



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2017 30 stp
Fakultet for Miljøvitenskap og naturforvaltning (MINA)

Estimering av sagtømmerandel ved hjelp av flybåren laserskanning

Estimating sawlog proportions using Airborne laser
scanning

Tallak Dieset
Skogfag

Forord

Denne masteroppgaven er siste ledd i skogbruksutdanningen min ved Norges Miljø- og Biovitenskapelige Universitet(NMBU). Oppgaven er skrevet ved Fakultet for Miljøvitenskap og naturforvaltning (MINA) og er en del av et samarbeidsprosjekt mellom NMBU, NIBIO, Viken Skog SA og Digpilot. Oppgaven baserer seg på produksjonsdata fra en hogstmaskin påmontert ekstra posisjoneringssystem, som har registrert data siden sommeren 2016.

Jeg vil benytte anledningen til å takke min veileder Knut Marius Hauglin for all den hjelp jeg har mottatt under skrivingen av denne masteroppgaven. Det har vært utrolig lærerikt for meg!

Jeg vil også takke Viken Skog SA og Røsåsen skogsmaskiner AS som har stått for avviklingen og gitt tilgang til felldata gjennom deres produksjonsfiler.

Tusen takk til Digpilot som har stått for teknikken i det nye posisjoneringssystemet som er benyttet i denne oppgaven.

Tusen takk til Norges Forskningsråd som har stått for finansieringen av datainnsamling, gjennom prosjektet «Sustainable Utilization of Forest Resources in Norway».

Til slutt ønsker jeg å takke alle som har bidratt og hjulpet meg på en eller annen måte gjennom masterskrivingen, spesielt ønsker jeg å takke Phd-student Solrun Karlsen Lie for korrekturlesing av oppgaven.

Norges miljø- og biovitenskapelige universitet.

Ås, 12. mai 2017

Tallak Dieset

Sammendrag

Flybåren laserskanning er den mest brukte metoden for å kartlegge skogressurser i dagens skogbruk i Norge. For å kunne drive et rasjonelt skogbruk er man avhengig av å vite mest mulig om de ressurser som står tilgjengelig. Detaljert beskrivelse av volum, middel-høyde, antall trær og grunnflatesum er eksempler på informasjon estimert med flybåren laserskanning. Det kreves fortsatt manuelle prøveflatetakseringer i felt for å kalibrere laserdataene, dette er både tidkrevende og dyrt. Det finnes lite informasjon om tømmerkvalitet eller et estimat for sagtømmer i skogbruksplaner. Marginene i skogbruket blir stadig lavere, og behovet for mer presis informasjon om hvilke ressurser som er tilgjengelig er viktig for å kunne utnytte dem på best mulig måte.

I denne studien har det blitt gjort et forsøk på å modellere sagtømmerandel og klassifisering av andelen sagtømmer ved hjelp av flybåren laserskanning. Modellene ble laget og testet med feltdata i form av produksjonsfiler som referansedata til flybåren laserskannerdata. Feltdata ble innhentet av en hogstmaskin i kommersiell drift. Hogstmaskinen ble utstyrt med ekstra posisjoneringsutstyr som logget nøyaktig posisjon til hogstmaskinaggregatet til enhver tid, ved felling ble posisjonen lagt inn som en egen parameter i produksjonsfilen. Feltdata ble delt opp i virtuelle prøveflater på 100m², 200m² og 400m², og sagtømmerandelene for prøveflatene ble beregnet. Det ble brukt laserdata med forskjellig punkttetthet, for å se om dette ville gjøre utslag i modellen. Det ble brukt to laserdatasett med 0.7 punkter pr. m² og 2 punkter pr. m². Random Forest ble brukt til å utvikle modeller for estimering og klassifisering av sagtømmerandel, modellene utviklet med Random Forest ble deretter kryssvalidert. Modellen klarte å estimere sagtømmerandel med en RMSE (Root mean square error) på 0.14 og RMSE% på 29.4. RMSE var generelt lavest for de modellene som ble kjørt på 400m² prøveflater. Det ble påvist stor forskjell i forhold til prøveflatestørrelse i modellene, der de mindre prøveflatene ga liten eller ingen sammenheng mellom sagtømmerandel og laserdata. Det var liten forskjell i forhold til punkttettheten i laserdata med 0.7pkt/m² og 2pkt/m². Det ble forsøkt å klassifisere datasettet i to klasser; under eller over 50% sagtømmer. Ved klassifisering med klasser på over eller under 50% sagtømmer viste den beste modellen en gjennomsnittsfel på 28.8%. Det ble funnet en moderat sammenheng mellom sagtømmerkvalitet og laserdata. Modellene kan hjelpe en beslutningstaker i skogbruket med å prioritere riktig bestand for avvirkning.

Nøkkelord: Flybåren laserskanning, laserdata, produksjonsfiler, sagtømmerandel, random forest

Abstract

Airborne laser scanning is the most commonly used method for mapping forest resources in today's forestry in Norway. In order to have a rational forestry, we are dependent on knowing the resources available. Detailed description of volume, mean height, number of trees and basal area are examples of information estimated with airborne laser scanning. There are still required manual sample plot inventories in the field to calibrate laser data, this is both time consuming and expensive. There is little information about timber quality or an estimate of sawlog volume in forest inventories. Forestry margins are getting lower and the need for more precise information about what resources are available is important in order to exploit them in the best possible way.

In this study, an attempt has been made to model the sawlog proportion and classify the proportion of logs by means of airborne laser scanning. The models were created and tested with field data in the form of production files as reference data for airborne laser scanner data. Field data was obtained by a harvester in a commercial operation. The harvester was equipped with additional positioning equipment which logged the exact position of the harvester head at any time. By felling a tree, the position was entered as a separate parameter in the production file. Field data was divided into virtual test areas of 100m², 200m² and 400m² plots, and the sawlog proportions for the test plots were calculated. Laser data with different density was used to see if this would change result in the model. Two laser data sets were used with 0.7 points per m² and 2 points per m². Random Forest was used to develop models for estimation and classification of the sawlog share, models developed with Random Forest were then cross validated. The model was able to estimate the sawlog share with an RMSE (Root mean square error) of 0.14 and RMSE% of 29.4. RMSE was generally the lowest for the models that were run on 400m² plots. A significant difference was observed in relation to the sample plot size in the models, where the smaller plots leave little or no correlation between sawlog proportions and laser data. There was a slight difference in the density of laser data with 0.7pkt/m² and 2pkt/m². There was an attempt to classify the data set into two classes; Below or above 50% sawlog. When classified with classes above 50% or less, the best model showed an average error of 28.8%. A moderate relationship between sawlog quality and laser data was found. The models can help a decision maker in the forestry to prioritize the correct stand for harvest.

Keywords: Airborne laser scanning, laser data, production files, sawlog proportion, random forest

Innholdsfortegnelse

Forord	I
Sammendrag	III
Abstract	V
Innholdsfortegnelse	VII
Introduksjon	1
Skogbruksplaner og bruk av fjernmåling.....	1
Flybåren laserskanning.....	2
Bruk av fjernmåling til modellering.....	2
Tømmerkvalitet i industrien og verdikjedens behov for ressursoversikt	4
Material og metode.....	5
Forskningsområdet.....	5
StanForD 2010.....	5
Datainnsamling.....	6
Laserdata	9
Laservariable	10
Databehandling	11
Datasett	14
Random Forest	15
Dataanalyse	16
Regresjonsmodell	16
Klassifisering.....	16
Kryssvalidering.....	17
Resultater	18
Regresjon.....	18
Klassifisering.....	23
400m ² prøveflater	23
200m ² prøveflater	24
100m ² prøveflater	25
Diskusjon	26
Konklusjon	29
Litteraturliste.....	30

Introduksjon

Målet med denne studien er å utforske mulighetene for å predikere og klassifisere tømmerkvalitet med Flybåren laserskanning (FLS) og hogstmaskindata som referansedata.

Skogbruksplaner og bruk av fjernmåling

Skogbruksplaner er et verktøy som skal bistå en forvalter når beslutninger skal tas, når et bestand skal hogges eller om det skal iverksettes tynning (Eid et al. 2004). Skogbruksplaner utarbeidet ved hjelp av flybåren laserskanning gir bestandsinformasjon som er i samsvar med kravene som blir stilt av skogforvaltere og skogplanleggere (Næsset 2007). Flybåren laserskanning som metode er økonomisk konkurransedyktig, da de totale kostnadene er funnet relativt lave i forhold til andre metoder (Eid et al. 2004). Den laser-baserte metoden gir bedre estimater for grunnflate, dominerende trehøyde og antall trær pr. hektar sammenlignet med fototolkning (Eid et al. 2004).

Tømmerkvalitet som en parameter i skogbruksplanen vil gi skogeierne større fleksibilitet ved slutthogst. Det kan for eksempel gjøre det enklere å gjøre gode prioriteringer av hvilke bestand som skal avvirkes (Bollandsås et al. 2011). At skogbruksplanen inneholder relevant informasjon er avgjørende for at en kan ta en så riktig beslutning som mulig (Eid et al. 2004). En feil avgjørelse kan resultere i et tap av nåverdi, for eksempel kan man tape betydelige summer hvis man på bakgrunn av feilinformasjon avvirker et bestand for tidlig eller for sent (Eid et al. 2004).

