

REGISTRERING AV DIAMETER I BRYSTHØYDE VED HJELP
AV DATA FRA BAKKEMONTERTE LASERSKANNER

REGISTRATION OF DIAMETER AT BREASTHEIGHT USING
DATA FROM TERRESTRIAL LASER SCANNING

JOACHIM KJELSTRUP

UNIVERSITETET FOR MILJØ- OG BIOVITTENSKAP
INSTITUTT FOR NATURFORVALTNING
MASTEROPPGAVE 30 STP. 2010



Forord

Gjennom 5 år med skogfag på høyskole og universitet har jeg lært mye om nytten og viktigheten av god kvalitet på taksering og inventering av skogressurser. Personlig tror jeg praktisk gjennomføring av manuell taksering alltid vil være nyttig i undervisningsøyemed i forhold til å øke forståelsen for hva som måles. Samtidig går den teknologiske utviklingen videre og åpner for nye takseringsmetoder. En nyvinning det er stor fokus på nå, er bruk av laserteknologi. Målet er å effektivisere takseringen uten at det går på bekostning av kvaliteten. Flybåren laser er allerede i kommersiell bruk, mens bakkelaser fortsatt er på forsøksstadiet. Potensialet i teknologien er stort, og personlig er jeg overbevist om at all skogtaksering vil bli utført ved hjelp av laserteknologi i overskuelig fremtid. Dette gjør temaet fremtidsrettet og spennende for skogfagsstudenter som undertegnede.

Denne masteroppgaven er koblet til prosjektet BALABU – Bakkemontert laser som verktøy for bedre utnyttelse av skogressursene (2009 – 2012). Forskningsarbeidet utføres gjennom et samarbeid mellom Universitet for miljø- og biovitenskap ved instituttet for naturforvaltning, og Norsk institutt for skog og landskap. I tillegg er kommersielle aktører som Viken Skog BA, Geoplan Øst og TreeMetrics Ltd. tilknyttet prosjektet. BALABU er støttet av Norges forskningsråd, skogtiltaksfondet og SkatteFUNN.

Jeg vil takke hovedveileder Terje Gobakken og hjelpeveileder Vegard Lien, begge ansatte ved Institutt for naturforvaltning ved Universitetet for miljø- og biovitenskap, for all hjelp de har bidratt med i arbeidet med denne oppgaven.

Joachim Kjelstrup

Ås, 15. desember 2010

Sammendrag

Målet med denne oppgaven er å teste hvor godt en bakkemontert laserskanner (TLS) registrerer diameter i brysthøyde (dbh) i produksjonsskog. Seks sirkulære prøveflater i Aurskog-Høland kommune, hver på 1000 m², ble totalklavet og skannet i perioden vinteren 2007 til høsten 2010. Den manuelle klavingen ble regnet som sann verdi, og var sammenlikningsgrunnlaget til laserskanningen. Sammenhengen mellom observert og predikert dbh har så blitt testet ved hjelp av enkel regresjonsanalyse. Estimert feilmargin på predikerte dbh-verdiene varierte fra $\pm 1,6$ til $\pm 13,4$ cm mellom de ulike flatene, og var på 5,2 cm for alle flatene sett under ett. Denne undersøkelsen indikerer at det er stor usikkerhet knyttet til predikering av dbh-verdier ved hjelp av TLS-data i produksjonsskog.

Summary

The aim of this paper is to test how well diameter at breast height (dbh) can be predicted by the use of TLS-data from a production forest. All the trees on six circular plots, each 1000 m², in Aurskog-Høland kommune was manually callipered and subsequently scanned with a terrestrial laser scanner in the period from winter 2007 to fall 2009. The manual caliper measurements were considered to represent true value, and were the basis for comparison with the TLS-data. The relationship between observed and predicted dbh were tested by simple linear regression analysis. The predicted values estimated margin of error varied from $\pm 1,6$ to $\pm 13,4$ cm among the individual plots, and was 5,2 cm for all plots combined. This survey indicates that there is substantial uncertainty related to predicting dbh in a production forest by the use of TLS-data.

Innholdsfortegnelse

1	Innledning.....	5
2	Material og metode.....	6
2.1	Feltdata	6
2.1.1	Manuell registrering	8
2.1.2	Registrering med bakkelasar	8
2.1.3	Generell beskrivelse av bakkemontert laserskanner (TLS).....	8
2.2	Databehandling	9
2.2.1	Geografisk posisjon.....	9
2.2.2	Begrensning av datamaterialet	9
2.2.3	Testgrunnlag	9
2.2.4	Generelt om prosessering av TLS-data	10
2.2.5	Valg av skannfil	10
2.3	Statistisk analyse.....	14
2.3.1	Regresjonsmodell	14
2.3.2	Forklaringsvariabler	15
2.3.3	Forutsetninger.....	16
	Normalfordeling	17
	Innbyrdes uavhengighet	17
	Konstant varians	17
	Ekstrem innflytelse.....	17
3	Resultater.....	18
3.1	Flate S3-03.....	19
3.2	Flate S3-05.....	21
3.3	Flate S3-13.....	23
3.4	Flate S3-14.....	25
3.5	Flate S3-15.....	27

3.6	Flate S3-26.....	30
3.7	Alle flater.....	32
4	Diskusjon.....	35
	Valg av skannfil	35
	Testgrunnlag.....	35
	Tilvekst.....	36
	Residualer.....	36
	Feilmargin	36
	Praktiske utfordringer ved bruk av TLS til registrering av dbh	37
4.1	Konklusjon.....	38
5	Referanseliste	39
6	Vedlegg	40
6.1	S3-03.....	40
6.2	S3-05.....	41
6.3	S3-13.....	42
6.4	S3-14.....	43
6.5	S3-15.....	44
6.6	S3-26.....	45

1 Innledning

Skogeiere planlegger og gjennomfører hogst og ulike skjøtselstiltak på bakgrunn av de ulike skogbestandenes tilstand. Tilstanden beskrives av ulike parametre som blant annet alder, bonitet, høyde, treantall og diameter i brysthøyde (dbh). Tradisjonelt utarbeidet man skogbruksplaner på bakgrunn av manuelle takseringer med systematisk prøveflatenett og utvelgelse av prøvetrær, men i de senere år har fjernmåling etablert seg som den foretrukne takseringsmetoden. Av de ulike fjernmålingsmetodene er det flybåren laserregistrering som er den mest utbredte. Metoden har vist seg å være like god, og i noen tilfeller også bedre enn manuell taksering når det gjelder presisjon. Og ettersom den dekker store geografiske områder på en rask og effektiv måte, er den et økonomisk akseptabelt alternativ. En ulempe ved denne metoden er imidlertid at den ikke gir direkte informasjon om trærnes dbh. I stedet estimeres dbh på bakgrunn av sammenhengen med andre parametre som høyde, alder, bonitet og trekrone. Dette fungerer bra i bestand som har stor dominans av bartrær, men ikke spesielt godt i bestand hvor det er et betydelig innslag av løvtrær (Næsset et al. 2004). Dbh brukes til estimering av blant annet tømmervolum og tilvekst, og er derfor av avgjørende betydning innen skogtaksering.

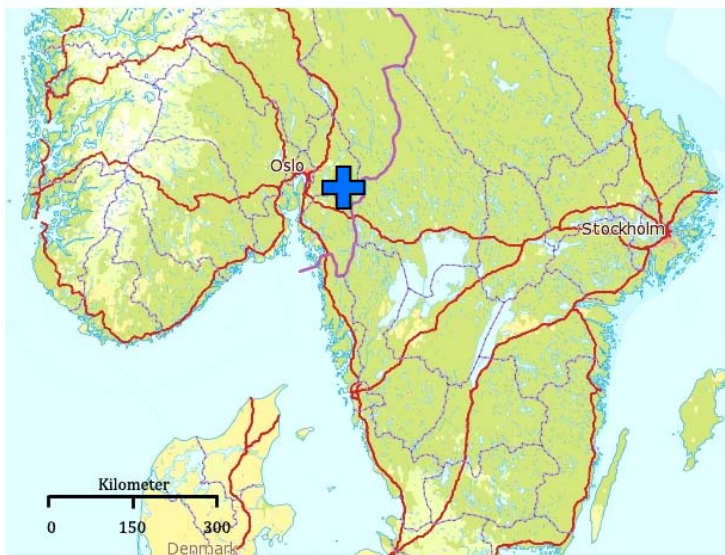
Bakkemontert laser er en teknologi som kan kartlegge objekters overflate i 3D med millimeterpresisjon, og som i utgangspunktet ble utviklet for bruk innen mekanisk ingeniørkunst. I de senere år har teknologien gjort store framskritt, og blir i dag brukt til et bredt spekter av ulike formål. I skognæringen er den interessant fordi den åpner for en direkte, objektiv og reproducerbar registrering av individuelle trærns dbh og stammeprofil i horisontalplanet. Selv om bakkelasen åpner for innhenting av detaljert informasjon som ikke er tilgjengelig ved flybåren laserskanning, så har den også klare begrensninger. I prinsippet skal teknologien kunne registrere objekter på mer enn 50 meters avstand, men rekkevidden begrenses kraftig i skog på grunn av skyggeeffekt fra stammer, kvist og buskvegetasjon. Denne effekten øker i styrke med økende skogtetthet. En annen ulempe er at bruk av bakkelasen er kostbart og lite effektivt sammenliknet med alternative takseringsmetoder (Leeuwen & Nieuwenhuis 2010).

Målsettingen med denne oppgaven er å teste hvor presist bakkemontert laser registrerer dbh sammenliknet med manuell klaving, på seks sirkulære prøveflater av 1000 m² i produksjonsskog i Aurskog-Høland kommune.

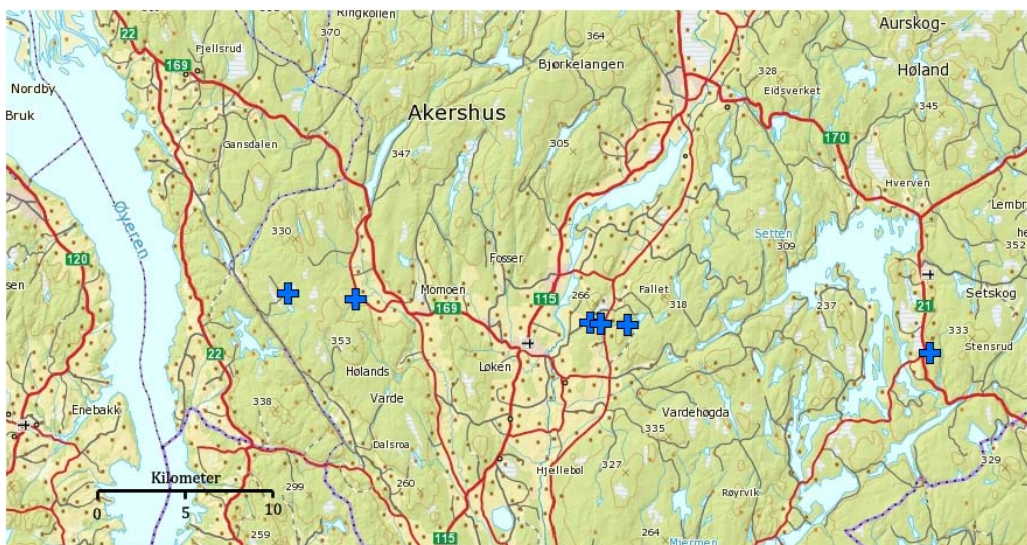
2 Material og metode

2.1 Felldata

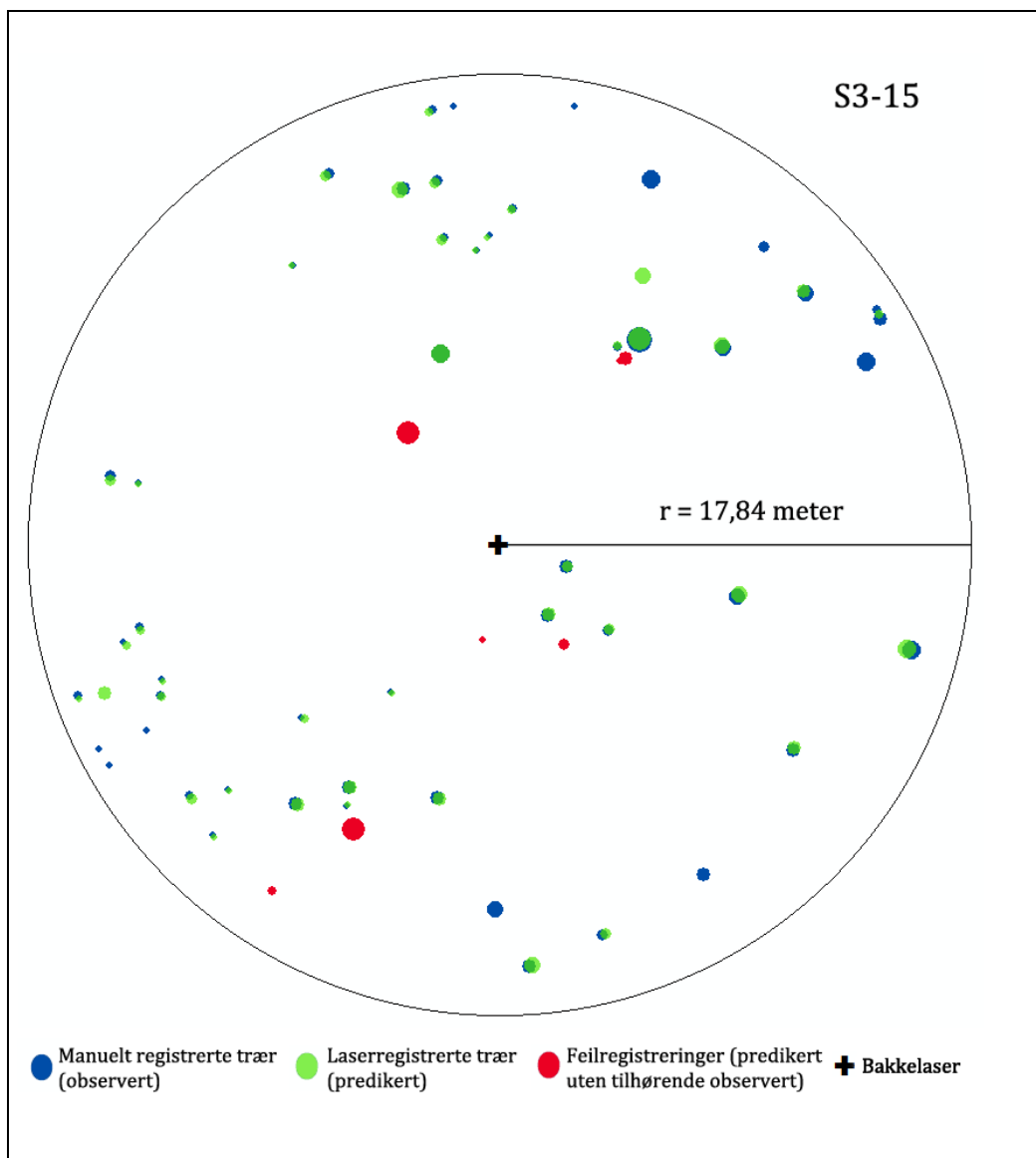
Alle data er hentet fra felt- og laserregistreringer i seks forskjellige prøveflater i Aurskog-Høland kommune i Akershus fylke. Prøveflatene består hovedsakelig av gran og furu, med noe innslag av løv. Flatene er sirkulære og har en radius på 17,84 meter, som gir en grunnflate i horisontalplanet på 1000 m². Den manuelle registreringen ble gjennomført i perioden desember 2007 til april 2008, mens bakkelaserregistreringen ble gjennomført i september/oktober 2009.



Figur 1. Oversiktskart, Aurskog-Høland merket med blått kryss. Kartgrunnlag: Statens kartverk.



Figur 2. Detaljert kart, prøveflater markert med blå kryss. Kartgrunnlag: Statens kartverk.



Figur 3. Grafisk framstilling av prøveflate S3-15.

Tabell 1. Prøveflateposisjoner, UMT sone 32N, datum WGS1984.

<i>Flate</i>	<i>Posisjon Nord</i>	<i>Posisjon Øst</i>
S3-03	6632494.844	629672.565
S3-05	6632489.303	632130.521
S3-13	6632454.214	640689.15
S3-14	6632439.349	641077.951
S3-15	6632466.676	642056.561
S3-26	6632475.917	653050.129

2.1.1 Manuell registrering

Først ble flatesenters geografiske posisjon målt inn med GPS. Deretter ble alle trær med brysthøydiameter over 5 cm registrert med vinkel og avstand til flatesenter ved hjelp av totalstasjon og prisme. I tillegg ble vinkelen til et annet flatesentrum registrert. Avstanden som ble målt og registrert er lik den horisontale avstanden mellom flatesentrum og senter av trærnes stamme. På alle trær hvor avstanden fra stammesenter til flatesenter var mindre eller lik 17,84 meter ble det registrert treslag (gran, furu eller løv), diameter i brysthøyde (dbh) med klavearmen vinkelrett mot flatesenter, og tilstand. Den manuelle registreringen ble gjennomført av Institutt for naturforvaltning (UMB, Universitetet for Miljø- og biovitenskap) i forbindelse med et forskningsprosjekt.

2.1.2 Registrering med bakkelas

De samme prøveflatene som ble manuelt registrert vinteren 2007/2008, ble skannet med bakkelas høsten 2009. Geoplan Øst hadde hovedansvaret for utstyr og arbeid med bakkelas og totalstasjon på prøveflatene. TLS-instrumentet var av typen Leica HD6000, som skanner 360° i horisontalt og 310° vertikalt og har en teoretisk rekkevidde på 79 m. Totalstasjonen var av typen Sokkia SET5F. Det ble gjort 5 skann på hver flate. Et fra flatesentrum, samt 4 perifere for å få tilfredsstillende dekning av alle trær på flaten. For hvert skann ble posisjonen til laseren, samt 3 målskiver, registrert med totalstasjon. Ved tåke eller regn ble det ikke gjort registreringer med bakkelas. Etter at prøveflaten var ferdig skannet ble det tatt 360° fotografi av flaten fra senter.

2.1.3 Generell beskrivelse av bakkemontert laserskanner (TLS)

Bakkemontert laserskanner (Terrestrial Laser Skanner, TLS) er en form for LiDAR (Light Detection and Ranging) teknologi. LiDAR fungerer ved å sende ut lyspulser. Når pulsene treffer et objekt, sendes det en refleksjon tilbake, og tiden det tar fra pulsen sendes til refleksjonen mottas kan med utgangspunkt i lysets hastighet brukes til å beregne avstand. Alle refleksjonspunktene registreres og danner en punktsky som beskriver de skannede objektene i tre dimensjoner, med X, Y og Z koordinater (Leeuwen & Nieuwenhuis 2010).

