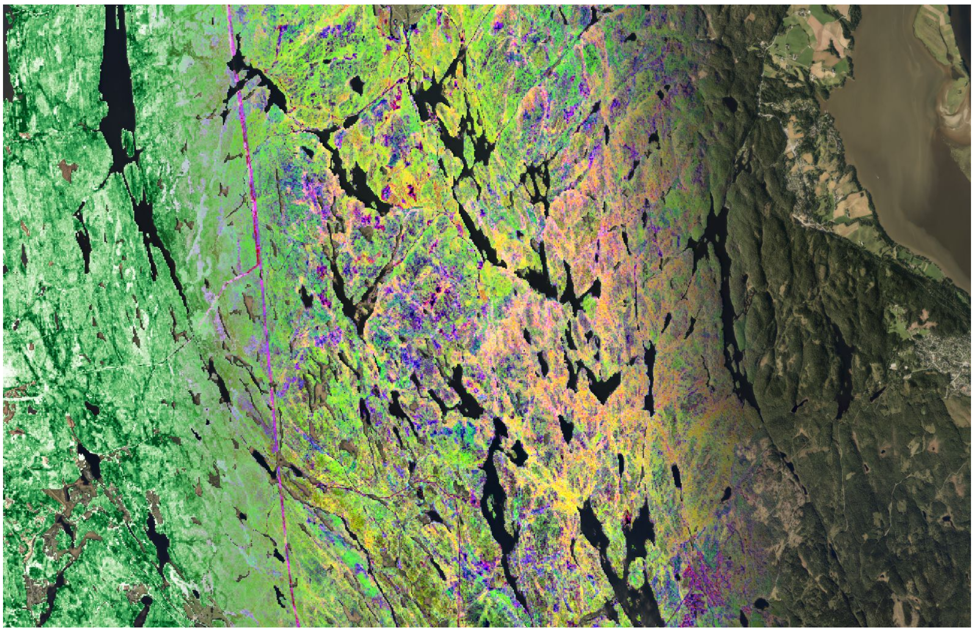
Final project report

Mapping natural forest by means of remote sensing

(Kartlegging av naturskog ved bruk av fjernmåling)

Hans Ole Ørka[1], Victor Strimbu[1], Jörg Haarpaintner[2], Anne Sverdrup-Thygeson[1] and Terje Gobakken[1]

[1]Norwegian University of Life Sciences, Faculty of Environmental Sciences and Natural Resource Management
[2]Norut - Northern Research Institute



Version: 9. April 2018

Preface

The work presented in this report is the result of the project “Kartlegging av naturskog ved hjelp av fjernmåling” (Mapping natural forest by means of remote sensing) with reference number 2017/4526. The project was funded by the Norwegian Environment Agency and conducted in autumn of 2017 by the Norwegian University of Life Sciences with contributions from Norut (Norut - Northern Research Institute) on processing of satellite radar data.

M-940|2018 Mapping natural forest by means of remote sensing Rapport NMBU EN 43 s.
Contact person at the client: Tomas Holmern

Front page: A collage of biomass, forest structure habitat index and orthophoto. Østmarka forest reserve appears as pink just right of the image center.

Table of content

Abbreviations

ALS	Airborne laser scanning
AOI	Area of interest
BOA	Bottom-Of-Atmosphere reflectance
CWD	Coarse woody debris
DAP	Digital aerial photogrammetry
FSHI	Forest structure habitat index
GINI	Gini coefficient: a measure of inequality
MEV	Mathiesen Eidsvoll Verk
NDH	National elevation model for Norway (See www.hoydedata.no)
NDI	Normalized Difference Index
NFI	National forest inventory
OA	Oslo and Akershus county
SAR	Synthetic Aperture Radar
TOA	Top-Of-Atmosphere reflectance

Sammendrag

Skogbehandling endrer strukturen på skogen og påvirker det biologiske mangfoldet. Skog som er forynget med planting etter flatehogst har mindre variasjon i alder og størrelse på trærne enn skog der det er benyttet selektive hogster og naturlig foryngning. Mengden død ved og artsmangfoldet vil også være høyere i slik skog og refereres gjerne til som naturskog. Naturskog forynget etter selektive hogster har en skogstruktur som ligner på det som finnes i urørte skoger. I slik gammelskog har trær i ulike aldre og størrelser og det er mye død ved. Gradienten fra skog som forvaltes intensivt til helt urørt skog er viktig i forvaltning av biologisk mangfold, spesielt skille mellom gammel kulturskog og naturskog.

Målet med denne studien har vært å kvantifisere mengden naturskog ved hjelp av fjernmåling og referansedata. Ulike definisjoner som er mer eller mindre strenge er brukt for å definere naturskog. Definisjonene danner grunnlaget for å produsert kart som viser naturskog over store områder, samt å estimere areal andelen naturskog for området. Ulike fjernmålingsdata ble evaluert med hensyn på mulighetene for å skille ut gammel naturskog fra andre skogtyper. Fjernmålingsdata som ble evaluert var tre-dimensjonale data etablert gjennom bildematching av flyfoto fra omløpsfotografieringen, samt data fra flybåren laserscanning. I tillegg til de flybårne dataene ble satellittdata vurdert. Studien evaluerte optiske data fra Landsat 8 og Sentinel-2, samt radardata fra Sentinel-1 og ALOS PALSAR. Studien ble gjennomført i Oslo og Akershus (OA) fylke i Sørøst-Norge. I fylket ble et mindre område (MEV) på 17 000 ha brukt til å undersøke konkrete forskningsspørsmål som det ikke var mulig å analysere for hele fylket. Referansedata for denne studien ble hentet fra landsskogtakseringen. For det mindre område (MEV) ble informasjonen om gammel naturskog hentet fra historiske bestandsregister. For begge områder ble analysen begrenset til skogarealet. I alt ble det brukt fjernmålingsdata fra 16 laserscannings-prosjekter, 8500 flybilder bilder, SAR-data fra Sentinel-1 og ALOS PALSAR-2, Landsat 8 og Sentinel-2 bilder. Effekter av punkttetthet i laserscanningene, prøveflatestørrelse og tidspunkt for opptak ble evaluert.

Klassifikasjonsnøyaktighetene som ble oppnådd ved bruk av data fra radarsatellittenene; Sentinel-1 og ALOS PALSAR-2 hadde begrenset verdi for å skille ut gammel gammel naturskog. Også nøyaktigheten ved bruk av optiske satellittdata fra Landsat 8 og Sentinel-2 ga lav nøyaktighet. Bruken av Sentinel ble imidlertid fullstendig belyst i denne studien, og ulike tilnærminger for å lage mosaikker bør vurderes ytterligere, sammen med bruk av multi-temporale datasett og teksturanalyse. Resultatene indikerte at laserscanning ga høyere nøyaktighet enn både optiske- og radarsatellitter. Videre var effekten av punkttetthet større ved bruk av små prøveflater. Bruk av 3D-data fra flyfoto resulterte i en kappa-koeffisient på 0,58, mens for samme datasett oppnådde ALS en kappa-koeffisient på 0,65. Det skal også bemerkes at nøyaktigheten ved bruk av flyfoto er litt lavere enn den beste Landsat 8 klassifiseringen, men samlet gir det mye bedre resultater.

For hele Oslo og Akershus fylke ble to indekser som beskriver skogstruktur beregnet fra laserdata og vurdert som en måte å identifisere naturskog på uten bruk av referansedata.

Resultatet ble vurdert visuelt og viste klare sammenhenger mellom de produserte kartene og eksisterende naturreservater.

For hele Oslo og Akershus ble naturskog klassifisert med logistisk regresjon, en meget robust klassifikator som er godt egnet for både å gjøre klassifisering og for å beregne arealandeler og usikkerhet basert på modell-assistert og modell-basert estimering. De optiske satellittbildene fra Landsat 8 og Sentinel-2 var bedre egnet til å klassifisere naturskog enn laserscanning for tre av de fem definisjonene som ble testet. For to av definisjonene var laserdata mye bedre egnet enn optiske data. Generelt var areal-estimater basert på fjernmåling, enten optiske satellittbilder eller laserdata mer presise enn tilsvarende estimater basert på kun referansedata (20 av 30 tilfeller). Det var også forskjeller mellom fjernmålingsdatene. Landsat 8 ser ut til å være beste, med 9/10 tilfeller hvor presisjonen ble forbedret, etterfulgt av Sentinel-2 med 8/10 tilfeller og flybåren laserskanning med 3/10 tilfeller.

I tillegg til å klassifisere naturskog ble det utviklet en biomassemodell basert på de 16 laser prosjektene i Oslo og Akershus og referansedata fra landsskogtakseringen. To variabler ble valgt for modellbygging og modellen forklarte 68% av variasjonen i biomasse og hadde en feil på 37% (RMSE). Et kart som viser biomasse ble laget for 98,8% av skogen i Oslo og Akershus fylke. Resultatene fra detter er viser en meget lovende tilnærming for å etablere biomassekart over store områder.

Samlet sett gir denne studien lovende resultater for videre kartlegging og overvåkning av naturskog. Landsskogtakseringens data er godt egnet for å etablere ulike definisjoner av naturskog, selv om det er noen mangler. Det er imidlertid en stor ulempe at det er begrenset tilgang til koordinatene til landskogtakserings prøveflater Den beste måten for å opprette referansedata for videre analyse eller validering av identifisert naturskog er å sjekke historisk behandling av bestand. Dette kan tolkes manuelt i eldre flyfoto eller satellittdata. I denne studien har kombinasjon av ulike datakilder ikke vært evaluert. Vi ser at optiske data gir bedre resultater når definisjonene er basert på alder, mens når naturskogen ble definert basert på struktur ble resultatene bedre med laserdata. Ved å bruke komplementære kilder som laserskanning og optiske satellitter eller radar- og optiske satellitter, vil det være mulig å forbedre resultatene. Generelt var resultatene basert på Landsat 8 mest lovende for klassifisering av naturskogen selv om flybåren laser ga modeller med lignende nøyaktighet eller enda bedre nøyaktighet.

Abstract

Forest management changes the forest structure and affects the biodiversity. Managed forests regenerated after clear-cuts show a small variation in age and sizes of the trees. Furthermore, the amount and diversity of dead wood are small. Forests regenerated after selective logging have a structure that resembles old-growth forests. Old-growth boreal forest displays a heterogeneous age and size structure together with a larger amount of dead wood. Such, naturally regenerated forest are often referred to as natural forests. The gradient of naturalness from old managed to old-growth is important in forest biodiversity management and the definition of natural forest are on that gradient.

The aim of this study has been to quantify the amount of natural forest based on different definitions of naturalness and to map natural forest over large areas. Quantifying and mapping natural forests are challenging tasks. Recent large-scale aerial photograph and airborne laser scanning data acquisitions and recent satellite launches in the Copernicus programme provide large amounts of spatial data free of charge. The performance of these remotely sensed datasets to classify and map the amount of natural forest was evaluated. The study was conducted in Oslo and Akershus (OA) county in south-eastern Norway. Within the county, a smaller area, Mathiesen Eidsvold Verk (MEV), of 17 000 ha was used to investigate specific research questions that were not possible to analyze for the entire county. Field reference for this study was provided from the Norwegian National Forest Inventory (NFI). For the MEV dataset, information about old managed forests and old natural forests were extracted from historical stand records of harvesting regimes. For both areas, the analysis was restricted to the forest area. In total data from 16 airborne laser scanning (ALS) projects, 8500 aerial images, Synoptic Aperture Radar (SAR) data from Sentinel-1 and ALOS PALSAR-2, and optical satellite imagery from Landsat 8 images and Sentinel-2 were used. Effects of ALS point density and sample plot size were also analyzed.

The classification accuracies obtained with Sentinel-1 and ALOS PALSAR-2 were of limited value for the separation to the old natural forest from the old manage forest and the classification accuracy obtained with Landsat 8 was also low. Furthermore, the accuracy obtained using Sentinel-2 was not sufficient for detecting natural forests. However, the use of Sentinel was not analyzed to its full potential in this study and different approaches for creating mosaics should also be considered together with multi-temporal datasets and texture analysis before concluding. The results indicated that ALS data provided higher accuracy of natural forest identification than optical and radar satellite imageries and that the effect of point density was larger on smaller plots. Using 3D data from aerial photographs resulted in a kappa-coefficient of 0.58, while for the same dataset the ALS obtained a kappa-coefficient of 0.65. It should also be noted that the performance when using aerial photographs is slightly lower than the best Landsat 8 classification, but overall it provides much better results.

For the entire OA two indices that are related to forest structure, were tested using for unsupervised detection of natural forests. A visual assessment showed a high degree of match between the maps derived from the two indices and existing forest reserves.

The natural forest was classified using logistic regression, a robust classifier, suitable for model-assisted and model-based area estimation. The multispectral (Landsat 8 and Sentinel-2) models were better than the models based on laser scanning in predicting natural forest according to three of the five definitions tested for natural forests. For the other two natural forest definitions, the models based on laser scanning data outperformed the multispectral models. In general, the estimates based on remotely sensed data were more precise than the corresponding field-based estimates (20/30 cases). Among the remotely sensed data types, the Landsat 8 seems to be the best, with the 9/10 cases where the precision was improved, followed by the Sentinel-2 with 8/10 cases and ALS with 3/10 cases. It is clear that optical data better capture definitions based on age while structural based definitions are better separated using laser scanning.

For the forest biomass models, two variables were selected in the variable selection. The model explained approximately 68% of the variation and had an RMSE of 37%. A biomass map covering 98.8% county's forests was developed.

Overall, the current research provides results that are promising for further mapping and monitoring of natural forest. The NFI data are suitable for using several definitions of natural forest. However, limited access to the coordinates of the NFI plots is a huge drawback. The best way to establish reference data for further analysis or validation of the current research is to check historical management practices in available recent and historical aerial and satellite imageries. Fusion of different data sources has not been focused in this study. Using complementary sources such as laser scanning and optical satellites or radar and optical satellites might provide an additional dimension that improves the obtained results. Remotely sensed data improve the precision of natural forest estimates. In general, the results based on Landsat 8 were most promising for discerning the natural forest. Although ALS provide models with similar accuracy or even better accuracy.