Det er gjort kostnads- og tapsanalyser av forskjellige takseringsmetoder for å finne ut hvor mye ressurser det er optimalt å legge inn i en skogbrukstakst. En kostnads- og tapsanalyse har som mål å minimere de totale kostnadene, kostnaden for målingene legges sammen med kostnaden av en feil beslutning ved ikke å ha den informasjonen tilgjengelig (Eid et al. 2004). Summerer man opp de totale kostnadene (takstkostnaden og nåverditapet) for laserskanning og fototolkning ser man at den totale kostnaden for fototolkning er dobbelt så høy som laserskanning (Eid et al. 2004). Det er derfor interessant å utforske andre områder som laserdata kan anvendes på da dette ikke vil føre til en merkostnad i en skogbrukstaksering. Hvis det er mulig å bruke FLS data til å predikere tømmerkvalitet vil kostnadene ved dette være veldig lave i forhold til andre metoder da vi allerede bruker FLS data til å predikere blant annet volum (Næsset 2004)

Flybåren laserskanning

Flybåren laserskanning er en metode for fjernmåling av skog. Teknologien som brukes heter LiDAR (light detection and ranging), innhenting av LiDAR-data blir ofte omtalt som flybåren laserskanning (Vauhkonen et al. 2014). LiDAR er en teknologi som sender ut laserpulser og registrerer tiden laserpulsen bruker på å treffe et objekt og reflektere pulsen tilbake til en mottaker (Vauhkonen et al. 2014). Ved å utstyre flyet med global navigation satellite system (GNSS) kan man regne ut nøyaktig posisjon x,y,z til det punktet som reflekterer laserpulsen (Bollandsås et al. 2011). Tidsforbruket laserpulsen brukte fra den ble sendt ut til den traff et objekt og reflektert tilbake til mottaker blir registrert. Posisjonen til punktet kan da bli beregnet ved å bruke posisjonen til sender (flyets GNSS posisjon), tidsforbruket og vinkelen pulsen ble sendt ut med (Bollandsås et al. 2011). Ved skogbrukstaksering brukes det en skanningsmekanisme som sprer laserpulsene i en stripe under flyet. Skanningsmekanismen består som regel av et roterende eller oscillerende speil (Nordkvist et al. 2013). Skanningsvinkelen og flyhøyden bestemmer hvor bred stripen blir, ved å skanne flere striper inntil hverandre får man dekket større området (Nordkvist et al. 2013). Flyhastigheten og stripebredden bestemmer hvor mange laserpulser det blir pr. m²(punkttetthet) (Nordkvist et al. 2013). Det blir gjort en bakkeklassifisering av laserpulsene. Laserpulsene har koordinatene og avstanden i forhold til mottakeren på flyet, disse blir beregnet slik at laserpulsene blir referert i forhold til bakkenivå. Høyde over bakken for alle laserpulser kan da bli beregnet. Laserpulsene blir delt inn i forskjellige klasser i forhold til hvor de treffer; bakke/terreng, snø, støy, bru, eller uklassifisert. Laserpulsene kan brukes til å lage såkalte laservariable som sier noe om skogegenskapene (Holmgren 2004). Laservariablene forklarer hvordan punktskyen ser ut, Høydedistribusjon av laserpulsene og tetthetsdistribusjonen av laserpulsene er eksempler på laservariable (Næsset 2002). Laserpulsene kan også brukes til å lage en 3D-modell av terrenget (Bollandsås et al. 2011).

Bruk av fjernmåling til modellering

Det har blitt gjennomført lignende studier der det er forsøkt å predikere sagtømmervolum og sagtømmerandel med laserdata, ingen av de tidligere studiene har i motsetning til denne studien hatt nøyaktige feltdata, noe som kan ha påvirket deres resultater. Bollandsås et al. (2011) brukte laservariable til å modellere sagtømmervolum og sagtømmerandel med faktisk sagtømmervolum fra hogstmaskindata som referansedata, de konkluderer med at nøyaktigheten i deres studie viser en generell anvendbarhet av denne type laser-basert fremgangsmåte. Dette må undersøkes nærmere da de tror tilleggsinformasjon som nøyaktig treposisjon ville kunne gi bedre resultater (Bollandsås et al. 2011). Peuhkurinen et al. (2008) brukte laserdata med sagtømmerandel som referansedata til å predikere sagtømmerandel for homogene bestand. De (Peuhkurinen et al. 2008) brukte en ikke-

parametrisk metode kalt k-nærmeste nabo (k-nearest neighbor(k-NN)) med en databank av prøveflater. For å estimere sagtømmerandel for en prøveflate søkte de (Peuhkurinen et al. 2008) på høyde-diameter distribusjonen til prøveflaten i databanken, og den som lignet mest ble brukt som sagtømmerandel for prøveflaten (Peuhkurinen et al. 2008).

I dagens skogbrukstaksering med flybåren laserskanning trenger man fortsatt å bruke georefererte prøveflate for å kunne utvikle modeller fra FLS-data (Næsset 2007). I fremtiden kan det være mulig å bruke tredata fra hogstmaskiner som referansedata til å kunne predikere tre-egenskaper med større nøyaktighet (Lindroos et al. 2015). Dette avhenger av at det utvikles metoder og teknologi for å kunne posisjonere hvert felte tre med en nøyaktighet på under én meter (Lindroos et al. 2015).

Dagens moderne hogstmaskiner er utstyrt med GNSS teknologi for å loggføre hvor maskinen har operert, og som hjelpemiddel for hogstmaskinfører for å finne ut hvor i terrenget han befinner seg. Posisjonen til hogstmaskinen blir lagret, men per i dag er det ikke teknologi til å loggføre nøyaktig posisjon for hvert felte tre standardutstyr. Det er tidligere gjort forsøk på å bruke posisjonen til hogstmaskinen som tre-posisjon sammen med FLS-data (Olivera et al. 2016), men dette blir for unøyaktig til å fungere som referanse data sammen med FLS. Treets relative posisjon i forhold til GNSS mottakeren på hogstmaskinen kan variere veldig, og informasjon om retning fra hogstmaskinen, kranlengden og vinklene på krana er ikke med. Lindroos et al. (2015) beregner denne nøyaktigheten til å være rundt 10-20 meter, da er feilmarginen til GNSS mottakeren og posisjonen til hogstaggregatet relatert til GNSS mottakeren medregnet.

I 2016 ble det satt i gang et prosjekt hvor forskjellige metoder for estimering av posisjonen til hogstmaskinaggregat skulle testes (Hauglin et al. 2017).

Det ble testet to systemer:

1. En vanlig GNSS mottaker ble montert direkte på hogstaggregatet.
2. Et lignende system som brukes i dag på gravemaskiner ble montert på hogstmaskinaggregatet.

System nr. 2 ga den mest nøyaktige posisjonen til hogstmaskinaggregatet, avleses posisjonen i det øyeblikket treet felles kan man bruke denne posisjonen som treets faktiske posisjon med en gjennomsnittlig nøyaktighet på 0.94 meter (Hauglin et al. 2017). Blir dette standard system på alle hogstmaskiner vil man kunne innhente mye informasjon og bruke det som referansedata for å utvikle nye modeller generert av FLS-data.

Tømmerkvalitet i industrien og verdikjedens behov for ressursoversikt

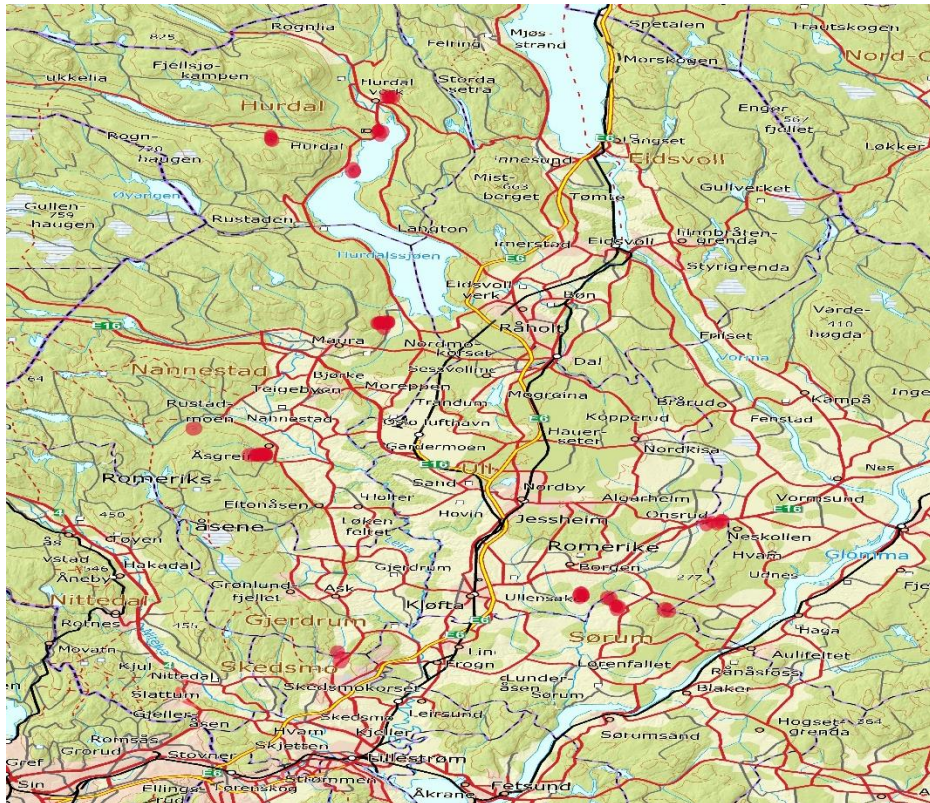
Tømmermarkedet er globalt. Hovedutfordringen er de store transportkostnadene for tømmeret. I Norge har flere store massevirkefabrikker blitt utkonkurrert eller lagt ned de siste årene. En betydelig andel av avvirkningen blir eksportert til utlandet for videreforedling. Sagbruksindustrien har dermed et økende behov for presise leveranser av tømmer for å kunne optimalisere sin produksjon (Forsberg 2003). For å få en optimal virkesstrøm er det noen problemer så må løses. I en vanlig avvirkning er det flere ulike produkter som blir produsert, og de ulike produktene skal vanligvis til forskjellige kunder. Å kunne estimere volum av forskjellige tømmerprodukter før avvirkning er en av hovedfaktorene til å kunne optimalisere virkesstrømmer (Barth & Holmgren 2013).

Jeg ønsket med denne studien å undersøke om det er mulig å predikere andelen av sagtømmer ved bruk av FLS-data. Sekundært ønsket jeg å undersøke hvordan bruk av forskjellige punkttetthet i laserdata og størrelsen på prøveflatene påvirket prediksjonene.

Material og metode

Forskningsområdet

Denne studien ble utført i Øvre Romerike (Figur 1). Skogen i Øvre Romerike er boreal og domineres av gran (*Picea Abies*), furu (*Pinus Sylvestris*) og bjørk (*Betula Pubescens*).



Figur 1: Studieområdet: Kartet viser lokaliseringen til de driftene som datasettet er hentet fra (drifter markert som røde punkter).

StanForD 2010

I denne studien er det brukt data i form av produksjonsfiler, laget med formatet til StanForD 2010, fra en hogstmaskin i kommersiell drift. StanForD 2010 er en dataprotokoll for hogstmaskiner og treprosesserere. StanForD står for «Standard for Forest Machine Data» og er utviklet av Skogforsk i Sverige (Olivera & Visser 2014). Den skal være en standard for utveksling av data til og fra hogstmaskiner. Standarden er global og brukes av alle de store maskinprodusentene for hogstmaskiner. StanForD 2010 er den oppdaterte StanForD-standard som kom på 1980-tallet. I StanForD 2010 brukes det en indeksing for å gi en identitet til hver maskin, avvirkningsobjekt, trær og stokker m.m. Systemet gir en teoretisk mulighet til å kunne spore hver eneste stokk som blir

produsert globalt. Informasjon om hver stokk blir lagret i «harvester production files» (HPR-filer). Dette gir muligheter for fleksibel styring av avvirkning, en har da mulighet til å endre produksjon ut fra hvilke produkter er produsert og som blir etterspurt, ved å endre apteringsfilene (Arlinger et al. 2012).