2.2 Databehandling

2.2.1 Geografisk posisjon

På bakgrunn av den manuelle registreringen ble alle trær posisjonert i et polart koordinatsystem med utgangspunkt i treets vinkel og avstand til flatesenter som origo. Samtidig ble eksakt geografisk posisjon, og vinkel mellom de ulike flatenes senter, registrert med henholdsvis GPS og totalstasjon. Med utgangspunkt i differansen mellom sann geografisk vinkel og målt vinkel med totalstasjon mellom de ulike flatesentrene, ble vinkelen til individuelle trær på de enkelte flatene justert tilsvarende. Deretter ble de polare koordinatene omformet til UTM koordinater, som er et kartesiansk koordinatsystem med en X og en Y koordinat. UTM sone 32N, datum WGS1984, ble valgt som koordinatsystem.

2.2.2 Begrensning av datamaterialet

Denne oppgaven tar kun utgangspunkt i senter-skannet på hver flate. Ved bruk av bakkelasere vil det dannes skyggesoner bak alle objekter som treffes av lyspulsene. Derfor vil, i tilfeller hvor flere trær står bak hverandre i rett linje fra flatesenter, kun det nærmeste treet bli registrert under skanning. Alle feltregistrerte trær hvor mer enn 50 % av grunnflaten i brysthøyde var dekket av skyggesone ble derfor fjernet fra datamaterialet, ettersom de ikke kunne forventes å bli registrert av bakkelaseren. Bestemmelse av skyggesone og fjerning av skyggetrær ble gjort i ArcMap. Små stammer er vanskeligere å finne under prosessering av TLS-data enn større, og det er derfor større usikkerhet knyttet til registrering av disse. Ettersom trær med dbh < 10 cm heller ikke representerer noen verdi som tømmer for skogeieren på kort sikt ble også disse fjernet fra datamaterialet.

2.2.3 Testgrunnlag

Utgangspunktet for de statistiske analysene er TLS-data, og data fra den manuelle registreringen. De manuelt registrerte trærne representerer observert sann verdi, mens TLS-dataene representerer predikert verdi. Ikke alle TLS-registrerte trær er posisjonert eksakt på samme sted som de manuelt registrerte, og det er mange tilfeller hvor ett predikert tre er nærmest flere observerte trær enn andre predikerte trær. Det fører til at flere observerte trær knyttes til samme predikerte tre, og det oppstår mange duplikater av predikerte trær i datamaterialet. For å fjerne duplikatene må det defineres hvilket observerte tre som skal knyttes til det predikerte treet. Det er ofte åpenbart hvilket observert tre som er det rette når

man ser en grafisk fremstilling av prøveflaten, men en slik metode er svært tidkrevende, og derfor lite egnet når man skal analysere store datamengder. I stedet valgte jeg å gjøre to analyser, en basert på hvilket av de observerte trærne som er nærmest i avstand, og en basert på hvilket som har minst avvikende dbh. De resterende observerte trærne ble fjernet fra datamaterialet i dbh-analysen ettersom det ikke finnes sammenlikningsgrunnlag. Det medfører at gjennomsnittstallet for observert dbh endres på bakgrunn av hvilket sammenlikningsgrunnlag som velges. En slik metode er langt mer effektiv ettersom den kun tar utgangspunkt i registrerte tallverdier, og derfor kan utføres i statistikkprogrammer. Det er også tilfeller av predikerte trær som ikke ligger nærmest noe observert tre. Disse betraktes som feilregistreringer og ekskluderes fra datamaterialet.

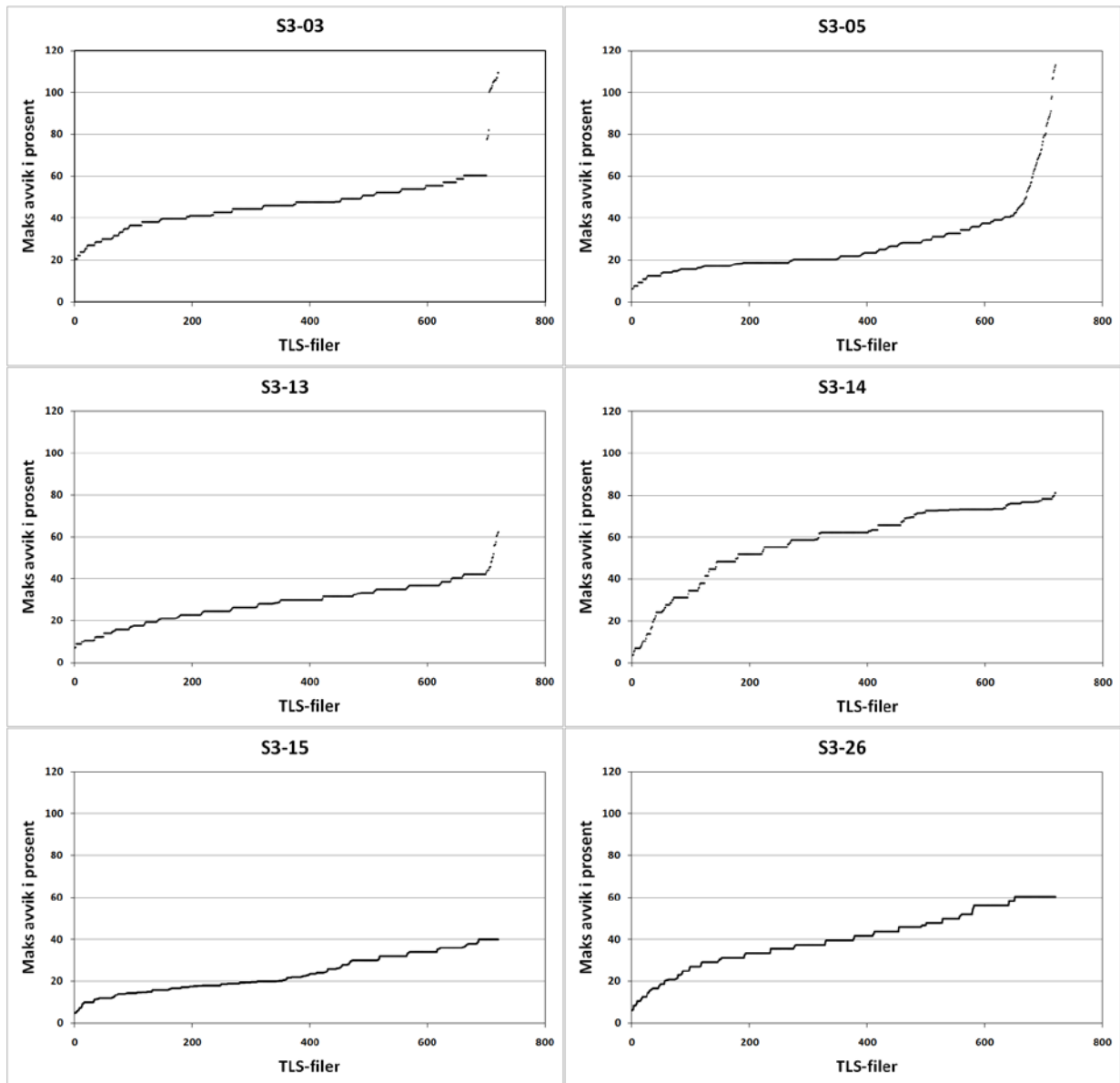
2.2.4 Generelt om prosessering av TLS-data

Registrering av enkeltrær og bestemmelse av dbh fra TLS-data blir normalt gjort ved å se hvor godt sirkel- eller sylinderformer passer i punktskyen. Dette arbeidet ble utført av TreeMetrics Ltd. i Irland. Alle deler av punktskyen som danner sirkler eller sylinderformer som tilfredsstiller kravet til passform, blir registrert som et tre. For å finne optimal passform brukes som regel minste kvadraters metode. En sylindrisk passform vil ikke fungere for en trestammes fulle lengde, ettersom både radius og akseorientering vil endre seg i høyden. Derfor er det viktig, ved bruk av minste kvadraters metode, å se på passform kun i segmenter av stammen. En utfyllende beskrivelse av denne metoden er beskrevet i Thies et al. (2004). Hough-transformasjon er et alternativ til minste kvadraters metode når det gjelder formgjenkjennelse, men egner seg ikke til detaljert målregistrering av for eksempel dbh (Leeuwen & Nieuwenhuis 2010).

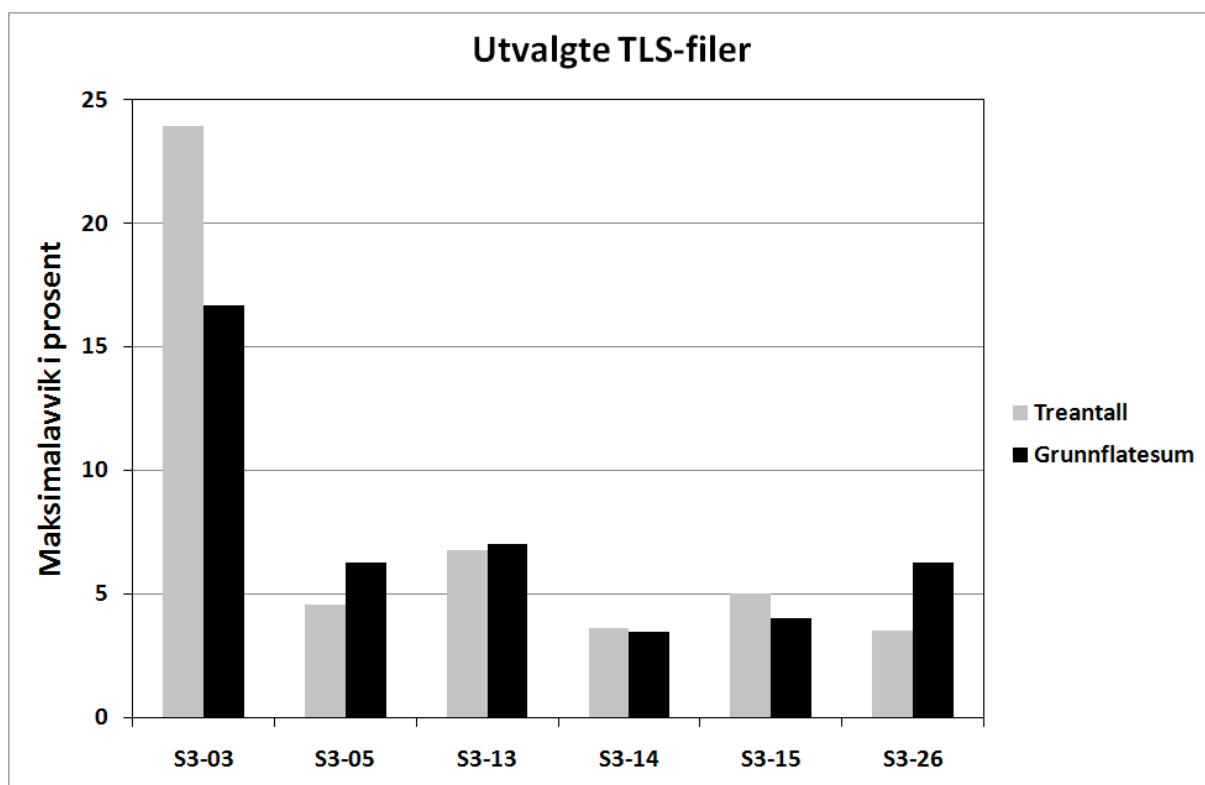
2.2.5 Valg av skannfil

Etter at dataene var prosessert av TreeMetrics, ble det levert 720 ulike skannfiler for hver prøveflate. De 720 ulike filene skilte seg fra hverandre ved ulike innstillinger, som blant annet krav til rundhet og krav til antall stammesegmenter stablet i høyden. For å finne den filen som best samsvarte med felldata på den enkelte flate, rangerte jeg filene i forhold til prosentvis avvik på grunnflatesum og treantall. De to variablene ble veid likt, og den høyeste avviksprosenten av grunnflatesum og treantall (maksimalavvik) ble valgt for hver fil. Deretter ble den filen med laveste maksimalavvik valgt som best treff på flaten. Dette tilsvarer verdien som ligger nærmest Y-aksen på hver flate i figur 4. Ulike innstillinger på skannfilene ga store

utslag på både grunnflatesum og treantall, og det var forskjellige innstillinger som ga best resultat på de ulike flatene.

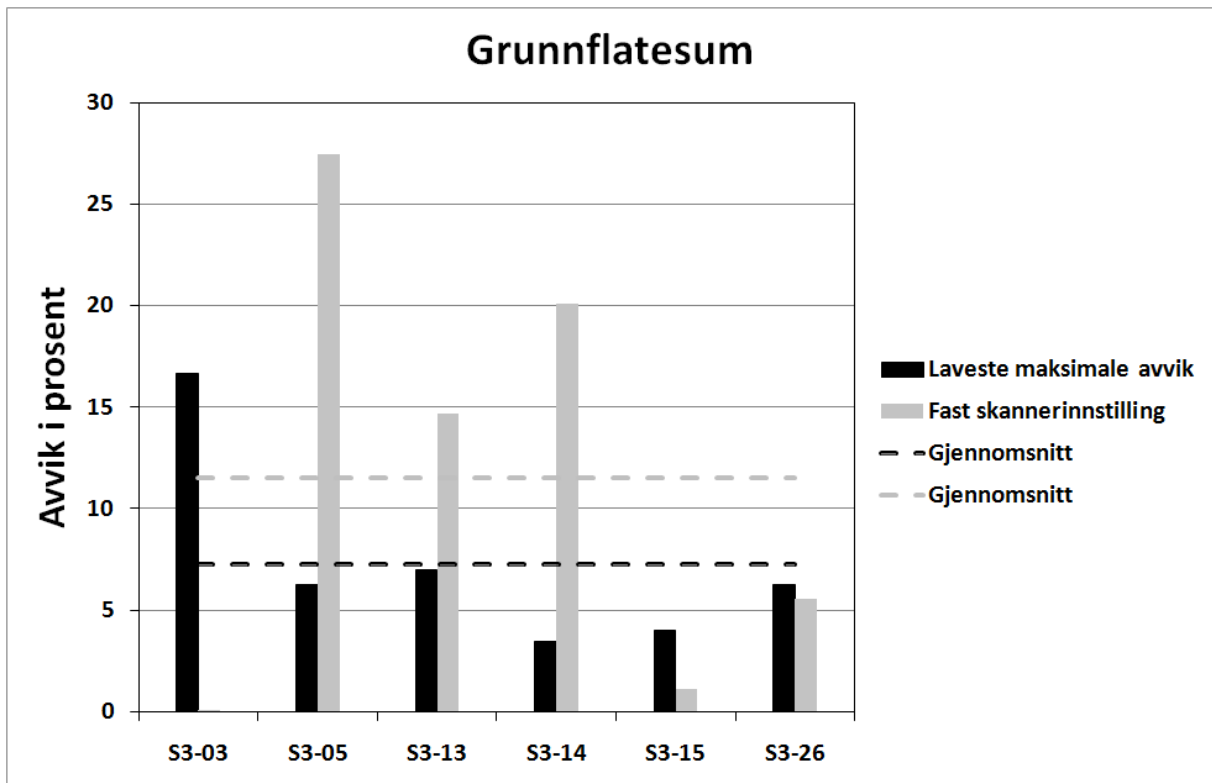


Figur 4. Maksimalavvik fra feltmålt grunnflatesum og treantall ved 720 ulike innstillinger på skannfilene for hver enkelt flate.

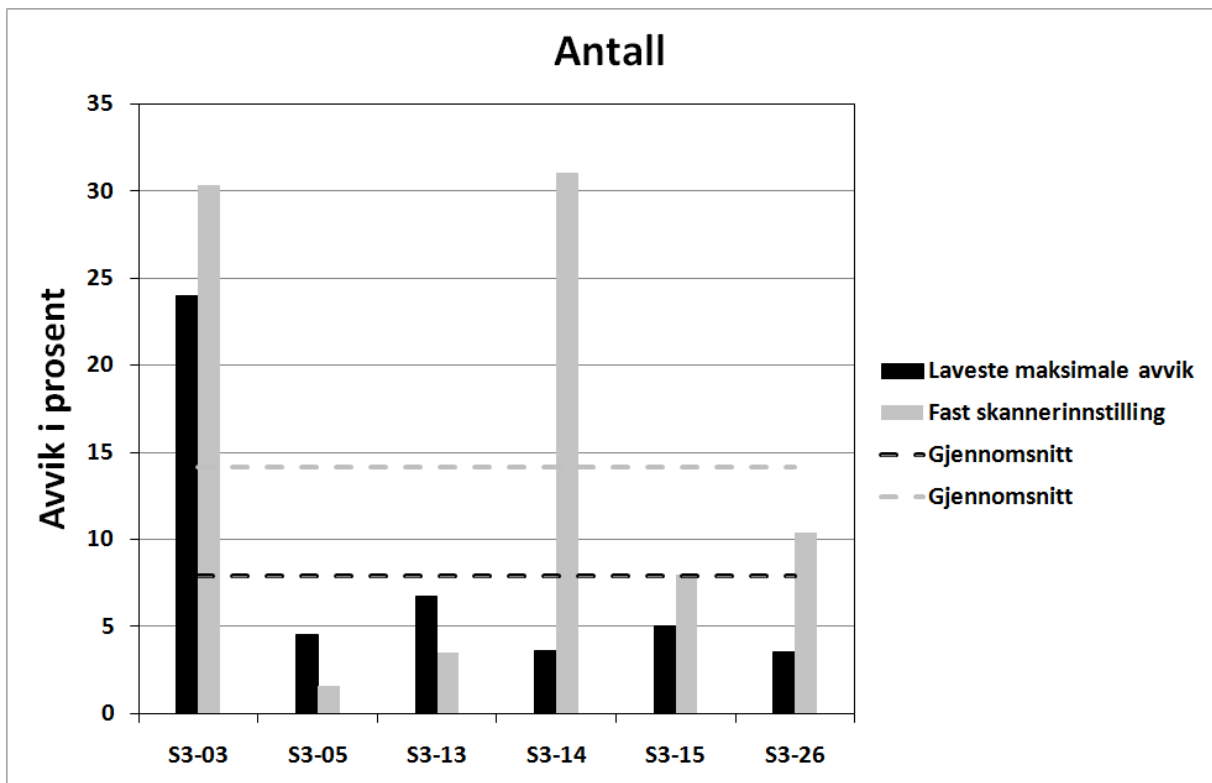


Figur 5. Avvik fra feltmålte verdier på skannfilene som ble valgt for videre analyse. Høyeste verdi av treantall og grunnflatesum for hver flate tilsvarer verdien som ligger nærmest Y-aksen på de ulike flatene i figur 1.