Introduction

Forests have large variation in structure and habitats. Forest management changes the forest structures and affects the biodiversity in these habitats. Former choice of harvesting methods have a large impact on the current stand structure. Managed forests regenerated after clear-cuts show a small variation in age and sizes of the trees. The amount and diversity of dead wood are also small. Forests regenerated after selective logging have a structure that resembles old-growth forests. Old-growth boreal forest displays a heterogeneous age and size structure together with a larger amount of dead wood. This gradient of naturalness from managed to old-growth is important in forest biodiversity management.

The amount of natural forest in Norway varies from 1.3%[1] to 25% depending on the definition used (Rolstad, Framstad, Gundersen, & Storaunet, 2002; Rolstad & Storaunet, 2015). Thus, there is both a need for better quantifying the amount based on different degrees of naturalness but also for mapping the degree of forest naturalness over large areas. Quantifying and mapping natural forests are challenging tasks. Current operational forest management inventories do not collect information on naturalness. The national forest inventory collect detailed information on forests on 250 m^2 field plots in a 3 by 3 km grid.

Recently, airborne laser scanning was used to separate areas regenerated after clear-cutting and areas regenerated after selective harvest (Sverdrup-Thygeson, Ørka, Gobakken, & Næsset, 2016). The accuracy of the classification was high and the ongoing nationwide ALS acquisition in Norway further motivates testing such methods for larger areas. However, when going beyond the municipality level different ALS acquisitions (e.g. different scanners or flight parameters) make such analyses difficult due to inconsistencies between the different datasets (Erik Næsset, 2009; Ørka, Næsset, & Bollandsås, 2010). Thus, testing other remote sensing sources with more uniform data, like optical aerial imagery, optical satellite imagery and radar satellite imagery, over large areas are of interest. Nevertheless, ALS data are in a unique position for mapping forest structure and effective methods for using ALS data from multiple projects in assessing forest naturalness would be of great importance.

The recent satellite launches in the Copernicus programme provide large amount of spatial data free of charge that might provide means for mapping and monitoring forest structure and naturalness. Especially, Sentinel-1, a Synthetic Aperture Radar (SAR) satellite and Sentinel-2, an optical satellite, provide data that might be beneficial for mapping forests. The Sentinel-1 mission consists of two satellites carrying a C-band imaging system operating in different imaging modes with varying resolutions and coverages. The advantage of the SAR satellites is that they are able to “see through” clouds because the wavelengths used penetrate clouds. SAR satellites are active microwave instruments that emit radio waves and receive the reflected backscatter signal. Being an active sensor, SAR systems are capable acquire data in the absence of the day light. The flexible image acquisitions that are not limited to sunlight and clear sky are important advantages for mapping and monitoring forest

[1] http://www.skogoglandskap.no/filearchive/skog_uten_inngrep.pdf

structure and naturalness. Sentinel-2 consists of two optical satellites with 13 spectral bands covering visible and infrared parts of the spectrum. The spatial resolutions are 60 m, 20 m and 10 m depending on the band. The Sentinel-1A was first launched in 2014 and the first Sentinel-2A satellite was launched in 2015. In April 2017, all four satellites were launched. In addition, the Landsat 8 satellite provides data with a spatial resolution of 30 m that might be suitable for characterization of forest naturalness.

The temporal resolution or the revisit period for the different optical satellite systems varies from daily acquisitions to many weeks between the acquisitions. How often we can get good optical data for the same area is important if for example leaf of and leaf on acquisitions are beneficial. Furthermore, we know that seasonal variation, sun angles and mosaicking are important for the image quality and if we have a large number of good images available we can also restrict variations between sun angles etc. for the images used.

Evaluating possibilities of applying recent development in remote sensing for characterizing forest structure and naturalness is of great interest from a forest monitoring perspective. Thus, the main objective of the research was to develop methods to identify natural forests, with emphasis on old natural forest. The specific objectives were:

1. Develop a framework for separation between natural forest, regenerated after selective logging and managed forests regenerated after clear-cuts.
2. Evaluating different classifications of natural forest.
3. Evaluate how seasonal variation, sun angles and mosaicking influence the results.
4. Develop probability maps of old natural forest and estimate the area of such forest within a county.

Material and methods

Study areas and reference data

The study was conducted in Oslo and Akershus (OA) county in south-eastern Norway. Approximately 70% of the land area in the county is forest. The forest is dominated by spruce (72%), with some pine (20%) and broadleaved species (8%). Within the county a smaller area was used to investigate specific research questions that were not possible to analyze for the entire county. This area consists of the privately owned forest; Mathiesen Eidsvoll Verk (MEV). The productive forest area was 17 000 ha. See Sverdrup-Thygeson et al. (2016) for details. The two areas were used independently in the study because the field reference data differed highly between the two areas. The location of the study areas appear in Figure 1.

For the main area (OA) the field reference for this study was provided from the Norwegian National Forest Inventory (NFI). The dataset included information about stem volume, biomass, stem diameters and recordings about the characteristics of the forests. Based on the recordings we established 5 binary variables representing different definitions (Table 1). The first definition is the one used by the NFI to classify “natural forest”, “conventional forest” and “plantations”. It requires no visible impact, native species covering minimum 0.5 ha, in addition minimum two of the following criteria must be met: 1) old age, 2) multilayered forest, and 3) dead wood. This definition is referred to as D1. The following two definitions (D2 and D3) are based on the stand age recorded in the NFI data. The D2 definition includes forest older than 160 years often used as a threshold for old forests, and the D3 definition has a lower

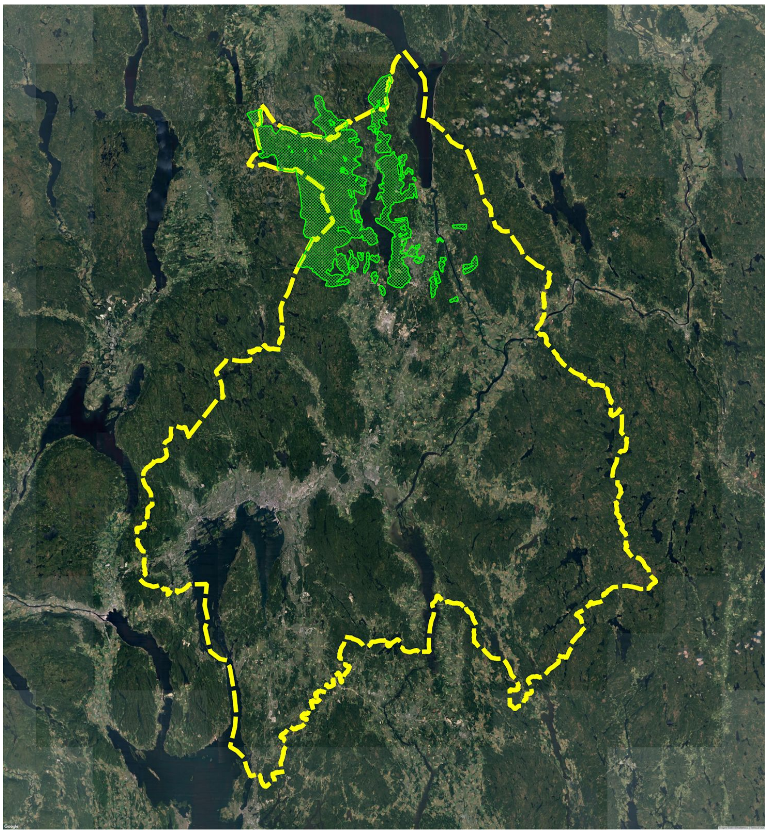


Figure 1. Map of the two study areas, Oslo and Akershus county (OA) in yellow and the private forest Mathiesen Eidsvoll Verk (MEV) in light green. Background map: Google satellite.

age threshold and includes forest older than 140 years. The last two definitions (D4 and D5) are based on the GINI coefficient. The GINI-coefficient is an index characterizing variations in the size distribution of trees based on diameter at breast height. It provides values between 0 and 1 where forest stands with a low GINI-coefficient displays low variation in tree sizes while forest stands with a large GINI-coefficient display large variation in tree sizes. Thus, the index describes one important factor for characterizing naturalness. When compared to other indices the Gini coefficient was found superior with respect to characterizing tree size diversity in boreal forests (Lexerød & Eid, 2006). As we aim to detect natural forest regenerated after selective cutting such an index provides means, based on structure, to separate stands regenerated based on selective logging and clear-cuts. However, it does not take into account the amount of dead wood or sizes of the trees. For definition D4 and D5 we applied only development class V as natural forest using two thresholds of the GINI-coefficient namely the 75th and 50th percentile of the GINI distribution. It is indicated that 65% of the development class V in the region was regenerated after selective logging (Rolstad & Storaunet, 2015). Thus, the 50th and 75th percentiles were found appropriate as thresholds for the GINI-coefficient. The relationships between the recordings of stand age, GINI and NFI classification of natural forest appear in Figure 2.

For the MEV dataset, information about old managed forests and old natural forests were extracted from historical stand records of harvesting regimes. See Sverdrup-Thygeson et al. (2016) for details about this dataset. Thus, the MEV dataset provided a strict definition based on observed management while the OA dataset was based on field observations of current forest conditions.

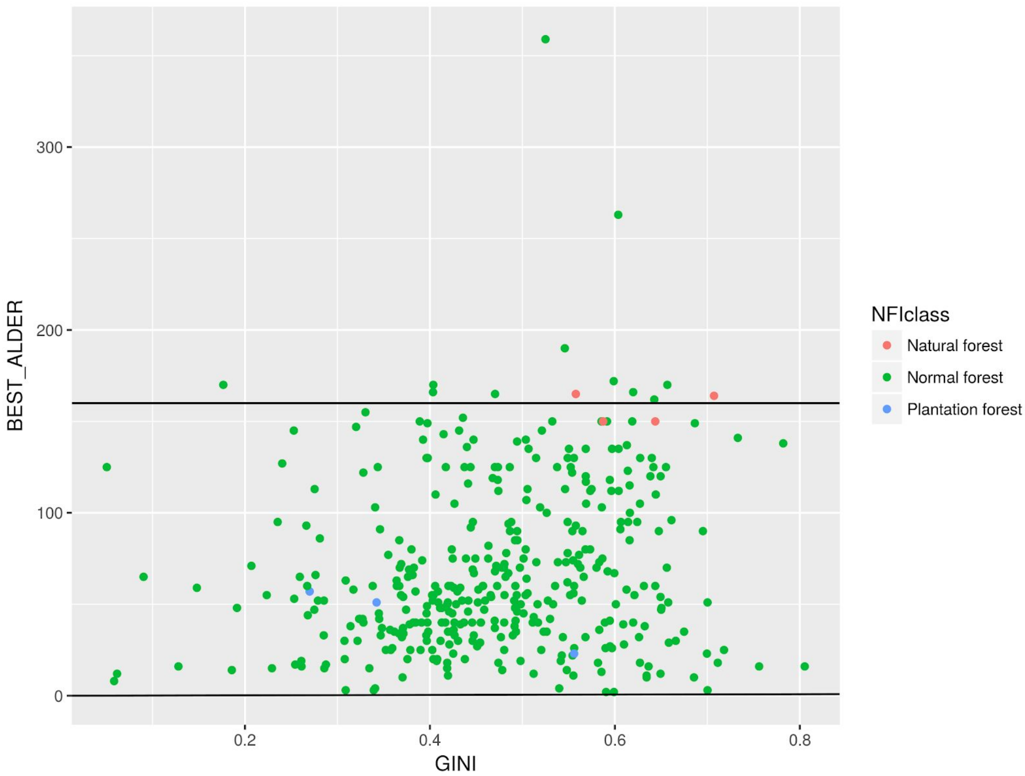


Figure 2. Relationship between NFI forest characteristics classification, stand age and Gini coefficient. Horizontal line represents a stand age of 160 often used to separate old natural and old managed forests.

For both areas, the analysis was restricted to the forest area. The forest area can be derived by remote sensing i.e. ALS data (Eysn, Hollaus, Schadauer, & Pfeifer, 2012), SAR data from e.g. Sentinel-1 (Dostálová, Hollaus, Milenković, & Wolfgang, 2016) or optical satellite data e.g. Sentinel-2 (Immitzer, Vuolo, & Atzberger, 2016). However, in the current study we used official AR5 land cover maps to delineate the forest area. The forest area within the forest mask in OA was 345456 ha.

Table 1. Overview of definitions of natural forest used in the current study.

Defini-tion	Short name	Thresholds	Area/ Dataset	Number of observations
D1	NFI	NFI definition of “natural forest” used for forest characteristics[2]. The non-natural forest classes is “conventional forest” and “plantations”	OA/NFI	4
D2	AGE160	Stand age >= 160 years old	OA/NFI	12
D3	AGE140	Stand age >= 140 years old	OA/NFI	32
D4	STRUCT25	The 25% most size diverse forests in development class V.	OA/NFI	26
D5	STRUCT50	The 50% most size diverse forests in development class V.	OA/NFI	52
D6	HIST	Based on historical recordings on management practices.	MEV	89

Remotely sensed data

Airborne laser scanning

In total 16 ALS projects were dowenloaded from www.hoydedata.no[3]. The data were further normalized to obtain height above ground. An overview of projects and point density are presented below (Table 2). Standard ALS metrics used in forest inventories in Norway today were computed. The metrics are described in Næsset (2004) and computed for all first and single return echoes above 1.3 m. The metrics are different statistical properties of the height distribution such as the mean or standard deviation of the laser echoes height. These measures were used for all analysis using ALS data. In total after removing the oldest acquisitions for overlapping areas the ALS data covered 99.8% of the entire forest area in the county.

Digital aerial photogrammetry

In total 8500 spectral images were acquired with a Vexcel UltraCam sensor on 9 June 2016, respectively, by Terratec AS, Norway in the project named Oslo-Østlandet. The acquisition covers nearly the entire OA. However, in this report only the use of the 330 images covering the MEV area is reported. The side and forward overlaps between images were 20% and 60%, respectively. In the current study the color infra-red (RGBI) 8 bit digital imagery was used. The images were acquired with a ground sample distance of 25 cm.

[2]http://www.skogoglandskap.no/filearchive/landsskogtakseringens_feltinstruks_2008_h0508.pdf (page 55).

[3] The dataset from Hurdal 2007 from hoydedata.no was not used because the version missed echo classification. NMBU have access to an updated version and that was used.