StanForD benytter Extensible Markup Language (XML). En XML-fil er delt opp i forskjellige lagringsenheter som inneholder enten analysert data eller ubearbeidede data (Bray et al. 1998). I utgangspunktet kan en XML-fil sammenlignes med en database, et verktøy for lagring av informasjon. Fordelen med at StanForD benytter seg av denne typen lagringsenhet er at dette er en enkel form for datalagring. Det er mulig å hente ut filer fra hogstmaskinene, som for eksempel produksjonsfiler(HPR-filer) og kvalitetsfiler, og man kan sende filer til hogstmaskinene som apteringsfiler, objektfiler (informasjon om oppdraget) eller geografiske filer som for eksempel kart over området (Skogforsk 2007). Siden XML er et veletablert format, slipper man å konvertere filene når data fra StanForD2010 skal brukes videre i andre systemer, noe som gjør systemutvikling enklere (Arlinger et al. 2012). For mer informasjon om XML filer og dets oppbygning se Bray et al. (1998).

Datainnsamling

Datainnsamling ble utført gjennom et samarbeidsprosjekt mellom NMBU, NIBIO, Viken Skog AS og Digipilot. Finansieringen av datainnsamlingen for dette samarbeidsprosjektet er finansiert av Norges Forskningsråd, med tittel «Sustainable Utilization of Forest Resources in Norway» med prosjektnummer 225329/E40. Fra en hogstmaskin (John Deere 1270E) i normal drift ble det samlet inn enkelttredata fra Juni 2016 til Mars 2017. Dataene ble lagret i formatet til StanForD2010 (Skogforsk 2007), for hver enkelt skogeier. Hvert enkelt felte tre ble registrert med en unik tre-ID, antall stokker, sortiment, volum per stokk, diameter i brysthøyde, lengde per stokk og om apteringssystemet satte sortiment automatisk eller om hogstmaskinføreren overstyrte systemet og satte sortiment manuelt ut fra andre hensyn. Det ble påmontert et nytt posisjoneringssystem kalt real time kinematic (RTK) basert integrert posisjoneringssystem, for å kunne måle nøyaktig posisjon til hogstaggregatet når hvert tre ble kappet. Det ble brukt to Topcon Hiper SR RTK-GNSS mottakere med eksterne antenner (Figur 2), fem trådløse vinkelsensorer(Figur 3) og to sensorer montert på hydraulikksylindere, se beskrivelse av Hauglin et al. (2017). GNSS-antennene var koblet til en mottaker som mottok korreksjonsdata i sanntid fra Kartverkets nett av basestasjoner via mobilnettet (Kartverket 2015). Antennene ble montert bak på hogstmaskinen og sensorene ble montert på kranen og på selve hogstmaskinkroppen (Figur 4). I tillegg til informasjon fra de monterte sensorene, ble informasjon fra en original sensor som registrerte kran-rotasjonen brukt (Hauglin et al. 2017).

Data fra hver enkelt sensor og RTK-GNSS mottakerne ble kontinuerlig tatt imot av en sekundær-datamaskin inne i hogstmaskinen. En kan med nøyaktige vinkler og sylinder-posisjon regne ut hvor langt det er fra GNSS mottakerne til hogstaggregatet. Setter man sammen informasjonen fra alle disse sensorene kan man regne seg frem til en nøyaktig posisjon for hogstaggregatet i forhold til de to GNSS antennene (Hauglin et al. 2017). Nøyaktig posisjon til hogstaggregatet blir sendt kontinuerlig til hoved-datamaskinen i hogstmaskinen. Da treet ble felt, ble dermed den nøyaktige posisjonen registrert i produksjonsfilene (Hauglin et al. 2017).

Nøyaktighetene på sensorene som ble montert varierte fra type sensorer. De trådløse vinkelsensorene var veldig nøyaktige; da de ble testet på gravemaskiner hadde de en nøyaktighet på < 5 cm for flytting av graveskuffen (testen ble utført av forhandler). De to hydrauliske sensorene hadde en nøyaktighet på <1cm ifølge spesifikasjonene til produktet. Den originale rotasjons-sensoren fra John Deere hadde en oppløsning på 1° (Hauglin et al. 2017).



Figur 2: John Deere 1270E med ekstra GNSS antenner



Figur 3: En trådløs vinkelsensor fra DigPilot.



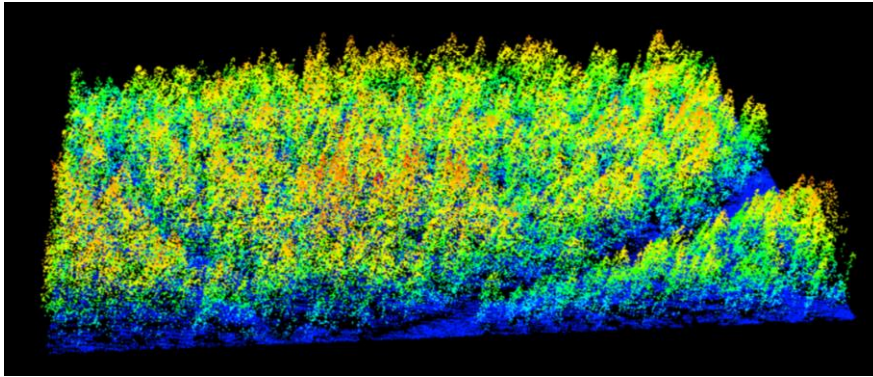
Figur 4: Plassering av sensorer montert på hogstmaskin, hvite sirkler viser GNSS antenner, blå ringer viser hydrauliske sylindere og røde ringer viser vinkelsensorer. Kilde: (Hauglin et al. 2017)

Laserdata

Laserdata ble hentet fra nettstedet www.hoydedata.no, hvor laserdata for store deler av hele Norge er tilgjengelig. I vårt studieområde finnes det flere laser-prosjekter med ulik punkttetthet. I denne oppgaven er det brukt laserdata fra to prosjekter ved navn LACHA033 og LACH0001. Begge prosjektene er utført av et kommersielt firma, Terratec AS, på oppdrag fra Statens Kartverk. LACHA033 ble skannet i 2013 med laserskanner ALS70, med åpningsvinkel ± 16 grader. Høyde over laveste terreng var 3124 meter med flyhastighet på 150 km/t. Minimumskravet til punkttetthet var $0,7\text{pkt/m}^2$.

LACH0001 ble skannet i 2015 med flere forskjellige laserskannere: Leica ALS70 HP, Leica ALS70 HM og Optech ALTM Titan MV. Skannevinkel var maksimum ± 20 grader fra loddlinjen. Punkttetthet ble satt til minimum 2pkt/m^2 , og i utvalgte områder ble det skannet med minimum 5pkt/m^2 . Skanningen

med forskjellig punkttetthet ble utført i forskjellige skanningsblokker (Figur 5). I denne oppgaven er det kun brukt laserdata fra skanningsblokken med minimum 2pkt/m².



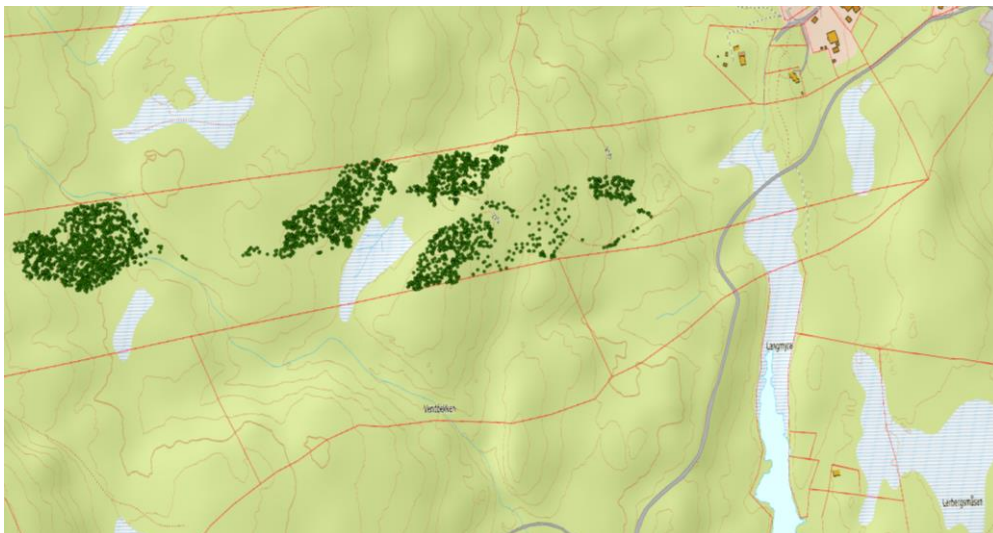
Figur 5: Skanningsblokk, eksempel på hvor detaljert en punktsky kan være. Her er minimumstettheten 2pkt/m².

Laservariable

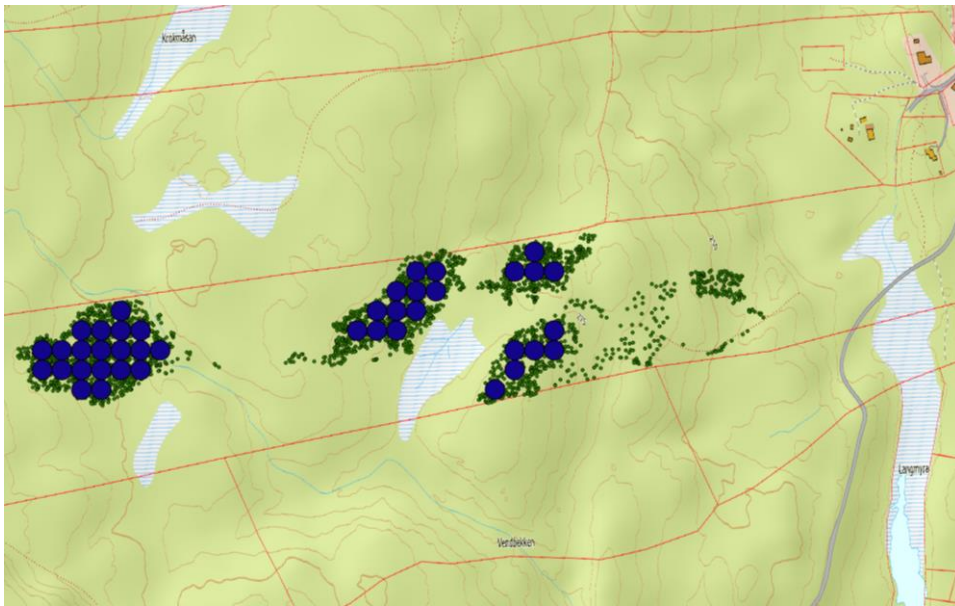
Laservariable som Hmax, Hmean, Hsd, Hcv, Hkurt, Hskewness, H10, H20, H30, H40, H50, H60, H70, H80, H90, D0, D1, D2, D3, D4, D5, D6, D7, D8, D9 ble beregnet for hver prøveflate. Hmax er laserekkoet registrert på det høyeste punktet på prøveflaten, Hmean er gjennomsnittlig ekko høyde, Hsd er standard avviket og Hcv er variasjonskoeffisienten. Hkurt og Hskewness er laservariable som beskriver høydefordelingen av laserekkoet (Evans et al. 2009). H10, H20,.....,H90 er høydelag og D0,D1,.....,D9 er tetthetsvariabler (Næsset 2002). H10-variabelen for eksempel betegner høyden over bakken hvor akkumuleringen av laserekko i vegetasjonen var 10%. Hvis verdien av H10 var 6m, var 10% av laserpulsene som lå innenfor prøveflaten under denne høyden. Ved H90 = 20m lå 90% av laserpulsene under 20meter over bakken. Tetthetsvariablene D0, D1, ..., D9 er andelen av laserekko som lå over høydelaget relativt til totalt antall ekko. For hver prøveflate ble høydeområdet med laserekko delt inn i lag med lik høyde, i dette tilfellet i 10 lag. Antall ekko som var over et lag ble talt opp og så delt på totalt antall ekkoer for prøveflaten. For eksempel, i D5 variabelen er det antall ekko som er over det femte høydelaget delt på totalt antall ekko (Bollandsås et al. 2011).