For å effektivisere arbeidet ville det vært ønskelig å velge samme innstilling på skannfilene fra alle flatene. Dette ble forsøkt ved å finne gjennomsnittet av innstillingene som ga lavest maksimalavvik, og bruke den som fast innstillingen på alle flater. Selv om denne metoden ga bedre resultat på enkelte flater, så ble den forkastet ettersom den i gjennomsnitt viste seg å gi større avvik fra både feltmålt grunnflatesum og treantall (figur 6 og 7).



Figur 6. Avvik fra feltmålt grunnflatesum for de to ulike metodene for valg av TLS-fil.



Figur 7. Avvik fra feltmålt treantall for de to ulike metodene for valg av TLS-fil.

2.3 Statistisk analyse

2.3.1 Regresjonsmodell

Hvor godt bakkelaserens prediksjon av diameter i brysthøyde samsvarer med observerte verdier ble testet ved hjelp av regresjonsanalyse, både for hver enkelt flate og for alle trær sett under ett. Analysen bygger på to variable; responsvariabel Y = predikert dbh, og forklaringsvariabel x = observert dbh. Det antas å være en tilnærmet lineær sammenheng mellom de to variablene. Det vil si at endringer i Y i stor grad kan forklares med endringer i x ved hjelp av følgende lineære funksjon: $Y = \alpha + \beta x$, hvor α er skjæringspunktet på y-aksen og β er stigningstallet. Utgangspunktet for mine analyser er følgende hypoteser:

$H_0: \beta = 0$, det er ingen sammenheng mellom observert og predikert dbh.

$H_1: \beta \neq 0$, det er sammenheng mellom observert og predikert dbh.

Men dersom det finnes en lineær sammenheng, så vil den ikke være deterministisk. En bestemt verdi for x vil med andre ord ikke gi en bestemt verdi for y , ettersom det alltid vil være en viss grad av unøyaktighet ved målinger. Derfor må modellen også inneholde et ledd som beskriver dette; $Y = \alpha + \beta x + \varepsilon$, hvor ε er feilleddet. Dette feilleddet innebærer at predikerte verdier ikke vil ligge nøyaktig på regresjonslinjen som estimerer sammenhengen mellom y og x . Avstanden fra predikerte verdier, y , til regresjonslinjen, \hat{y} , kalles residualer. For å finne den regresjonslinjen som best estimerer sammenhengen mellom x og y så tilpasses linjen slik at residualenes kvadratsum (Sum of Squared Errors, SS_E) blir minst mulig. Dette kalles minste kvadraters metode (MKM).

$$\varepsilon = y - \hat{y}$$

$$SS_E = \text{Sum of Squared Errors} = \sum (y - \hat{y})^2$$

$$MKM = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

SS_E minimeres ved å estimere hvilke verdier av a og b som resulterer i den regresjonslinja gir lavest SS_E . Dette gjøres ved hjelp av normallikningene for a og b . Resultatet av minste kvadraters metode kalles 'den sanne regresjonslinja': $\hat{y} = \hat{a} + \hat{b}x$, hvor \hat{a} og \hat{b} er estimater av α og β . (Hocking 1976).

2.3.2 Forklaringsvariabler

Jeg vil også teste *hvor godt* regresjonsmodellen beskriver sammenhengen mellom observert og predikert dbh. For å gjøre det har jeg sett på hvor stor variasjon det er i predikert dbh i forhold til hva regresjonsmodellen estimerer. Et mål på hvor mye responsdataene varierer er den totale variasjonen (Total Sum of Squares, SS_T):

$$SS_T = \text{Total Sum of Squares} = \sum (y_i - \bar{y}_i)^2$$

Men den totale variasjonen har to kilder. Variasjonen i x som forklares gjennom regresjonsmodellen (Regression Sum of Squares, SS_R), og variasjonen i residualene (Sum of Squared Errors, SS_E), som er uforklart:

$$SS_R = \text{Regression Sum of Squares} = \sum (\hat{y}_i - \bar{y}_i)^2$$

$$SS_E = \text{Sum of Squared Errors} = \sum (y_i - \hat{y}_i)^2$$

Ved hjelp av SS_R og SS_T kan vi beregne den multiple regresjonskoeffisienten, R^2 , som gir oss et mål på hvor mye av den totale variasjonen i regresjonsmodellen som er forklart av regresjonslinjen:

$$R^2 \approx \frac{SS_R}{SS_T} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

Dersom all variasjon i predikert dbh kunne forklares ved hjelp av regresjonslinjen, så ville det innebære at bakkelaseren ga eksakt samme dbh-verdier som den manuelle registreringen. Men selv om TLS-teknologien har potensial til å måle objekters overflate med millimeternøyaktighet, så vil en slik presisjon være uoppnåelig. Spørsmålet er da hvor stor andel av variasjonen er ikke forklart av regresjonslinjen? Eller hvor stor forskjell er det i faktisk og estimert verdi på responsvariabel? Forskjellen kalles *residual*, og er grunnlaget for å uttrykke standardfeilen ved predikert dbh i denne undersøkelsen. Standardfeilen beregnes ved å ta roten av residualenes varians, som er residualenes standardavvik (Root Mean Squared Error, RMSE). RMSE har den nyttige egenskapen at den uttrykkes i samme enhet som predikerte verdier, og kan derfor brukes som et direkte mål på laserskanningens feilmargin:

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

I tillegg til usikkerheten som er knyttet til regresjonslinja som helhet, så er det også usikkerhet knyttet til estimeringen av regresjonskoeffisienten \hat{b} . Denne usikkerheten uttrykkes som middelfeilen eller standardfeilen til \hat{b} :

$$SE(\hat{b}) \approx \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{(N - 2) \sum_{i=1}^N (X_i - \bar{X})}$$

$SE(\hat{b})$ brukes videre til å finne t-verdi. T-verdien er lik regresjonskoeffisienten delt på egen standardfeil:

$$t = \frac{\hat{b}}{SE(\hat{b})}$$

Formelen angir hvor mange standardfeil den observerte regresjonskoeffisienten befinner seg fra 0. Etersom min nullhypotese er $H_0: \beta = 0$ så vil risikoen for å forkaste en H_0 som er sann bli mindre jo større t-verdien er. Kritisk t-verdi angir grensen for når H_0 forkastes, og bestemmes av krav til signifikansnivå og antall frihetsgrader. I denne undersøkelsen er signifikansnivået satt til 95 %. Antall frihetsgrader er lik antall observasjoner, minus antall forklaringsvariable. Med disse forutsetningene er kritisk t-verdi som regel $\approx 2,0$ i mine analyser. Med en t-verdi $\geq 2,0$ vil jeg kun stå i fare for å forkaste en rett $H_0 \leq 5\%$ av gangene. Eller det er 95 % sannsynlighet for at det er riktig å forkaste H_0 ved kritisk t-verdi. (Hocking 1976).

2.3.3 Forutsetninger

For at resultatet av en regresjonsanalyse skal være statistisk gyldige, er det noen forutsetninger som må være oppfylt (Meloun & Militky 2001). Disse forutsetningene baserer seg på sannsynlighetsfordelingen til residualene:

- De skal være normalfordelte
- De skal være innbyrdes uavhengige
- De skal konstant varians
- Det må kontrolleres om det finnes enkelte residualer som har ekstrem innflytelse

Om disse forutsetningene er oppfylt kan observeres i et residualplot. Det finnes en rekke statistiske tester som analyserer om disse forutsetningene er oppfylt, men i denne oppgaven nøyer jeg meg med å se på residualplottene.

Normalfordeling

Normalfordelingen er kontrollert ved å observere om residualene danner et tilnærmet klokkeformet histogram. Denne formen indikerer normalfordeling.

Innbyrdes uavhengighet

Innbyrdes uavhengighet innebærer at verdien på variabel x for enhet N i stor grad er upåvirket av verdien på x for enhet $N-1$. Dette er spesielt viktig der når man opererer med tidseriedata, men i denne undersøkelsen representerer x -verdier individuelle trærers dbh, og det er ingen grunn til at dbh-verdien på ett tre skal påvirke dbh-verdien på et annet.

Konstant varians

Residualene skal ha konstant og tilnærmet lik varians uavhengig av forklaringsvariablenes (X) eller målevariabelens (Y) verdi. Dette undersøkes ved å se om det tegner seg et mønster i residualplottet. Dersom det er konstant varians vil det ikke være noe tydelig mønster, og man har da *homoskedastisitet*. I motsatt fall har man *heteroskedastisitet*, som indikerer variansen påvirkes av variablene i modellen.

Ekstrem innflytelse

Enkeltresidualer som ligger mer enn 3 standardavvik fra estimert verdi vil ha spesielt stor innflytelse på hvordan regresjonslinjen legges, og bør kontrolleres som mulige *utliggere*. Dersom slike residualer er et resultat av en feilmåling har jeg fjernet de, men i motsatt fall er de beholdt i datamaterialet. Hvorvidt det er snakk om en feilmåling må vurderes på bakgrunn av detaljene rundt hvert enkelt tilfelle. Et eksempel er om predikert dbh-verdi overstiger kjente dimensjoner i norsk målestokk. Da er det overveiende sannsynlig at det er snakk om en feilmåling.

3 Resultater

Selv om det ble gjennomført 5 skann på hver prøveflate, så er kun ett skann fra flatesenter som er benyttet i denne oppgaven. Regresjonsanalysen viser at det er signifikant sammenheng mellom observert og predikert dbh på alle flater, bortsett fra flate S3-03 når avstand brukes som seleksjonskriterium, og flate S3-14 for begge seleksjonsmetoder. Som ventet forklarer en lineær regresjonsmodell sammenhengen mellom observert og predikert dbh bedre når duplikatene selekteres etter minst avvikende dbh, i stedet for nærmeste tre i avstand. Derfor er tallene fra analysene hvor avstand ble brukt som seleksjonskriterium kun presentert i tabell 2, mens resultatene fra analysene hvor dbh ble brukt som seleksjonskriterium er presentert i større detalj. Residualplottene viser stor forskjell avhengig av om residualene plottes mot observert eller predikert dbh. Plottet mot observerte verdier oppfyller residualene kravene til sannsynlighetsfordeling, og det finnes ikke noe tydelig mønster på noen av flatene. Det er derimot ikke tilfellet når de plottes mot predikerte verdier, og det er disse resultatene som er presentert videre.

Tabell 2. Resultater fra regresjonsanalyse. Tallene i parentes er fra regresjonsanalyse etter fjerning av utligger(e).

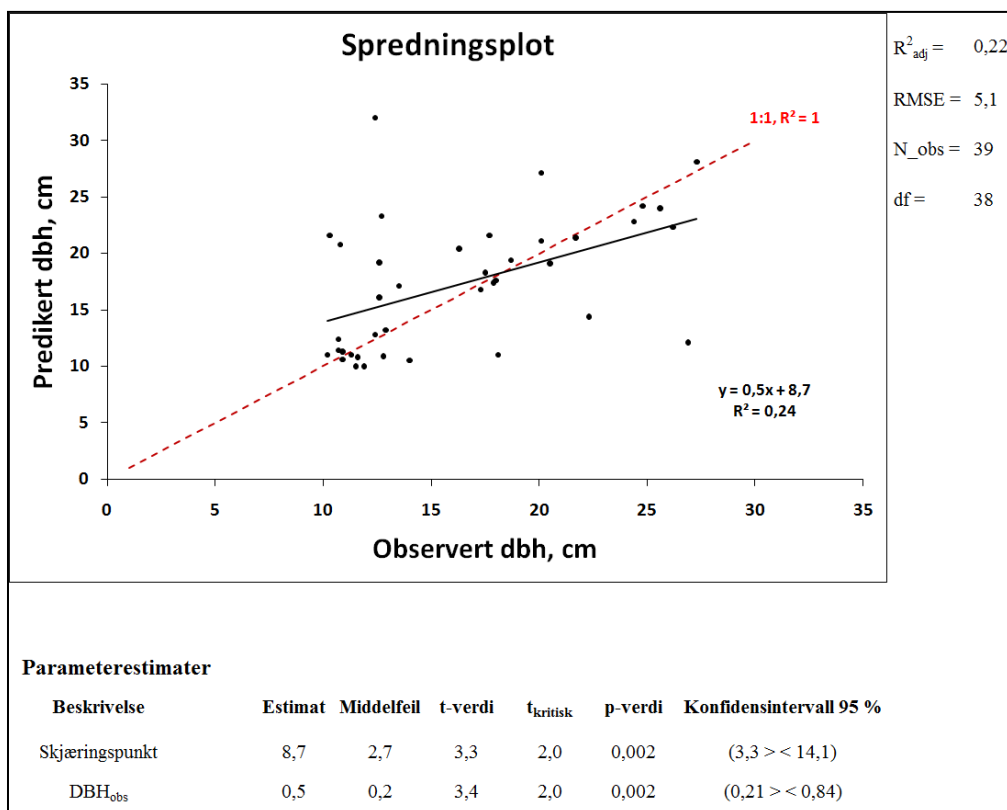
Flate	Duplikat seleksjon	Antall trær	Observert dbh gjennomsnitt, cm	Predikert dbh gjennomsnitt, cm	Standardavvik cm	R ² _{adj}	RMSE cm
S3-03	Avstand	39	17,1	17,5	6,9	0,04	5,7
	Dbh	39	16,1	17,1	5,7	0,22	5,1
S3-05	Avstand	48	21,4	22,9	5,8	0,56	5,7
	Dbh	48	21,5	22,9	3,6	0,82	3,6
S3-13	Avstand	37	24,5	23,8	6,8	0,37	6,2
	Dbh	37	23,8	23,4	4,3	0,71	4,2
S3-14	Avstand	19	25,8	24,6	17,7	0,06	8,6
	Dbh	19	26,8	24,6	14,2	-0,02	13,4
S3-15	Avstand	41	21,8	21,9	2,0	0,82	3,4
	Dbh	41 (40)	21,6 (21,3)	21,0 (21,5)	3,2 (1,8)	0,85 (0,96)	2,9 (1,6)
S3-26	Avstand	34	22,5	22,9	6,4	0,45	5,8
	Dbh	34	22,4	22,9	5,2	0,59	5,1
Alle	Avstand	217	21,8	22,1	7,7	0,32	7,0
	Dbh	217 (215)	21,4 (21,4)	22,0 (21,7)	5,9 (5,5)	0,55 (0,58)	5,7 (5,2)

3.1 Flate S3-03

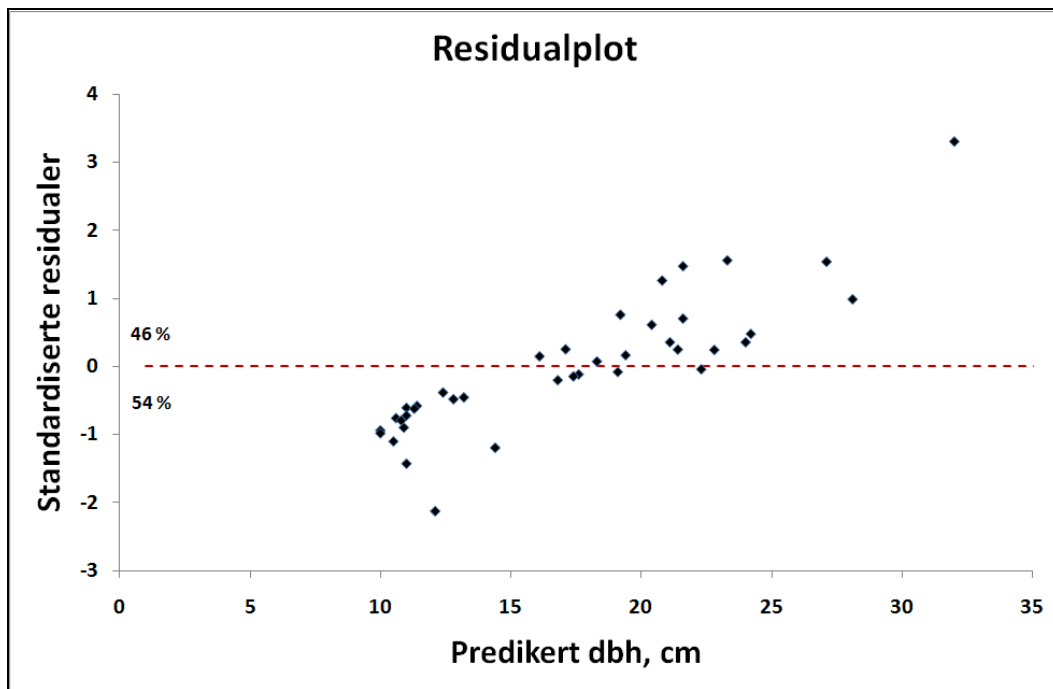
Etter fjerning av trær i skygge og trær med dbh < 10 cm, så var observert treantall redusert til 63 på flate S3-03. Av disse ble 25 fjernet på grunn av kobling til duplikate predikerte trær. Bakkelaseren registrerte 50 trær på samme flate. Av disse ble 12 fjernet som feilregistreringer på grunn av manglende kobling til manuelt registrerte trær. Gjennomsnittlig predikert dbh ble 17,1 cm med en feilmargin på $\pm 5,1$ cm.

Regresjonsmodellen som beskriver sammenhengen mellom predikert og observert dbh på flate S3-03 ble som følger: $DBH_{predikert} = 8,7 + 0,5 \times DBH_{observert}$

H_0 forkastes og H_1 beholdes. Estimaten for β har en t-verdi som er større enn kritisk t-verdi. Det betyr at β er statistisk signifikant forskjellig fra 0, og at det er en signifikant sammenheng mellom observert og predikert dbh. På denne flaten er $R^2_{adj} = 0,22$, som innebærer at kun 22 % av variasjonen i predikert dbh forklares av regresjonslinja. Dermed er 78 % av variasjonen uforklart, og den lineære regresjonsmodellen er dårlig egnet til å forklare sammenhengen mellom observert og predikert dbh på denne flaten.

