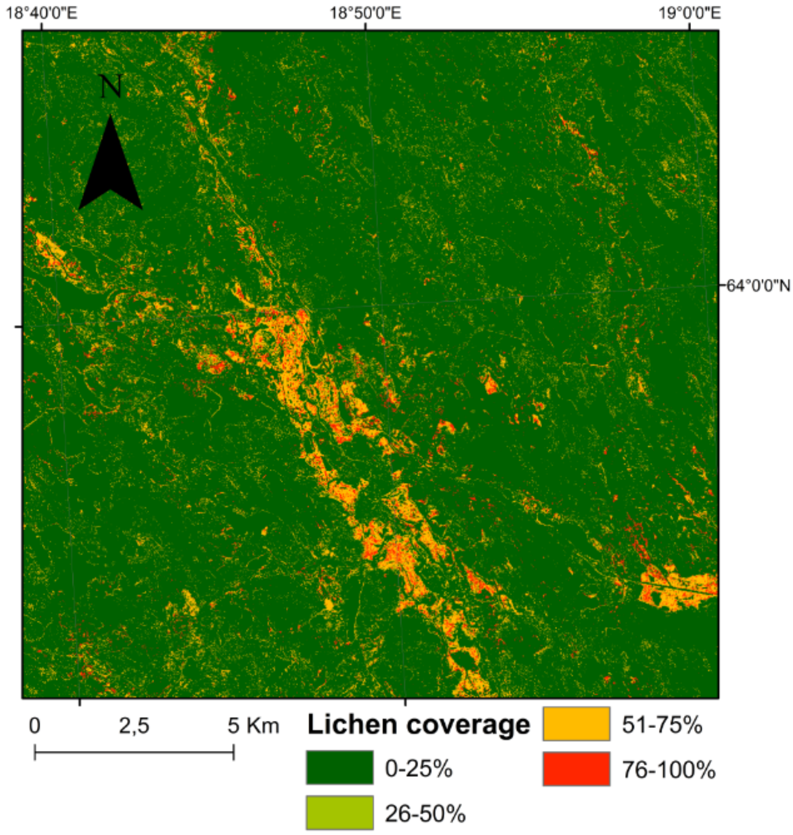
Appendix figure 15. Lichen coverage map of the study area, classified with the Random Forest algorithm according to classification scheme 4, using the Sentinel-2 image from 2015-08-19 and ALS-derived variables (DEM, forest height, canopy density and wetness index).



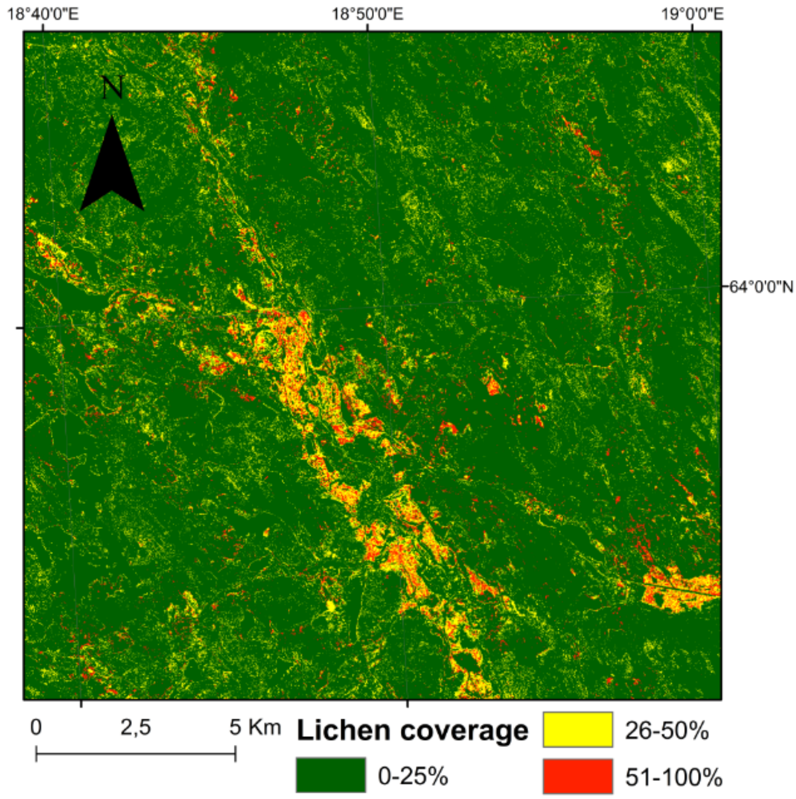
Appendix figure 16. Lichen coverage map of the study area, classified with the Random Forest algorithm according to classification scheme 6, using the Sentinel-2 image from 2015-08-19 and ALS-derived variables(DEM, forest height, canopy density and wetness index).



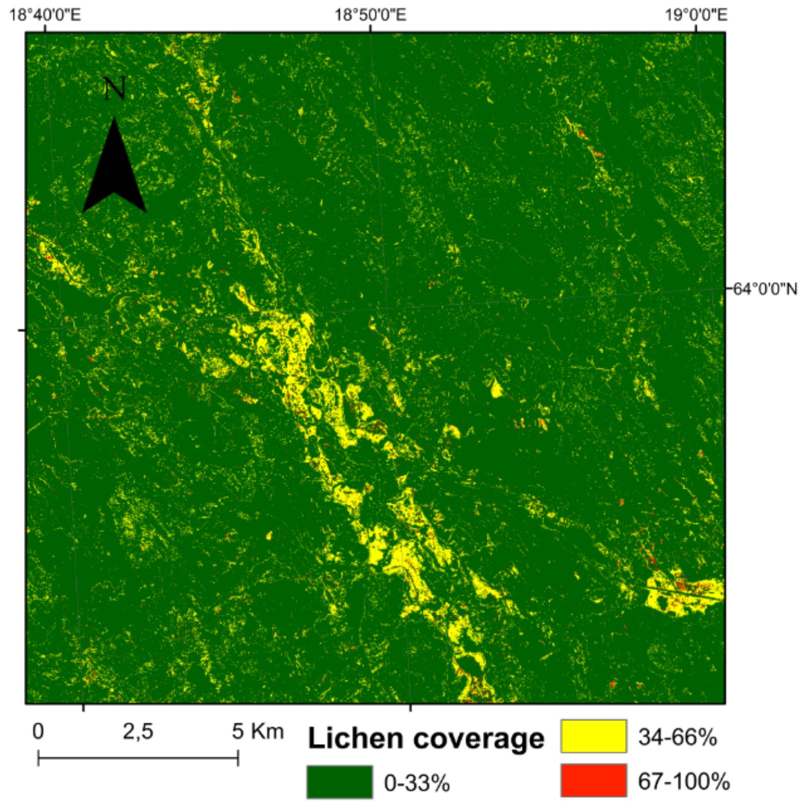
Appendix figure 17. Map of part of the study area. The Sentinel-2 image from 2015-08-19 is shown in true colour.



Appendix figure 18. Detail lichen coverage map of part of the study area, classified with the Random Forest algorithm according to classification scheme 2, using the Sentinel-2 image from 2015-08-19.



Appendix figure 19. Detail lichen coverage map of part of the study area, classified with the Random Forest algorithm according to classification scheme 4, using the Sentinel-2 image from 2015-08-19.



Appendix figure 20. Detail lichen coverage map of part of the study area, classified with the Random Forest algorithm according to classification scheme 6, using the Sentinel-2 image from 2015-08-19.