Databehandling

Feltdata ble innsamlet av en John Deere 1270E hogstmaskin og lagret i produksjonsfiler for hver enkelt skogeier. Som forklart ovenfor er det en sekundær-datamaskin som kontinuerlig regner ut nøyaktig posisjon til hogstmaskinaggregatet via det påmonterte RTK-baserte posisjoneringssystemet. Når hvert tre blir felt blir posisjonen hentet ut fra den sekundære-datamaskinen og sendt til hoved-datamaskinen som legger posisjonen til treet i produksjonsfilen. Posisjonen til hvert enkelt tre ble plottet inn i et GIS-system, i denne oppgaven ble QGIS benyttet, for å få en oversikt over hogstfeltet (Figur 6).



Figur 6: Enkelttre plottet på kart fra posisjonert hogstmaskindata. Hvert punkt representerer et felt tre.



Figur 7: Prøveflatene ble systematisk lagt ut på hogstfeltet, her i størrelse 400m² pr prøveflate.

Prøveflatene ble lagt ut virtuelt i tre størrelser: 100m², 200m² og 400m². På grunn av lite data ble prøveflatene lagt systematisk så tett som mulig (Figur 7). Hver prøveflate inneholdt en liste over alle trær som berørte prøveflaten, listen over trærne inneholdt produksjonsdata fra hogstmaskinen. Prøveflatene ble lagret i en shapefil (.shp) for hver flatestørrelse for at det senere kunne klippes ut laserdata fra eksakt samme område. Sagtømmer ble definert ut fra sortimentslisten for hver skogeier. Følgende sortimenter ble definert som sagtømmer (ProductKey i parentes): Spesial(3135), Sagt(3136), Vänerbl(3141) og Sagt(3142) (Tabell 1). Det er i denne oppgaven valgt å se på de sortimentene som har høyest verdi. Sagtømmervolum og sagtømmerandel i prosent kunne da regnes ut for alle prøveflater.

Tabell 1 Sortimentstabell: Alle sortimenter har et unikt produktnummer (Productkey).

ProductKey	ProductName
3135	Spesial
3136	Sagt
3137	Pall410
3138	Fritt
3139	Masse
3140	Vrak
3141	Vänerbl
3142	Sagt
3143	Pall410
3144	Fritt
3145	Masse
3146	Energi
3147	TorrVik
3148	Vrak
3149	Fritt
3150	BjoMasse
3151	BjorkVed
3152	Vrak
3153	Fritt
3154	BlandVed
3155	TorrPri
3156	BlandBio
3157	Kobling
3158	Vrak
2147483647	NA

Laserdataene kom i blokker over større områder. Laserdata fra hvert hogstfelt ble klippet ut i programmet Rstudio. Kode for dette ble gjort tilgjengelig for meg av veileder. Området som ble klippet ut kan beskrives i en shapefil. Shapefilene med prøveflater ble lest inn og alle laserekkoene som lå innenfor radiusen til en prøveflate ble klippet ut, vi fikk da ut laserekko for de samme prøveflatene som feltdatane ble delt opp i. Laservariablene ble beregnet for hver prøveflate samtidig som vi beregnet ekkohøyde relativt til bakken.

Datasett

Datsettene brukt i denne oppgaven inneholder data fra flere prosjekter og teknologier. Dette er på grunn av at hogstmaskinen har samlet inn feltdata fra et stort område. I tillegg har noe av utstyret blitt skiftet ut og erstattet, noe som har påvirket nøyaktigheten på dataene som er registrert. Jeg har derfor valgt å dele dataene opp i forskjellige datasett. Feltdata fra hogstmaskin er delt opp i «alldata» og «2017data», fordi det ble erstattet en sensor i januar 2017. Vi vet ikke når i 2016 sensoren ble defekt, derfor er det stor usikkerhet i disse dataene. Data fra 2017 er kontrollert to ganger. Ved første måling den 17.01.2017 ble det målt 20 trær med en gjennomsnittsfel på 0,46meter, andre måling ble utført den 07.03.2017 og viste en gjennomsnittsfel på 0,96meter for 10 trær. Det var noen områder som ikke hadde dekning for laserdata fra 2015 med 2pkt pr m². Det ble laget forskjellige datasett med laserdata med lavtetthet(LT) (0,7pkt pr m²) og middelstetthet(MT) (2pkt. pr m²). Jeg ønsket å se på om variasjon i prøveflatestørrelse ville påvirke resultatet, og lagde datasett med 100m², 200m² og 400m² prøveflater. Det ble i alt 12 forskjellige datasett, størrelsen på datasettene ble derfor forskjellige (Tabell 2). Det ble avvirket 33881 trær og prøveflatene ble lagt dit der hele bestand var avvirket (Tabell 3). Gjennomsnittlig antall trær per prøveflate lå på 6.8 - 7.3 trær for 100m², 13.4 – 16.0 trær for 200m² og 26.9 - 32.0 trær for 400m² prøveflater (Tabell 4). Gjennomsnittlig sagtømmerandel i datasettene lå mellom 41.09% og 48.26%(Tabell 5), mens antall prøveflater med over 50% sagtømmer varierte i forhold til størrelsen på datasettene (Tabell 6).

Tabell 2: Antall prøveflater i hvert datasett.

	MT_alldata	MT_2017data	LT_alldata	LT_2017data
100m ²	997	673	1504	707
200m ²	401	271	617	286
400m ²	143	97	218	101

Tabell 3: Totalt antall trær i hvert datasett.

	MT_alldata	MT_2017data	LT_alldata	LT_2017data
100m ²	7319	4564	11929	4941
200m ²	5832	3620	9894	3939
400m ²	4200	2605	6983	2780

Tabell 4: Gjennomsnittlig antall trær pr. hektar, gjennomsnittlig antall trær pr. prøveflate i parentes for hvert datasett

	MT_alldata	MT_2017data	LT_alldata	LT_2017data
100m ²	734.1 (7.3)	678.2 (6.8)	793.2 (7.9)	698.9 (7.0)
200m ²	727.2 (14.5)	667.9 (13.4)	801.8 (16.0)	688.6 (13.8)
400m ²	734.3 (29.4)	671.4 (26.9)	800.8 (32.0)	688.1 (27.5)

Tabell 5: Gjennomsnittlig sagtømmerandel pr. prøveflate i hvert datasett.

	MT_alldata	MT_2017data	LT_alldata	LT_2017data
100m ²	42.72 %	41.53 %	43.00 %	41.09 %
200m ²	46.88 %	46.48 %	46.35 %	45.98 %
400m ²	48.26 %	47.66 %	47.48 %	47.31 %

Tabell 6: Andel i % av prøveflater med mer enn 50% sagtømmer i hvert datasett.

	MT_alldata	MT_2017data	LT_alldata	LT_2017data
100m ²	40.0 %	39.1 %	40.4 %	38.2 %
200m ²	27.9 %	40.6 %	41.0 %	39.2 %
400m ²	44.1 %	39.2 %	45.0 %	37.6 %

Random Forest

Random Forest er en ikke-parametrisk metode som kan brukes til estimering og klassifisering. Metoden benytter seg av et stort antall klassifiserings- og regresjons-trær. Et regresjons-tre må ikke forveksles med et vanlig tre, det heter tre fordi man sender data gjennom flere noder der datasettet blir forsøkt delt opp med forskjellige regler. Når informasjonen har gått gjennom alle nodene vil strukturen kunne etterligne et tre. Hvert tre gir et estimering- eller klassifiseringsresultat og det endelige resultatet er beregnet som en middelværdi ved regresjon eller som en stemme i klassifiseringstilfellet (Nordkvist et al. 2013). Fordelen med Random Forest er at man kan bruke et stort antall uavhengige variabler, inkludert de som er korrelert med hverandre. Random Forest har vist seg å gi nesten like gode resultater som ren regresjon, uten å bygge modeller av forholdet mellom avhengige og uavhengige data (Nordkvist et al. 2013).

Vi har valgt å bruke Random Forest da denne metoden ikke er like sensitiv som lineær regresjon for uavhengige data. En annen fordel er at prediksjonene alltid vil være innenfor variasjonen som finnes i modelldataene. Prøveflatene er lagt så tett som mulig og inntil hverandre for at datasettet skulle bli tilstrekkelig stort, da vi ikke har ubegrenset med data fra hogstmaskinen.

Dataanalyse

Jeg brukte statistikkprogrammet R, versjon 3.3.2 til å kjøre de statistiske modellene. Det ble brukt to modeller; en regresjonsmodell og en klassifiseringsmodell. For regresjonsmodellen var responsvariablen observert sagtømmerandel og for klassifiseringsmodellen var responsvariablen en sagtømmerklasse. Alle laservariablene (Hmax, Hmean, Hsd, Hcv, Hkurt, Hskewness, H10, H20, H30, H40, H50, H60, H70, H80, H90, D0, D1, D2, D3, D4, D5, D6, D7, D8, D9) ble brukt som forklaringsvariable for både regresjons og klassifiseringsmodellene.

Regresjonsmodell

Regresjonsmodellen ble generert med Random Forest-metoden. Random Forest deler datasettet i to, der den ene delen blir brukt som et treningssett som Random Forest utvikler en modell med. Det andre settet blir brukt til å teste modellen; predikert sagtømmerandel blir sammenlignet med observert sagtømmerandel.