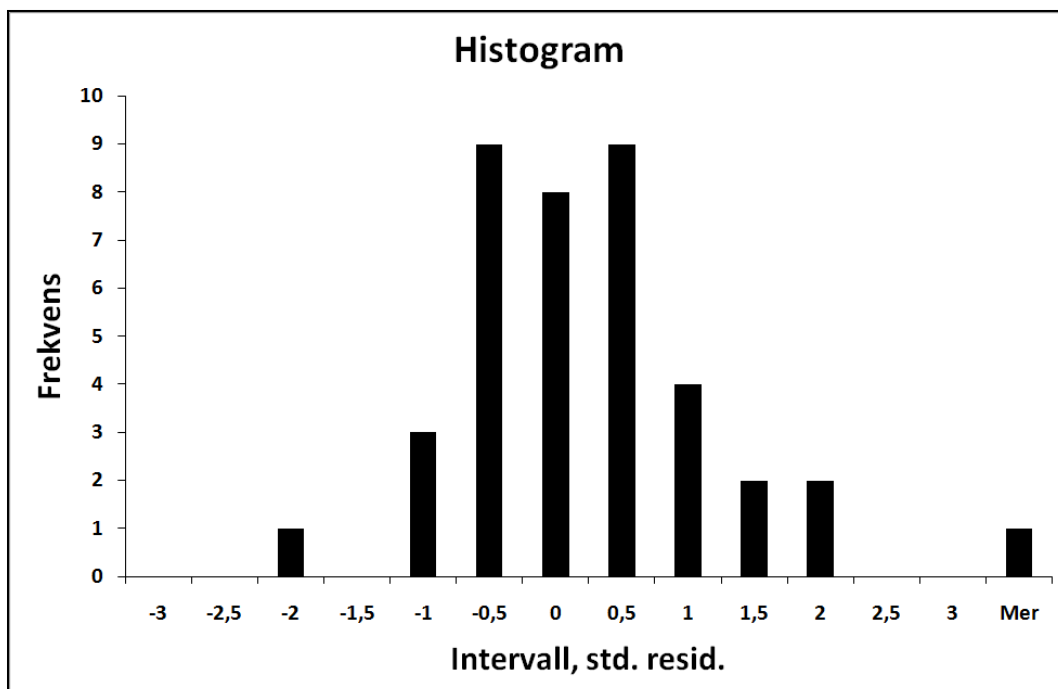


Figur 8. Resultater fra den lineære regresjonsmodellen for sammenhengen mellom observert og predikert dbh på flate S3-03.



Figur 9. Forskjeller mellom estimert og predikert dbh på flate S3-03.

Det er et tydelig mønster som tilsier at regresjonslinja overestimerer små trær med dbh < 15 cm, mens den underestimerer større trær med dbh > 23 cm. Mønsteret indikerer autokorrelasjon, og det må stilles spørsmålstegn ved gyldigheten til analyseresultatene.



Figur 10. Gruppevis fordeling av standardiserte residualer basert på antall standardavvik fra estimert verdi.

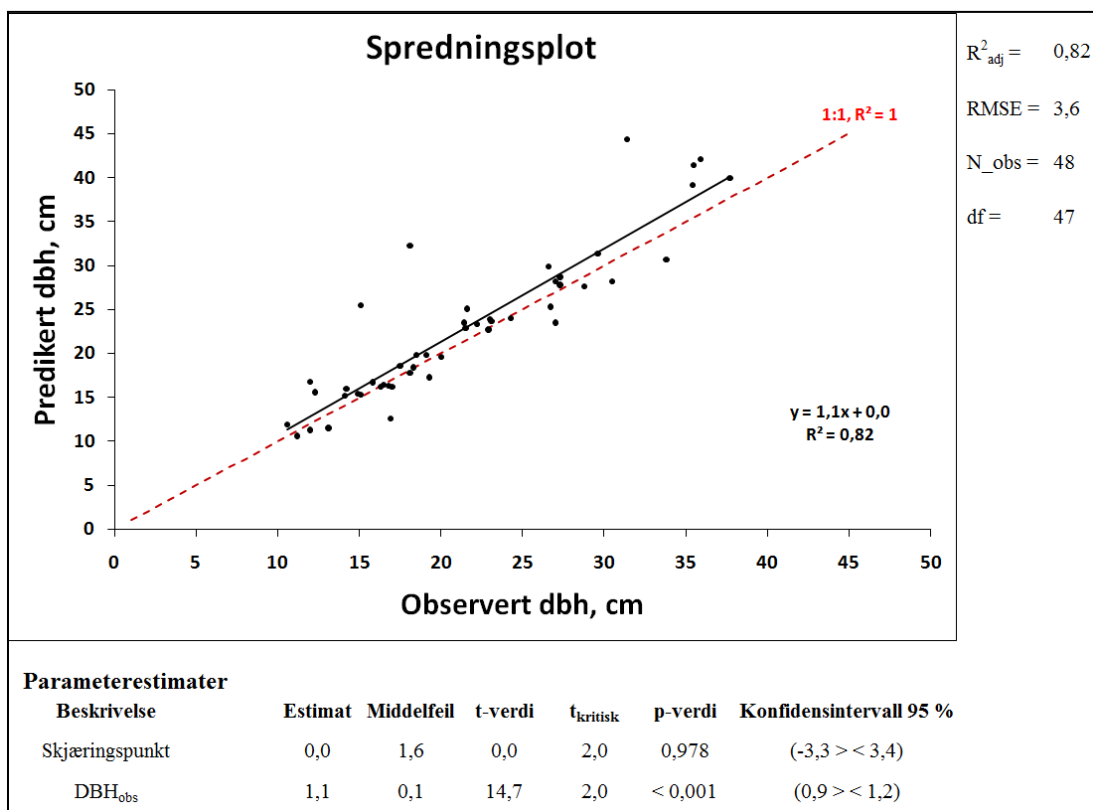
Histogrammet i figur 10 viser en tilnærmet klokkeform, og forutsetningen om normalfordeling synes å være oppfylt på flate S3-03.

3.2 Flate S3-05

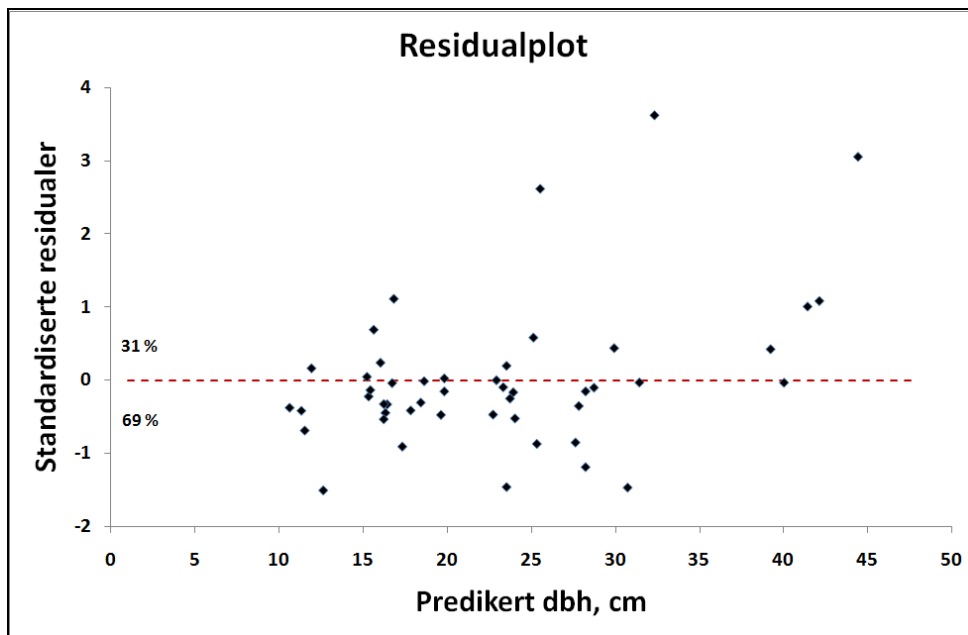
Etter fjerning av trær i skygge og trær med dbh < 10 cm, så var observert treantall redusert til 64 på flate S3-05. Av disse ble 16 fjernet på grunn av kobling til duplikate predikerte trær. Bakkelaseren registrerte 60 trær på samme flate. Av disse ble 12 fjernet som feilregistreringer på grunn av manglende kobling til manuelt registrerte trær. Gjennomsnittlig predikert dbh ble 22,9 cm med en feilmargin på $\pm 3,6$ cm.

Regresjonsmodellen som beskriver sammenhengen mellom predikert og observert dbh på flate S3-05 ble som følger: $DBH_{predikert} = 0 + 1,1 \times DBH_{observert}$

H_0 forkastes og H_1 beholdes. Estimaten for β har en t-verdi som er større enn kritisk t-verdi. Det betyr at β er statistisk signifikant forskjellig fra 0, og at det er en signifikant sammenheng mellom observert og predikert dbh. På denne flaten er $R^2_{adj} = 0,82$ som innebærer at 82 % av variasjonen i predikert dbh forklares av regresjonslinja. Dermed er kun 18 % av variasjonen uforklart, og den lineære regresjonsmodellen er godt egnet til å forklare sammenhengen mellom observert og predikert dbh på denne flaten.

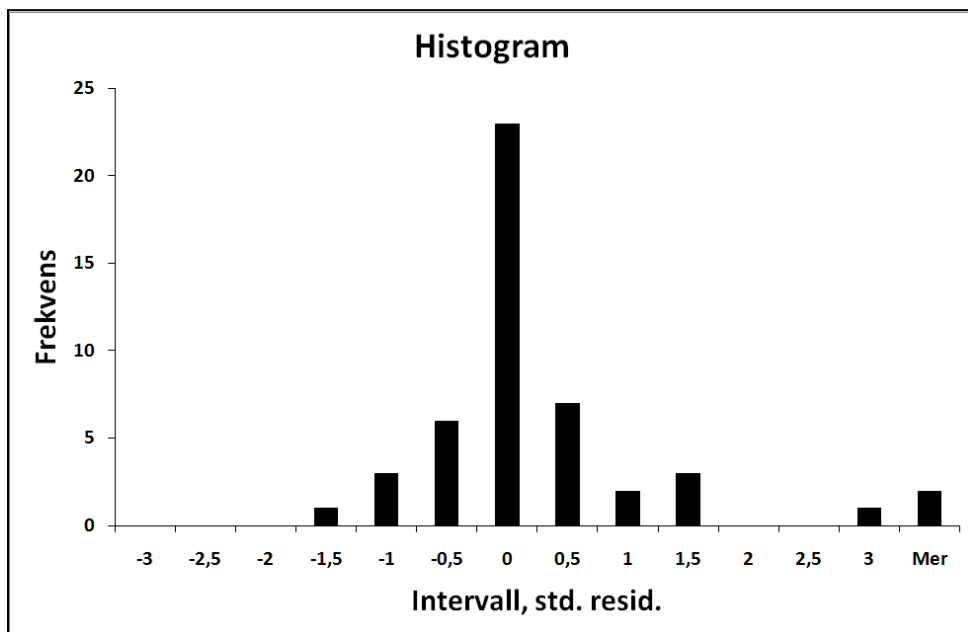


Figur 11. Resultater fra den lineære regresjonsmodellen for sammenhengen mellom observert og predikert dbh på flate S3-05.



Figur 12. Forskjeller mellom estimert og predikert dbh på flate S3-05.

Regresjonslinja overestimerer flere trær enn den underestimerer for alle predikerte verdier. Det er en tendens til at residualene sprer seg ut i vifteform med økende *predikerte* dbh-verdier. Mønsteret indikerer heteroskedastisitet, og det må stilles spørsmålstegn ved gyldigheten til analyseresultatene på flate S3-05.



Figur 13. Gruppevis fordeling av standardiserte residualer basert på antall standardavvik fra estimert verdi.

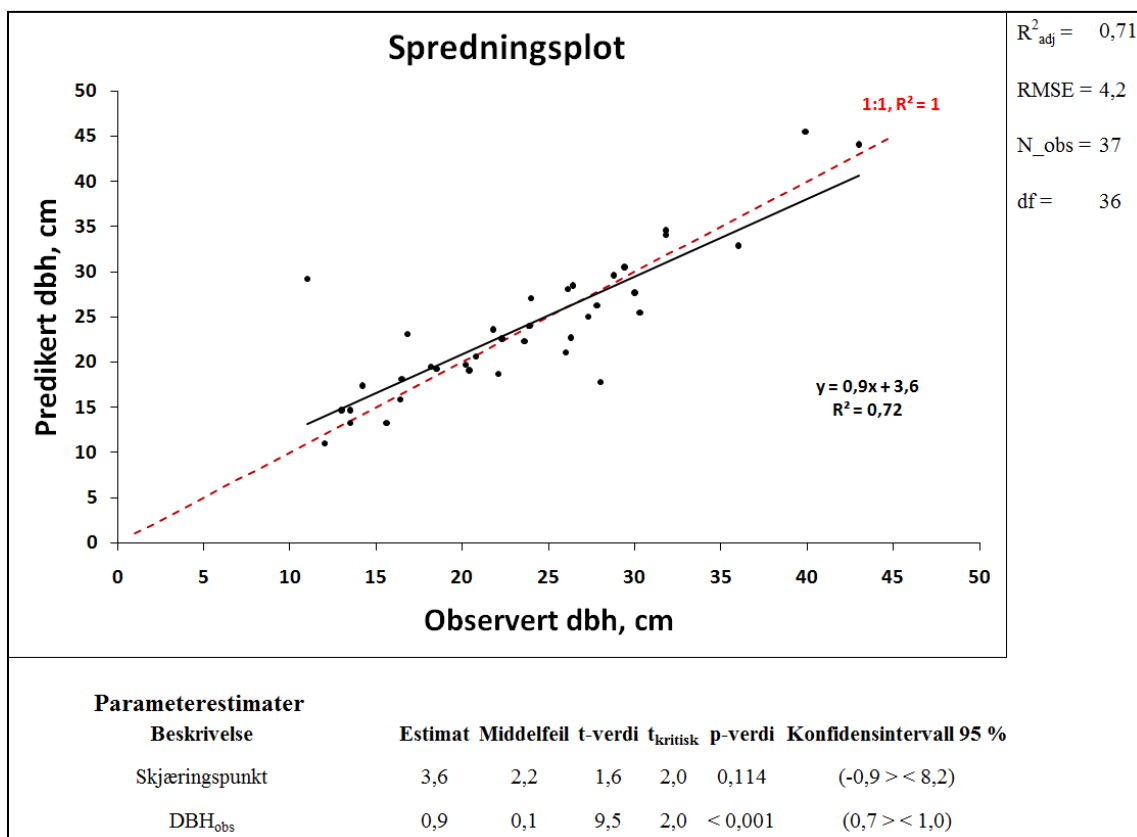
Histogrammet i figur 13 viser en tilnærmet klokkeform, og forutsetningen om normalfordeling synes å være oppfylt på flate S3-05.

3.3 Flate S3-13

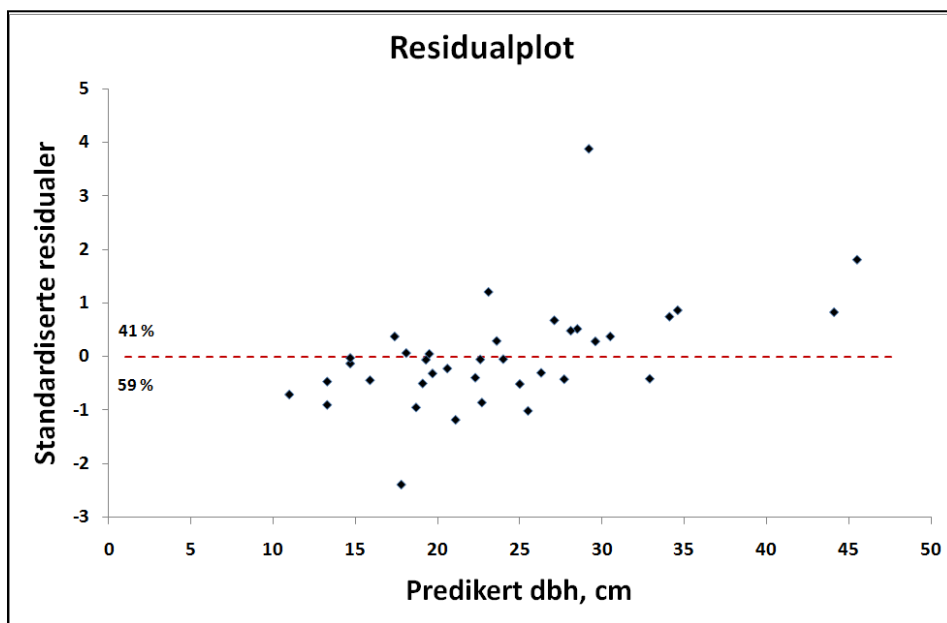
Etter fjerning av trær i skygge og trær med dbh < 10 cm, så var observert treantall redusert til 57 på flate S3-13. Av disse ble 20 fjernet på grunn av kobling til duplikate predikerte trær. Bakkelaseren registrerte 53 trær på samme flate. Av disse ble 16 fjernet som feilregistreringer på grunn av manglende kobling til manuelt registrerte trær. Gjennomsnittlig predikert dbh ble 23,4 cm med en feilmargin på $\pm 4,2$ cm.

Regresjonsmodellen som beskriver sammenhengen mellom predikert og observert dbh på flate S3-13 ble som følger: $DBH_{predikert} = 3,6 + 0,9 \times DBH_{observert}$

H_0 forkastes og H_1 beholdes. Estimaten for β har en t-verdi som er større enn kritisk t-verdi. Det betyr at β er statistisk signifikant forskjellig fra 0, og at det er en signifikant sammenheng mellom observert og predikert dbh. På denne flaten er $R^2_{adj} = 0,71$ som innebærer at 71 % av variasjonen i predikert dbh forklares av regresjonslinja. Dermed er 29 % av variasjonen uforklart, og den lineære regresjonsmodellen er egnet til å forklare sammenhengen mellom observert og predikert dbh på denne flaten.

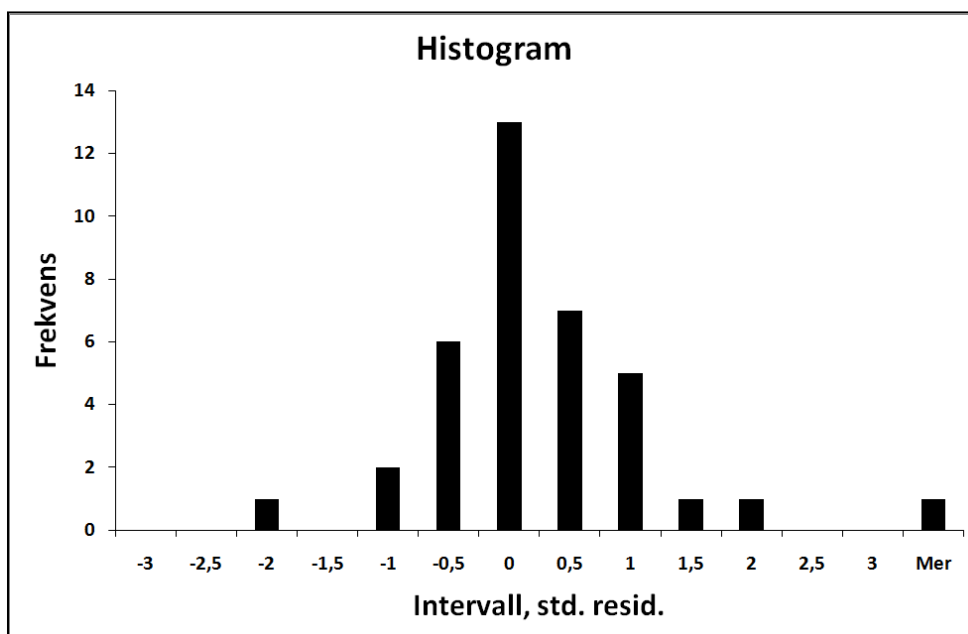


Figur 14. Resultater fra den lineære regresjonsmodellen for sammenhengen mellom observert og predikert dbh på flate S3-13.