The sensor location and orientation during image acquisition were recorded using a GNSS and an inertial navigation system. A photogrammetric point cloud and canopy height model were constructed from the aerial images using SURE Photogrammetric software (Rothermel, Wenzel, Fritsch, & Haala, 2012) which adopts a matching algorithm similar to Semi-Global Matching (SGM) proposed by Hirschmuller (2008) and Rothermel et al. (2012). The software was chosen over alternative ones because of the ability to efficiently process large datasets and the ability to ensure larger height variations compared to other photogrammetric software (Haala, 2014). It is likely that larger height variations can provide more valuable information on the forest canopy.

The input data for the generation of the dense point cloud were: 1) non-orthorectified 8 bit RGBI UltraCam imagery; and 2) the aerial triangulation provided by the data vendor. The processing was performed with default settings and resulted in the production of a point cloud with a point density of approximately 33.1 points m^{-2}.

Heights above the ground surface were calculated for all digital aerial photogrammetry (DAP) point cloud by subtracting the ALS TIN heights from the height values of all points recorded. The same metrics as for ALS was also calculated for DAP data.

Table 2. Overview of ALS projects used.

Location	Year	Point density[4]	Forest area (ha)
Asker	2012	21.34	5530
Aurskog-Høland	2005	2.48	71546
Bærum	2012	23.03	11449
Bærum	2013	9.31	748
Eidsvoll	2007	1.06	29693
Enebakk	2012	3.77	7210
Follo	2014	7.32	43279
Hurdal	2007	13.18	40853
Fet (NDH)	2016	16.77	9858
Oslo Nordmarka (NDH)	2016	5.04	20994
Nittedal	2012	21.04	4232
Oslo Byggesonen	2014	13.97	4560
Oslo Kommuneskogen	2010	19.20	12596
Romerike	2013	3.11	23473
Romeriksåsene	2013	1.34	57607
Ullensaker og Nes	2010	1.17	61497

[4] Average point density in grid cells inside forest mask.

Radar data (Sentinel-1 and ALOS PALSAR-2)

SAR data from Sentinel-1 and ALOS PALSAR-2 were processed to create mosaics based on the average backscatter from multi-temporal acquisition (See figure 3). The backscatter is in dB and the averages are based on both the ascending and descending paths for the time interval. The backscatter values were slope corrected. After mosaicking there were some seasonal patterns from ascending and descending orbits but the results was judged to be close to optimal results.

The spatial resolution of all mosaics were 20 m and they all constituted of 3 bands, namely the co-polarized band, cross-polarized band and the Normalized Difference Index (NDI). Thus, the 3 bands for ALOS-2 were rendered in false color with RGB = [HH(dB), HV(dB), NDI = (HH-HV)/(HH+HV)] and for Sentinel-1 with RGB = [VV(dB), VH(dB), NDI= (VV-VH)/(VV+VH)].

From ALOS-2 two yearly mosaics were created from 2015 and 2016 with freely available data from JAXA[5]. SAR lay-over, SAR shadow, and very low and high incidence angles were masked out. From Sentinel-1 two yearly mosaics from 2015 and 2016 were created in addition to a bi-yearly mosaic from both years. Similarly, seasonal mosaics based on data from the winter (December to March) and summer (June to September) were created for both years and together. In all images SAR shadow, SAR overlay, and very low and high projection angles (complementary to incidence angles) were filtered out. For MEV monthly mosaics were created.

The seasonal mosaics were created to better provide data for the different analysis of the project. Summer mosaics are anticipated to be more correlated with biomass than yearly and winter mosaics. While the winter mosaics likely will provide a better differentiation between conifers and broadleaved trees.

[5] http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm

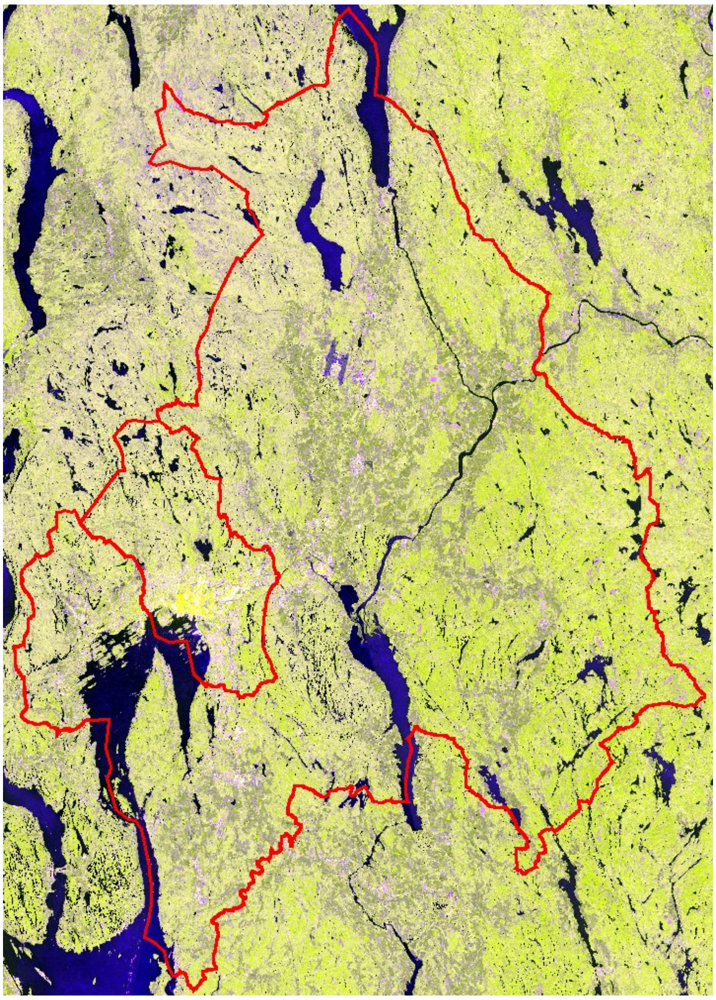


Figure 3.Averaged Sentinel mosaic of all data of the summer months (June-September) for 2015 and 2016. The RGB colors are the VV (vertical polarization emitted , vertical received) coplarized backscatter in dB in the red band, VH (vertical polarization emitted, horizontal received) coplarized backscatter in dB in the green channel and the normalized difference index NDI (NDI = (VV-VH)/(VV+VH)) of the two polarizations. Green areas show strong VH backscatter which means high volume scattering, meaning strong vegetated areas like forest, black and dark blue areas are generally water, bright areas (~white) are generally strong backscatter f.e. buildings, grey areas are low vegetated areas, pinc/purple areas are very low/no vegetated areas/bare land/rocks.

Landsat 8

Landsat 8 images covering the study area were identified using the LIBRA Landsat 8 browser[6]. The selection criteria for scenes to be included in the analysis were that they were acquired in 2016, that the image was Tier 1 (i.e. data that meet geometric and radiometric quality requirements), that the cloud cover in the area of interest (AOI) was < 50% and that the scene cover > 50% of the AOI. The images were downloaded and processed to top-of-atmosphere (TOA) reflectance. The first step of the processing chain was to process the “raw” scaled radiance using coefficients stored in the metadata (Chander, Markham, & Helder, 2009). The second step was to conduct a linear transformation that accounts for

[6] https://libra.developmentseed.org/

solar elevation and seasonally variable earth-sun distance (Chander et al., 2009). The processing to TOA reflectance was conducted using the Google Earth Engine[7] (GEE) that offers possibilities for cloud storage and processing of large amounts of satellite imagery.

From the Landsat images the tasseled cap transformation (Baig, Zhang, Shuai, & Tong, 2014), i.e. brightness, greenness and wetness and normalized difference vegetation index (NDVI) were computed. In addition, as per Gizachew et al. (2016), the following were computed: Enhanced vegetation index (EVI), Soil adjusted vegetation index (SAVI), Modified soil adjusted vegetation index (MSAVI), Normalized difference moisture index (NDMI), Normalized burn ratio (NBR) and Normalized burn ratio-2 (NBR2).

To detect clouds and cloud shadows we used the fmask[8] algorithm (Zhu, Wang, & Woodcock, 2015; Zhu & Woodcock, 2012). The algorithm can be run in different programing languages like Matlab, python and in the GEE. The raster produced by fmask include the following classes; clear land pixel (value=0), clear water pixel (value =1),

Two images that were acquired in the beginning of the growing season were used to create a mosaic covering more than 99% of the forested area in OA (see table 3 and figure 4). The images were resampled to a pixel size of 250 m^2, matching the area of the NFI plots.

Table 3. Landsat 8 mosaics

Scene ID	Date	Mosaic proportion (%)
LC81970182016147LGN00	May 26, 2016	93.55
LC81970182016131LGN00	May 10, 2016	6.45

[7] https://earthengine.google.com/
[8] https://github.com/prs021/fmask

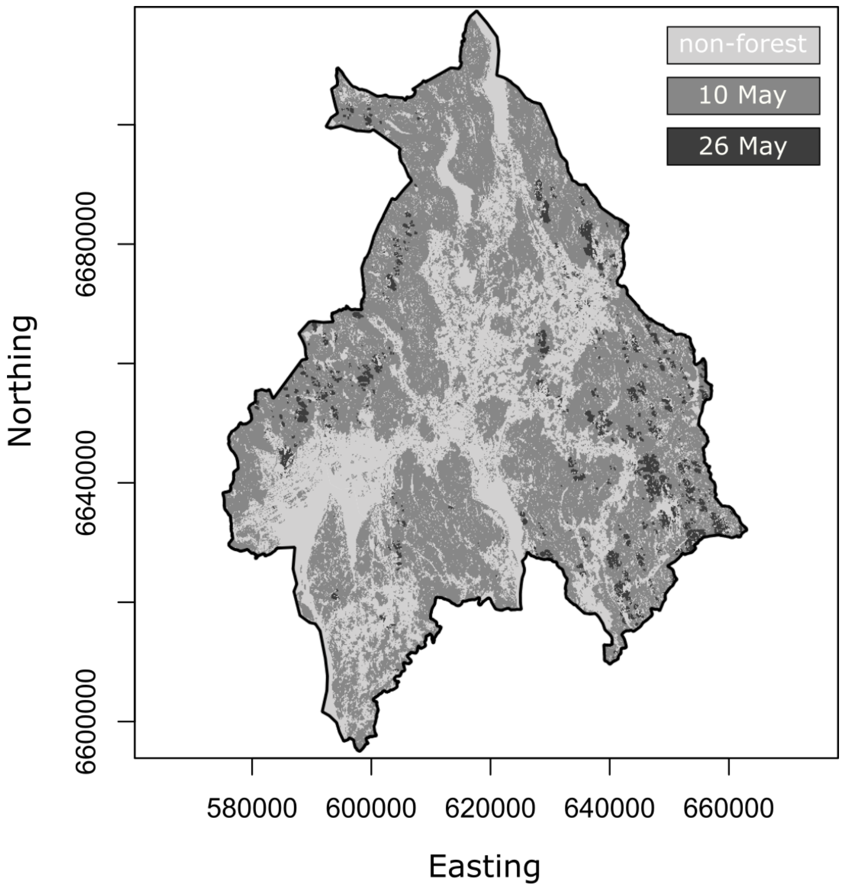


Figure 4. Landsat 8 mosaic coverage. The darker shades of grey show the extent of the two Landsat 8 images acquired in May 2016. The area in light gray represents non-forest and was excluded from this study

Sentinel-2

Sentinel-2 images were downloaded using a bash script provided by the European Space Agency and the Copernicus Open Access Hub[9]. The script was run on a Raspberry pi with a external hard drive connected to run the downloads. The criteria for selecting images were the same as for Landsat 8. The downloaded images were Level 1C[10], top-of-atmospheric reflectance.

Further processing included classification of clouds and cloud shadow using the fmask algorithm (Zhu et al., 2015) and converting to Level 2A using the SNAP[11] software with the sen2cor[12] plugin. However, the convention was just run on a subset of the data because issues related to installing the software (documented on large amount of user forums) on different servers and the time for processing the images.

The mosaic was created using ArcGIS with the processing products. The images were masked and resampled to fit the rest of the project and all the bands were merged into one

[9] https://scihub.copernicus.eu/userguide/5APIsAndBatchScripting
[10] https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/processing-levels/level-1
[11] http://step.esa.int/main/toolboxes/snap/
[12] http://step.esa.int/main/third-party-plugins-2/sen2cor/

single image. Although two images were initially selected for generating the mosaic, only the one with the smallest cloud coverage was used. This was related to the performance of the fmask algorithm on that image.

The images used for the mosaic were:
S2A_USER_PRD_MSIL2A_PDMC_20160817T201311_R008_V20160816T104022_20160816T104025.SAFE from August 17, 2016

To overcome the problems with high cloud coverage additional cloud free mosaics were created for Sentinel-2 data using the GEE. Seasonal mosaics for the growing season (1. of May to 1. of September) and for the whole year (1. of January to 31. of December) were created. Cloud free pixels were selected by a minimum reflectance. The pixel value for each band was scaled by the sum of the pixel values for all bands.

Potential of remotely sensed data to identify natural forest

Using the MEV data, we investigated different remote sensing sources including all SAR mosaics described above, Landsat 8 satellite images and Sentinel-2 mosaics. The image classifications were evaluated in terms of seasonality.

In addition we investigated the use of 3D data from ALS and DAP. For the ALS we thinned the data to evaluate the effect of point density. We also reduced the sample plot size to investigate the effects of plot size.

All classifications were carried out using the same training and test dataset as used by Sverdrup-Thygeson et al. (2016) and the random forest algorithm (Breiman, 2001). Thus, the classifications were not optimized in any way but the results should give a relative value of the different remote sensing sources and the impact of seasonality etc. The accuracies of the classifications were assessed by computing the error-matrix and the kappa statistic. The kappa statistic – also referred to as the Cohens kappa – is a measure of the overall accuracy, and is well suited for comparison of different models solving the same classification problem. There are several ways of interpreting the kappa value – Landis and Koch (1977) consider 0-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate and 0.61-0.80 as substantial, and 0.81-1 as almost perfect.

Unsupervised detection of natural forests

For the entire OA two indices that are related to forest structure were tested. First, the forest structure habitat index (FSHI) presented by Coops et al. (2016/8) was created based on the highest ALS echo, the lowest density metrics and the coefficient of variation of the echo height distribution. To handle the challenges of different ALS projects the values were scaled for each project before merging. The FSHI is visualized in three dimension describing the forest height, forest density and forest complexity. The index is presented and some examples of forest reserves is highlighted. In total, the FSHI displays the combination of height, density and complexity of the forest.