Størrelsen r^2 brukes ofte ved lineær regresjon for å si noe om hvor godt en regresjonsmodell passer til modelldataene. En lignende verdi kan regnes ut for en Random Forest modell, og i denne oppgaven er r^2 regnet ut for et sett med flater som

$$r^2 = \rho(\text{predikerte sagtømmerandeler}, \text{observerte sagtømmerandeler})^2$$

der ρ er pearsons korrelasjon.

Klassifisering

Det ble bruk Random Forests innebygde metode for klassifisering. For klassifisering brukte vi en gjennomsnittlig klassefeil for å sammenligne resultatene. Vi delte opp datasettet i to sagtømmerklasser, og prøvde å predikere hvilken klasse en prøveflate tilhørte. Alle prøveflater med over eller lik 50% sagtømmer fikk klassen «1» mens alle prøveflater med mindre enn 50% sagtømmer fikk klassen «0». Resultatet av klassifiseringen ble enten riktig eller feil. Vi fikk dermed en matrise på 2x2 der alle predikerte prøveflater var sammenlignet mot observerte prøveflater. Andelen som ble klassifisert feil ble delt på antall observasjoner. Dette tallet kalte vi «gjennomsnittlig klassefeil», tallet viser hvor god klassifikasjonen har vært.

Kryssvalidering

For å validere regresjons- og klassifiseringsmodellen brukte vi kryssvalidering. Ved kryssvalidering bruker man et datasett der man holder ute én og én prøveflate og lager en modell med Random Forest med det resterende datasettet (Kohavi 1995). Laservariablene for prøveflaten som ble holdt ute ble deretter testet i modellen, og modellen estimerte en sagtømmerandel eller en sagtømmerklasse for prøveflaten. Dette ble gjort for alle prøveflater i datasettet, slik at det ble laget en modell for hver prøveflate som den deretter ble testet i. Predikert sagtømmerandel blir sammenlignet med observert sagtømmerandel for regresjon og for klassifisering sammenlignes predikert klasse mot observert klasse. Modellene ble på den måten testet og det var mulig å kvantifisere hvor god sammenhengen mellom laserdataene og tømmerkvaliteten var.

RMSE (Root mean square error) er en gjennomsnittsfel og et sammenligningstall som sier noe om hvor godt modellen klarer å predikere. RMSE er kvadratroten til gjennomsnittet av de kvadrerte residualene, og viser forskjellen mellom predikert verdi og observert verdi over antall prøveflater (Bollandsås et al. 2011). RMSE% oppgis som prosent av gjennomsnittet av de observert verdiene.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - \hat{S}_i)^2}$$

$$RMSE\% = \left(\frac{RMSE}{\frac{\sum_{i=1}^n S_i}{n}} \right) * 100$$

\hat{S}_i = Gjennomsnitt av predikerte verdier.

S_i = Gjennomsnitt av observerte verdier.

n = antall prøveflater

Resultater

Regresjonsmodeller

Regresjonsmodellene viste en moderat sammenheng mellom laserdata og tømmerkvalitet (Tabell 7). Beste resultater ble oppnådd med 400m² store prøveflater og datasettet LT_alldata med en RMSE på 0.14 (Tabell 8), RMSE% ble 29.4 (Tabell 9). Generelt ga modellene best resultater med 400m² prøveflater, uansett kombinasjon mellom lasertetthet og hogstmaskindata. Gjennomsnittlig RMSE og RMSE% for de ulike datasettene med 400m² var henholdsvis 0.15 og 31.4. For 200m² prøveflater økte RMSE og RMSE% jevnt over for alle datasettene i forhold til 400m². Den modellen som ga best resultat for 200m² prøveflater, var MT_alldata med henholdsvis 0.16 og 35.0 for RMSE og RMSE%. Det skiller generelt veldig lite mellom de ulike datasettene for prøveflatene med størrelse 200m². Datasettene med prøveflater på 100m² skiller seg vesentlig ut med størst variasjon og de dårligste resultatene i forhold til RMSE og RMSE%. Datasettet med best resultat for 100m² prøveflater var MT_alldata, med en RMSE og RMSE% på 0.22 og 50.5. Det var en klar sammenheng mellom prøveflatestørrelse og RMSE og RMSE%. Sammenligner vi MT_alldata med MT_2017data ser vi at MT_alldata er jevnt over bedre for alle prøveflatestørrelser, det samme gjelder for lavpunktstetthet. LT_alldata har jevnt over bedre resultater enn LT_2017data (Tabell 8). Lav- og middelpunktstetthet hadde liten sammenheng, og hvis man sammenligner MT_alldata og LT_alldata skiller det få prosentpoeng. Det samme med MT_2017data mot LT_2017data, men med litt mer variasjon.

Tabell 7: r² viser hvor godt modelldataene passer i modellen.

	MT_alldata	MT_2017data	LT_alldata	LT_2017data
100m ²	0.14	0.14	0.10	0.07
200m ²	0.15	0.18	0.11	0.10
400m ²	0.14	0.21	0.17	0.15

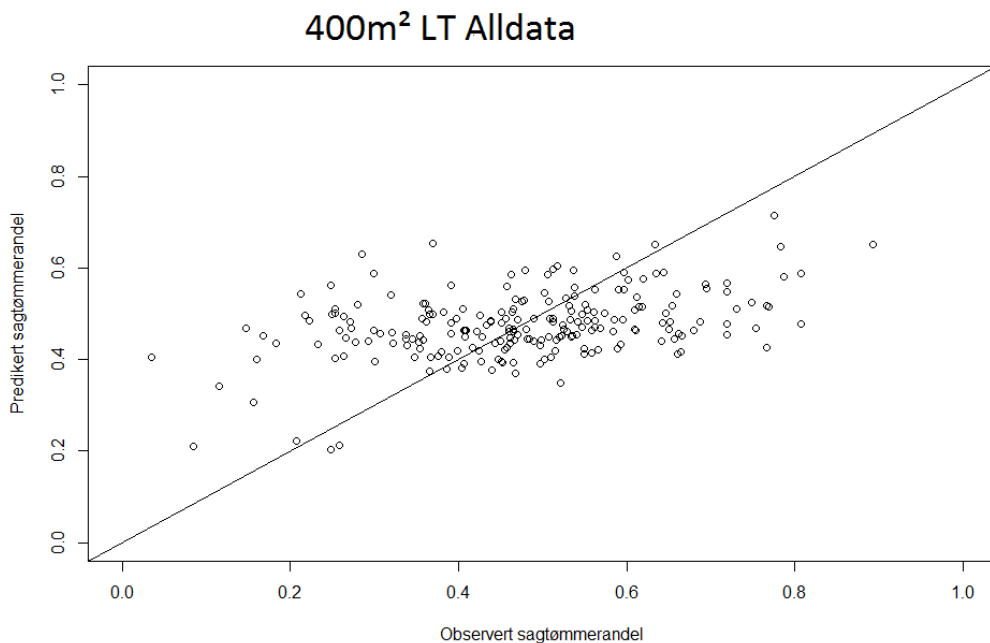
Tabell 8: RMSE viser forskjellen mellom predikert sagtømmerandel og observert sagtømmerandel for de ulike datasettene.

	MT_alldata	MT_2017data	LT_alldata	LT_2017data
100m ²	0.22	0.23	0.22	0.24
200m ²	0.16	0.17	0.17	0.18
400m ²	0.15	0.15	0.14	0.16

Tabell 9: Gjennomsnittlig avvik fra observert verdi, for alle datasett (RMSE%).

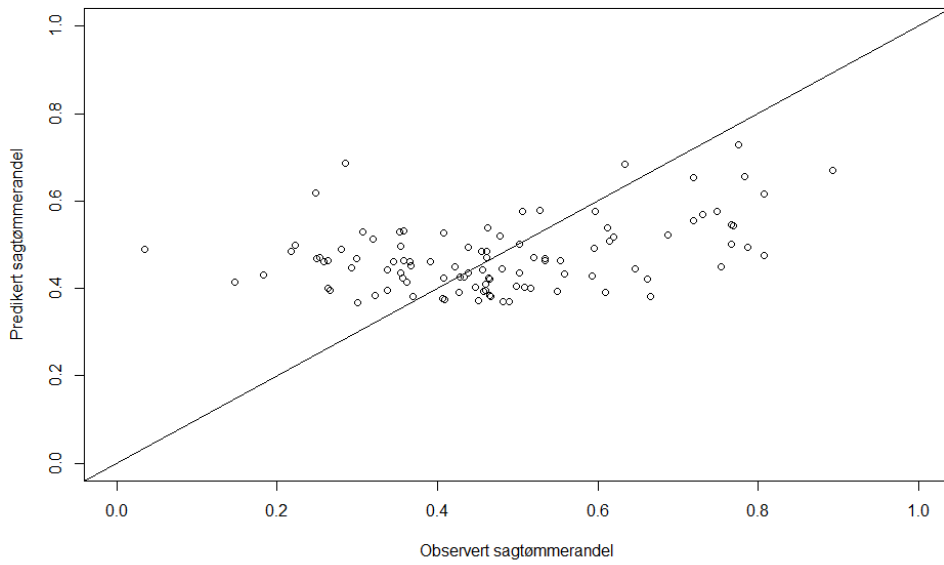
	MT_alldata	MT_2017data	LT_alldata	LT_2017data
100m ²	50.5	55.2	50.8	57.5
200m ²	35.0	37.2	35.9	38.3
400m ²	30.7	32.3	29.4	33.2

Sammenhengen mellom predikerte verdier basert på laserdata og observert sagtømmerandel var relativt svak (Figur 8). Vi ser at modellen overpredikerer ved lave observerte sagtømmerandeler og underpredikerer ved høy observert sagtømmerandel. For 400m² med LT_alldata kan vi se en ørliten tendens av at prediksjonen av sagtømmer følger regresjonslinjen (Figur 8). Resterende datasett viser at det er en svak sammenheng mellom sagtømmerandel og laserdata (Figur 9-14). En god regresjonsmodell vil ligge tettere og mer jevnt med regresjonslinjen.