Figur 15. Forskjeller mellom estimert og predikert dbh på flate S3-13.

Det samme mønsteret som på flate S3-03, med overestimering av lave predikerte verdier og overestimering av de høye, kan observeres i residualplottet på denne flaten. Tendens er ikke like tydelig, men indikerer autokorrelasjon, og det må stilles spørsmålstegn ved gyldigheten til analyseresultatene på flate S3-13.



Figur 16. Gruppevis fordeling av standardiserte residualer basert på antall standardavvik fra estimert verdi.

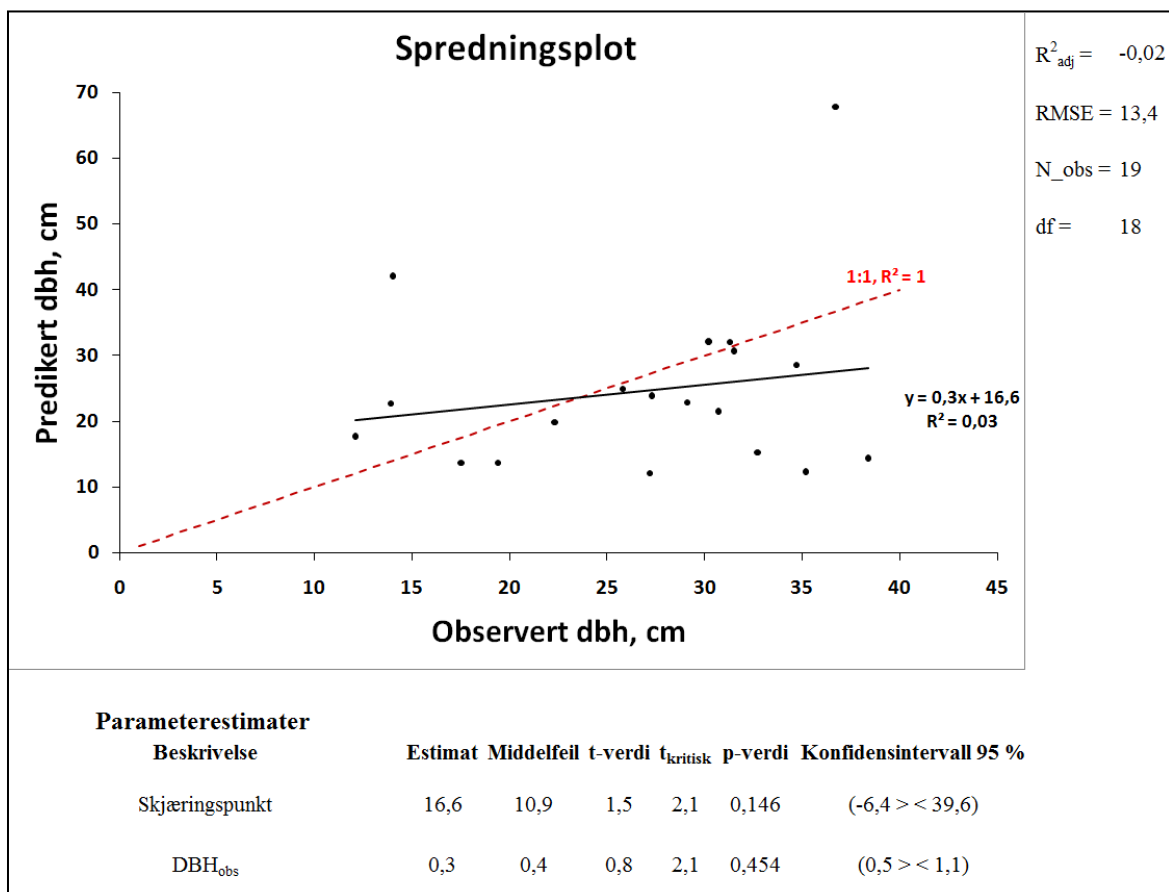
Histogrammet i figur 16 viser en tilnærmet klokkeform, og forutsetningen om normalfordeling synes å være oppfylt på flate S3-13.

3.4 Flate S3-14

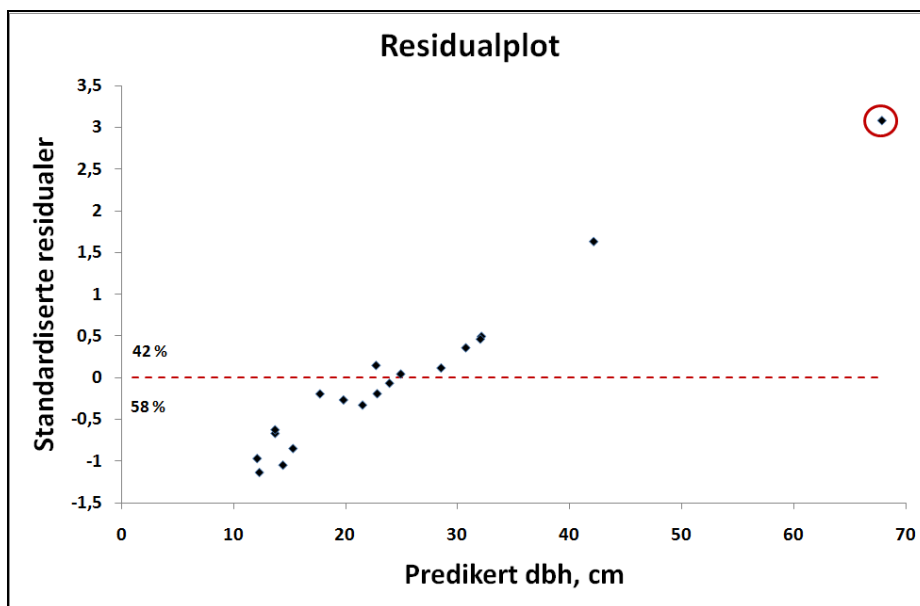
Etter fjerning av trær i skygge og trær med dbh < 10 cm, så var observert treantall redusert til 29 på flate S3-14. Av disse ble 10 fjernet på grunn av kobling til duplikate predikerte trær. Bakkelaseren registrerte 33 trær på samme flate. Av disse ble 14 fjernet som feilregistreringer på grunn av manglende kobling til manuelt registrerte trær. Gjennomsnittlig predikert dbh ble 24,6 cm med en feilmargin på $\pm 13,4$ cm.

Regresjonsmodellen som beskriver sammenhengen mellom predikert og observert dbh på flate S3-14 ble som følger: $DBH_{predikert} = 16,6 + 0,3 \times DBH_{observert}$

H_0 beholdes og H_1 forkastes. Estimaten for β har en t-verdi som er mindre enn kritisk t-verdi. Det betyr at β ikke er statistisk signifikant forskjellig fra 0, og at det ikke er en signifikant sammenheng mellom observert og predikert dbh. På denne flaten er $R^2_{adj} = -0,02$. Det betyr at tilnærmet all variasjon er uforklart ved denne regresjonslinjen.

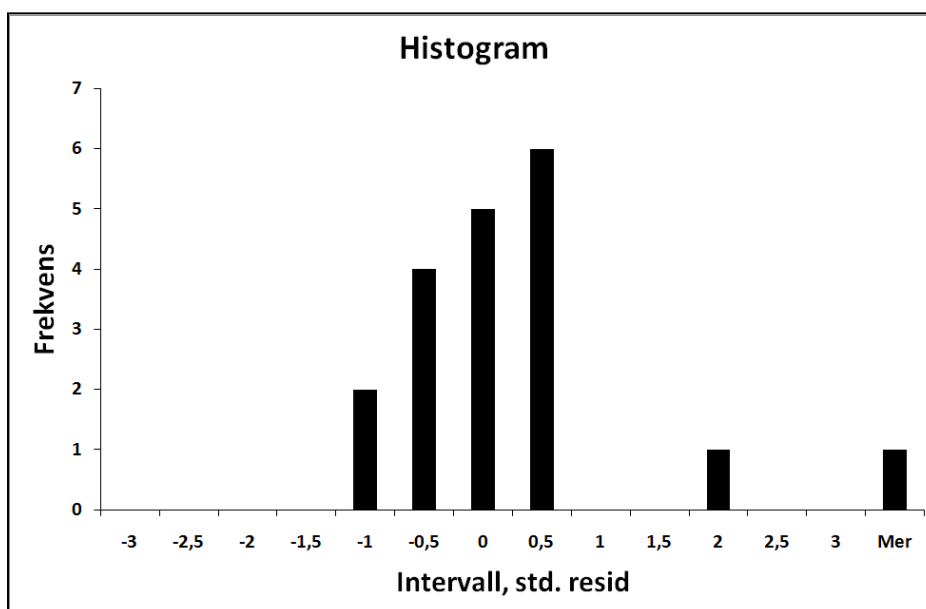


Figur 17. Resultater fra den lineære regresjonsmodellen for sammenhengen mellom observert og predikert dbh på flate S3-14.



Figur 18. Forskjeller mellom estimert og predikert dbh på flate S3-14. Utligger markert med rød ring.

Residualen som er markert med rød ring i figur 18 er åpenbart en utligger ettersom en dbh på nesten 70 cm ikke er observert på denne prøveflaten. Det er en feilregistrering av bakkelaseren. Jeg forsøkte å kjøre ny regresjonsanalyse etter å ha fjernet utliggeren, men dette forbedret ikke resultatet nevneverdig. Residualene viser tegn på autokorrelasjon, som på flate S3-03 og S3-13.



Figur 19. Gruppevis fordeling av standardiserte residualer basert på antall standardavvik fra estimert verdi.

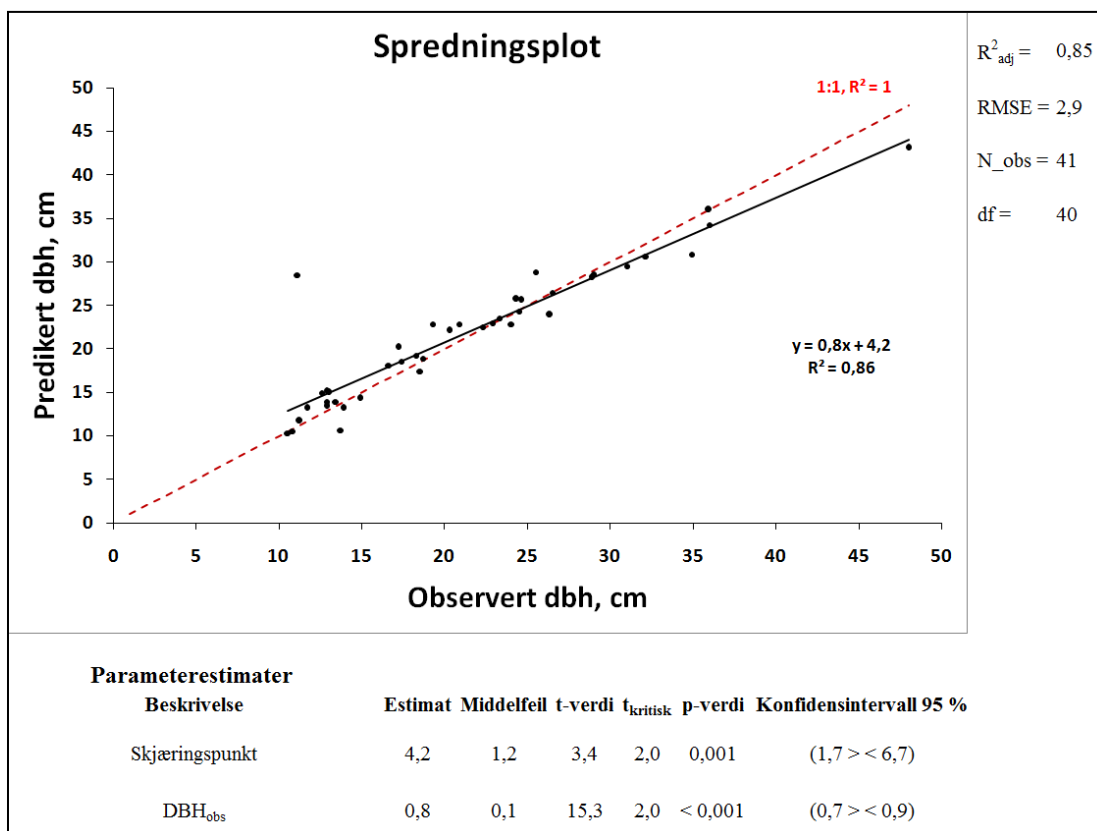
Histogrammet i figur 19 viser en ingen klokkeform, og forutsetningen om normalfordeling er ikke oppfylt på flate S3-14.

3.5 Flate S3-15

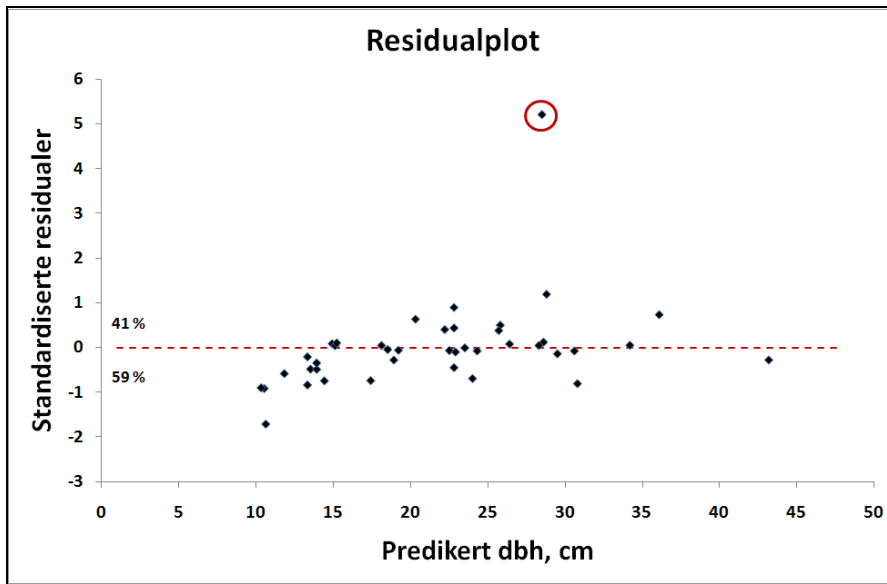
Etter fjerning av trær i skygge og trær med dbh < 10 cm, så var observert treantall redusert til 50 på flate S3-15. Av disse ble 11 fjernet på grunn av kobling til duplikate predikerte trær. Bakkelaseren registrerte 48 trær på samme flate. Av disse ble 7 fjernet som feilregistreringer på grunn av manglende kobling til manuelt registrerte trær. Gjennomsnittlig predikert dbh ble 21,0 cm med en feilmargin på $\pm 2,9$ cm.

Regresjonsmodellen som beskriver sammenhengen mellom predikert og observert dbh på flate S3-15 ble som følger: $DBH_{predikert} = 4,2 + 0,8 \times DBH_{observert}$

H_0 forkastes og H_1 beholdes. Estimaten for β har en t-verdi som er større enn kritisk t-verdi. Det betyr at β er statistisk signifikant forskjellig fra 0, og at det er en signifikant sammenheng mellom observert og predikert dbh. På denne flaten er $R^2_{adj} = 0,85$ som innebærer at 85 % av variasjonen i predikert dbh forklares av regresjonslinja. Dermed er kun 15 % av variasjonen uforklart, og den lineære regresjonsmodellen er godt egnet til å forklare sammenhengen mellom observert og predikert dbh på denne flaten.

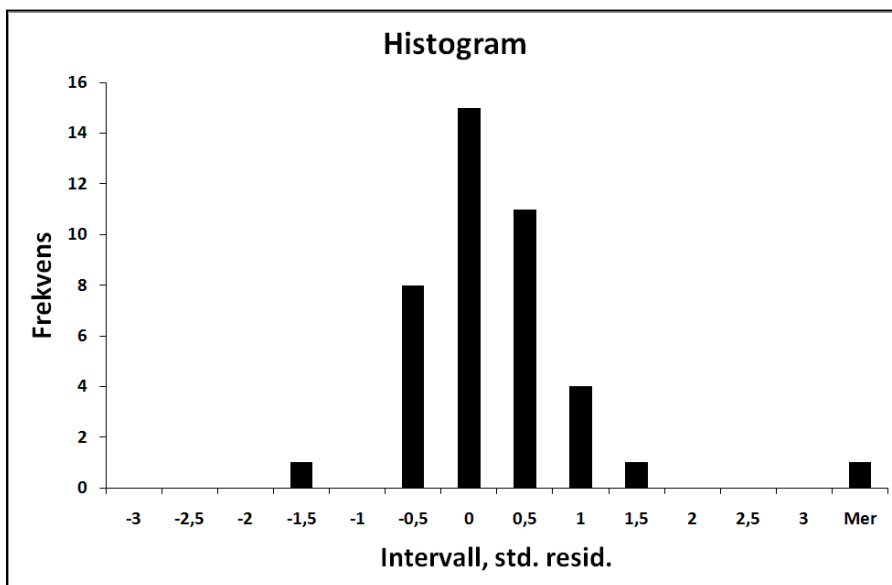


Figur 20. Resultater fra den lineære regresjonsmodellen for sammenhengen mellom observert og predikert dbh på flate S3-15.