Furthermore, the GINI coefficient (Gini, 1921) was computed and predicted to produce a map showing forest size diversity. The GINI coefficient have values between 0 and 1 and values. A homogeneous forest in terms of size of the trees will have a GINI close to zero, whereas a forest with heterogenous tree sizes will have a GINI close to one. The reasoning was the same as for using the GINI-coefficient for defining natural forest. The modeling was carried out using a mixed effects model where the project was a random effect.

Natural forest in Oslo and Akershus

Natural forest classification and estimation

The natural forest was classified using logistic regression, a robust classifier, suitable for model-assisted and model-based area estimation (McRoberts, 2010). Logistic regressions were selected and fitted on the NFI field data for five natural forest definitions (D1 through D5) and three remotely sensed data available wall-to-wall: Landsat 8, Sentinel-2 and ALS (15 models in total). The logistic models were selected using the best subset method in the bestglm R package, with AIC information criterion and limiting the maximum number of variables to five.

The general form of the logistic regression is:

$$\hat{y}_i = f(X_i, \beta) = \frac{\exp\left(\sum_{j=1}^{J} \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^{J} \beta_j x_{ij}\right)}$$

$$x_{ij} = \begin{cases} 1, j = 1 \\ \text{auxiliary variable}, j > 1 \end{cases}$$

J , number model parameters

n , number of field observations

N , total number of observations in the population

$y_i \epsilon \{0,1\}$, observed probability of natural forest

$\hat{y}_i \epsilon [0,1]$, predicted probability of natural forest

The area proportion of natural forest was estimated using three different estimators:

1. HT: Horvitz-Thompson direct estimator using the field observation only. This is the benchmark estimator.
2. MA: model-assisted estimator using a probabilistic sample of field observations and wall-to-wall auxiliary variables derived from remotely sensed data.
3. MB: model-based estimator using the wall-to-wall auxiliary variables derived from remotely sensed data.

The estimators of area proportion of natural forest as well as estimators of variance are provided below:

Direct (HT) estimator:

$$\hat{\mu}_{HT} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\widehat{Var}\left(\hat{\mu}_{HT}\right) = \frac{1}{n(n-1)}\sum_{i=1}^{n} \left(y_i - \hat{y}\right)^2$$

Model-assisted estimator:

$$\hat{\mu}_{MA} = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i + \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i$$

$$\widehat{Var}\left(\hat{\mu}_{MA}\right) = \frac{1}{n(n-1)}\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$$

Model-based estimator:

$$\hat{\mu}_{MB} = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i$$

$$\widehat{Var}\left(\hat{\mu}_{MB}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} Z_i^T V_{\hat{\beta}} Z_j$$

Z_i is the first-degree gradient vector of length J (number of β parameters in the model):

$$z_{ij} = \frac{\partial f\left(X_i, \hat{\beta}\right)}{\partial \beta_j}$$

$V_{\hat{\beta}}$ is the estimated variance-covariance matrix of the model parameters

Since the double summation in $\widehat{Var}\left(\hat{\mu}_{MB}\right)$ would pose computational difficulties due to large N, the following approximation strategy was used:

Take a sample of $m<N$ observations

Calculate $\widehat{Var}_{aprox}\left(\hat{\mu}_{MB}\right) = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} Z_i^T V_{\hat{\beta}} Z_j$

Repeat k times and take the average

We used $k=100$ and $m=10000$

Note that the results are reported as the proportion of natural forest for each definition, expressed in percentage of the total forest area. The total area of natural forest in OA is straightforward to derive by multiplying the estimated proportion to the total forest area. The uncertainty is reported in terms of estimated standard error (SE - square root of estimated variance)

Forest biomass map

Based on the 16 ALS projects and the NFI plots a random mixed effects model was created. After removing NFI plots covering multiple land cover classes (e.g. split plots) and some outliers likely related to difference in timing between field survey and ALS acquisition. Forest management such as thinning or harvest will for example influence largely either the management are carried out before or after field inventory. There could also be long time between the recordings and ALS, e.g. the oldest acquisition is from 2007 while field inventory might be carried out in 2017.

Results and discussion

Potential of remotely sensed data to identify natural forest

SAR

The classification accuracies obtained with Sentinel-1 and ALOS PALSAR-2 were of limited value for the separation to the old natural forest from the old manage forest. The largest accuracy obtained was a kappa value of 0.25. The months of February and March produced the highest accuracies for both years (Figure 5). The masked version of ALOS PALSAR-2 provided similar results. However, based on the accuracies obtained using the MEV dataset the Sentinel-1 and ALOS PALSAR-2 data for separation of the two forest types seem to be of limited value. The limited value might largely be due to the forest type definitions here and different approaches, like texture analysis could probably still improve the results. A study to use Sentinel-1 data for land cover classification including differentiation between deciduous and conifer forests in Troms County showed high accuracies[13] using the full monthly time-series and we believe that the potential of Sentinel-1 has not fully been investigated here yet.

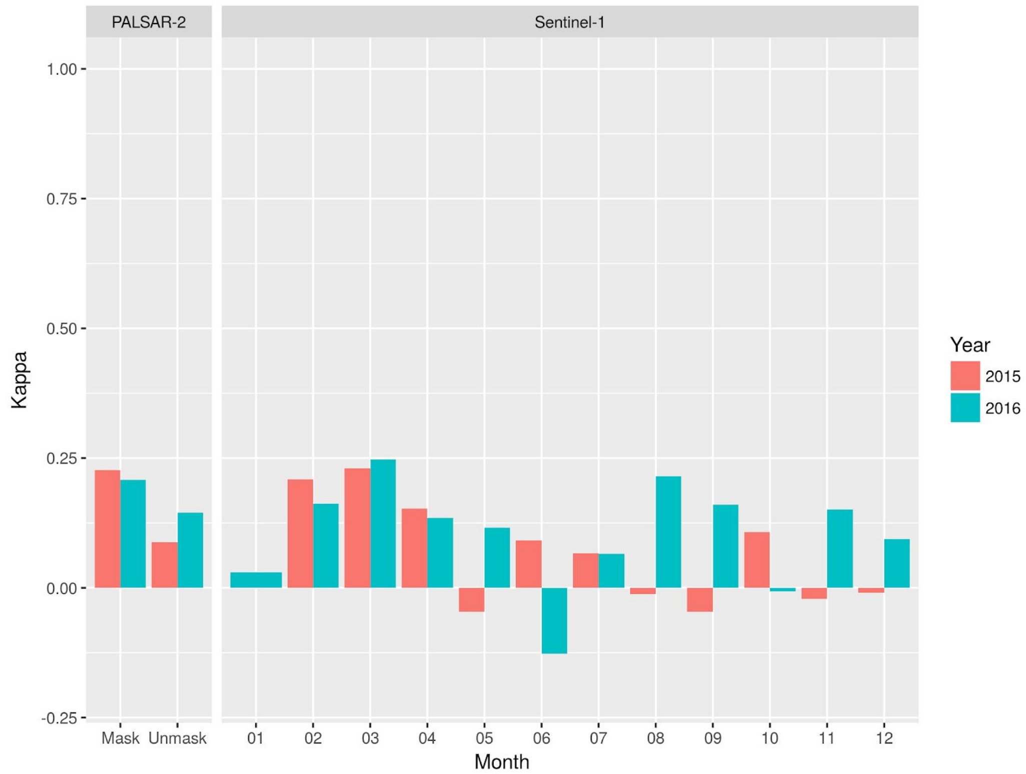


Figure 5. Accuracy of natural forest identification in MEV during the year based on SAR mosaics from 2015 and 2016.

[13] http://www.ifram.no/db.343156.no.html?lid=436.9d85d5fe2a30b13501bb4975ad1ed032

Landsat 8

The classification accuracy obtained with Landsat 8 for the separation of old natural forest and old manage forest was low. Accuracies in terms of kappa varied between 0.15 and 0.63, with a mean of 0.31. The highest accuracies were obtained using images from the autumn (Figure 6). Landsat 8 has higher accuracies than SAR data. However, compared to the results by Sverdrup-Thygeson et al. (2016) the accuracies are low.

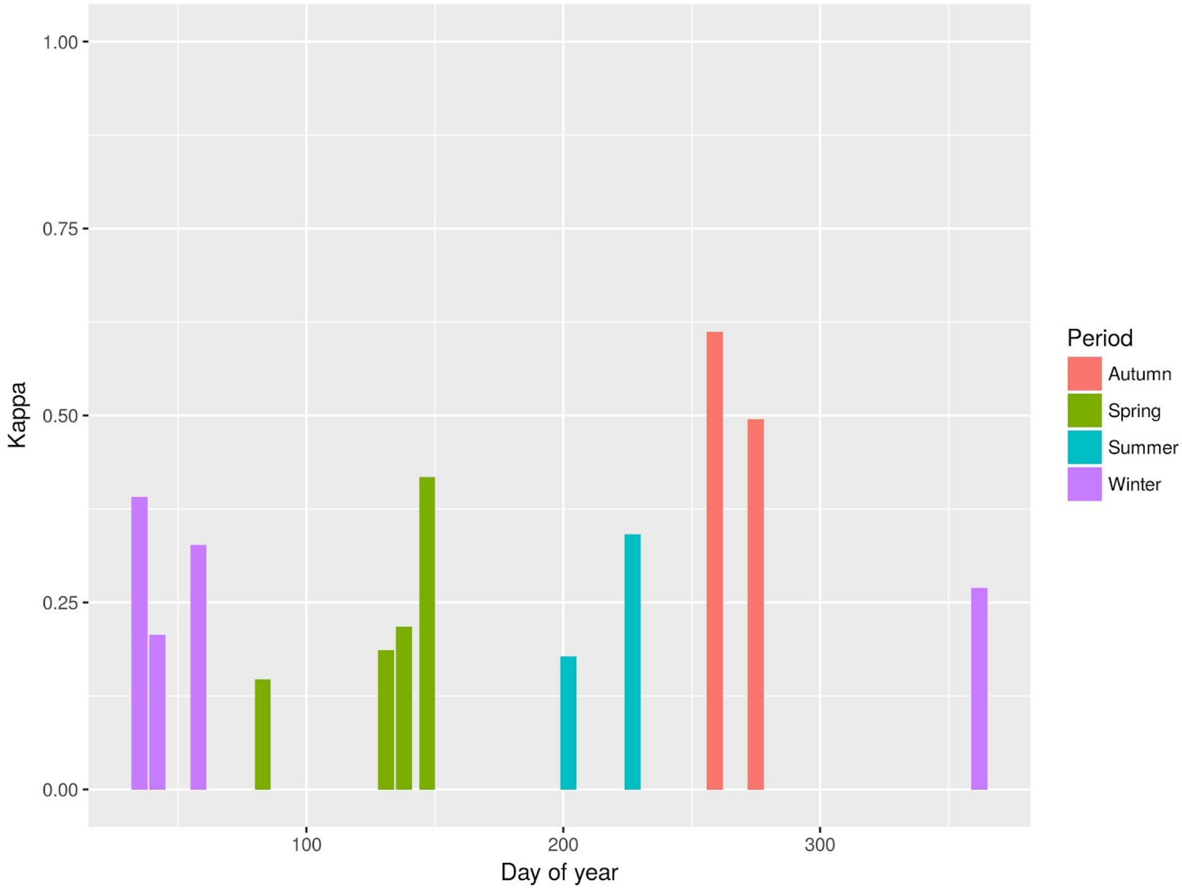


Figure 6. Accuracy of natural forest identification in MEV during the year based on Landsat 8 imagery.

Sentinel-2

The two mosaics created using Sentinel-2 were tested for the MEV area. The two mosaics reached accuracies of 0.20 and 0.23 in terms of kappa-coefficient for the test set. The accuracy was not sufficient for detecting natural forests. The use of Sentinel was not analyzed to its full potential in this study. Thus, different approaches to create mosaics should also be considered together with multi-temporal datasets and texture analysis. Thus, potential of Sentinel should be investigated further.

ALS

Effects of point density and sample plot area were analyzed in the MEV area based on selected sample plots from the study by Sverdrup-Thygeson (2016). Point density was reduced using a random binomial selection with 10 iterations. The classification was carried out using the random forest algorithm and for training and test datasets. The results showed small effects of point density and sample plot size (Figure 7). Earlier research using logistic regression indicated that a plot size of 1000 m^2 provided the highest kappa value, but the difference was not significant (Ørka, Sverdrup-Thygeson, Næsset, & Gobakken, 2014). The current trial indicated that the effect of point density was larger on smaller plots and that large plots were more robust to decrease in plot density. Overall the results indicated that ALS provided higher accuracy of natural forest identification than optical and radar satellite imageries.

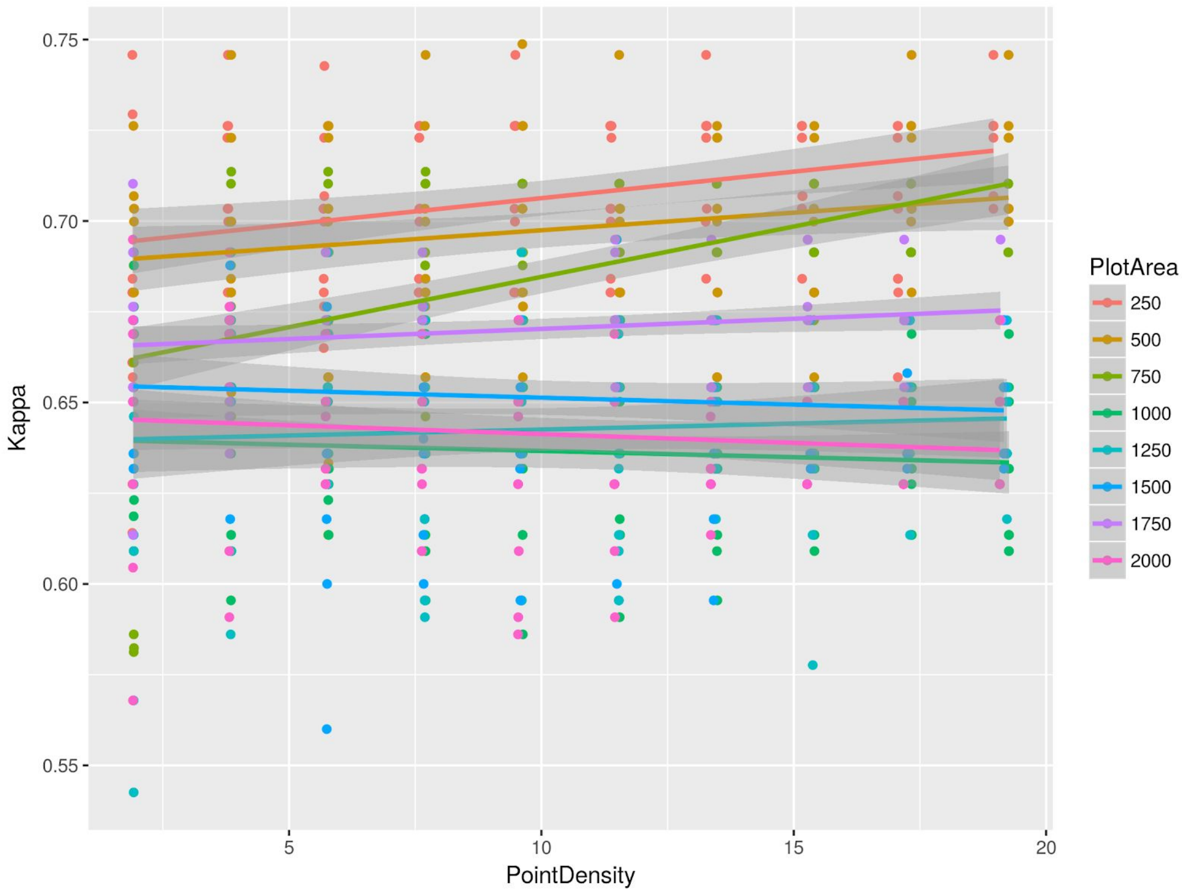


Figure 7. Effect of point density (PointDensity) and sample plot area (PlotArea) on classification of old natural forest and old managed forest.

DAP

The coverage of aerial images was not entirely overlapping with the ALS data. Thus, ALS was run with the same data for comparison. The DAP data resulted in a kappa-coefficient of 0.58, while for the same dataset the ALS obtained a kappa-coefficient of 0.65. The result was somewhat surprising based on the greater details in the ALS compared to the DAP data (Figure 8). Thus, the analysis was also carried out using boosted regression trees (De'ath, 2007) for comparison. The accuracies were lower for both data sources. The accuracies obtained with the boosted regression tree method were 0.51 and 0.61 for DAP and ALS, respectively. The accuracy statistics for the two methods (Figure 9) favor the ALS, but also show the potential of DAP data for mapping natural forests. It should also be noted that the performance for DAP is slightly lower than the best Landsat 8 classification, but overall it provides much better results.

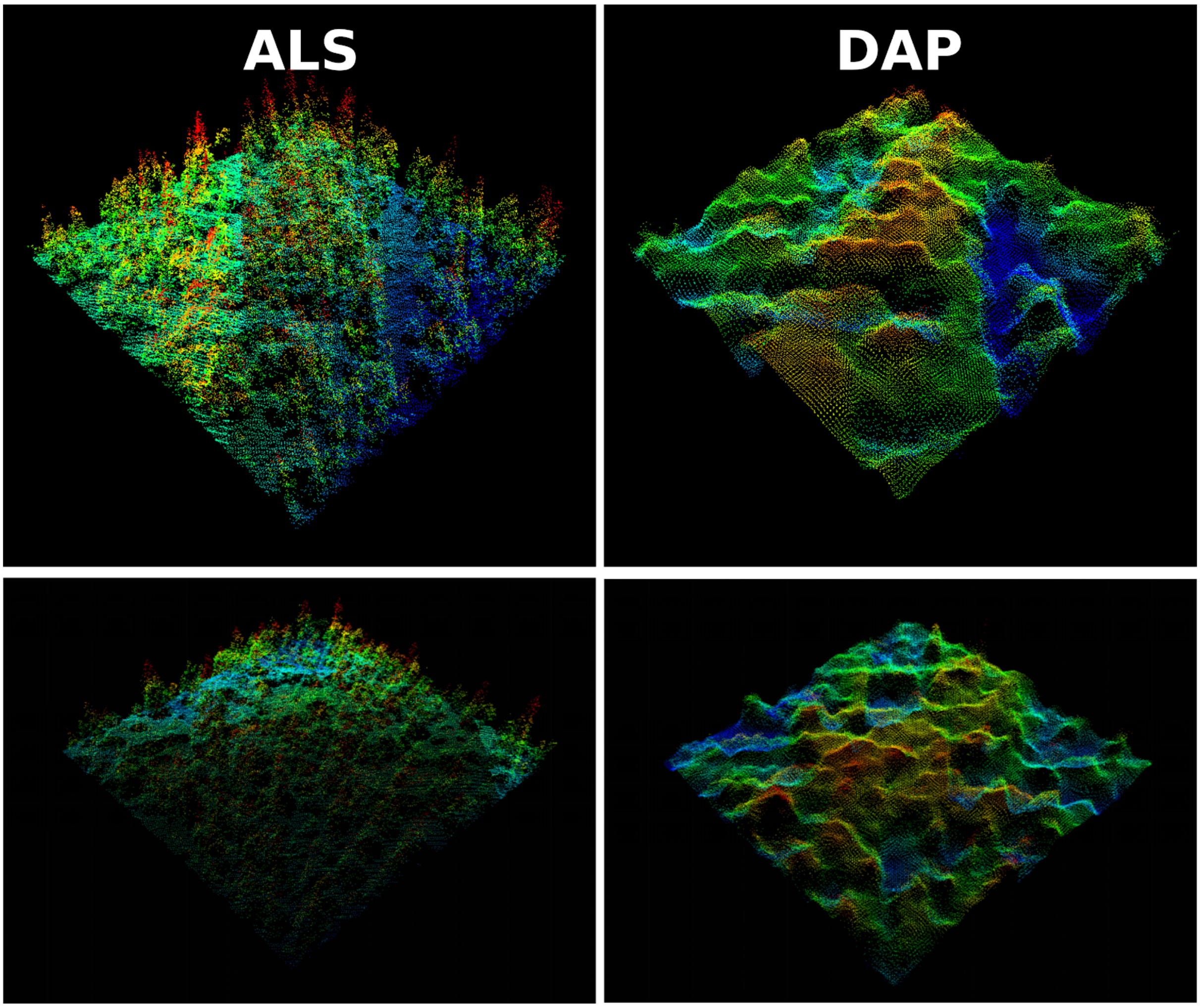


Figure 8. Comparison of details in ALS and DAP for plot 190 (top) and plot 53 (bottom) in the MEV study area. Plot 190 is a 80 year old managed forest on low site productivity, while plot 53 is a 123 years old old natural forest with medium site productivity. Both located approximately 600 m a.s.l.

Unsupervised detection of natural forests

The FSHI provided a visual product of forest structure. The corrected map did not show any severe differences between ALS projects. Thus, the major concern of using multiple ALS projects seem to be overcome at least for visual assessment. The map is simple to develop and provides important information on the forest structure. A visual assessment showed high degree of match between the map and forest reserves. For example, the map could easily be used to draw the borders of Østmarka forest reserve. (Figure 10). The similar color combination of the FSHI is also found on several locations and at many of these, there exists forest reserves. However, other areas e.g. Marifjell (Figure 10) displays a different FSHI because of the lower tree heights and more sparse forest. Two other reserves are shown as examples namely Hattersen and Prestehesten (Figure 10). The general impression is that forest reserves are mostly located within the medium density, medium-high complexity and

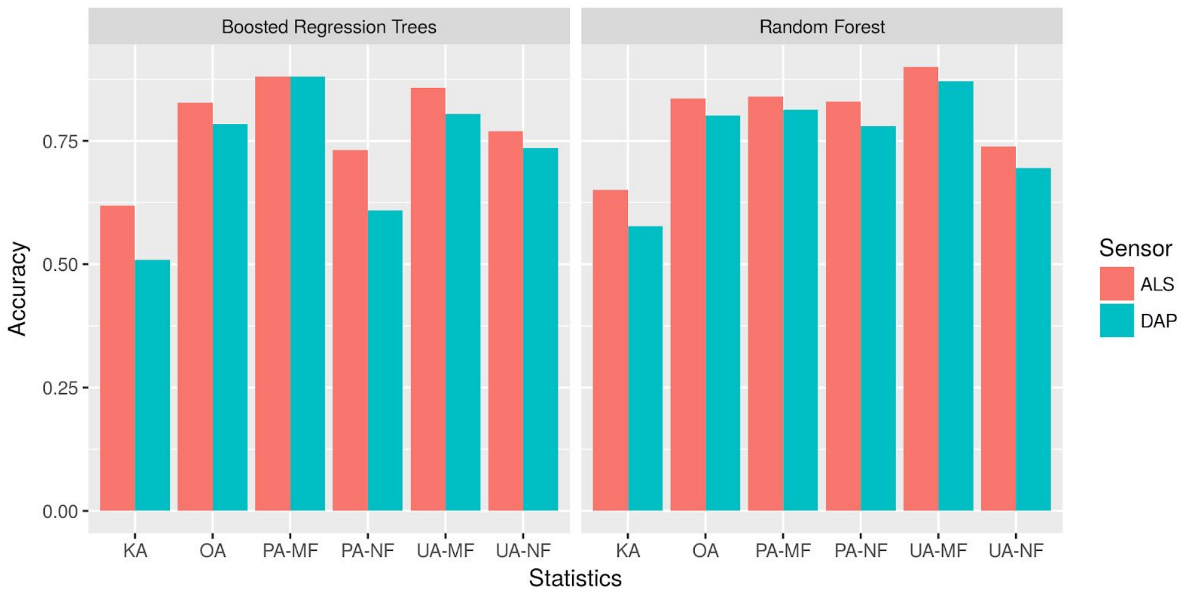


Figure 9. Accuracies for DAP and ALS using two statistical methods. The accuracy statistics are KA=kappa-coefficient, OA=overall accuracy, PA=producer's accuracy, UA=User's accuracy, while MF refers to Manage Forest and NF to Natural Forests.

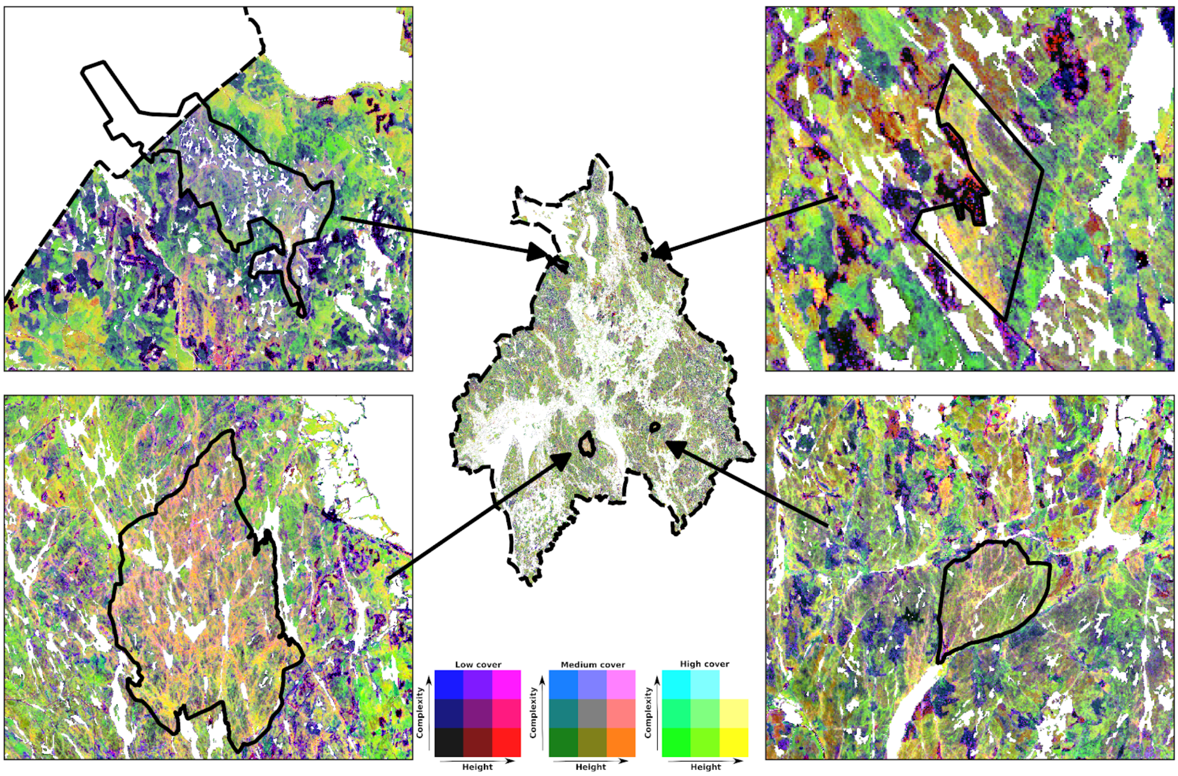


Figure 10. Corrected ALS based forest structural index for OA for four selected forest reserves. Upper left: Marifjell, upper right: Hattersen, lower left: Østmarka, lower right Prestehesten.

medium-high forest. The FSHI seems to be a promising tool for forest structural information. However, the ability to predict presence of natural forest was not further investigated, but this should be done in further research.

The GINI-coefficient was predicted using a random effects model. However, the explained variance was only 27%. However, the relative RMSE was 24%. There is a large variation in forest conditions that can show large size diversities. However, the produced map displayed similar trends as the FSHI (Figure 11).