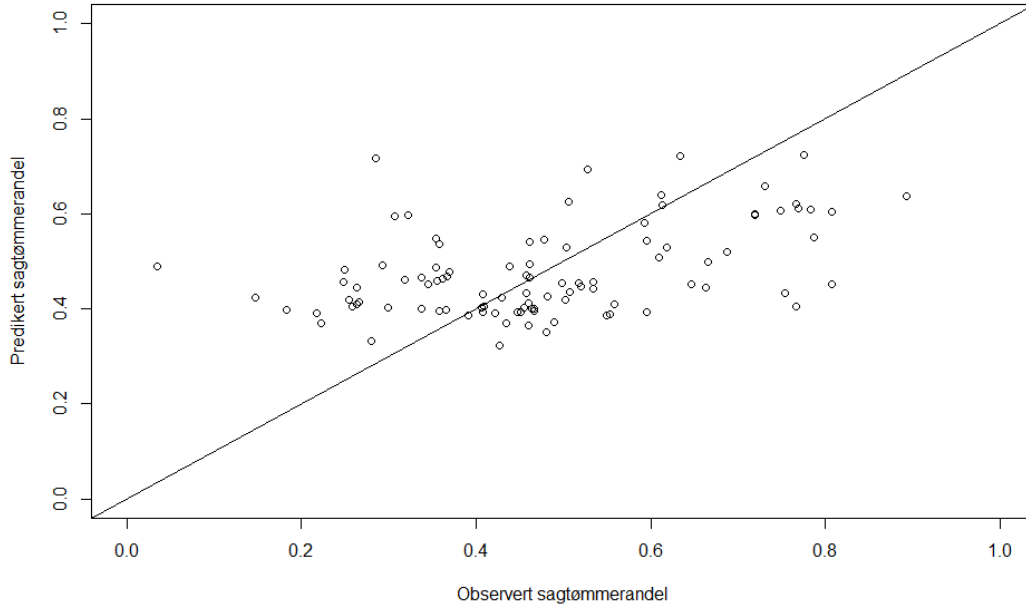
Figur 8: 400m² prøveflater Lavtethet(LT) alldata Observert sagtømmerandel mot predikert sagtømmerandel. RMSE% = 29,4

400m² LT 2017data

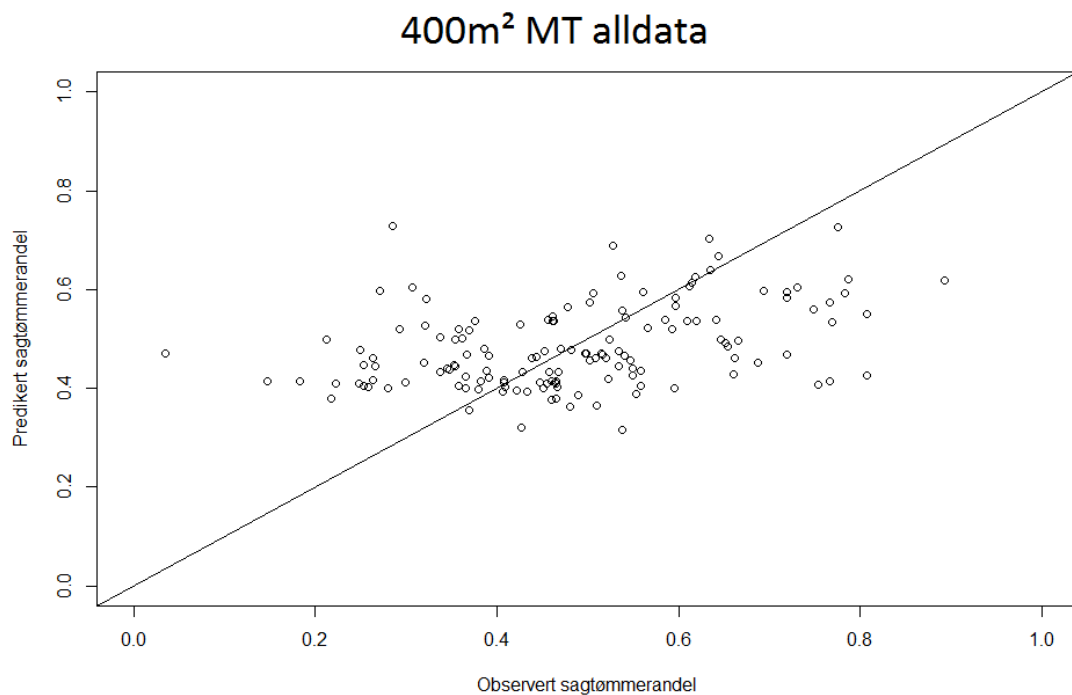


Figur 9: 400m² Lavtetthet(LT) 2017data. Observert sagtømmerandel mot predikert sagtømmerandel, RMSE% 33,2

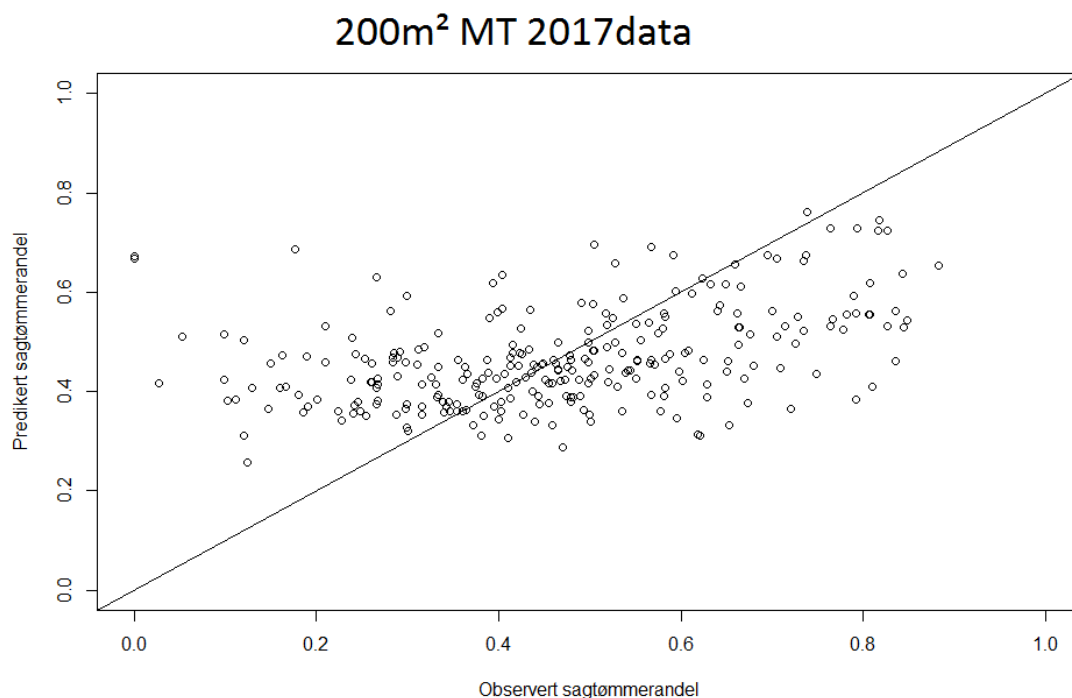
400m² MT 2017data



Figur 10: 400m² middelstetthet(MT) 2017data. Observert sagtømmerandel mot predikert sagtømmerandel, RMSE% 32,3

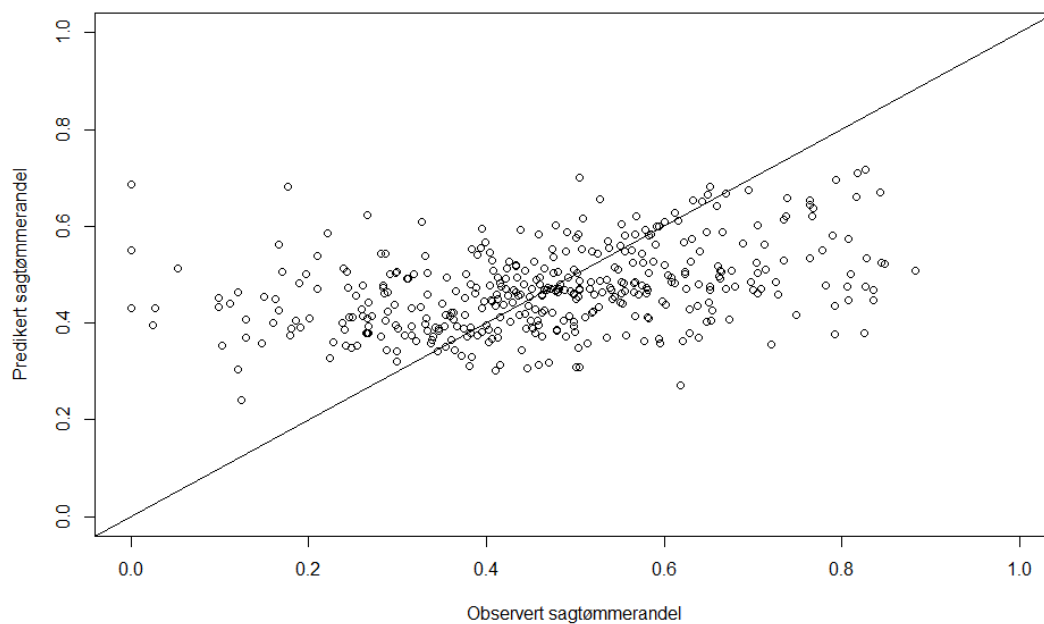


Figur 11: 400m² middelstetthet(MT) alldata. Observervert sagtømmerandel mot predikert sagtømmerandel, RMSE% 30,7



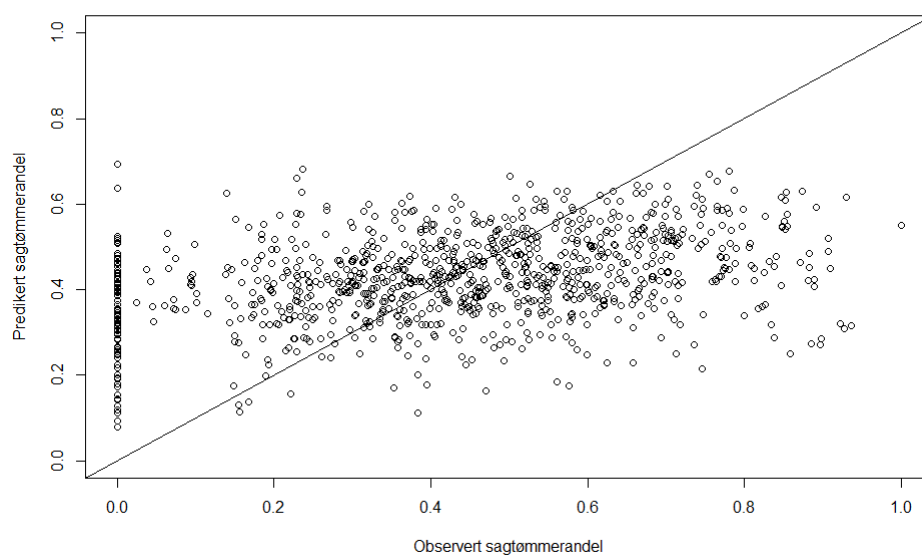
Figur 12: 200m² middelstetthet(MT) 2017data. Observervert sagtømmerandel mot predikert sagtømmerandel, RMSE% 37,2

200m² MT alldata



Figur 13: 200m² Middelstetthet(MT) alldata. Observert sagtømmerandel mot predikert sagtømmerandel, RMSE% 35,0

100m² MT alldata



Figur 14: 100m² middelstetthet(MT) alldata. Observert sagtømmerandel mot predikert sagtømmerandel, RMSE% 50,5

Klassifisering

Resultatet av klassifisering med kryssvalidering viste at datasettet med 200m² prøveflater og MT_2017data gav best resultat, med en gjennomsnittlig klassefeil på 28.8% (Tabell 10). For prøveflater med 400m² gav datasett MT_alldata best resultat, med en gjennomsnittlig klassefeil på 30.07%. På prøveflatene med 100m² var det datasettet med MT_2017data som ga best resultater med en gjennomsnittlig klassefeil på 33.88%. Datasettet med LT_alldata kom dårligst ut med en gjennomsnittlig klassefeil på 37.90%, 40.03% og 39.91% for henholdsvis 100m², 200m² og 400m² prøveflater. Størrelsen på prøveflatene forklarte mye av variasjonen, forskjellen mellom 100m² og 400m² var på mange prosentpoeng.