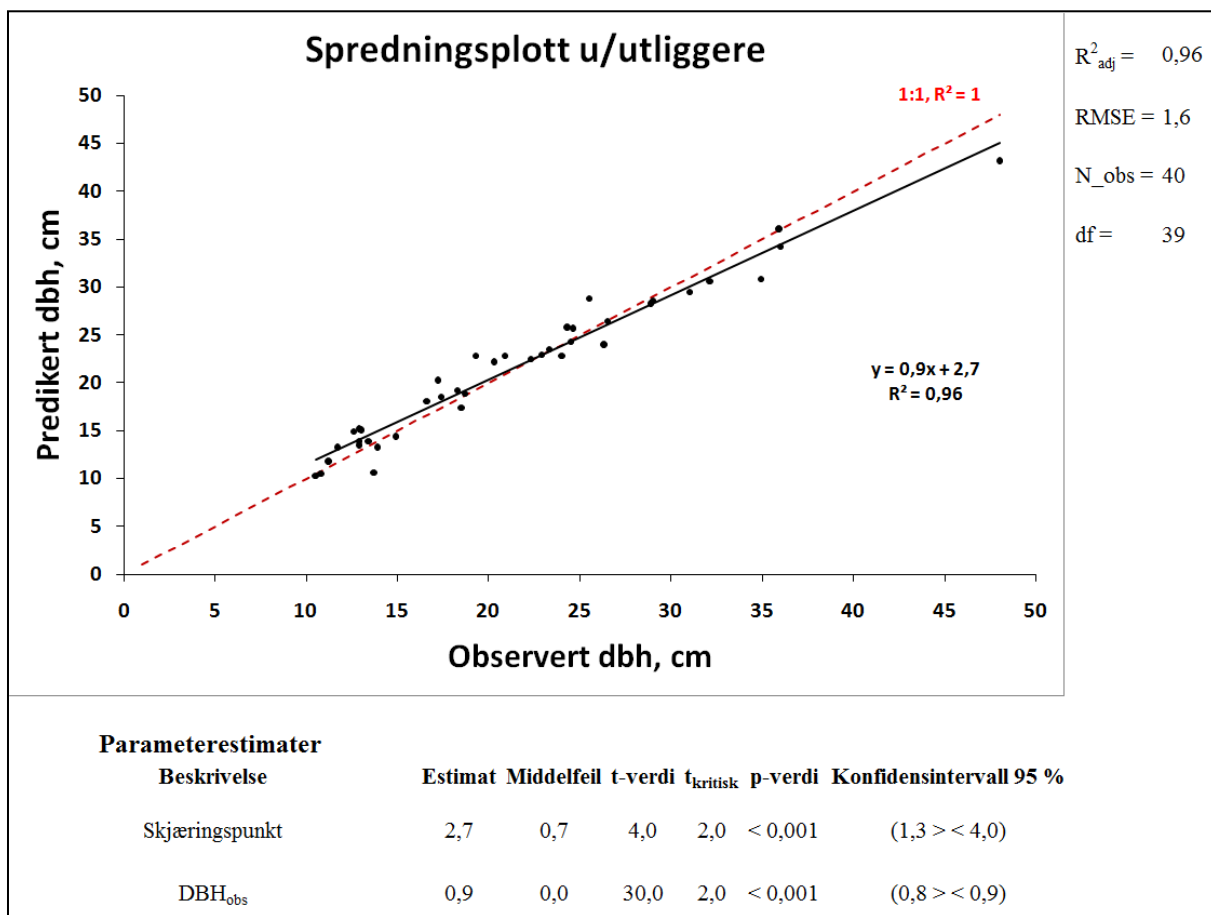
Figur 21. Forskjeller mellom estimert og predikert dbh på flate S3-15. Utligger markert med rød ring.

Residualen som er markert med rød ring i figur 21 er ikke bekreftet som en feil i datamaterialet. Men ettersom den er mer enn 5 standardavvik fra estimert verdi, så valgte jeg å fjerne den og gjøre en ny regresjonsanalyse (se figur 23). Det er også her tegn til autokorrelasjon, men ikke like sterkt som på flate S3-03 og S3-13. Det er også en tendens til et svakt oppadgående buet mønster, noe som kan tyde på at forholdet mellom predikert og observert dbh er påvirket av mer enn to faktorer, og derfor ikke er lineært.



Figur 22. Gruppevis fordeling av standardiserte residualer basert på antall standardavvik fra estimert verdi.

Histogrammet i figur 22 viser en tilnærmet klokkeform, og forutsetningen om normalfordeling synes å være oppfylt på flate S3-15.



Figur 23. Resultater fra den lineære regresjonsmodellen for sammenhengen mellom observert og predikert dbh på flate S3-15 etter å ha fjernet en utligger.

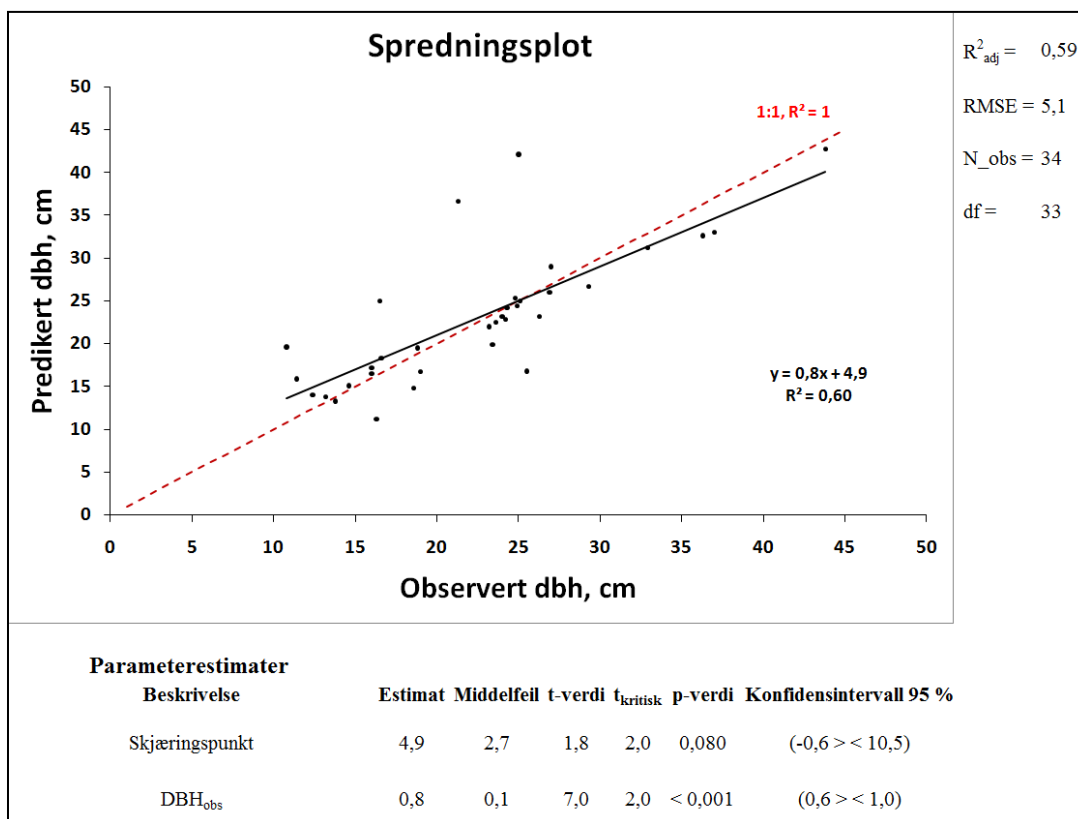
Etter fjerningen av utliggeren var gjennomsnittlig predikert dbh 21,5 cm med en feilmargen på $\pm 1,6$ cm. Regresjonsanalysen viste en sterkere sammenheng mellom observert og predikert dbh. R^2_{adj} og t-verdien økte, mens RMSE ble redusert.

3.6 Flate S3-26

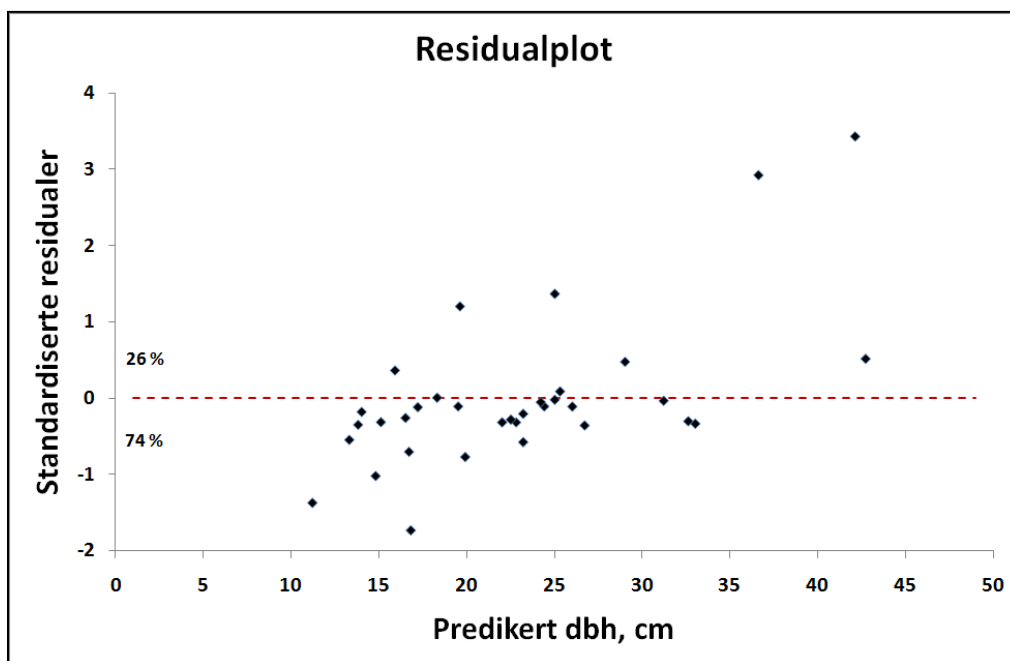
Etter fjerning av trær i skygge og trær med dbh < 10 cm, så var observert treantall redusert til 48 på flate S3-26. Av disse ble 14 fjernet på grunn av kobling til duplikate predikerte trær. Bakkelaseren registrerte 45 trær på samme flate. Av disse ble 11 fjernet som feilregistreringer på grunn av manglende kobling til manuelt registrerte trær. Gjennomsnittlig predikert dbh ble 22,9 cm med en feilmargin på $\pm 5,1$ cm.

Regresjonsmodellen som beskriver sammenhengen mellom predikert og observert dbh på flate S3-26 ble som følger: $DBH_{predikert} = 4,9 + 0,8 \times DBH_{observert}$

H_0 forkastes og H_1 beholdes. Estimaten for β har en t-verdi som er større enn kritisk t-verdi. Det betyr at β er statistisk signifikant forskjellig fra 0, og at det er en signifikant sammenheng mellom observert og predikert dbh. På denne flaten er $R^2_{adj} = 0,59$ som innebærer at 59 % av variasjonen i predikert dbh forklares av regresjonslinja. Dermed er 41 % av variasjonen uforklart, og den lineære regresjonsmodellen er lite egnet til å forklare sammenhengen mellom observert og predikert dbh på denne flaten.

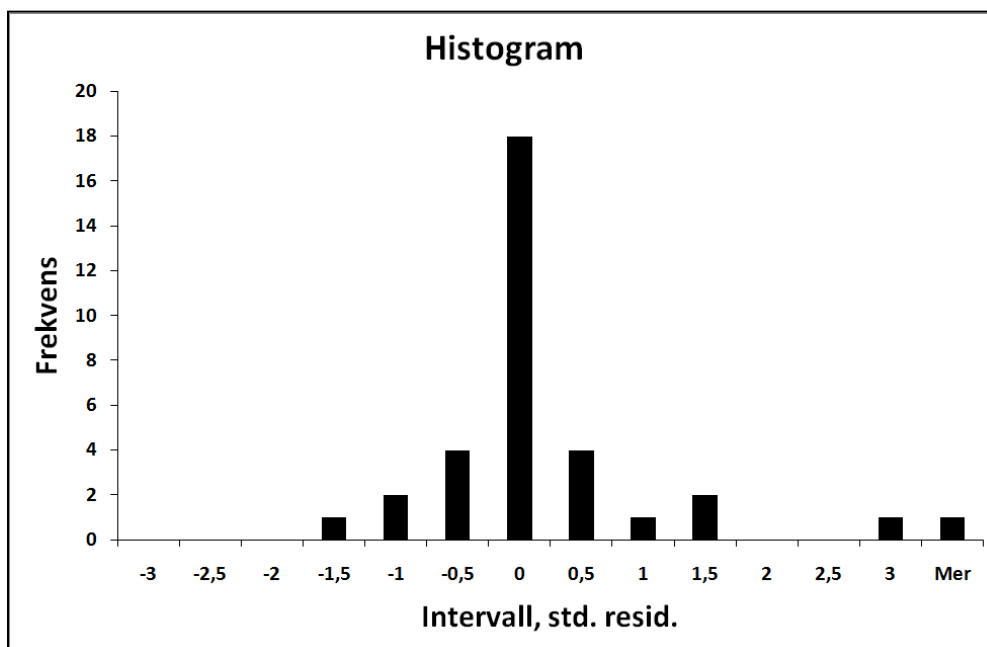


Figur 24. Resultater fra den lineære regresjonsmodellen for sammenhengen mellom observert og predikert dbh på flate S3-26.



Figur 25. Gruppevis fordeling av standardiserte residualer basert på antall standardavvik fra estimert verdi.

Residualplottet for flate S3-26 viser tegn til autokorrelasjon ved at regresjonslinja overestimerer små dimensjoner og underestimerer store.



Figur 26. Gruppevis fordeling av standardiserte residualer basert på antall standardavvik fra estimert verdi.

Histogrammet i figur 26 viser en tilnærmet klokkeform, og forutsetningen om normalfordeling synes å være oppfylt på flate S3-26.

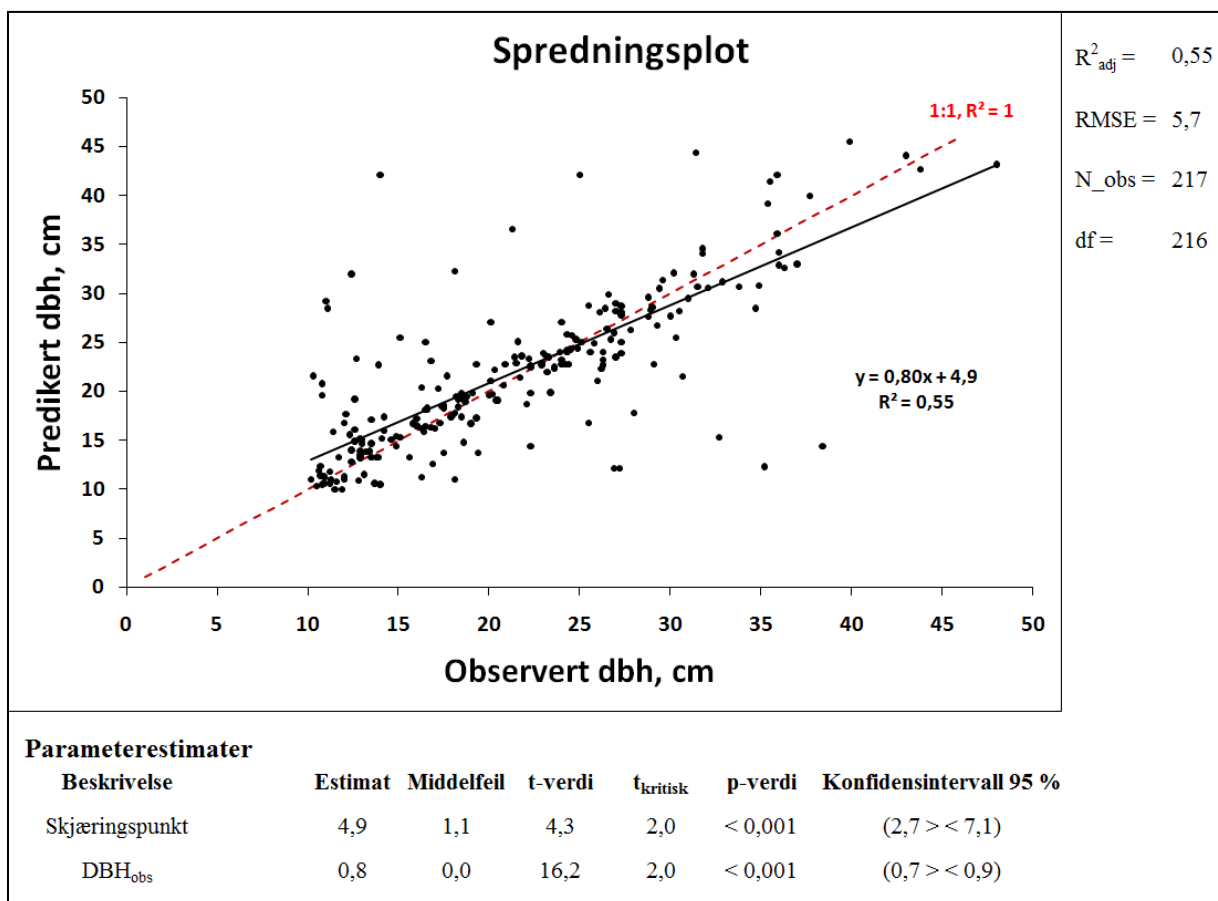
3.7 Alle flater

Gjennomsnittlig predikert dbh for alle trær ble 22,0 cm med en feilmargin på $\pm 5,7$ cm.

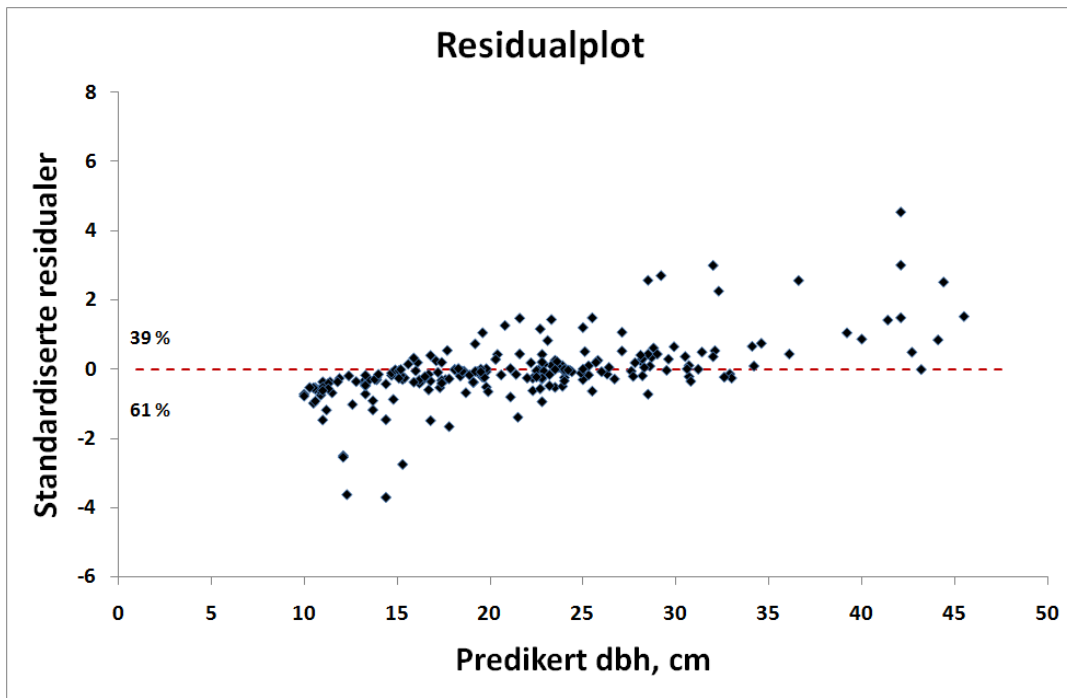
Regresjonsmodellen som beskriver sammenhengen mellom predikert og observert dbh for alle flater sett under ett ble som følger:

$$DBH_{predikert} = 4,9 + 0,8 \times DBH_{observert}$$

H_0 forkastes og H_1 beholdes. Estimaten for β har en t-verdi som er større enn kritisk t-verdi. Det betyr at β er statistisk signifikant forskjellig fra 0, og at det er en signifikant sammenheng mellom observert og predikert dbh. For alle flater er $R^2_{adj} = 0,55$ som innebærer at 55 % av variasjonen i predikert dbh forklares av regresjonslinja. Dermed er 45 % av variasjonen uforklart, og den lineære regresjonsmodellen er lite egnet til å forklare sammenhengen mellom observert og predikert dbh på alle flater sett under ett.