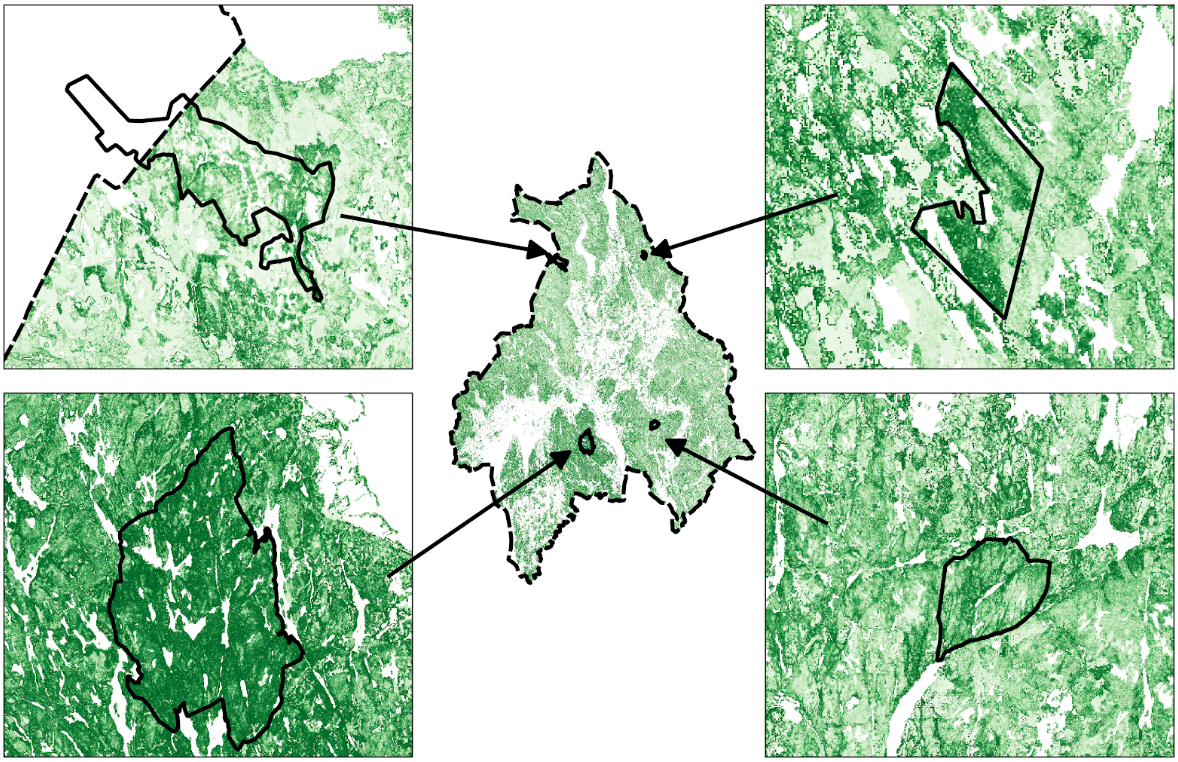


Figure 11. The size diversity in forest by GINI-coefficient predicted by ALS for OA for four selected forest reserves Greener areas have higher GINI-coefficient. Upper left: Marifjell, upper right: Hattersen, lower left: Østmarka, lower right Prestehesten.

Natural forest in Oslo and Akershus

Natural forest classification and estimation

The quality of the logistic models can be judged by the area under the receiver operating characteristic (ROC) - AUC (Figure 12). The ROC curve plots the true/false positive rate as the natural forest probability threshold varies between 0 and 1. In addition, figure 13 is a visual impression of how well the logistic model separates natural from non-natural forest. The multispectral (Landsat 8 and Sentinel-2) models were better than ALS in predicting natural forest according to definitions D1, D2, and D3. For the other two natural forest definitions (D4, D5), the ALS models outperformed the multispectral models.

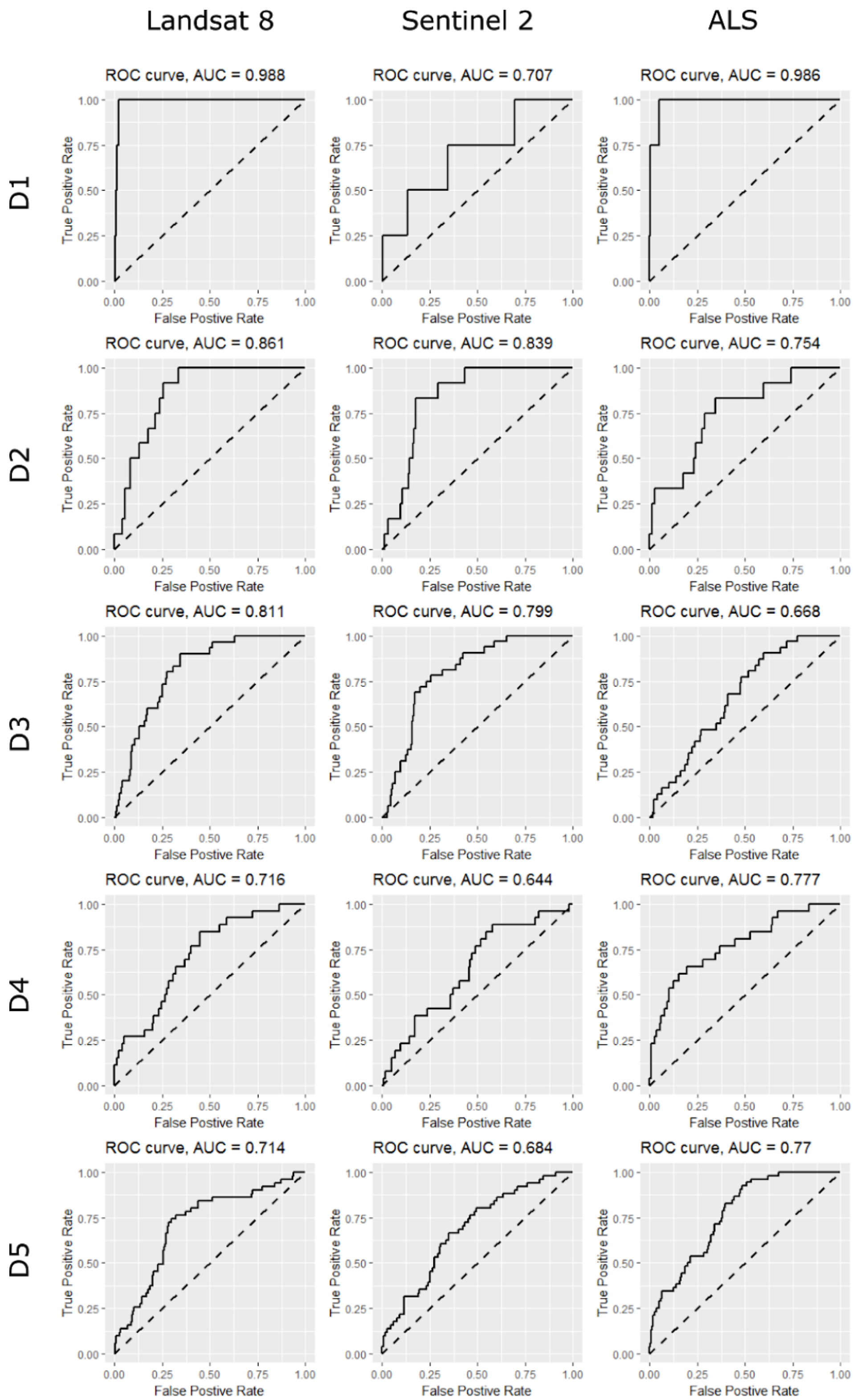


Figure 12. Receiver operating characteristic (ROC) curves of the logistic models for five different natural forest definitions (D1-D5) and three types of remotely sensed data. The definitions used are: D1: NFI definition of "natural forest" used for forest characteristics. D2: NFI stand age >= 160 years old, D3: NFI stand age >= 140 years old, D4: The 25% most size diverse forests in development class V based on GINI, D5: The 50% most size diverse forests in development class V based on GINI.

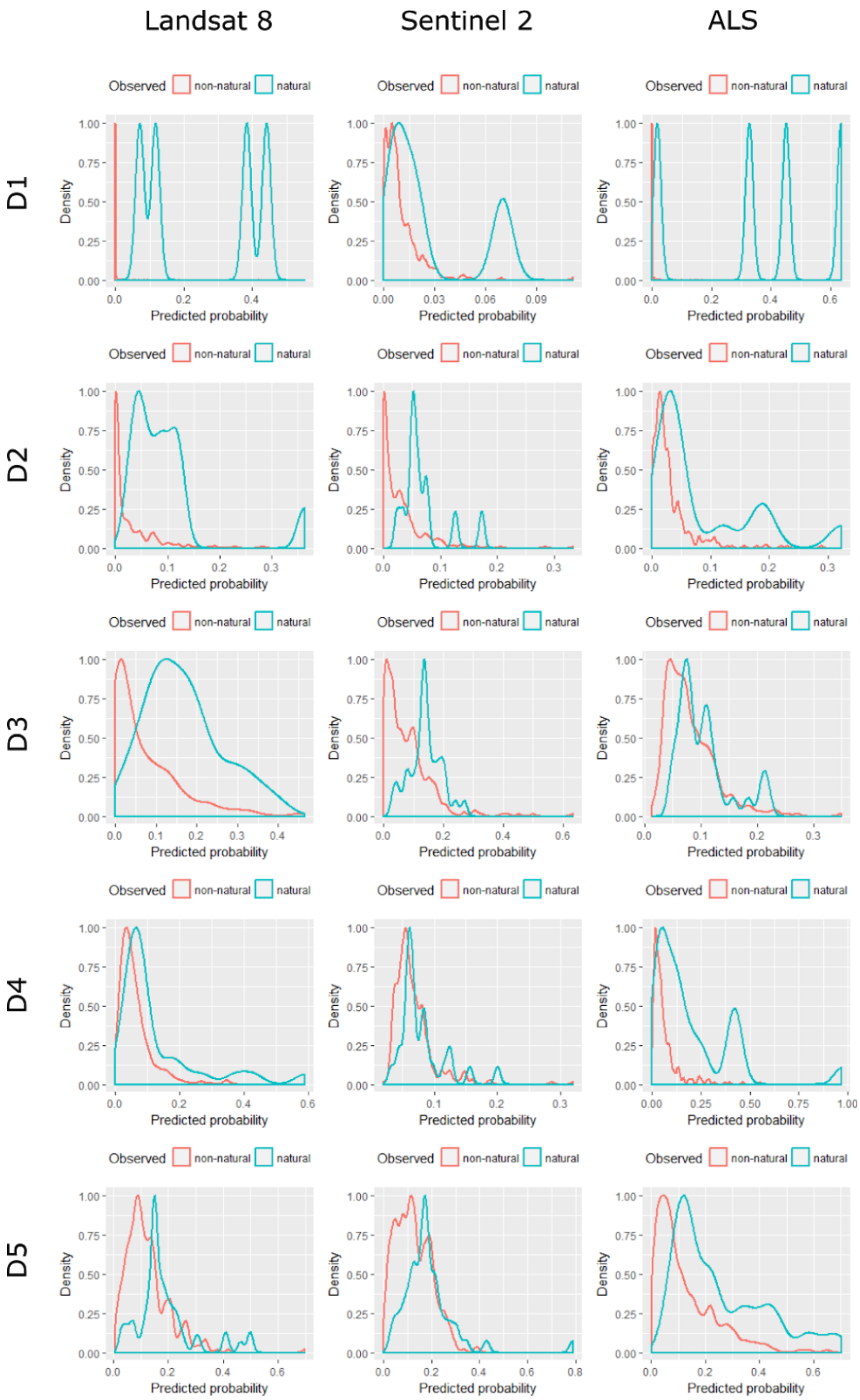


Figure 13. Predicted natural forest probabilities for observed natural or non-natural forest using five different natural forest definitions (D1-D5) and three different sources of remotely sensed data. The definitions used are: D1: NFI definition of “natural forest” used for forest characteristics. D2: NFI stand age >= 160 years old, D3: NFI stand age >= 140 years old, D4: The 25% most size diverse forests in development class V based on GINI, D5: The 50% most size diverse forests in development class V based on GINI.

The natural forest area estimation results are reported in table 4. In general, the estimates based on remotely sensed data were more precise than the corresponding field based estimates (20/30 cases). Among the remotely sensed data types, the Landsat 8 seem to be the best, with the 9/10 cases where the precision was improved, followed by the Sentinel-2 with 8/10 cases and ALS with 3/10 cases. The mean difference between the remotely sensed based SE and field-based SE expressed in percentage of the field based SE was -3.24%, -0.09% and +2.48 for Landsat 8, Sentinel-2 and ALS respectively. The MA and MB estimates of natural forest proportion were similar, indicating that the models had little bias.

Table 4. Natural forest area estimation results for different definitions (D1-D5). HT - NFI field plots only, MA - model assisted, MB - model based. SE values for estimates based on remotely sensed data are bolded if below the corresponding NFI field based estimator. The definitions used are: D1: NFI definition of "natural forest" used for forest characteristics. D2: NFI stand age >= 160 years old, D3: NFI stand age >= 140 years old, D4: The 25% most size diverse forests in development class V based on GINI, D5: The 50% most size diverse forests in development class V based on GINI.

Sensor	Natural forest definition	Estimator	Estimated proportion (%)	SE
	D1	HT	1.01	0.50
	D2	HT	3.06	0.87
	D3	HT	7.92	1.36
	D4	HT	6.59	1.25
	D5	HT	12.94	1.69
Landsat 8	D1	MA	1.37	**0.45**
		MB	1.37	0.56
	D2	MA	2.95	**0.84**
		MB	2.95	**0.82**
	D3	MA	7.79	**1.30**
		MB	7.79	**1.28**
	D4	MA	6.73	**1.21**
		MB	6.73	**1.22**
	D5	MA	12.71	**1.64**
		MB	12.71	**1.60**

Table 4. continued

Sensor	Natural forest definition	Estimator	Estimated proportion (%)	SE
Sentinel-2	D1	MA	0.96	**0.49**
		MB	0.98	**0.48**
	D2	MA	3.61	**0.85**
		MB	3.61	0.93
	D3	MA	7.87	**1.35**
		MB	7.87	**1.28**
	D4	MA	7.26	**1.24**
		MB	7.41	1.39
	D5	MA	13.12	**1.63**
		MB	13.31	**1.68**
ALS	D1	MA	1.76	**0.41**
		MB	1.76	0.55
	D2	MA	3.43	0.88
		MB	3.43	0.93
	D3	MA	8.60	1.41
		MB	8.60	1.45
	D4	MA	8.56	**1.22**
		MB	8.74	1.38
	D5	MA	15.93	**1.67**
		MB	15.93	1.83

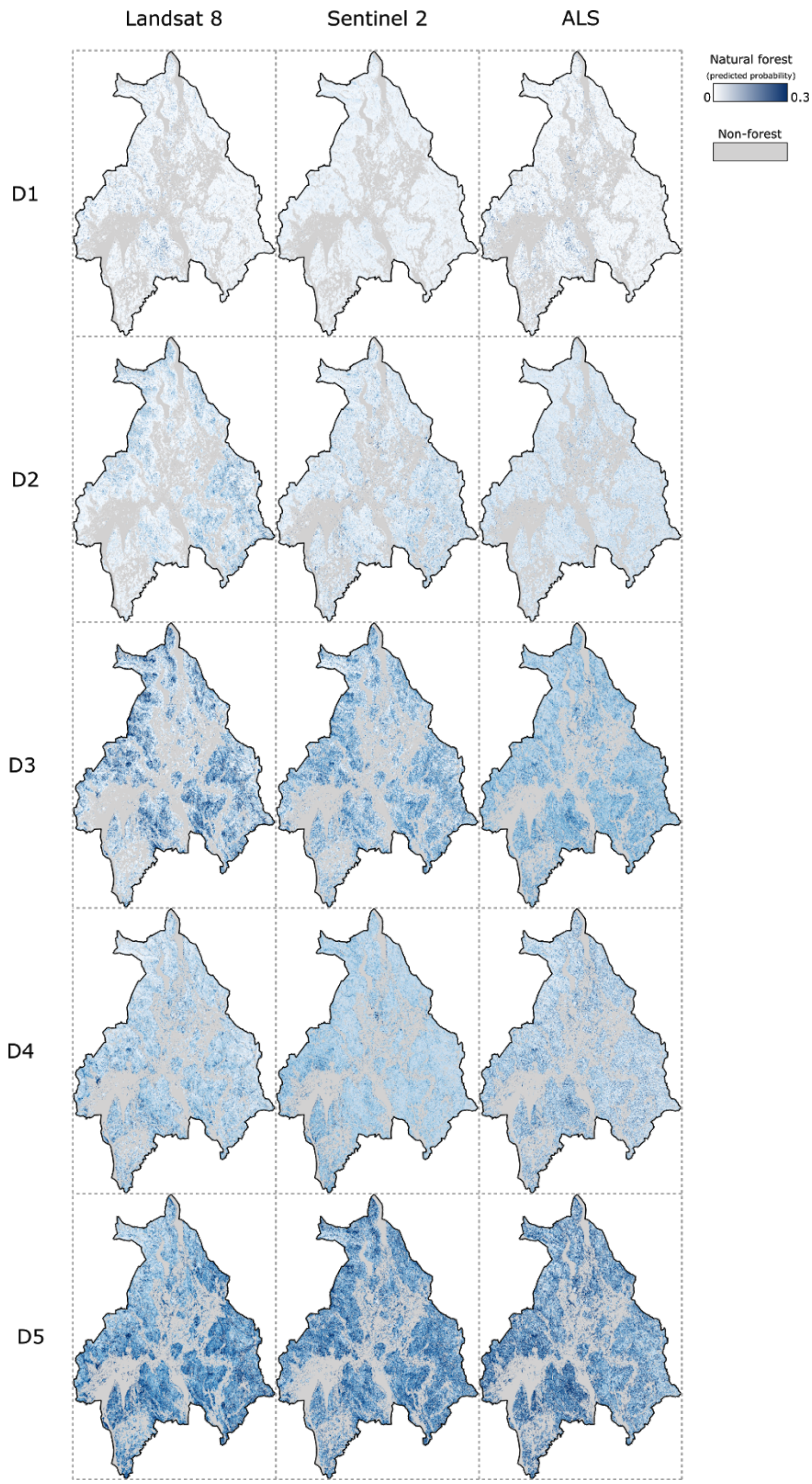


Figure 14. Natural forest prediction maps according to five natural forest definitions (D1-D5) using Landsat 8, Sentinel-2, and ALS. The definitions used are: D1: NFI definition of “natural forest” used for forest characteristics. D2: NFI stand age >= 160 years old, D3: NFI stand age >= 140 years old, D4: The 25% most size diverse forests in development class V based on GINI, D5: The 50% most size diverse forests in development class V based on GINI.

Forest biomass map

Two variables were selected in the variable selection namely the mean height of ALS echoes and the lowest density variable. The model explained approximately 68% of the variation and had a RMSE of 37%. Figure 15 (left) shows the predicted and observed values for the modeling plots and Figure 15 (right) shows the biomass map.

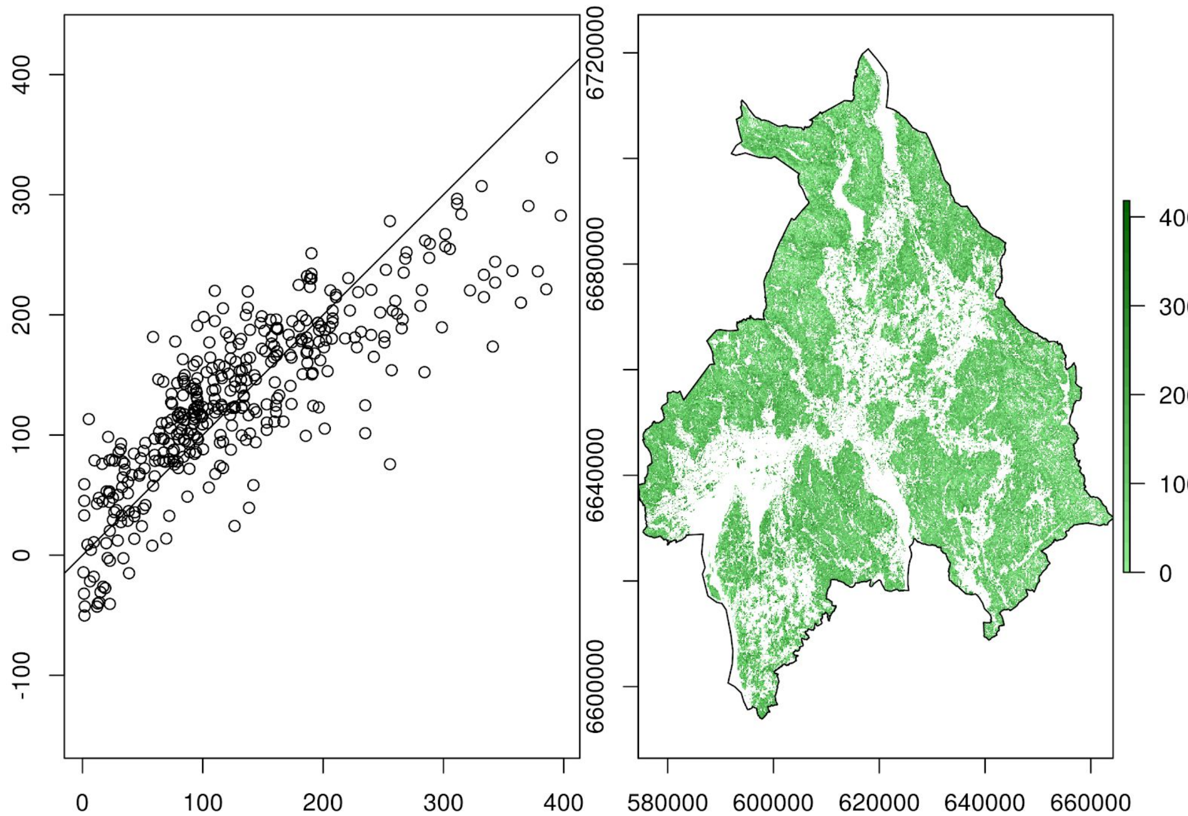


Figure 15. Observed and predicted above ground biomass (AGB) based on ALS on the plots used for modeling (left) and predicted AGB for OA (right). AGB in tons/ha.

General discussion

Overall, the current research provides results that is promising for further mapping and monitoring of natural forest. There are still a number of challenges that need to be solved. Furthermore, there are several ways of improving the methods and classifications for all datasets used. The current study screened different sensors and satellites and identified advantages and disadvantages. We also tested different definitions for natural forest. The definitions were mainly based on forest structure and/or age, but some also included dead wood characteristics or management history. The different definitions provide different proportions of natural forest. Since, the criteria for defining forest as natural forest change from where strict to including half of the forest in the mature development class. Independent of this the maps can be used to located the most interesting natural forest areas within the county could be identified for further investigation. The structural index maps produced also hold valuable information for identifying natural forest and revealed that structure of natural forest varies by location. Below we discuss different aspects of the current research.

Selection of reference data source for mapping natural forest

The NFI data have recordings of natural forest according to their definition. The number of observations was small, but they also represent a very strict definition. However, the NFI data are suitable to derive several definitions that might match the criteria of detecting natural forest that have not been regenerated after clear-cutting, but might have been managed by selective logging in the past. Both large age and size diversity are shown as examples.

The amount of dead wood is used by the NFI in their definition of natural forest. Remotely sensed data have been used for estimating the amount of dead wood. Maltamo et al. (2014) reviewed the existing research concerning coarse woody debris (CWD) characterization by means of ALS data. In general, the accuracy of ALS-based CWD models varies considerably, from accurate predictions in different nature conservation areas to hardly statistically significant models in managed forests in Hedmark county, Norway, with a very low amount of CWD. Thus, the amount of dead wood was not assessed in the current study because as in Hedmark county the amount of CWD was very low.

The best way to establish reference data for further analysis or validation of the current research is to check historical management practices in available aerial imageries. Using recent and historical imagery, it is easy to check if current old forests have regenerated from selective cuttings or clear-cuts. The availability of such images varies in Norway and the authors have no information about how the coverage of such aerial images are. In Sweden, such images have been used in an automated process to identifying clear-cut areas (Ahlkrona et al., 2017).

If definitions based on current state are used, additional data can be added from known natural forest sites such as natural reserves and other recordings in national databases, e.g. Naturbase[14]. Environmental registrations carried out as part of the operational forest management planning inventory in Norway (See Norwegian Agriculture Agency for further information[15]) might also be used as reference data. Further work should be devoted to evaluate if these environmental data could be used.

Definitions of natural forest

Natural forest can be defined in a number of different ways, depending on focus and degree of strictness (Rolstad et al., 2002). One definition, also adopted in the NiN system, characterizes natural forest like this: Naturskog er skogsmark med skogbestand framkommet ved naturlig foryngelse av stedegent genmateriale, der menneskelig påvirkning har funnet sted i så liten utstrekning, for så lang tid tilbake, eller er utført på en slik måte, at skogsmarkssystemets naturlige struktur, sammensetning, og økologiske prosesser dominerer[16].

---

[14] http://www.miljodirektoratet.no/no/Tjenester-og-verktoy/Database/Naturbase/
[15] https://www.landbruksdirektoratet.no/no/eiendom-og-skog/skog-og-miljoregistreringer/miljoregistreringer#om-miljoeregistrering-i-skog-
[16] https://www.artsdatabanken.no/Pages/181979

In the NiN system, the variable 7SD Skogbestandsdynamikk separates natural forest ("Naturskog", 7SD-0) from managed forest ("Normalskog", 7SD-NS). The categorization is based on a combination of criteria targeting absence of signs of previous logging (7SB-HS hogststubbeandel), amount and quality of dead wood (4DG Stående død ved, 4DL Liggende død ved), forest structure (9TS Tresjiktstruktur) and large trees (4TS Store trær).

In this report we use six different definitions of natural forest. The first definition (D1) is identical to the very strict definition used by NFI. This definition can be approximated in the NiN system by the variable 7SD Bestandsdynamikk. This definition is too strict to be useful in our context. The ecological relevance of the strict rule of no signs of previous logging allowed (i.e. no cut stumps) can also be questioned.

Definition 2 and 3 both use different stand age criteria to define natural forest, either stand age above 160 years (D2) or above 140 years (D3). The use of stand age as the sole criteria for defining natural forest is not ideal, for several reasons: (1) Maximum tree age differs between tree species and forest site index (2) Stand age is a mean weighed by basal area, meaning that large trees count more, although they are not necessarily the oldest trees. In NiN, these definitions would have to be approximated either by using 7SD Skogbestandsdynamikk, or by characterizing tree age using 4TG (Gamle trær).

Definition D4 and D5 target the larger size variation in natural forest, with two different cutoff values (D4: 25% most size diverse, D5: 50% most size diverse) within the oldest development class (hogstklasse 5). Again, the best possible description of these definitions in NiN would probably be by using 7SD Skogbestandsdynamikk.

The final definition (D6) is not based on age or structural characteristics as such, but rather on site-specific information on previous forest management (from old management maps). As this definition is not based on field characteristics, using NiN to describe it seems less relevant – although NiN do have variables describing the occurrence of selective logging (7SB Skog-bruk).

Several evaluations and reports have expressed a need for a more precise characterization of the variation along this gradient and such work is now being initiated (pers. comm. Rune Halvorsen). Hopefully, this work will provide a more precise and more ecologically relevant definition of natural forest, to be used in further work on mapping of natural forest.

In general, several present initiatives relate to forest with high conservation value (where natural forest is one component), and these initiatives would gain from good coordination (Naturtyper av nasjonal forvaltningsinteresse, e.g. Evju et al. (2017) og Evju et al. (2017), Kriterier for naturverdi i skog (Framstad, Halvorsen, Storaunet, & og Sverdrup-Thygeson, n.d.), NiN-kartlegging i verneområder, utvalgskartlegging i NiN, Blågrønn infrastruktur (Framstad, Bryn, Dramstad, & Sverdrup-Thygeson, 2017), MiS-kartlegging med NiN (Landbruksdirektoratet, 2017)).

Selection of remotely sensed data source for mapping natural forest

From the current studies ALS data seem superior to map natural forest on local scales. The large number of projects with different settings is a drawback for the use of ALS on large scales. However, different methods to scale the data were tested and found relevant especially for the estimation of biomass, GINI and to create a homogenous FSHI. The initial methods used to minimize the differences between different ALS acquisitions and sensor work relatively well and can easily be used to create such maps covering all laser scanned areas in Norway. Nevertheless, this is a challenging task that needs further research. The results from this study are very positive and point on a potential data source for large scale forest remote sensing. With the NDH data acquisitions covering the entire forested areas in Norway ALS will be a great data source. None of the satellites could offer the same cover of the county as ALS.

The horizontal metrics used by Sverdrup-Thygeson et al. (2016) were not used in the current research due to time restrictions. The method of segmentation of trees are time consuming. Furthermore, the processed data needed to be extracted for NFI plots before prediction and this was not possible within the time frame of the current project. However, this should be evaluated in further research. Similarly, only first returns was used adding information from subsequent returns, i.e. intermediate and last of many returns, might have improved the results.

Point clouds from image matching for the entire OA was not tested. The results from MEV are promising. The advantage of such data is that they will be more homogeneous over large areas. The drawback is that they do not describe the forest structure that well because the point clouds only describe the outer canopy surface while ALS also penetrate the canopies and capture internal variation and understory. There are some challenging issues in applying DAP for large areas (Rahlf, Breidenbach, Solberg, Næsset, & Astrup, 2017). However, when the NDH data are getting old and outdated and if national ALS acquisitions are not repeated, DAP data might be a good alternative since a national program for repeated aerial photographs was established already in 2006[17].

In the current analysis, SAR was not showing promising results. Other studies[18] suggest that there is a large potential for better use of SAR data (Haarpaintner et al. 2016). However, this could not be fully investigated here. C-band SAR from Sentinel-1 saturates at biomass well below 100t/ha (Mitchard et al. 2009) and it can therefore not distinguish forests of different ages by the backscatter alone. Texture analysis could give information suitable for distinguishing between homogeneous forest areas (trees of same age) and heterogenous natural forests with gaps from fallen trees, but has not been investigated here since we doubt that this works satisfactory and automatically over large areas.Sentinel-1 data should also be tested for classification in OA for comparison with optical satellite imagery. Although, SAR data penetrate clouds there are still holes in the mosaics especially in steep terrain.

[17] https://www.kartverket.no/geodataarbeid/flyfoto/nasjonalt-program-for-omlopsfotografering/
[18] http://www.ifram.no/db.343156.no.html?lid=436.9d85d5fe2a30b13501bb4975ad1ed032

Landsat 8 and Sentinel-2 did not provided very good results in the MEV area. However, on Landsat 8 image provided results close to the one obtained with ALS. Thus, selection of imagery are an important point for obtaining high accuracies in these optical images. For further remote sensing of natural forest the strength of optical data are in the historical records and the ability to detect changes (Cohen, Fiorella, Gray, & Helmer, 1998; Santoro, Pantze, Fransson, Dahlgren, & Persson, 2012; Solberg et al., 2014). The less accurate results of Sentinel-2 can be attributed to the creation of the mosaic, the time of year of the acquisitions and the resolution. The spectral characteristics of Sentinel-2 should be at least as good as for Landsat 8. Thus, it is recommended to further include Sentinel-2 in analysis of natural forests.

Fusion of different data sources has not been focused in this study. Using complementary sources such as ALS and optical satellites or SAR and optical satellites might provide an additional dimension that improve the obtained results.

Estimating natural forest by remote sensing

Remotely sensed data improve the precision of natural forest estimates. In this case study however, the improvement was rather modest for some of the data sources. In general, the results based on Landsat 8 were most promising for discerning the natural forest. Although ALS provide models with similar accuracy or even better accuracy. Thus, it was a surprising result that Landsat 8 provided lower SEs, knowing that ALS would provide information of the vertical forest structure, which intuitively should be good indication of how “natural” a forested area is. This result can be explained by the fact that the ALS dataset was rather heterogeneous, being comprised of 16 different ALS acquisitions. This may be further addressed by calculating robust ALS metrics that are more stable across different acquisitions. In contrast to the many ALS acquisitions, one Landsat scene covered approximately 93% of OA. Thus, a mosaic of many Landsat scenes would be needed, challenges might occur also for Landsat data. An invariant advantage of using georeferenced remotely sensed data is the potential to produce wall-to-wall maps, in this case of natural forest areas. Furthermore, the potential of fusing different data sources has not been explored in this study.

Lessons learned

“Cloud free optical satellite data is an illusion” - the amount of clouds in Norway limits the possibilities to get one cloud free scene during a year. Thus, mosaics need to be created. This is one of the points that need to be improved such that several mosaics can be tested in the classification. Initially only one year was selected for doing optical satellite analysis in the current research. In light of the obtained results, multiple dates and historical data should be used in further studies. One limiting factor in this regard is that NFI coordinates are not generally available for use in such system.

“Processing Sentinel-2 data is time consuming” - with no running processing chain for Sentinel-2 data it was too time consuming to follow a “use it all strategy” and thus mosaics were needed. To make an efficient processing chain simple tools to conduct these tasks are necessary. Downloading images and processing them using standard algorithms and standard serves are not effective and thus cloud computing is needed. One example on a tool that provide means to generate cloud free mosaics for short and long periods are the google earth engine. A yearly Sentinel-2 cloud free mosaic was generated and download in less than 1 hour. It should also be noted that in Sweden for example there exists a mosaicking service providing optical satellite image mosaics[19]. Table 5 summarizes some experiences on time consumption, resource consumption (storage etc.) and pros and cons for Landsat and Sentinel-2 processing for comparison.

“No coordinates” - the agreement for using data from the NFI was signed and access was granted. However, the coordinates of the NFI sample plots were not given. Coordinates are a must in remote sensing and lacking coordinates is a huge drawback for checking of outliers etc. It also limits the flexibility in testing different approaches and examine new ideas as they appear when working with a project. However, NIBIO provided personnel to run provided code against the remotely sensed data and supported when needed.

Further development

In Sweden an interesting work on old natural forest has been carried out and a change detection analyses were carried out to identify clear-cuts. The clear-cuts were derived from optical satellite imagery ranging from current Sentinel-2 and Spot images to historical Landsat images dating back to 1972 and historical orthophotos from the 1960s (Ahlkrona et al., 2017). Identifying clear-cuts in these images could be used to update our probability maps with information on clear-cuts. The models could include information about detected clear-cuts and thus limit the area of wrong classifications.

The current research could be improved by developing better procedures for mosaicing Sentinel-2 and Landsat 8. In addition, data from multiple years should be combined. Another area that is not threatened in the current research is the use of object based image analysis.

[19] http://saccess.lantmateriet.se

Table 5. Statistics and experiences from the Sentinel-2 and Landsat 8 processing

Task	Sentinel-2	Landsat 8
Number of images	117	24
Size (GB)	1110 GB	23 GB
Download	Download using script provided by ESA. Approximate 2 weeks.	Using LIBRA Easy to download all the bands together Download time: ≈ 10 - 20 min/image
Processing TOA/BOA	Using Sen2cor Processing time: ≈ 12h / folder.SAFE (old format) ≈ 1,5 h/folder.SAFE (new format)	Using Google Earth Engine Processing time: ≈ 9 - 15 min/image (save in drive) + 2 min/image (downloading)
Cloud masking	Using Fmask code for Matlab Processing time: ≈ 1-2 h/folder.SAFE	Using Google Earth Engine Processing time: ≈ 5 - 10 min/image (save in drive) + 1 min/image (downloading)
Advantages processing	Code for download all the Sentinel-2 images. Standalone version of the software	In general less images Easier and faster to process (TOA and FMask), Online software with more documentation and support.
Disadvantages processing	Some problems with the software (installation and some errors) Long processing times, which have made impossible to process all the images. Difficult to process more than one image at the same time	Manual download of the images With Google Earth Engine for download the images previously is necessary to save them in drive or the cloud
Overall performance	Higher resolution and different storage formats. Routines for BOA convention and cloud detection are not as good as for Landsat 8. Processing times and number of images are much larger. Offers larger potential when technices mature. Downloading availability of BOA and cloud detection on demand could improve analysis for end-users.	Simple format and downloading routines. Straightforward conversion to TOA and mature algorithms for detecting clouds and analysis.

Conclusions

The current work provides interesting results regarding the possibility to extract information on forest naturalness and identifying old natural forest. The key conclusions were:

1. Creating corresponding probability maps and area estimates of natural forest are straightforward using logistic regression.
2. DAP and ALS provide the highest classification accuracies on local scales, while specific Landsat 8 images also provide good classification results. On a county scale, Landsat 8 is most promising.
3. Sentinel-1 and Sentinel-2 did not produce high accuracies. However, pre-processing and variable extraction could be improved in further research.
4. ALS provides the highest accuracies in natural forest definitions based on forest structure.
5. ALS based maps of forest structure show large visual correspondence with existing forest reserves.
6. An ALS-based forest biomass map covering 98.8% county's forest was created from 16 different ALS-projects with a relatively high accuracy.

Acknowledgement

We would like to thank Dr. Marek Marian Pierzchala and MSc. Ana Maria de Lera for data processing and suggestions to the project.

References

Ahlkrona, E., Giljam, C., Wennberg, S., Jönsson, C., Alvarez, M., Bovin, M., … Renström, M. (2017). Kartering av kontinuitetsskog i boreal region. *Metria AB På Uppdrag Av Naturvårdsverket*.

Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters* , *5*(5), 423–431.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, *113*(5), 893–903.

Cohen, W. B., Fiorella, M., Gray, J., & Helmer, E. (1998). An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat imagery. *Engineering and Remote …*. https://doi.org/10.1.1.496.1031

Coops, N. C., Tompaski, P., Nijland, W., Rickbeil, G. J. M., Nielsen, S. E., Bater, C. W., & Stadt, J. J. (2016/8). A forest structure habitat index based on airborne laser scanning data. *Ecological Indicators*, *67*, 346–357.

De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, *88*(1), 243–251.

Dostálová, A., Hollaus, M., Milenković, M., & Wolfgang, W. (2016). FOREST AREA DERIVATION FROM SENTINEL-1 DATA. *ISPRS Annals of Photogrammtry and*

Remote Sensing Spatial Information Science, *III-7*, 227–233.
Evju, M., Blom, H., Brandrud, T. E., Bär, A., Johansen, L., Lyngstad, A., … Aarrestad, P. A. (2017). Naturtyper av nasjonal forvaltningsinteresse. Revidert forslag til vurdering av lokalitetskvalitet. *NINA rapport / Norsk institutt for naturforskning*, *1428*, 95.
Evju, M., Blom, H., Brandrud, T. E., Bär, A., Johansen, L., Lyngstad, A., … Aarrestad, P. A. (2017). Verdisetting av naturtyper av nasjonal forvaltningsinteresse. Forslag til metodikk. *NINA Rapport*, *1357*, 172.
Eysn, L., Hollaus, M., Schadauer, K., & Pfeifer, N. (2012). Forest Delineation Based on Airborne LIDAR Data. *Remote Sensing*, *4*(3), 762–783.
Framstad, E., Bryn, A., Dramstad, W., & Sverdrup-Thygeson, A. (2017). Blågrønn infrastruktur. Landskapsøkologiske sammenhenger for å ta vare på naturmangfoldet. *NINA Rapport / Norsk Institutt for Naturforskning*, *1410*.
Framstad, E., Halvorsen, R., Storaunet, K. O., & og Sverdrup-Thygeson, A. (n.d.). Kriterier for naturverdi i skog. *NINA Rapport / Norsk Institutt for Naturforskning*, *1447*.
Gini, C. (1921). Measurement of Inequality of Incomes. *The Economic Journal of Nepal*, *31*(121), 124–126.
Gizachew, B., Solberg, S., Næsset, E., Gobakken, T., Bollandsås, O. M., Breidenbach, J., … Mauya, E. W. (2016). Mapping and estimating the total living biomass and carbon in low-biomass woodlands using Landsat 8 CDR data. *Carbon Balance and Management*, *11*(1), 33.
Haala, N. (2014). Dense image matching. *EuroSDR Publication Series, Official Publication*, (64), 115–145.
Hirschmuller, H. (2008). Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 328–341.
Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing*, *8*(3), 166.
Landbruksdirektoratet. (2017). Veileder for kartlegging av MiS-livsmiljøer etter NiN. *VEILEDER VERSJON 1.0, JUNI 2017*, 26.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.
Lexerød, N. L., & Eid, T. (2006). An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management*, *222*(1), 17–28.
Maltamo, M., Kallio, E., Bollandsås, O. M., Næsset, E., Gobakken, T., & Pesonen, A. (2014). Assessing Dead Wood by Airborne Laser Scanning. In *Forestry Applications of Airborne Laser Scanning* (pp. 375–395). Springer Netherlands.
McRoberts, R. E. (2010). Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sensing of Environment*, *114*(5), 1017–1025.
Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research / Issued Bimonthly by the Nordic Forest Research Cooperation Committee* , *19*, 164–179.
Næsset, E. (2009). Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, *113*, 148–159.
Ørka, H. O., Næsset, E., & Bollandsås, O. M. (2010). Effects of different sensors and leaf-on

and leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning. *Remote Sensing of Environment*, *114*(7), 1445–1461.
Ørka, H. O., Sverdrup-Thygeson, A., Næsset, E., & Gobakken, T. (2014). Mapping old natural forest habitat using airborne laser scanning. *RSGIS4HQ Conference, Vienna*.
Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., & Astrup, R. (2017). Digital aerial photogrammetry can efficiently support large-area forest inventories in Norway. *Forestry*, 1–9.
Rolstad, J., Framstad, E., Gundersen, V., & Storaunet, K. O. (2002). Naturskog i Norge - Definisjoner, økologi og bruk i norsk skog-og miljøforvatltning. *Aktuelt Fra Skogforskningen*, *1*, 1–53.
Rolstad, J., & Storaunet, K. O. (2015). VEDLEVENDE RØDLISTE-SOPPER OG NORSK SKOGBRUK. *Oppdragsrapport Fra Skog Og Landskap*, *5*, 1–38.
Rothermel, M., Wenzel, K., Fritsch, D., & Haala, N. (2012). Sure: Photogrammetric surface reconstruction from imagery. In *Proceedings LC3D Workshop, Berlin* (Vol. 8). ifp.uni-stuttgart.de. Retrieved from http://www.ifp.uni-stuttgart.de/publications/2012/Rothermel_etal_lc3d.pdf
Santoro, M., Pantze, A., Fransson, J. E. S., Dahlgren, J., & Persson, A. (2012). Nation-Wide Clear-Cut Mapping in Sweden Using ALOS PALSAR Strip Images. *Remote Sensing*, *4*(6), 1693–1715.
Solberg, S., Bjørkelo, K., Gjertsen, A. K., May, J., Pierzchala, M., Timmermann, V., & Solheim, H. (2014). Satellittkartlegging av hogst og skader. *Oppdragsrapport Fra Skog Og Landskap*, *6*. Retrieved from http://www.skogoglandskap.no/filearchive/oppdragsrapport_06-2014_satellittkartlegging_av_hogst_og_skogskader.pdf
Sverdrup-Thygeson, A., Ørka, H. O., Gobakken, T., & Næsset, E. (2016). Can airborne laser scanning assist in mapping and monitoring natural forests? *Forest Ecology and Management*, *369*, 116–125.
Zhu, Z., Wang, S., & Woodcock, C. E. (2015). Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, *159*, 269–277.
Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, *118*(Supplement C), 83–94.