Tabell 10: Gjennomsnittlig klassefeil etter klassifisering med kryssvalidering av alle 12 datasett

	MT_alldata	MT_2017data	LT_alldata	LT_2017data
100m ²	35.41 %	33.88 %	37.90 %	38.19 %
200m ²	34.91 %	28.78 %	40.03 %	31.82 %
400m ²	30.07 %	30.93 %	39.91 %	32.67 %

400m² prøveflater

For 400m² prøveflater, LT_alldata ble 82 prøveflater klassifisert riktig i klassen 0 mens 38 prøveflater ble klassifisert feil i klasse 1, 49 prøveflater med over 50% sagtømmer ble klassifisert feil i klasse 0 og 49 prøveflater ble klassifisert riktig i klasse 1 (Tabell 11). Klassefeilen for de to klassene ble 31.66% og 50% for klassene 0 og 1. Gjennomsnittlig klassefeil ble 39.91%. Datasettet med LT_2017data forklarte forskjellen bedre og klassefeil for klassene 0 og 1 ble henholdsvis 20.6% og 52.6%. Gjennomsnittlig klassefeil ble 32.67% for LT_2017data.

Tabell 11: Klassifisering av 400m², LT_alldata og LT_2017data

400m ² , LT alldata				400m ² ,LT 2017data			
Klasse	0	1	Klassefeil	Klasse	0	1	Klassefeil
0	82	38	31.67 %	0	50	13	20.63 %
1	49	49	50.00 %	1	20	18	52.63 %
Gjennomsnitt av klassefeil			39.91 %	Gjennomsnitt av klassefeil			32.67 %

For datasettene med 400m² prøveflatestørrelse, MT_alldata og MT_2017data (Tabell 12) var gjennomsnitt av klassefeil på 30.07% og 30.93%. Med middelpunktthet ble klassefeilen for klasse 0 lavere med 22.5% og 18.6% for de respektive datasettene. 400m², MT_alldata hadde den laveste klassefeilen for klasse 1 av samtlige datasett med en klassefeil på 39.68%.

Tabell 12: Klassifisering av 400m², MT_alldata og MT_2017data

400m ² , MT alldata				400m ² , MT 2017data			
Klasse	0	1	Klassefeil	Klasse	0	1	Klassefeil
0	62	18	22.50 %	0	48	11	18.64 %
1	25	38	39.68 %	1	19	19	50.00 %
Gjennomsnitt av klassefeil			30.07 %	Gjennomsnitt av klassefeil			30.93 %

200m² prøveflater

Klassifiseringen av datasettene på prøveflater med 200m² viste at det var liten forskjell mellom 200m² og 400m² bortsett fra for datasettet LT_alldata som hadde den høyeste gjennomsnittlige klassefeilen av alle datasett på 40.03% (Tabell 13). MT_2017data var best i klassen med en gjennomsnittlig klassefeil på 28.78% (Tabell 14)

Tabell 13: Klassifisering av 200m², LT_alldata og LT_2017data

200m ² , LT alldata				200m ² , LT 2017data			
Klasse	0	1	Klassefeil	Klasse	0	1	Klassefeil
0	260	104	28.57 %	0	143	31	17.82 %
1	143	110	56.52 %	1	60	52	53.57 %
Gjennomsnitt av klassefeil			40.03 %	Gjennomsnitt av klassefeil			31.82 %

Tabell 14: Klassifisering av 200m², MT_alldata og MT_2017data

200m ² , MT alldata				200m ² , MT 2017data			
Klasse	0	1	Klassefeil	Klasse	0	1	Klassefeil
0	174	57	24.68 %	0	129	32	19.88 %
1	83	87	48.82 %	1	46	64	41.82 %
Gjennomsnitt av klassefeil			34.91 %	Gjennomsnitt av klassefeil			28.78 %

100m² prøveflater

For datasettene med 100m² prøveflater var gjennomsnittlig klassefeil høy (Tabell 15). LT_alldata og LT_2017data hadde høyest klasse 1 feil, men det samme nivået på klasse 0 feil som de andre datasettene (Tabell 15). Datasettet LT_alldata hadde flest prøveflater av alle datasett med 1504 prøveflater.

Tabell 15: Klassifisering av 100m², LT_alldata og LT_2017data

100m ² , LT alldata				100m ² , LT 2017data			
Klasse	0	1	Klassefeil	Klasse	0	1	Klassefeil
0	706	190	21.21 %	0	341	96	21.97 %
1	380	228	62.50 %	1	174	96	64.44 %
Gjennomsnitt av klassefeil			37.90 %	Gjennomsnitt av klassefeil			38.19 %

Klassifiseringen av MT_alldata og MT_2017data på 100m² prøveflater hadde en gjennomsnittlig klasse feil på henholdsvis 35.41% og 33.88% (Tabell 16). De to datasettene med middelpunkttetthet sammenfalt bedre med 100m² prøveflater enn datasettene med lavpunkttetthet (Tabell 15).

Klassevis var det klasse 0 som hadde minst feil, mens klasse 1 var den klassen som fikk flest feil.

Tabell 16: Klassifisering av 100m², MT_alldata og MT_2017data

100m ² , MT alldata				100m ² , MT 2017data			
Klasse	0	1	Klassefeil	Klasse	0	1	Klassefeil
0	465	133	22.24 %	0	320	90	21.95 %
1	220	179	55.14 %	1	138	125	52.47 %
Gjennomsnitt av klassefeil			35.41 %	Gjennomsnitt av klassefeil			33.88 %

Diskusjon

Hovedproblemstillingen i denne oppgaven var å undersøke mulighetene for å utvikle modeller for å estimere sagtømmerandel ved bruk av FLS.

Det ble brukt laserdata fra to prosjekter med forskjellig punkttetthet med datainnhenting fra 2013 og 2015. Selv om prosjektenes datainnhenting hadde foregått ved ulike tidspunkt anså vi det slik at skogbildet ikke ville endre seg radikalt fra 2013 til 2015 og at det ville være interessant å se på forskjellen mellom de to. Det var nærliggende å tro at laserdata med middeltetthet ville prestere bedre enn laserdata med lavtetthet. Regresjonsmodellene viste liten forskjell mellom middels- og lavpunktthet. RMSE for LT_alldata sammenlignet med RMSE for MT_alldata for alle prøveflatestørrelser viste liten forskjell (Tabell 8). Det samme gjelder for LT_2017data sammenlignet med MT_2017data. Det var store forskjeller i forhold til størrelse på prøveflatene. RMSE% var jevnt over 5 prosentpoeng høyere med 200m² i forhold til 400m² store prøveflater. De små prøveflatene på 100m² kom veldig dårlig ut med RMSE% på over 50 for alle modellene som ble laget for 100m² prøveflater.

Klassifiseringen ga litt andre resultater og viste at laserdata med middels punkttetthet presterte bedre hvis man sammenligner de forskjellige kombinasjonene med laserdata og hogstmaskindata. Datasettet MT_alldata presterte bedre enn LT_alldata for alle prøveflatestørrelser, det samme gjorde MT_2017data mot LT_2017data (Tabell 10). Ved klassifisering er det tydelig at det er en fordel med høyere oppløsning på laserdata, det var stor forskjell mellom middels- og lavpunktthet. I klassifiseringen ser vi at det er stor forskjell innad i klassene, for klasse 0 er feilen relativ lav, for klasse 1 er feilen relativt høy. Det betyr at modellen treffer bedre når sagtømmervolumet er under 50%. Når sagtømmerandelen er over 50% så blir det mer tilfeldig om modellen klarer å klassifisere riktig. Det vil være et stort antall flater som faktisk har stor sagtømmerandel som vil bli klassifisert med lav sagtømmerandel.

Hvorfor klassifiseringen er så skjevfordelt i forhold til feil i klasse 0 og 1 er usikkert. En forklaring kan være at gjennomsnittlig observert sagtømmervolum ligger under 50%. En kan anta at laservariablene for en flate med 48% sagtømmer er relativ lik en flate med 52% sagtømmer. Når det er et større antall flater med under 50% sagtømmer vil det være mange flere flater å sammenligne laservariablene med.

Det var ingen restriksjoner på hvilke bestand som skulle avvirkes fra min side, Viken Skog SA sto for avvirkningsoppdraget og hogstmaskinen i denne oppgaven var i normal kommersiell drift. Dette var

en utfordring at sensorene på hogstmaskinen hadde en tendens til å ryke og hogstmaskinen fortsatte avvirkning, noe som gjorde at vi fikk inn ubrukelige tredata uten posisjonering. Dette medførte at jeg fikk ett noe mindre datasett enn først antatt. Selv om resultatene fra Hauglin et al. (2017) viste gode posisjoner med en gjennomsnittfeil for enkelttreposisjon på 0.94 meter i 2016 vet vi at nøyaktigheten har variert mye frem til en recalibrering og ny type sensor i 2017. Som Lindroos et al. (2015) har kommet frem til må nøyaktigheten til hogstmaskindata for enkelttreposisjon være < 1 meter for å kunne brukes som georeferert data sammen med FLS. Hogstmaskindataene fra 2017 kan vi med høy sikkerhet si at treposisjonene har i underkant av 1 meter i gjennomsnittfeil.

Dette kan forklare noe av de mindre gode resultatene for 100m² prøveflater. 100m² har en relativ liten radius på 5.64 meter. Dette kan gjøre store utslag for om et tre blir registrert inne eller utenfor prøveflaten når vi hadde en feilmargin på treposisjon ved siste måling på gjennomsnittlig 0,96 meter. Det vil gjøre mindre utslag på større prøveflater. Ved større prøveflater blir det en mindre andel av trær som ligger i grensen om de er med på flaten eller utenfor avhengig av i hvilken geografisk retning vi har posisjoneringsfeilen.