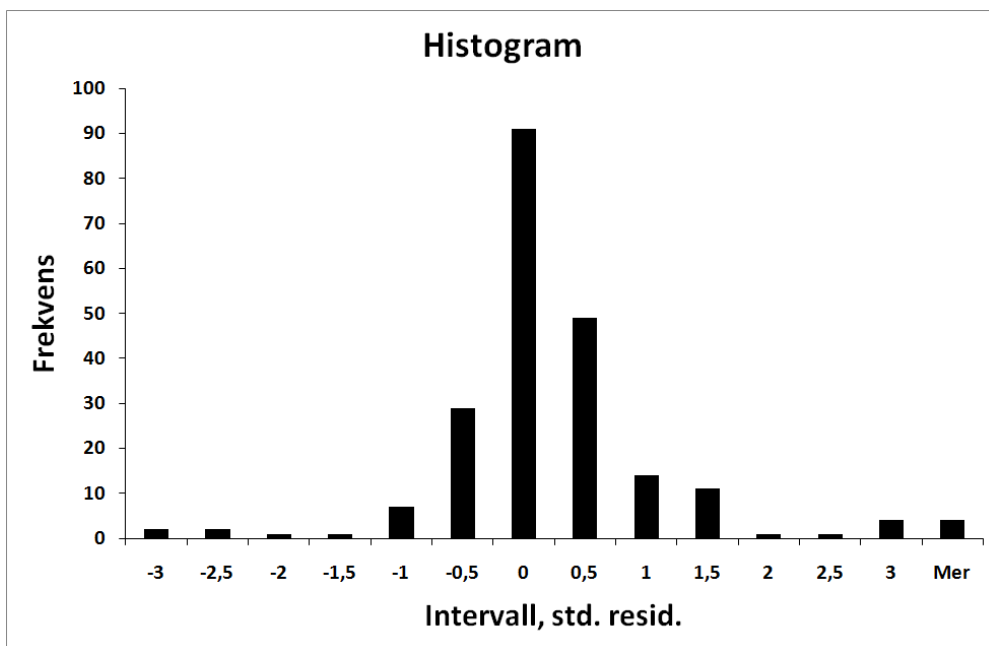


Figur 27. Resultater fra den lineære regresjonsmodellen for sammenhengen mellom observert og predikert dbh for alle flater sett under ett.



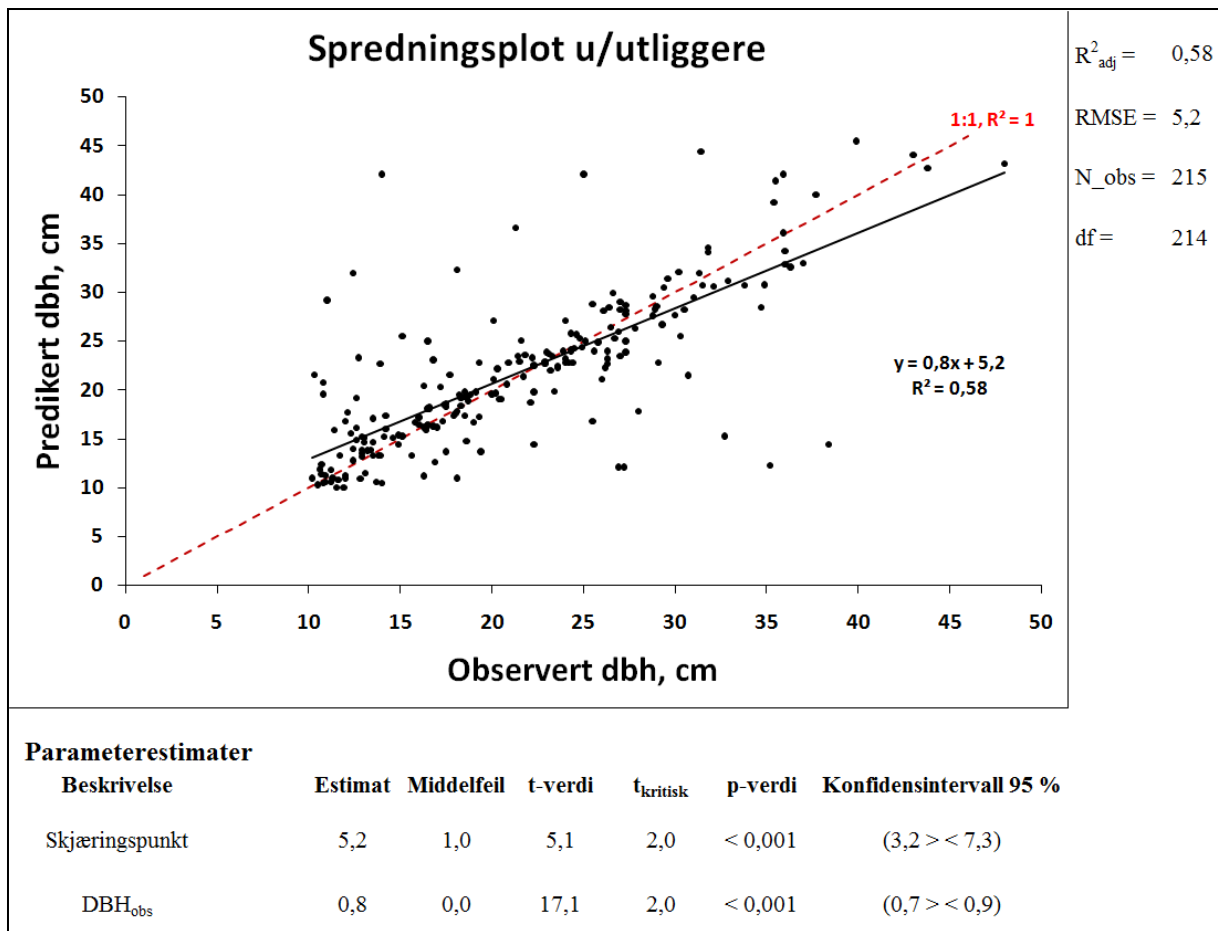
Figur 28. Forskjeller mellom estimert og predikert dbh for alle flater sett under ett.

Residualplottet for alle flater viser på lik linje med flere av enkeltflatene tegn på auto-korrelasjon, og gyldigheten til resultatene fra regresjonsanalysen må trekkes i tvil.



Figur 29. Gruppevis fordeling av standardiserte residualer basert på antall standardavvik fra estimert verdi.

Histogrammet i figur 29 viser en tilnærmet klokkeform, og forutsetningen om normalfordeling synes å være oppfylt for alle flater sett under ett.



Figur 30. Resultater fra den lineære regresjonsmodellen for sammenhengen mellom observert og predikert dbh for alle flater sett under ett, etter fjerning av to utliggere.

Etter fjerningen av utliggeren var gjennomsnittlig predikert dbh 21,7 cm med en feilmargen på $\pm 5,2$ cm. Regresjonsanalysen viste en marginalt sterkere sammenheng mellom observert og predikert dbh. Det var en svak økning i R^2_{adj} og t-verdi, mens RMSE ble redusert med 0,5 cm.

4 Diskusjon

Valg av skannfil

Det er liten sammenheng mellom treffprosent på grunnflatesum og treffprosent på treantall fra TLS-data sammenliknet med den manuelle registreringen. Selv om grunnflatesum er tilnærmet lik feltmålte verdier kan treantall avvike betydelig, og omvendt (figur 6 og 7). Ved manuelt å gå igjennom noen av flatene, og sammenlikne posisjon, treantall og størrelse fra feltregistrering med de utvalgte TLS-filene, fant jeg flere tilfeller hvor feltmålte trær ble feilregistrert av bakkelaseren. Det var tilfeller der bakkelaseren registrerte to små trær som ett stort, ikke registrerte trær som ikke stod i skygge eller registrerte trær som ikke eksisterer. Dette kan forklare den manglende sammenhengen mellom grunnflatesum og treantall i skannfilene. Slike feil kan medføre at det er tilfeldig hvilke TLS-filer som gir best treff på grunnflatesum og treantall i forhold til manuell registrering. Med unntak av åpenbare utliggere vil dette være skjult i den statistiske analysen ettersom den kun forholder seg til tallverdier, og skal man avdekke slike feil må det produseres en grafisk fremstilling av prøveflaten for hver eneste skannfil og manuelt sammenlikne posisjon, treantall og størrelse med feltmålte trær. Dersom det leveres 720 skannfiler for hver prøveflate som i denne oppgaven vil det være en svært tidkrevende jobb. Kvalitetssikringen av hvilken skannfil som best representerer reelle prediksjoner kan derfor ikke sies å være spesielt god. Men mangel på gode alternativer, og krav til begrensnig av arbeidsmengde førte til at valget falt på den filen med laveste maksimale avvik.

Testgrunnlag

For å fjerne duplikater blant predikerte trær måtte jeg velge hvilket observerte tre som skulle brukes som sammenlikningsgrunnlag. Alternativene var nærmeste tre i avstand og det treet med minst avvikende dbh. Av de to alternativer som ble testet, så var det naturlig minst avvikende dbh som resulterte i best sammenheng mellom observert og predikert verdier. Det betyr ikke nødvendigvis at minst avvikende dbh er et mer korrekt sammenlikningsgrunnlag. Det kan godt være at avstand er et bedre seleksjonskriterium, men metoden som er brukt her vil det være umulig å stadfeste. Jeg har heller ikke lyktes i å finne liknende undersøkelser hvor denne problematikken er nevnt. Forekomsten av duplikater var relativt høy i forhold til endelig treantall som ble testet på prøveflatene, og den usikre seleksjonsmetoden kan være en mulig årsak til svake sammenhenger og høye feilmarginer.

Tilvekst

Trærne som ble manuelt klavet vinteren 2007/2008 gikk gjennom to vekstperioder før de ble skannet høsten 2009. Tallene for observert dbh skulle derfor vært justert med en tilvekstfunksjon, men det er ikke gjort i denne studien. Dersom dette hadde vært gjort ville de observerte verdiene vært noe høyere, og det kunne bedret sammenhengen mellom observert og predikert dbh noe, men neppe i betydelig grad.

Residualer

En overvekt av residualene på hver enkelt flate er negative. Det betyr at de positive residualene er større enn de negative, noe som medfører at regresjonslinja i stor grad overestimerer predikert dbh. Forutsetningene for at resultatene fra regresjonsanalysen skal være gyldige henger sammen med sannsynlighetsfordelingen av residualene. Dersom kravene til sannsynlighetsfordeling er innfridd skal det ikke kunne observeres noe mønster i residualplottene. Mine analyser viser derimot tydelige mønster når residualene plottes mot predikerte verdier, men ikke mot observerte verdier. En mulig årsak til dette kan være at feilregistreringer av bakkelaseren oftere for høye, enn for lave. Dersom lysstrålen forstyrres av kvister som ligger tett inntil en trestamme vil bakkelaseren kunne registrere denne kvisten som en del av stammen, og dermed også en høyere dbh enn den som er reell. Følgene kan bli en kunstig høy feilregistrering på enkelte trær, med flere slike tilfeller vil en stor overvekt av de største residualene vil gå i én retning. Slike forstyrrelser og medfølgende feilregistreringer vil ikke skje i motsatt retning. En forstyrrelse forårsaket av en kvist vil ikke føre til at bakkelaseren registrerer en mindre dbh enn den som er reell.

Blant residualene er det mange som ligger mer enn 2 standardavvik fra regresjonslinja, og på grunn av relativt få observasjoner hadde de stor innvirkning på regresjonslinja. Det kan hende det ville vært mer hensiktsmessig med en mindre streng definisjon av hva som regnes som utliggere i mine regresjonsanalyser.

Feilmargin

Feilmarginen på predikert dbh i mine analyser strekker seg fra $\pm 1,6$ cm til $\pm 13,4$ cm på de enkelte flatene, og for alle flater sett under ett var den $\pm 5,2$ cm. Dette er høyere enn forventet. Flate S3-15 er den eneste flaten med feilmargin ned mot det som kan sies å være akseptabelt. Den ekstremt store feilmarginen som ble vist på flate S3-14 er sannsynligvis en følge av enkelte store feilregistreringer i kombinasjon med lavt treantall. Mine resultater viser noe større feilmarginer enn det jeg har funnet i sammenliknbare forsøk. Thies & Spiecker fant

i 2004 en gjennomsnittlig feilmargin på -4,1 % når de sammenliknet et senterskann med manuell klaving. Dessverre er feilmarginen oppgitt som gjennomsnittlig prosentvis avvik fra manuell klaving, og er derfor ikke direkte sammenliknbart med min feilmargin. Men ved å gå inn i tabellen deres og se på dbh verdien er det tydelig at det er snakk om mindre avvik i deres undersøkelse enn i min. Skogområdet de har testet i skiller seg også fra denne undersøkelsen ved å hovedsakelig bestå av løvtrær. Sammenliknet med grantrær har løvtrær mindre tett kvistverk ned mot bakkenivå, og potensialet for forstyrrelser i TLS-dataene er nok mindre. Allikevel beskrives de det som et 'worst case scenario' i forhold til at det skiller seg kraftig fra en ensaldret og oversiktlig plantasjeskog ved å ha stor bredde i diameterdistribusjon, flersjiktning og aldersfordeling. (Henning & Radtke 2006) fant et samlet standardavvik på 2,1 cm ved å sammenlikne TLS-data med manuell klaving på 105 Virakfurutrær (lat. *Pinus taeda*) i Virginia, USA. Dette er også lavere feilmargin enn i mine analyser, men deres tester ble utført i plantasjeskog hvor trærne stod på rekker med fast avstand mellom seg, og buskvegetasjon var kjemisk fjernet. Forutsetningene for å gjenkjenne stammer i TLS-dataene var altså langt bedre i deres undersøkelse. Forsøksdesignet skiller seg mye fra mitt i begge tilfeller, og jeg har ikke lyktes i å finne en studie som beskriver utfordringene jeg møtte når det gjelder å velge blant mange skannfiler fra samme flate, og de problemer som oppstår i forhold til duplikater i datamaterialet.

Praktiske utfordringer ved bruk av TLS til registrering av dbh

Oppsett av totalstasjon, TLS-instrument og målskiver er en tidkrevende prosess, spesielt dersom det skal gjennomføres fem forskjellige skann på hver prøveflate. I denne oppgaven er det kun tatt utgangspunkt i ett skann fra flatens senter. Det begrenser hvor stor andel av stammen som registreres av laserskanneren. Kun den siden av stammen som vender mot bakkelaseren vil bli registrert og resten av stammens tverrsnitt må estimeres under dataprosesseringen. Dette er en mulig feilkilde ettersom stammen antas å være sirkulær under prosessering, noe den svært sjelden er i virkeligheten. Dersom det gjennomføres fem skann fra forskjellige punkter på prøveflaten så vil en større andel av stammen registreres og potensialet for feilregistreringer reduseres (Thies & Spiecker 2004). Men bruk av TLS innen skogtaksering forutsetter en praktisk gjennomførbarhet som ikke avviker mye fra de allerede etablerte metodene. Under arbeidet med å skanne prøveflatene i denne undersøkelsen så var det ikke mulig å gjennomføre 5 skann på mer enn 2-3 flater pr normal arbeidsdag. Det er svært tidkrevende sammenliknet med alternative takseringsmetoder, og utelukker denne metoden i et kommersielt perspektiv. Dersom TLS-registrering skal være aktuell som

takseringsmetode må den ha unike fortrinn som gjør informasjonen mer verdifull sammenliknet med alternativene. Et mulig fortrinn er at TLS-teknologien åpner for en objektiv registrering av tømmerkvalitet i form av stammeprofiler. Tømmerkvalitet har så langt kun vært gjenstand for en subjektiv vurdering, og slik sett vil en objektiv registrering representere ny og verdifull informasjon for skogeieren.

4.1 Konklusjon

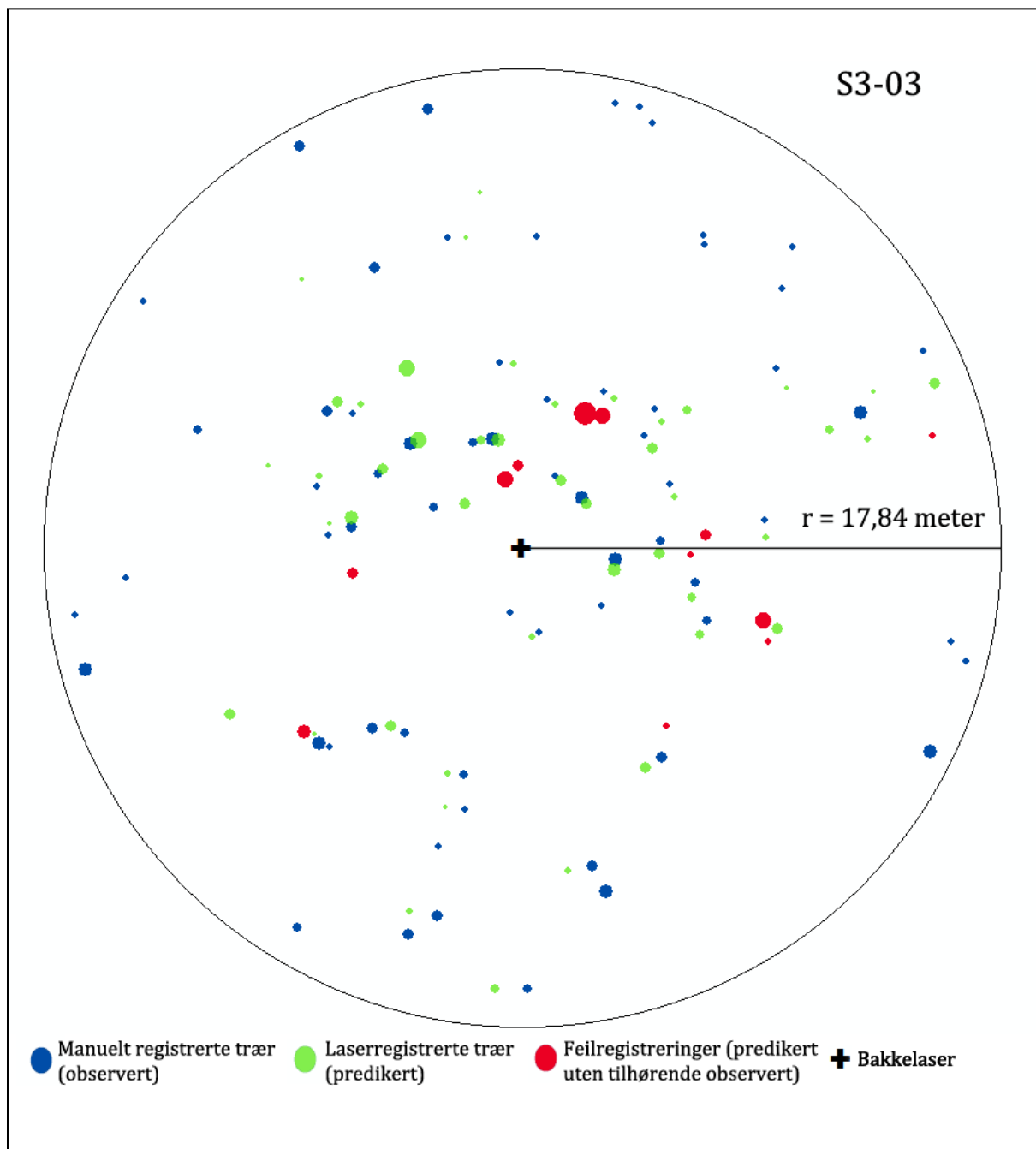
Resultater fra regresjonsanalyser er ofte bedre egnet som beskrivende enn som konkluderende. Hva som er *for stor* feilmargin kan være vanskelig å definere, men den er større enn den som er funnet i liknende studier. TLS-registrering vil ikke være et egnet alternativ som takseringsmetode med de feilmarginene som er funnet i denne oppgaven. Men problemet ligger ikke i teknologien, men i forutsetningene for valg av datagrunnlag. For å øke relevansen på regresjonsanalysene må kvaliteten på datagrunnlaget økes gjennom bedre metoder for valg av skannfil.

5 Litteraturliste

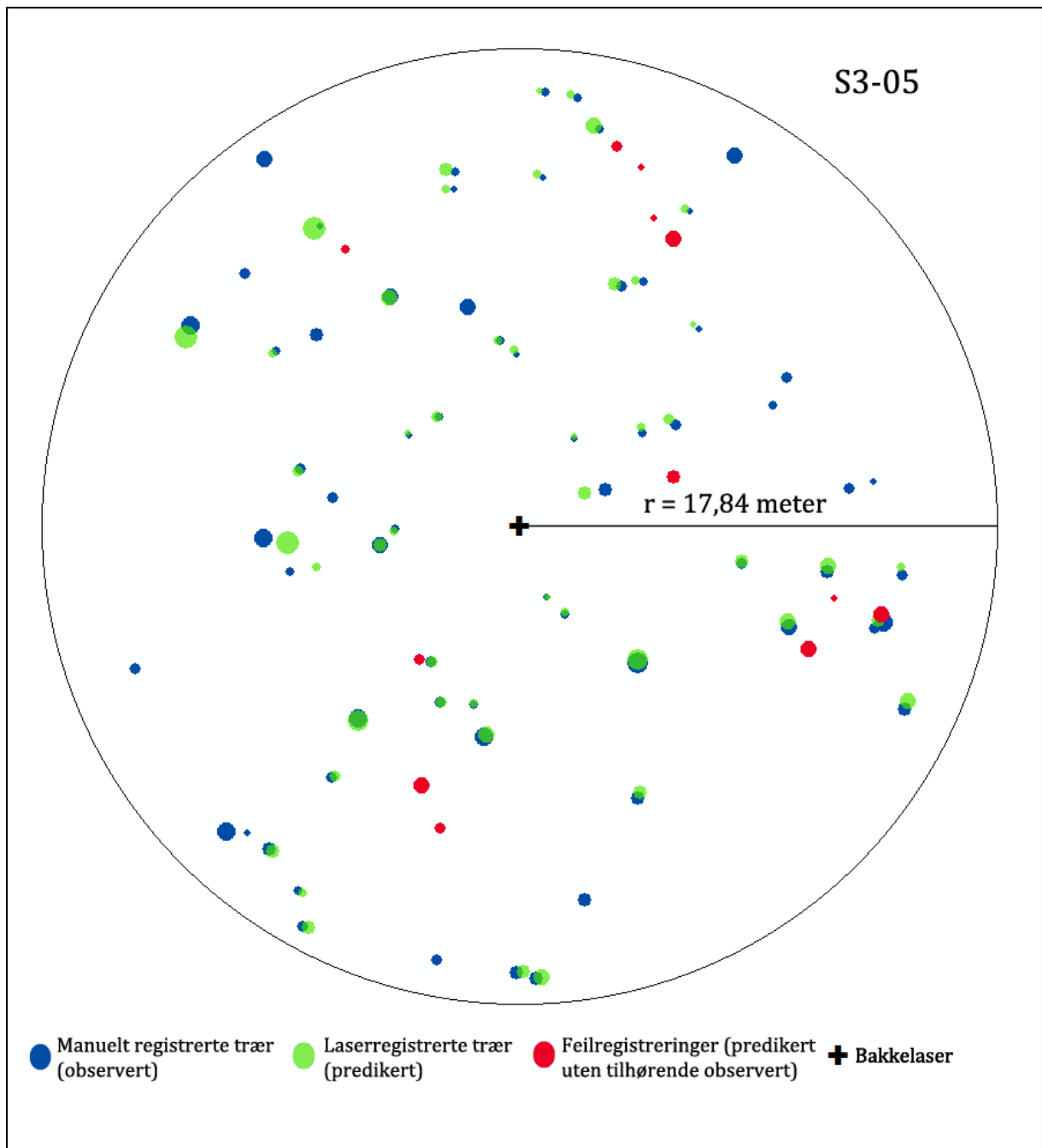
- Henning, J. G. & Radtke, P. J. (2006). Detailed stem measurements of standing trees from ground-based scanning lidar. *Forest Science*, 52 (1): 67-80.
- Hocking, R. R. (1976). Analysis and selection of variables in linear regression. *Biometrics*, 32 (1): 1-49.
- Leeuwen, M. & Nieuwenhuis, M. (2010). Retrieval of forest structural parameters using LiDAR remote sensing. *European Journal of Forest Research*, 129 (4): 749-770.
- Meloun, M. & Militky, J. (2001). Detection of single influential points in OLS regression model building. *Analytica Chimica Acta*, 439 (2): 169-191.
- Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, Hannu, Juha, Maltamo, M., Nilsson, M., Olsson, H., et al. (2004). Laser scanning of forest resources: the nordic experience. *Scandinavian Journal of Forest Research*, 19 (6): 482-499.
- Thies, M., Pfeifer, N., Winterhalder, D. & Gorte, B. G. H. (2004). Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser scanning of standing trees. *Scandinavian Journal of Forest Research*, 19 (6): 571-581.
- Thies, M. & Spiecker, H. (2004). Evaluation and future prospects of terrestrial laser scanning for standardized forest inventories. *In Proceedings of the ISPRS working group VIII/2 Laser-Scanners for Forest and Landscape Assessment*.

6 Vedlegg

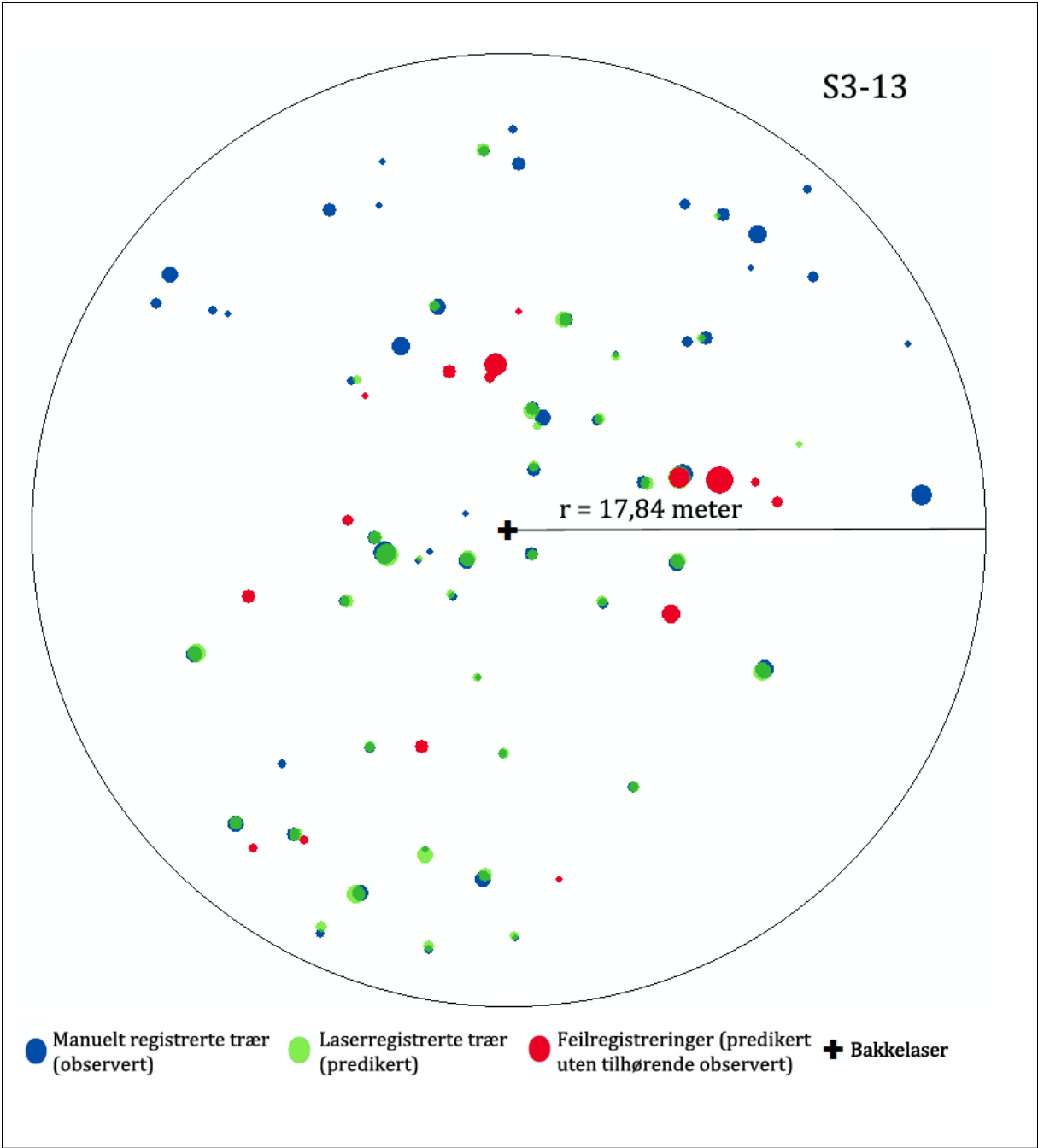
6.1 S3-03



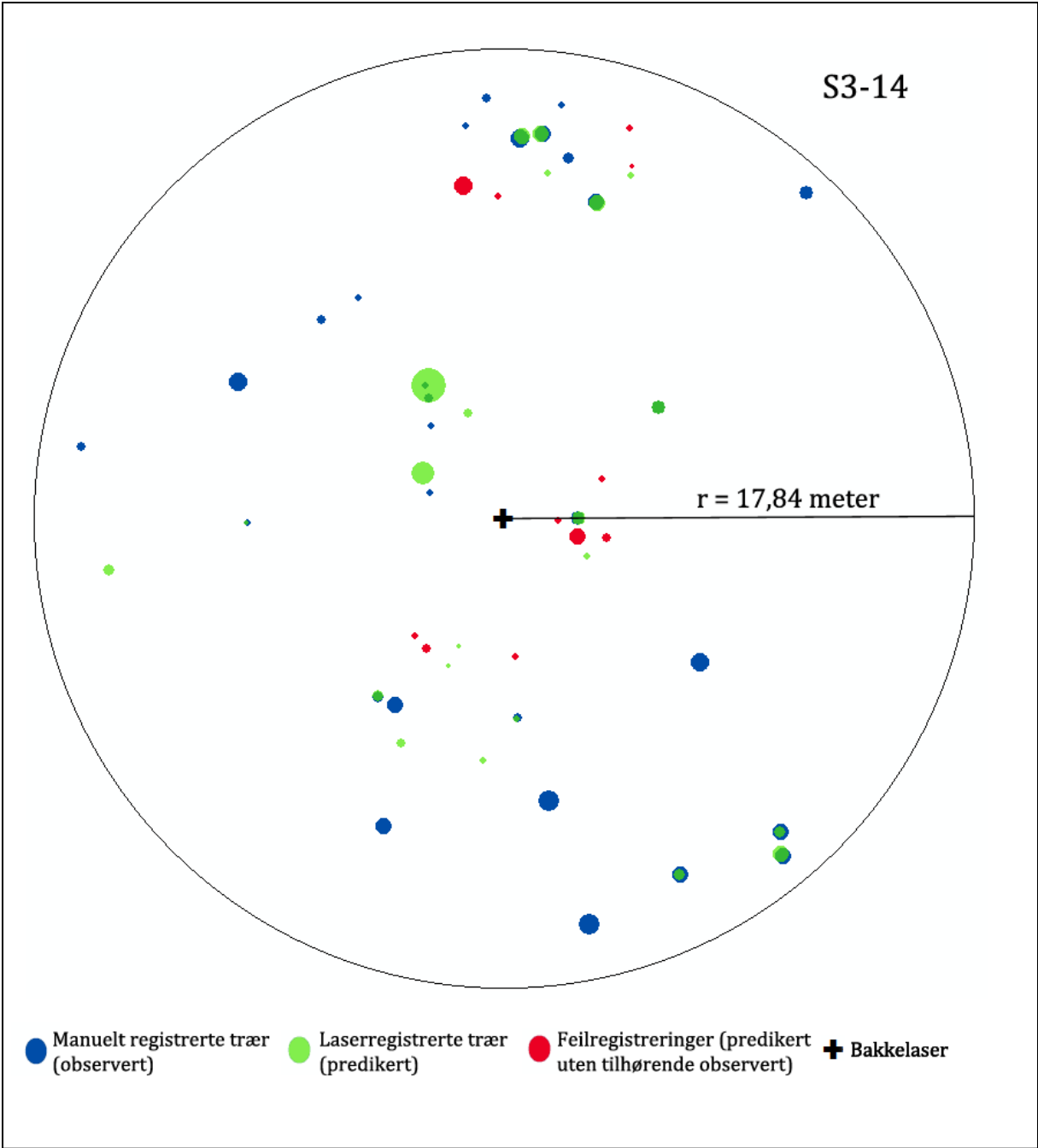
6.2 S3-05



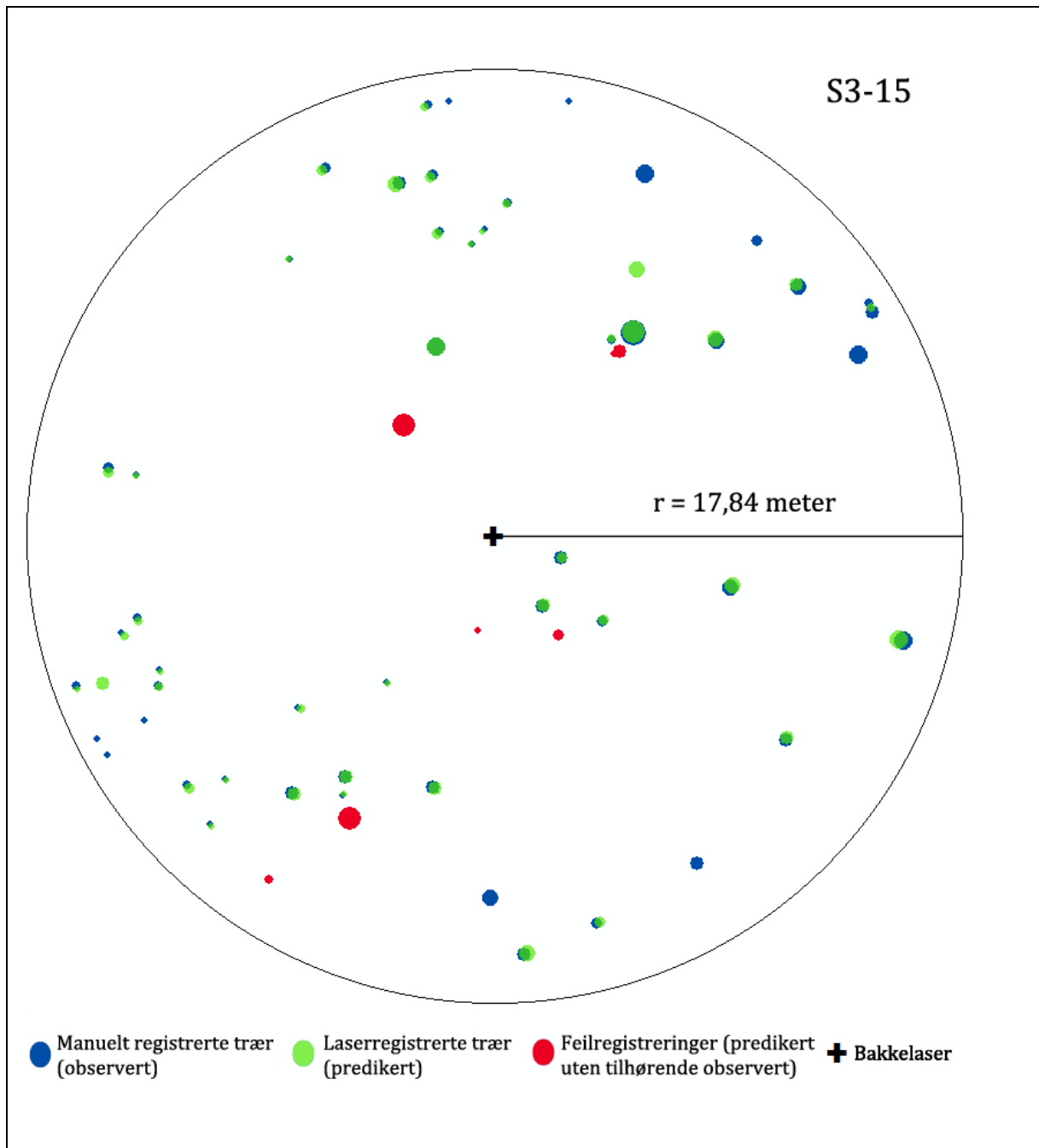
6.3 S3-13



6.4 S3-14



6.5 S3-15



6.6 S3-26