I denne studien ble sagtømmerandel definert etter fire sortimenter. Det var spesial og sagtømmer for Furu, sagtømmer og Vänerblokk for Gran. Det var ingen andre sortimenter som ble vurdert som sagtømmer, vi ønsket i denne studien å fokusere på sortimentene som var økonomisk viktige. Det kan diskuteres om for eksempel pallevirke som var et sortiment også burde vært klassifisert som sagtømmer. Pallevirke er et sortiment som blir skåret på sagbruk, og vil kunne ligne noe på vanlig sagtømmer. Kvaliteten til Pallevirke er dog mye lavere enn kvalitetskravene til sagtømmer, men over kvalitetskravene til massevirke (Nybakk & Birkeland 2009). Pallevirke har en pris som ligger mellom sagtømmer og massevirke og kan ligne veldig på sagtømmer visuelt. Det kan være et problem når man skal utvikle modeller for sammenhengen mellom laserdata og tømmerkvalitet. Det kan tenkes at resultatene i denne studien ville blitt annerledes hvis pallevirke var en del av sagtømmerandelen. Sagtømmerandelen ville blitt høyere og det kan tenkes at modellene ville fått en bedre sammenheng, dette ble ikke testet.

Datasettet i denne studien er en god blanding av forskjellige bestand og skogtyper fra hele Øvre Romerike. Bestandene som ble avvirket ble ikke stratifisert og det kan tenkes at modellene ville vist en bedre sammenheng hvis datasettet hadde vært delt opp i mer homogene bestand. Et rent granbestand vil ha en helt annen laserpunkt struktur enn et blandingsbestand. Det samme gjelder rene furu og lauvbestand, de vil ha en helt annen struktur i punktskyen. Det ville vært interessant å se om vi ville fått andre resultater ved å kjøre modeller på homogene bestand. Utvikling av andre laservariable kan også være en måte man kan forbedre modellene på. Utvikling av laservariable som

sier noe om kvistsettingen kan være med å gi bedre sammenhenger mellom tømmerkvalitet og laserdata.

En annen mulighet for å se på sammenhengen mellom laserdata og tømmerkvalitet er å bruke enkelttremodeller. Dette ble ikke undersøkt i denne studien, men en kan tenke seg at med et datasett som i denne oppgaven kunne man brukt enkelttremodeller istedenfor arealbaserte modeller (Koch et al. 2014). Et datasett med posisjonerte enkelttrær er en forutsetning for enkelttrematching med laserdata. Bruk av enkelttremodeller har i utgangspunktet vært forbeholdt forskningen i Norge, da kostnadene med datainnsamling er høye. Ved å bruke hogstmaskindata kombinert med ekstra treposisjonering som i denne oppgaven kan denne utfordringen løses. Spørsmålet er om målingene er nøyaktige nok til at enkelttremodeller kan anvendes.

Resultatene i denne studien er på prøveflatenivå, det var for lite felldata til å kunne gjøre dette på bestandsnivå. Resultatene viser et gjennomsnitt av alle prøveflatene i datasettet. For å kunne si noe om hvor godt modellen treffer på bestandsnivå måtte vi ha inkludert mye mer data, slik at vi kunne holdt ute hele bestand når modellen kjøres med kryssvalidering. Hadde vi hatt nok felldata kunne man ha aggregert resultatene opp på bestandsnivå. Forskjellene innad i bestand kunne da blitt utlignet med hverandre og gjennomsnittsfeilen for et bestand kunne blitt lavere. Den beste modellen, ble kjørt på 400m² prøveflater og datasettet LT_alldata, hadde en RMSE% på 29.4 og RMSE på 0.14. Det datasettet som presterte best både i regresjonsmodellen og klassifiseringen var på 400m² prøveflater, med MT_alldata, der RMSE og RMSE% var på 0.15 og 30.7, gjennomsnittsfeilen i klassifiseringen var på 30.07%. Til sammenligning predikerte Bollandssås et al. (2011) en sagtømmerandel i prosent av totalvolum med en RMSE på 0.13 på bestandsnivå. Peuhkurinen et al. (2008) modellerte sagtømmerandel for rene gran og furu bestand og fikk en RMSE% på 28.3 og 27.2 for henholdsvis furu og gran. I forhold til mine resultater er Peuhkurinen et al. (2008) sine resultater litt lavere, men man må ta med seg at disse resultatene også er på bestandsnivå i forhold til denne studien som er på prøveflatenivå. Til sammenligning med andre variabler har RMSE for volum, grunnflateveid middel høyde og grunnflateveid middel diameter i gjennomsnitt blitt predikert til 11-14%, 3-6% og 9-13% på bestandsnivå (Næsset 2007).

Konklusjon

Forutsatt at flybåren laserskanning fortsatt vil bli brukt i fjernmåling av skog, vil det være nyttig å implementere modeller for tømmerkvalitet. Det økende behovet for rasjonalisering i skogbruket vil føre til færre beslutningstakere og et økende behov for mer informasjon om skogen i skogbruksplanene.

Det ble i denne oppgaven funnet at sagtømmerandel kan estimeres med en gjennomsnittssfeil på 29.4% - 33.1% på prøveflatenivå. Å kunne estimere sagtømmerandel med en gjennomsnittssfeil på ca. 30% eller bedre vil gjøre det lettere å kunne prioritere bestand for slutthogst. Skogeiere vil bedre kunne regne ut lønnsomheten av et inngrep da sagtømmerandelen er en vesentlig faktor for om en hogst vil være lønnsom eller ikke. Noen nullområder vil kunne bli lønnsomme eller aktuelle ved bedre økonomi i driften. For skogbruksnæringen vil man ha en bedre ressursoversikt og kunne vite hvilke områder det er sannsynlig å få tak i de produktene som er etterspurt. Det er avgjørende for både tømmerkjøper og skogeier at det finnes avsetningsmuligheter for alle produkter som blir produsert i en hogst. En har muligheten til å begrense produksjonen av massevirke ved at man prioriterer hogst i bestand der sannsynligheten er stor for en høy sagtømmerandel. Dette bør være av interesse for skognæringen, da kostnaden ved disse modellene vil være liten så lenge FLS er den primære måten for fjernmåling av skog.

Litteraturliste

- Arlinger, J., Nordström, M. & Möller, J. J. (2012). *StanForD 2010: Modern Communication with Forest Machines*: Skogforsk.
- Barth, A. & Holmgren, J. (2013). Stem taper estimates based on airborne laser scanning and cut-to-length harvester measurements for pre-harvest planning. *International Journal of Forest Engineering*, 24 (3): 161-169.
- Bollandsås, O. M., Maltamo, M., Gobakken, T., Lien, V. & Næsset, E. (2011). Prediction of Timber Quality Parameters of Forest Stands by Means of Small Footprint Airborne Laser Scanner Data. *International Journal of Forest Engineering*, 22 (1): 14-23.
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E. & Yergeau, F. (1998). Extensible markup language (XML). *World Wide Web Consortium Recommendation REC-xml-19980210*. <http://www.w3.org/TR/1998/REC-xml-19980210>, 16: 16.
- Eid, T., Gobakken, T. & Næsset, E. (2004). Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. *Scandinavian Journal of Forest Research*, 19 (6): 512-523.
- Evans, J. S., Hudak, A. T., Faux, R. & Smith, A. (2009). Discrete return lidar in natural resources: Recommendations for project planning, data processing, and deliverables. *Remote Sensing*, 1 (4): 776-794.
- Forsberg, M. (2003). Behov av informationssystem för kundorienterad virkesstyrning. *Skogforsk, Arbetsrapport* (530): 18.
- Hauglin, M., Hansen, E. H., Næsset, E., Busterud, B. E., Gjevestad, J. G. O. & Gobakken, T. (2017). Accurate single-tree positions from a harvester: A test of two global satellite-based positioning systems. *Scandinavian Journal of Forest Research* (just-accepted): 1-24.
- Holmgren, J. (2004). Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research*, 19 (6): 543-553.
- Kartverket (2015). *Grunndata fra GNSS-stasjoner*: Kartverket. Tilgjengelig fra: <http://www.kartverket.no/posisjonstjenester/Grunndata/> (lest 09.05.2017).
- Koch, B., Kattenborn, T., Straub, C. & Vauhkonen, J. (2014). Segmentation of forest to tree objects. I: Maltamo, M., Næsset, E. & Vauhkonen, J. (red.) *Forestry Applications of Airborne Laser Scanning*, s. 89-112: Springer.
- Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Ijcai: Stanford, CA. 1137-1145 s.
- Lindroos, O., Ringdahl, O., La Hera, P., Hohnloser, P. & Hellström, T. H. (2015). Estimating the Position of the Harvester Head—a Key Step towards the Precision Forestry of the Future? *Croatian Journal of Forest Engineering*, 36 (2): 147-164.
- Nordkvist, K., Sandström, E., Reese, H. & Olsson, H. (2013). Laserskanning och digital fotogrammetri i skogsbruket.(2. uppl.). *Arbetsrapport / Sveriges lantbruksuniversitet, Institutionen för skoglig resurshushållning och geomatik (1401-1204)*, 1401-1204. 117 s.
- Nybakk, E. & Birkeland, T. (2009). Fordelingsaptering versus tradisjonell verdiaptering i ulike skogtyper i Norge. *Oppdragsrapport fra Skog og landskap*, 18/09: 27s.
- Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80 (1): 88-99.
- Næsset, E. (2004). Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. *Scandinavian Journal of Forest Research*, 19 (6): 554-557.

- Næsset, E. (2007). Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research*, 22 (5): 433-442.
- Olivera, A. & Visser, R. (2014). Integration of harvester data and geospatial information. *Harvesting technical note HTN07-03. Future Forest Research Ltd. Rotorua, New Zealand.*
- Olivera, A., Visser, R., Acuna, M. & Morgenroth, J. (2016). Automatic GNSS-enabled harvester data collection as a tool to evaluate factors affecting harvester productivity in a Eucalyptus spp. harvesting operation in Uruguay. *International Journal of Forest Engineering*, 27 (1): 15-28.
- Peuhkurinen, J., Maltamo, M. & Malinen, J. (2008). Estimating species-specific diameter distributions and saw log recoveries of boreal forests from airborne laser scanning data and aerial photographs: a distribution-based approach. *Silva Fennica*, 42 (4): 625-641.
- Skogforsk. (2007). *Standard for Forest Data and communications*: Skogforsk. Tilgjengelig fra: http://www.skogforsk.se/contentassets/b063db555a664ff8b515ce121f4a42d1/stanford_ma_indoc_070327.pdf.
- Vauhkonen, J., Maltamo, M., McRoberts, R. E. & Næsset, E. (2014). Introduction to forestry applications of airborne laser scanning. I: *Forestry Applications of Airborne Laser Scanning*, s. 1-16: Springer.



Norges miljø- og biovitenskapelig universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway