



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2023 30 stp
Fakultet for realfag og teknologi

Treslagsklassifisering med Sentinel-2 tidsserier, laserdata og klimatiske variabler ved bruk av random forest

Tree species classification with Sentinel-2 time series, laser data and climatic variables through random forest

Helga Sofie Gisholt Jønland
Geomatikk

Forord

Denne masteroppgaven er skrevet for Norsk institutt for bioøkonomi (Nibio), og jeg har gjennom dem mottatt datagrunnlag og veiledning i masterarbeidet.

Det å skrive masteroppgave har vært et spennende prosjekt, der jeg har fått lov til å grave meg ned i analyser på egne premisser og med et team i ryggen som har hjulpet meg med å begrense de endeløse mulighetene jeg startet med. Jeg vil først og fremst takke mine veiledere: Ivar Maalen-Johansen (hovedveileder) for hjelp med å forme arbeidet til en oppgave, Dagrun Aarsten (biveileder) for innspill, ideer og gode tilbakemeldinger og ikke minst en stor takk til Johannes Breidenbach (biveileder) for tett oppfølging og for å ha tatt meg inn i varmen på Nibio.

Videre har jeg vært så heldig å få gjesteplass i møter med SR16-gjengen på Nibio, som har kommet med gode innspill på utfordringer og ideer til retninger å utforske underveis i prosessen. Marius Hauglin har også vært til stor hjelp med forståelse og uthenting av data.

Videre vil jeg takke mamma og pappa for å ha hjulpet meg med korrekturlesning og for moralsk støtte. Takk også til min viktigste støttespiller og søster Hanna for gode råd underveis.

Jeg vil til slutt takke for mye god støtte og sårt trengte avbrekk fra Vendel-Agathe, Mathilde, Anna og Mathilde, og mine medsamsvorne geomatikkmasterskrivere for et utløp for frustrasjon og for mange gode vafler.

Ås, 15. mai 2023

Helga Sofie Gisholt Jønland

Oppsummering

Formålet med dette oppgaven er å se på mulighetene for å bruke multispektrale satellittbilder fra Sentinel-2, klimatiske variabler og laserdata for å skille gråor og eik fra andre løvtrær og ulike granarter fra hverandre. Klassifiseringsmuligheter er undersøkt med random forest, og fremgangsmetoden er basert på norsk skogsinventar fra Landsskogstakseringen. Oppgaven vil ta for seg virkningen av områdeavgrensninger for klassifisering, utjevning av observasjonsantall for ulike treslag, og en variabelanalyse.

Oppgaven presenterer en lovende modell for å skille ulike typer granarter for et begrenset geografisk område, og to mindre presise RF-klassifiseringsmodeller for deteksjon av eik og gråor blant løvtreslag. Felles for alle modellene er viktigheten av klimatiske variabler og NDVI for løvtreslagsklassifiseringene, og liten forbedring i klassifiseringene ved tilførsel av laserdata.

Abstract

The motivation behind this thesis is to explore the possibility for using multispectral satellite images from Sentinel-2, climatic variables, and laser data to identify oak and grey elder within deciduous tree species and differences between types of spruce. Classification is explored with the machine learning algorithm random forest, and the method is based on the Norwegian forest inventory data. Impact of narrowing down search areas in modelling, equalization of observation count between tree species and variable analysis will be analyzed in this paper.

This thesis presents a promising model for classification of different spruce species for a restricted geographical area, and two less precise RF classification models for detection of oak and grey elder among deciduous trees. Climatic variables were important in classification for all models and no NDVI were important in deciduous classification. No models improved significantly after adding laser variables.

Innholdsfortegnelse

1 Introduksjon	1
2 Teori.....	2
2.1 Sentinel-2.....	2
2.2 NDVI.....	2
2.3 Landsskogstakseringen.....	3
2.4 Random Forest	4
2.5 Statistikk	5
2.6 Kryssvalidering.....	6
3 Metode	7
3.1 Datapresentasjon	7
3.2 Variabler og deres korrelasjon	9
3.3 Manglende verdier	11
3.4 Variabelutvelgelse	12
3.5 Treslag med flere enn 15 observasjoner	13
3.6 Outliere.....	14
3.7 Jevne ut antall observasjoner per treslag	16
3.8 Klassifisering av treslag.....	16
3.8.1 Granklassifisering	17
3.8.2 Eikedeteksjon	20
3.8.3 Gråordeteksjon.....	21
3.8.4 Forskjellen mellom løvtreslag.....	22
3.9 Utjevning for variasjon i antallsreduerte treslag og kryssvalidering	22
4 Resultater	23
4.1 Treslagsutvelgelsen	23
4.2 Gran	25
4.3 Eik	28
4.4 Gråor.....	30
4.5 Fordeling i to løvtreklasser	32
5 Diskusjon	33
5.1 Datagrunnlaget.....	33
5.2 Metode	34
5.2.1 Interpolasjon	34
5.2.2 Observasjoner – antall og fordeling	34
5.2.3 Variabelutvelgelse og korrelasjoner	34
5.2.4 Kryssvalidering og utjevning for observasjonsreduksjon	35

5.3 Random forest som klassifiseringsverktøy	35
5.4 Vurdering av resultater	36
5.4.1 Gran	36
5.4.2 Eik	37
5.4.3 Gråor.....	38
5.4.4 Klassifisering med fordelte løvtrær på 2 klasser	38
5.5 Forslag til videre arbeid	40
6 Konklusjon	41
7 Referanser	42
Vedlegg 1	44
Vedlegg 2	46

1 Introduksjon

Skognæringen er en av de viktigste næringene i Norge, og har lange tradisjoner (Breidenbach, Granhus, et al., 2020). Det å kartlegge ressursene gir et bedre grunnlag for bærekraftig drift. Dette innebærer alt fra oversikt over innholdet i norsk skog til måling av miljøpåvirkninger (van Aardt & Wynne, 2007), biodiversitet, dyrehabitat og overvåkning av uønskede arter (Fassnacht et al., 2016; Grabska et al., 2019). Treslagsdeteksjon er en del av forskningsarbeidet i Landsskogstakseringen, som gjøres for å holde oversikt over de norske skogsområdene. Her blir storskala data fra blant annet fjernmåling brukt i kartlegging og analyser av forekomster av ulike arter og fenomener (Breidenbach, Granhus, et al., 2020)

Satellittdata er tidligere brukt for å skille furu, gran og løvtrær fra hverandre, som gjort i skogressurskartet SR16, utviklet av NIBIO (Norsk institutt for bioøkonomi). Ved hjelp av satellittbilder fra ulike tidspunkter på året er det mulig å skille disse tre overordnede klassene med en total nøyaktighet på 90% (Breidenbach, Waser, et al., 2020). Både multispektrale bilder med refleksjonsverdier utover synlige bølgelengder, og tidsaspektet i tidsseriene kan bidra med å fange opp ulikheter i treslagene (Grabska et al., 2019).

Sentinel-2 tidsseriedata danner et grunnlag for undersøkelsene gjort i denne oppgaven, der muligheten for å skille ut Eik og Gråor fra andre løvtrær er undersøkt, samt muligheten for å skille mellom ulike granarter. I tillegg er klimatiske variabler og laserdata undersøkt i kombinasjon med tidsseriene for å se om disse har noen effekt i en *random forest*-klassifisering.

2 Teori

2.1 Sentinel-2

Sentinel-2 (S2) er en satellitttype, bestående av to satellitter (Sentinel-2A og Sentinel-2B) (1 ESA, 2023), som hyppig passerer over Norge. MSIn (Multispectral Instrument) fanger opp bølgelengder i VNIR og SWIR (Visible near infrared og Short-wave infrared) mellom 443nm og 2190nm fordelt på 13 bånd som vist i Tabell 1 (2 ESA, 2023).

Tabell 1 Oversikt med informasjon om bånd fra Sentinel-2 og hvilke som er brukt i datasettet (2 ESA, 2023).

Båndnavn	Bølgelengde [nm]	Oppløsning [m]	I datasettet
B02	490	10	x
B03	560	10	x
B04	665	10	x
B08	842	10	-
B05	705	20	-
B06	740	20	-
B07	783	20	-
B8A	865	20	x
B11	1610	20	x
B12	2190	20	x
B01	443	60	-
B09	940	60	-
B10	1375	60	-

Bånd B8A, B11 og B12 representerer nærinfrarødt lys. Det er dermed forventet å få høy refleksjon i områder med vegetasjon. I analysen er det ikke bruk for begge 8-båndene da disse representerer cirka den samme delen av det elektromagnetiske spekteret, men med ulik oppløsning. I og med at datasettet brukt i denne oppgaven har bånd med 20 meters oppløsning, er alle bånd gjort om til 20 meters piksler for at det skal samsvare. Derfor er også B8A tatt i bruk, heller enn å velge det tilsvarende båndet B8 med høyere oppløsning.

2.2 NDVI

NDVI står for *Normalized Difference Vegetation Index* og sier noe om friskheten til vegetasjon, og regnes ut basert på forholdet mellom reflektert lys i de røde og nærinfrarøde bølgelengdene fra vegetasjon. NDVI regnes ut med følgende formel:

$$NDVI = \frac{(NIR - Rød)}{(NIR + Rød)}$$

I utregningen i datasettet er bånd B8A brukt for nærinfrarød (NIR) reflektert verdi, og B04 er brukt for reflektert rød verdi i utregning av NDVI. Det vil si at for hver av ukene til alle observasjonene er NDVI regnet ut slik:

$$NDVI = \frac{(B8A - B04)}{(B8A + B04)}$$

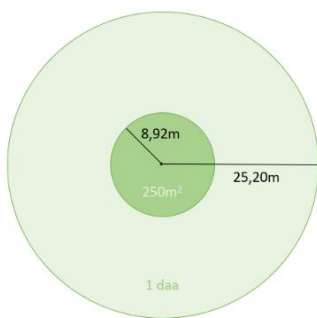
NDVI-en kan variere mellom -1 og 1 og vegetasjonen vil være friskere med økt NDVI. En plate eller et tre med NDVI mellom -1 og 0 er gjerne død (Huang et al., 2021).

2.3 Landsskogstakseringen

Landsskogstakseringen er et prosjekt drevet av Nibio, der det gjøres feltobservasjoner på permanent etablerte prøveflater over hele landet. Dette benyttes til kartlegging av tilstand og beregninger av skog, og som referansedata i forskning rundt dette. Innhenting av data til Landsskogstakseringen gjøres jevnlig for å sikre oppdaterte data. På 1/5 av de 22 008 permanente prøveflatene gjennomføres det en retaksering hvert år. Dermed vil alle flatene retakseres innen fem år (Viken, 2021).

Ikke alle disse prøveflatene inneholder utelukkende skog. Ved å sette dette som et kriterium, reduseres derfor antall flater til 11 726. Videre vil dette antallet reduseres ytterligere for å se på flater der treslagssammensetningen er mest mulig homogen uten at antall observasjoner blir for liten. I denne oppgaven benyttes terskelverdien 75% volumdominans for ett treslag på hver prøveflate. Ved å filtrere vekk flater der ingen treslag har mer enn 75% volumdominans, ender man opp med 7977 observasjoner til videre analyse.

Data fra feltarbeidet i Landsskogstakseringen legges i en tabell der hvert tre er knyttet opp mot en flateID tilhørende prøveflaten det befinner seg innenfor. Det er på bakgrunn av disse verdiene volumet for hvert treslag innad i hver prøveflate er beregnet.

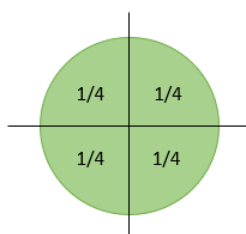


Figur 1 Illustrasjon av dimensjoner på en prøveflate i Landsskogstakseringen.

En prøveflate består av en indre og en ytre sirkel. Dataene i denne oppgaven er kun registrert i den indre sirkelen som måler 250m² (Viken, 2021). En hel prøveflate er illustrert i Figur 1.

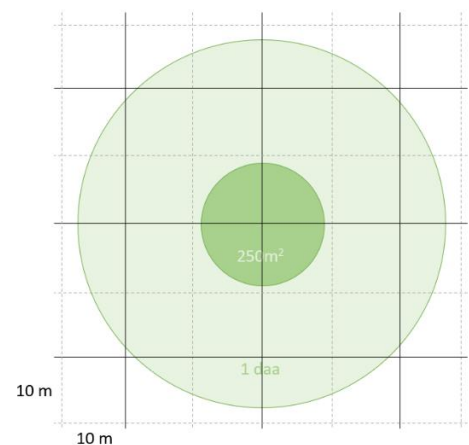
Videre kan man se hvordan pikslene fra Sentinel-2 (S2) ser ut over en prøveflate i Figur 2. Romlig oppløsning for satellittdataene brukt i denne oppgaven er på 20m, og de heltrukne linjene viser et eksempel på hvordan det kan se ut over en prøveflate.

Pikselverdier fra S2-data vil for de fleste prøveflater ikke treffe med kun én piksel per flate. Med en oppløsning på 20 meter vil derfor hver prøveflate være representert av mellom 1 og 4 verdier. Her vektet hver piksel ut ifra arealet den opptar på prøveflata. I Figur 3 er eksempelvis en indre prøveflate der krysningen mellom fire piksler kommer akkurat i midten av sirkelflaten. Da vil alle fire piksler oppta



Figur 3 Vekting av pikselverdier innad en indre prøveflate.

like mye areal, og hver piksel kan vektet likt i utregningen av reflektert verdi for hvert av båndene. Resultatet blir én reflektert verdi per prøveflate for hvert S2-bånd over 22 uker.



Figur 2 Piksel dimensjoner over en prøveflate.

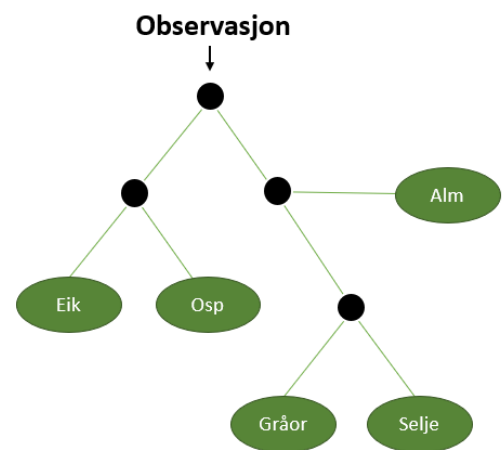
2.4 Random Forest

I denne oppgaven er random forest-klassifisering (RF-klassifisering) benyttet for å studere forskjell på ulike treslag. Denne klassifiseringsmetoden er mye brukt da den er svært robust i møte med utfordringer som overtilpasning og stort variabelantall (Belgiu & Drăguț, 2016; Breidenbach, Waser, et al., 2020).

En RF-modell benytter beslutningstrær for å klassifisere objekter. Et beslutningstre kan se ut som i Figur 4, der 5 av treslagene skiller ut ifra beslutningene i hver av de svarte prikkene. Her brukes en sammensetning av input-variabler i beslutningene for å plassere objekter i en kategori (Breiman, 2001).

Fra datasettet hentes derfra, for hvert beslutningstre, inn et tilfeldig *bootstrap*-datasett. Her velges et likt antall observasjoner som i det originale datasettet, men det vil være noen observasjoner som er med i *bootstrap*-datasettet flere enn én gang. Dette gjør at noen av observasjonene ikke er med i utvalget. Disse såkalte *Out-of-Bag*-observasjonene brukes til å teste beslutningstreet som er trent og bygget opp av *bootstrap*-datasettet. Slik klassifiseres alle OOB-observasjonene i hvert beslutningstre. Den kategorien flest av beslutningstrærne har plassert en observasjon i, blir det tildelte klassen i klassifiseringen. Ved å benytte OOB-observasjonene i klassifiseringen i hvert beslutningstre hindrer man dermed overtilpasning (Ljumović & Klar, 2015; Matthew W., 2011).

I RF-klassiferingen gjort i denne oppgaven er det brukt 1000 beslutningstrær. Denne er økt fra 500, da det er mange variabler i datasettet i forhold til observasjoner. Økt antall beslutningstrær betyr at modellen får brukt flere variabelkombinasjoner i beslutningene for hvert beslutningstre i klassifiseringen (Belgiu & Drăguț, 2016).



Figur 4 Beslutningstre med fem klasser representert.

2.5 Statistikk

Tabell 2 viser et eksempel på en fiktiv forvirringsmatrise, med oversikt over klassifiseringene gjort av for eksempel en RF-modell. Her telles de observasjonene som er klassifisert riktig opp, og plasseres på diagonalen. På samme måte telles observasjoner som er klassifisert som et av de andre treslagene opp, og plasseres deretter. Eksempelvis er det én observasjon av treslagstype 1 i Tabell 2 som er feilaktig klassifisert som treslag 2.

Tabell 2 Forvirringsmatrise og tilhørende nøyaktigheter for et fiktivt treslagsklassifiseringseksempel.

		Predicted				
		Treslag1	Treslag2	Treslag3	Totalt	
Actual	Treslag1	15	1	0	16	UA ₁
	Treslag2	3	17	5	25	UA ₂
	Treslag3	2	4	12	18	UA ₃
	Totalt	20	22	17	n = 59	
		PA ₁	PA ₂	PA ₃	OA	

For å evaluere klassifiseringsevnen til en modell gjennom forvirringsmatrisen, kan tre ulike typer nøyaktighetsmål regnes ut. UA (User's accuracy) beskriver hvor stor andel av de faktiske treslagene i hver kategori som er riktig klassifisert. PA (Producer's accuracy) beskriver i hvor stor grad de observasjonene som er klassifisert av modellen i hver kategori stemmer for det faktiske treslaget. OA (Overall accuracy) beskriver total prosentandel riktig klassifiserte observasjoner i modellen.

Formlene for utregning av disse nøyaktighetene er illustrert under med utregninger for forvirringsmatriseeksemplet i Tabell 2:

User's accuracy (UA):

$$UA_1 = \frac{15}{16} = 94\%, \quad UA_2 = \frac{17}{25} = 68\%, \quad UA_3 = \frac{12}{18} = 67\%$$

Producer's accuracy (PA):

$$PA_1 = \frac{15}{20} = 75\%, \quad PA_2 = \frac{17}{22} = 77\%, \quad PA_3 = \frac{12}{17} = 71\%$$

Overall accuracy (OA):

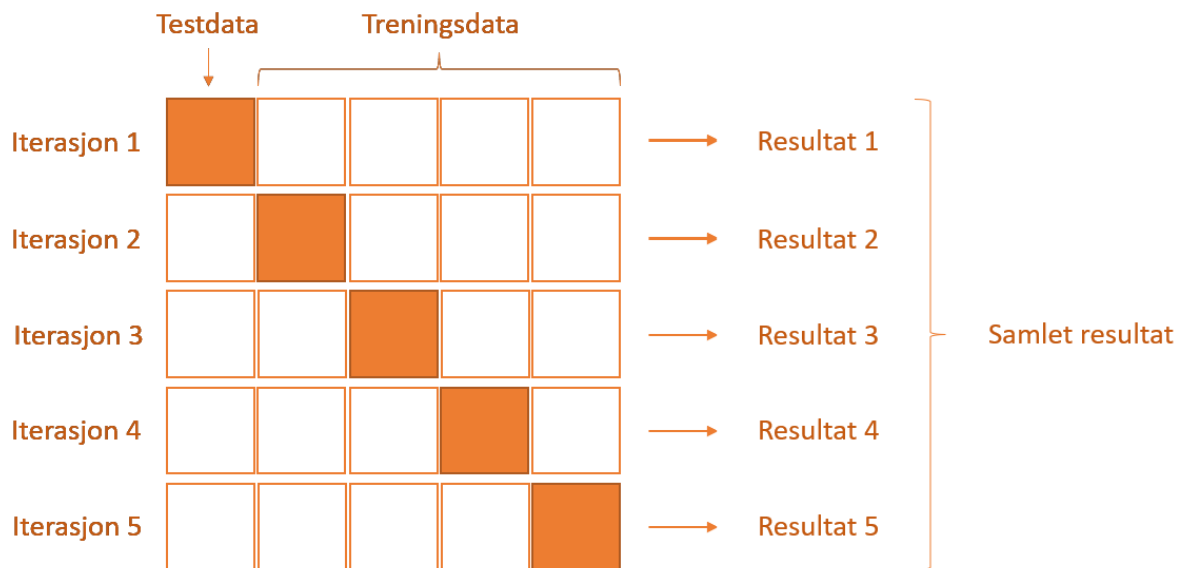
$$OA = \frac{15 + 17 + 12}{59} = 75\%$$

Kombinasjonen av de tre typene nøyaktighet er en god indikator på hvor godt modellen klarer å klassifisere et sett med observasjoner (Story & Congalton, 2008). Dette er målene som videre brukes i denne oppgaven for å vurdere RF-klassifiseringenes kvalitet.

Standardavvik (sd) er også brukt for å uttrykke variasjonen blant modellenes klassifiseringsevner. Her blir standardavviket (regnet ut med standardavvik-funksjonen for *sample variance* i R) for hver av nøyaktighetene omgjort til prosent. I eksemplet over vil et standardavvik på 2% for UA₁ inkludere verdier mellom 91,75% og 95,75%.

2.6 Kryssvalidering

Kryssvalidering er en metode som brukes for å sikre troverdige resultater i en modell. Det brukes også til å finjustere parametere og motarbeide overtilpasning.



Figur 5: Illustrasjon av en kryssvalidering med 5 iterasjoner.

Som Figur 5 viser, deles i en kryssvalidering datasettet inn i et gitt antall deler, der hvert av delene brukes som testdata i RF-klassifiseringen. Testdataene er ikke brukt i modellbygging er derfor uavhengige fra modellen. Det samlede resultatet er regnet ut basert på alle iterasjonene med formelen:

$$\frac{1}{k} * \sum_{i=1}^k Resultat_i$$

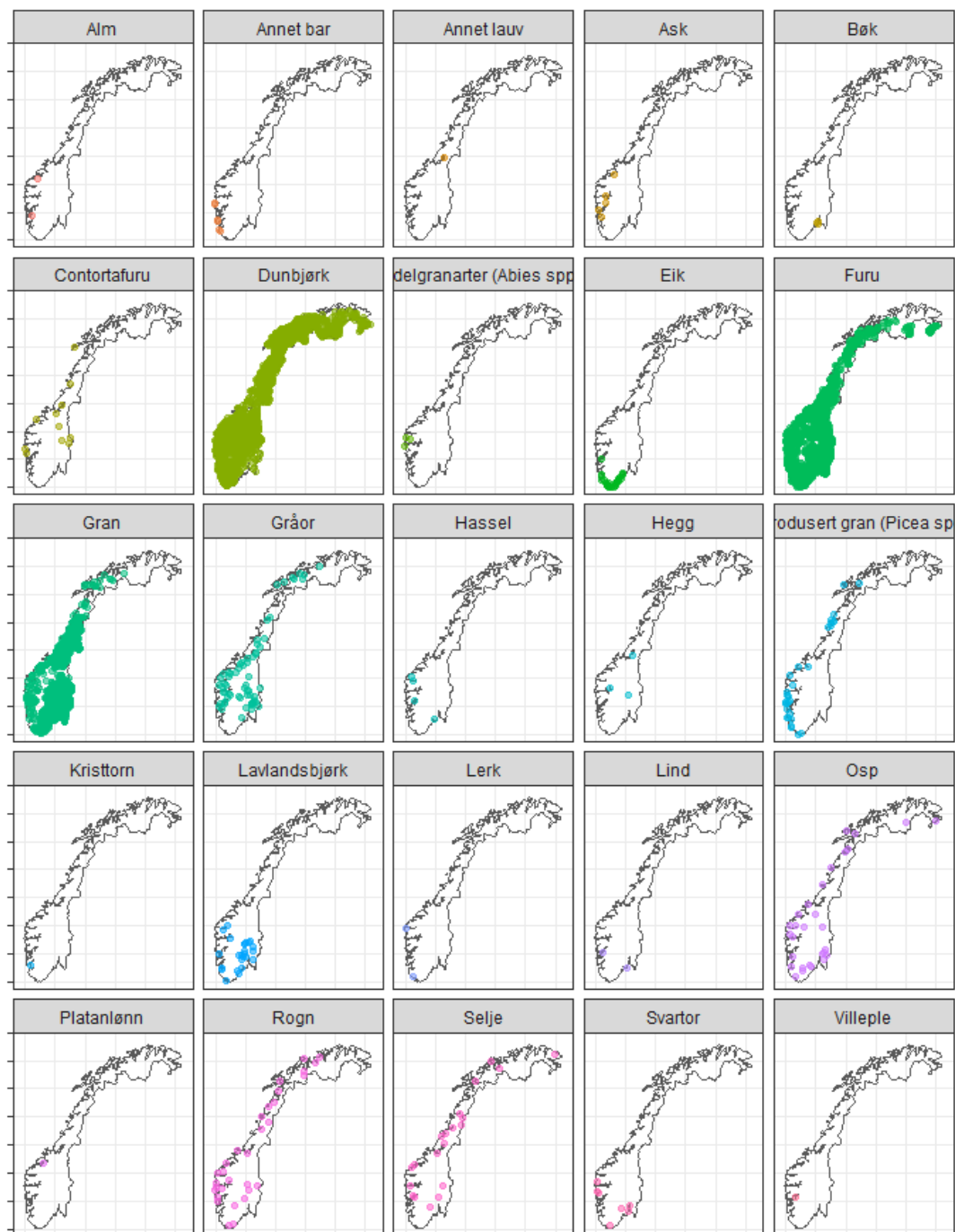
Der i er iterasjonene og k er antall iterasjoner. Hver av iterasjonene vektet altså likt, og danner til sammen et resultat for hele klassifiseringen (Berrar, 2018).

Det innebygde alternativet til dette i random forest er bagging (bootstrap aggregation). Dette gjøres til forskjell fra vanlig kryssvalidering for hvert tre, heller enn for hele modellen av gangen (Lisein et al., 2015).

3 Metode

3.1 Datapresentasjon

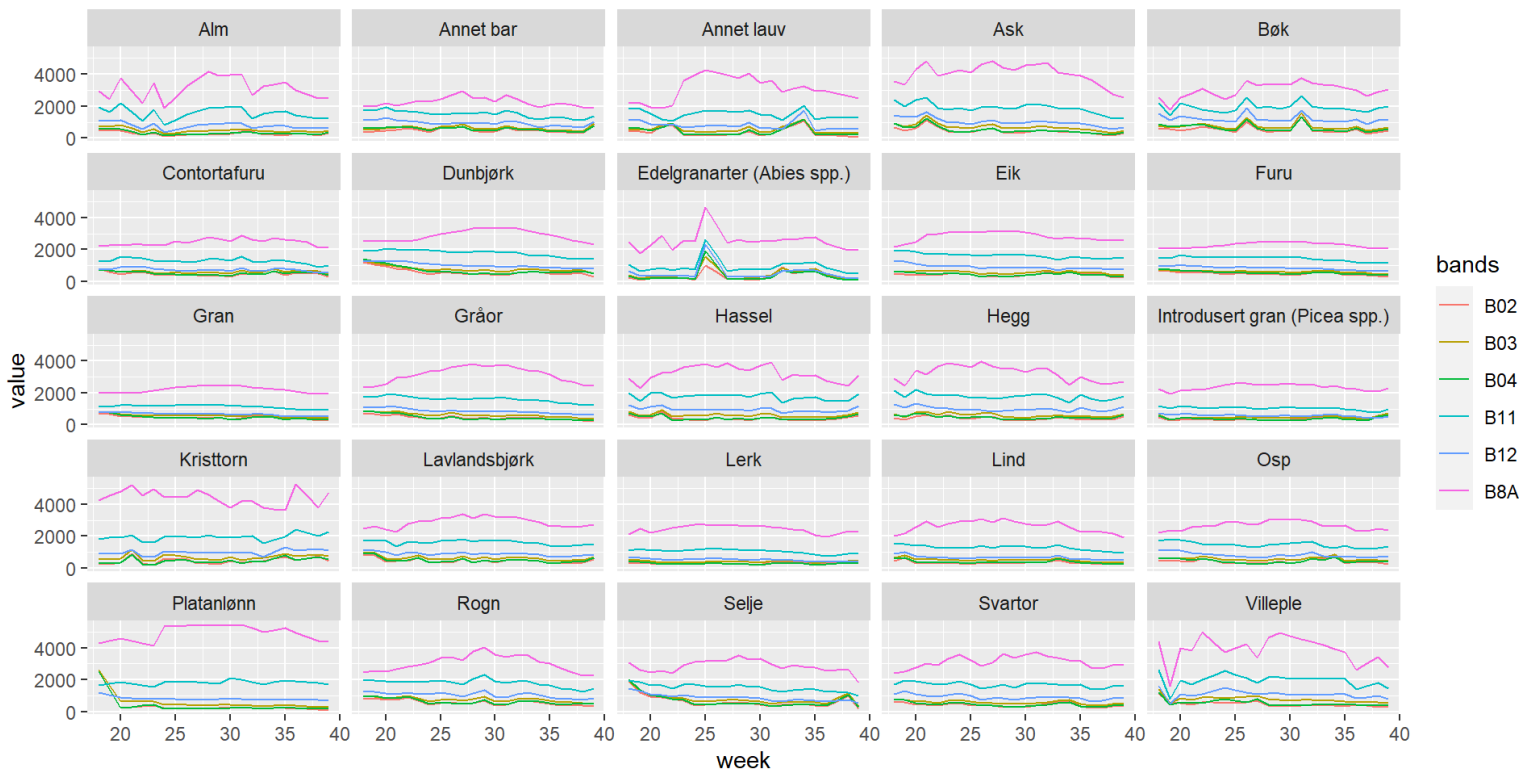
Figur 6 viser hvor i landet de ulike treslagene dominerer prøveflatene i Landsskogstakseringen med mer enn 75% av volumet blant trærne innen hver prøveflate. Her er det stor forskjell på antallet skogsflater i hvert av treslagene. Det er også veldig ulikt hvor i landet hvert av treslagene vokser. Introdusert gran vokser for eksempel langs kyst på vestsida av landet, og eik vokser i stor grad langs sørlandskysten ut ifra det man kan lese av figuren. Dunbjørk brukes i denne oppgaven om arten vanlig bjørk (*Betula Pubescens*) som inneholder både dunbjørk og fjellbjørk. Lavlandsbjørk er på tilsvarende vis brukt om det som også kalles Hengebjørk (*Betula Pendula*) (Grindeland, 2022).



Figur 6 Geografisk fordeling av dominante prøveflater for hvert av treslagene i Landsskogstakseringen.

Sentinel-2 (S2) passerer over Norge opptil flere ganger i uka, og for å finne én verdi per uke er meridianverdien plukket ut for et 7 dagers vindu per uke for uke 18 til 39 i 2019 (Hauglin). Dette tilsvarer perioden fra og med mai og til og med ut september. Disse ukene er det størst endringer, spesielt i løvtreslag. Her er det også minst fare for forstyrrelser i form av snø.

Ved å fremstille gjennomsnittsverdiene til hvert av båndene for de ulike treslagene kan det visualiseres hvordan treslagene endrer seg over sesongen som vist i Figur 7. Satellittdataene er hentet ut med avgrensninger for prøveflatene fra Landsskogstakseringsflatene (Hauglin), som beskrevet i kapittel 2.3.



Figur 7 Båndverdiene for hvert treslag over de 22 observasjonsukene.

Verdiene i S2-datasettet er atmosfærisk korrigert, og lagret på 16 bits (3 ESA, 2023). Filene brukt i denne oppgaven er på *signed integer*-format, og maksimal reflektert verdi er dermed på 32 768.

Den store forskjellen i glattheten på kurvene i Figur 7 kommer av forskjellen i antall observasjoner av hvert treslag. For treslag med få observasjoner der én av disse har en outlier i én av ukene, vil dette gi store utslag på gjennomsnittet som er plottet over. Eksempelvis har kurvene for alle bånd tilhørende edelgran, som bare har tre prøveflater med volumdominans i datasettet, en tydelig topp i uke 25.

I Figur 8 er NDVI fremstilt over de 22 ukene representert i datasettet for hvert av treslagene. Generelt har bartrær lavere NDVI enn løvtrær. NDVI er regnet ut som beskrevet i kapittel 2.2.

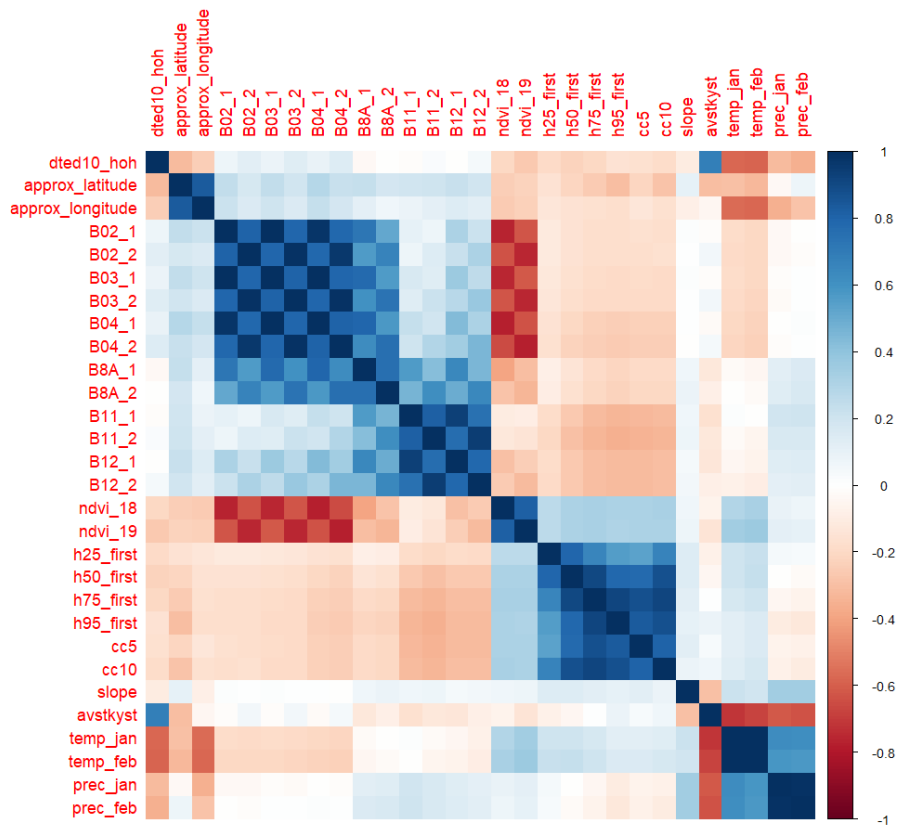


Figur 8 NDVI for hvert treslag over de 22 observasjonsukene.

3.2 Variabler og deres korrelasjon

Input i en RF-klassifisering bestemmes av hvilke variabler man ønsker å ta med i beregningene. Derfor er det viktig å være bevisst betydningen av disse variablene. Utgangspunktet for klassifiseringen er tidsseriedata fra Sentinel-2 i perioden mai til ut september. Ellers er variabler for temperaturer, nedbør, helningsgrad, høyde over havet, avstand til kyst etc. tilgjengelig for prøveflatene. Disse variablene kommer ikke fra satellittdata, men er mulig å måle uten feltobservasjoner. I tillegg til disse klimatiske variablene finnes det landsdekkende laserdata som ble ferdigstilt for hele landet i 2022 (Kartverket, 2022). Dette er også tilgjengelig i analysene. I Landsskogstakseringen finnes også informasjon om hvert enkelt tre innad hver indre prøveflate. Dette er for blant annet høyde, volum, alder etc. Disse variablene utelates fra klassifiseringsgrunnlaget da de vil være vanskelig å benytte utenfor prøveflatene fra Landsskogstakseringen, og er mindre relevante i sammenheng med oppløsningen fra S2-dataene. Enkeltrærnes volum er likevel benyttet i utregningen av treslagsdominans per prøveflate.

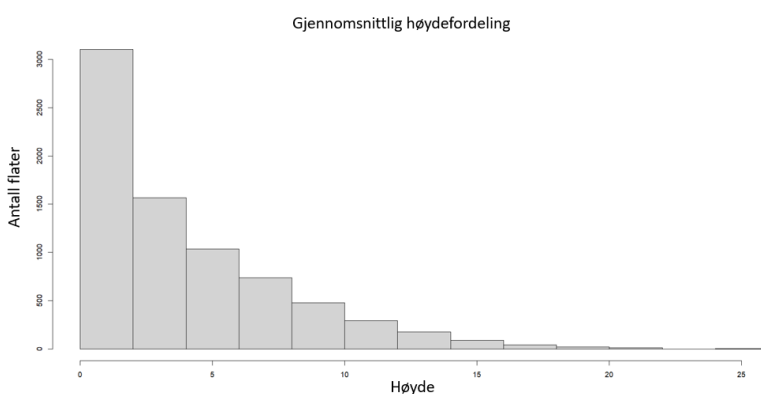
Mange av de tilgjengelige variablene vil ha høy korrelasjon da for eksempel temperaturen i januar og temperaturen i februar vil kunne se ganske likt ut for prøveflatene. Det samme vil være tilfellet for tidsseriene. Noen bånd vil kunne se relativt like ut for hver prøveflate, og for de forskjellige ukene innad i hvert bånd. Sterkt korrelerte variabler vil kunne påvirke klassifiseringsens stabilitet (Zhang & Yang, 2020), variabelviktighetsplot og klassifiseringsresultatene (Janitza et al., 2018).



Figur 9 Korrelasjon mellom et utvalg variabler.

Figur 9 visualiserer korrelasjon mellom et utvalg variabler gjennom R2-tallet. Verdier nærmere 1 og -1 tilsier at variablene er sterkt korrelerte.

Det er naturligvis høy korrelasjon mellom temperaturer, men det er også eksempelvis relativt høy negativ korrelasjon mellom temperaturer og avstand til kystlinje da det generelt sett er kaldere innover i landet.



Figur 10 Histogram over fordeling av gjennomsnittlig høyde på første-etterer på prøveflatene (Hmean_first).

Laserdataene er beskrevet gjennom 8 variabler. cc5 og cc10 beskriver trærnes dekningsgrad på hhv. 5 og 10 meter over bakkenivå. Hmean_first, h25_first, ... og h95_first beskriver høydene for første-etterer. Figur 10 viser fordelingen av gjennomsnittshøyden på første-etterer per prøveflate. De resterende første-etterer-variablene er kvantiler for fordelingen av høyde på første-etterer innad i hver prøveflate.

3.3 Manglende verdier

Variablene som benyttes i denne oppgaven er S2-tidsserievariabler, klimatiske variabler og laservariabler. En RF-klassifisering vil ikke benytte et datasett med manglende verdier (NA-verdier) for noen av variablene i noen av observasjonene som input. For å se nærmere på hvilke observasjoner og variabler dette gjelder, telles antall NA-verdier opp for hver mulige inputvariabel i Tabell 3.

Tabell 3 Variabler som mangler verdier for observasjoner og antallet observasjoner som mangler for hver av variablene.

	Variabelnavn	Antall observasjoner
1	hmean_first	426
2	h25_first	426
3	h50_first	426
4	h75_first	426
5	h90_first	426
6	h95_first	426
7	cc5	426
8	cc10	426
9	slope	212
10	avstkyst	212
11	dtw	3089

Det er én variabel som skiller seg tydelig fra resten av variablene som inneholder NA-verdier. *Dtw* (*depth to water*) kan ha betydning for vekstforholdene i ulik grad for ulike treslag (Mueller-Dombois, 1964). I Tabell 4 inneholder en oversikt over antall av hvert treslag som inneholder slike NA-verdier for én eller flere av variablene i Tabell 3. Manglende verdier er overlappende for hver av variablene med samme antall manglende verdier.

Ved å fjerne observasjonene som mangler verdier for *dtw*-variabelen i Tabell 3, vil antallsfordelingen på hvert treslag se ut som i kolonnen *Observasjoner uten NA-obs. for dtw* i Tabell 4. For de resterende variablene vil en fjerning av observasjoner med manglende verdier være fordelt blant treslagene som i kolonnen *Observasjoner uten NA-obs. for laser, slope og avstkyst* i samme tabell. Variabelen *dtw* fjernes fra variabelutvalget for datasettet brukt i klassifiseringene da antallet observasjoner er redusert med en betydelig prosentandel for mange av treslagene. Tabell 4 viser også ubalansen i antall observasjoner for hvert treslag.

Tabell 4 Fordeling av antall trær per treslag før og etter observasjonene med NA-verdier er utelatt for dtw og for de resterende variablene.

Treslag		Antall observasjoner total	Observasjoner uten NA-obs. for laser, slope og avstkyst	Observasjoner uten NA-obs. for dtw
1	Alm	2	2	1
2	Annet bar	6	6	6
3	Annet lauv	1	1	1
4	Ask	5	5	3
5	Bøk	2	2	2
6	Contortafuru	11	11	7
7	Dunbjørk	2498	2126	1112
8	Edelgranarter (Abies spp.)	3	3	1
9	Eik	28	26	25
10	Furu	2918	2763	2004
11	Gran	2284	2240	1583
12	Gråor	63	62	39
13	Hassel	4	4	3
14	Hegg	3	3	2
15	Introdusert gran (Picea spp.)	30	29	19
16	Kristtorn	1	1	1
17	Lavlandsbjørk	25	23	19
18	Lerk	2	2	2
19	Lind	2	2	2
20	Osp	29	27	21
21	Platanlønn	1	1	1
22	Rogn	31	26	17
23	Selje	20	16	12
24	Svartor	7	7	4
25	Villeple	1	1	1

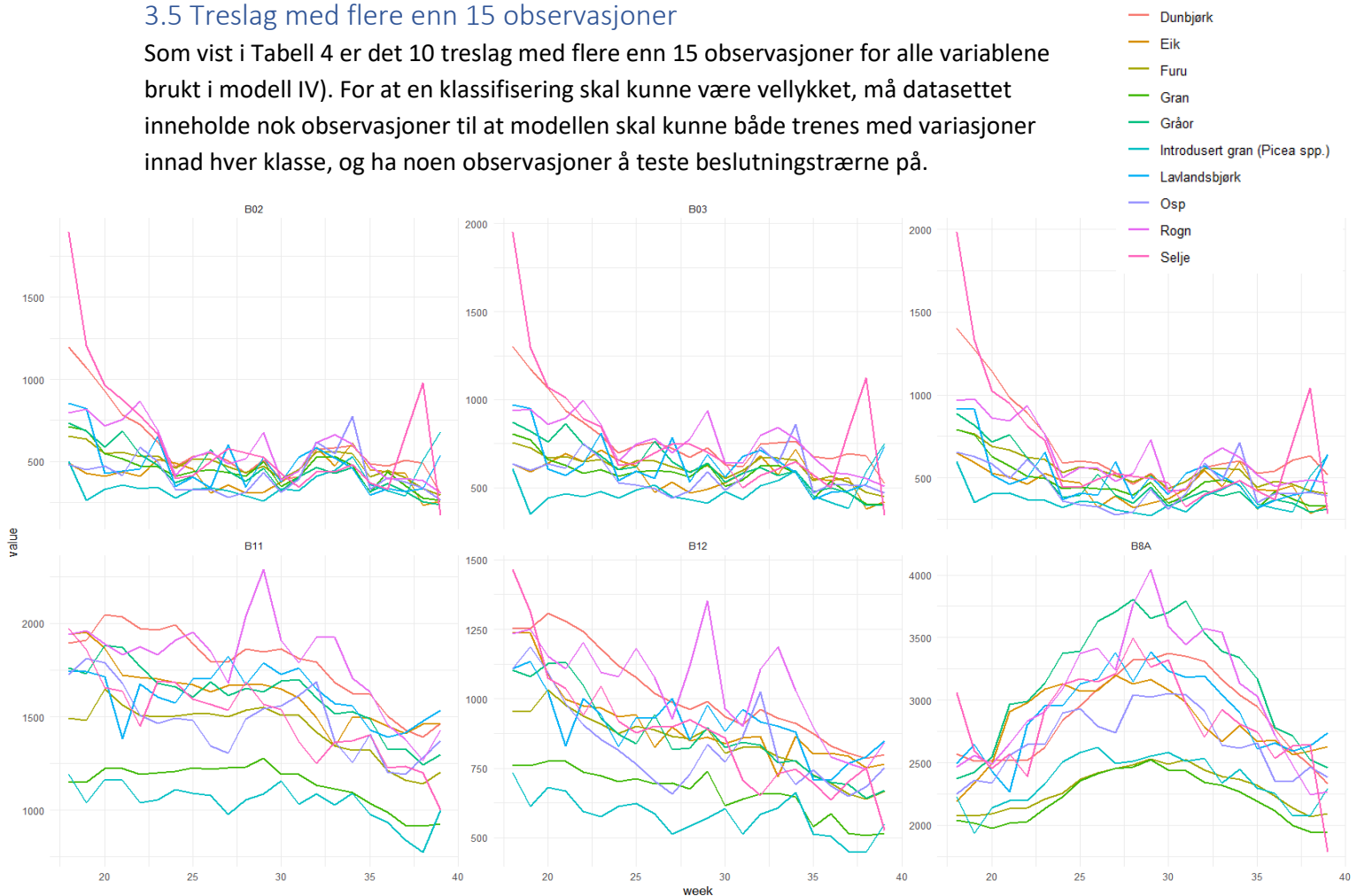
3.4 Variabelutvelgelse

For å se om laserdata har noen effekt i RF-klassifiseringen av treslag, benyttes et sett med variabler med og et uten. S2-tidsseriene er konstant for alle modellene, da dette er utgangspunktet for klassifiseringene. På samme måte som for laserdata legges klimatiske variabler til i utvalgte modeller, der *dtw* er utelatt. Ved å lage et alternativt input-variabelsett med og uten disse, vil viktigheten av de klimatiske variablene avgjøres. Til slutt legges alle variabler til en modell for å se kombinert virkning i RF-klassifiseringene. Det settes på denne måten opp fire modeller med følgende variabler:

- I) Sentinel-2 tidsserier
- II) Sentinel-2 tidsserier med helning, avstand til kyst, høyde over havet, temperaturer og nedbør
- III) Sentinel-2 tidsserier med laservariabler
- IV) Sentinel-2 tidsserier med helning, avstand til kyst, høyde over havet, temperaturer, nedbør og laservariabler

3.5 Treslag med flere enn 15 observasjoner

Som vist i Tabell 4 er det 10 treslag med flere enn 15 observasjoner for alle variablene brukt i modell IV). For at en klassifisering skal kunne være vellykket, må datasettet inneholde nok observasjoner til at modellen skal kunne både trenes med variasjoner innad hver klasse, og ha noen observasjoner å teste beslutningstrærne på.

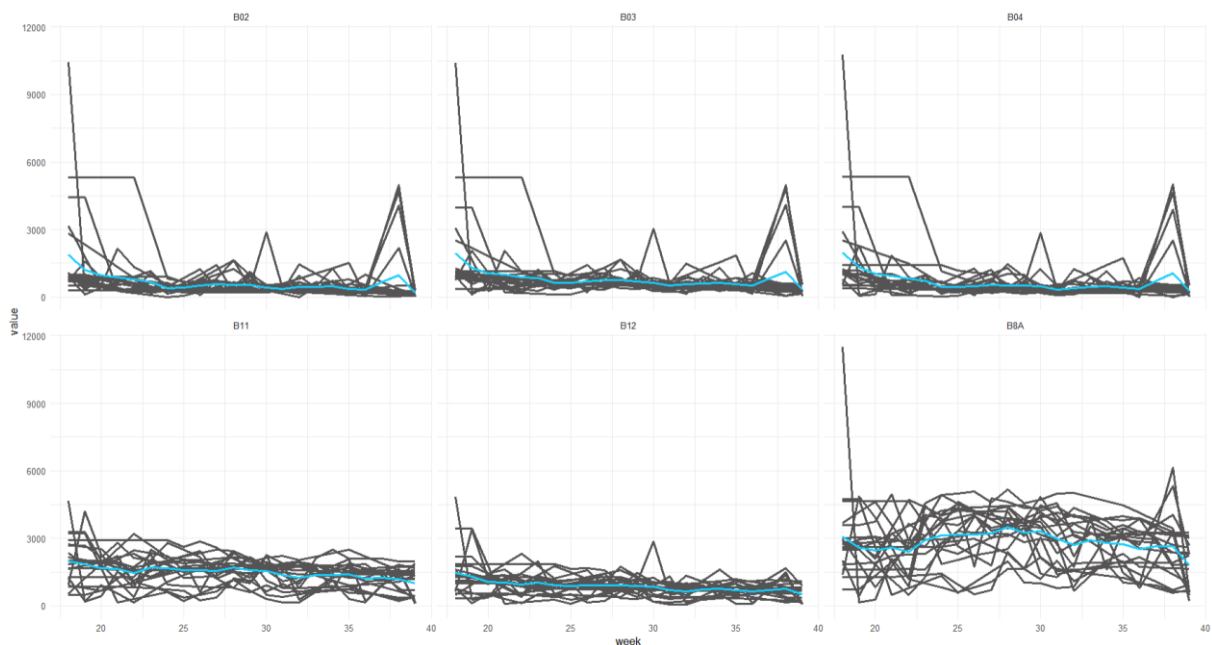


Figur 11 Gjennomsnittsverdi per uke for de 6 båndene for 10 utvalgte treslag.

I Figur 11 har hvert bånd en graf med gjennomsnittsverdien for hvert av treslagene med mer enn 15 dominante prøveflater per uke langs x-aksen. Dette viser endringen i reflekterte verdier for hvert bånd over sesongen mai til ut september.

I bånd B8A ser man en tydelig forskjell mellom løvtrær og bartrær. Furu og gran er videre veldig forskjellige i bånd B11 og B12. Dette støtter opp under tidligere forskning rundt mulighetene for å skille furu, gran og løvtrær i stor grad (Breidenbach, Waser, et al., 2020). Videre i oppgaven vil også disse tre grupperingene hentes ut for seg når ytterligere klassifiseringsmuligheter skal undersøkes.

I uke 18 skiller selje seg tydelig fra de andre treslagene i Figur 11. For å se nærmere på hvorfor selje skiller seg ut denne uka, er hver prøveflate plottet for seg i Figur 12. Den turkise linja er gjennomsnittet som er brukt i fremstillingen i Figur 11.



Figur 12 Prøveflater dominert av selje, plottet hver for seg over tidsserien, med gjennomsnittet markert i turkis.

Noen outlierer i bånd 2, 3, 4 og 8A trekker i stor grad opp gjennomsnittet i ukene der selje skiller seg tydelig fra de andre treslagenes gjennomsnitt. Disse outlierne kan skyldes flere ting, blant annet snø eller skydekke som ikke er plukket opp når det er gjort romlige interpoleringer for skydekke i bildene fra Sentinel-2.

3.6 Outliere

I samtlige av enkeltflateplottene for hvert av treslagene finnes det outlierer i minst ett av båndene, der noen er tydeligere enn andre. De større avvikene kan skyldes feil i innsamling, men når det ikke bare er en og annen av dem, kan det tyde på at det er et naturlig fenomen.

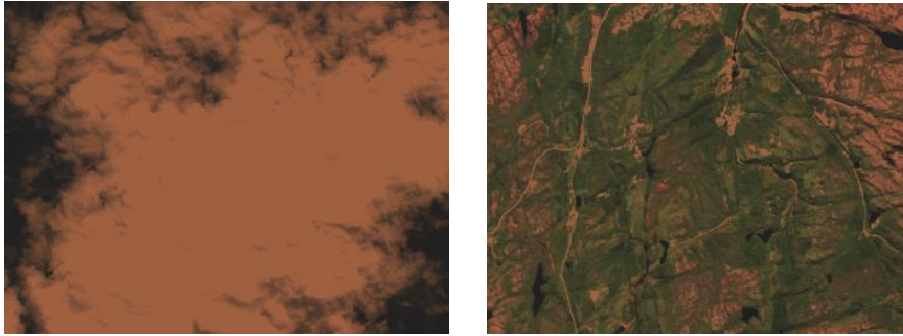
Mange av de typiske avvikene ligger på mellom 2000 og 4000 i reflektert verdi i RGB-båndene. Til sammenlikning kan et skydekke gi verdier tilsvarende de i Tabell 5.

Tabell 5 Typiske verdier for en piksel med tett skydekke.

Bånd	Verdi – skydekke
2	15648
3	14272
4	13482
8A	12449
11	6126
12	4190

For Tabell 5 er det plukket ut et punkt med tett skydekke som befinner seg over et skogsområde. Området er vist for to ulike tidspunkt, ett med og ett uten skydekke, i Figur 13.

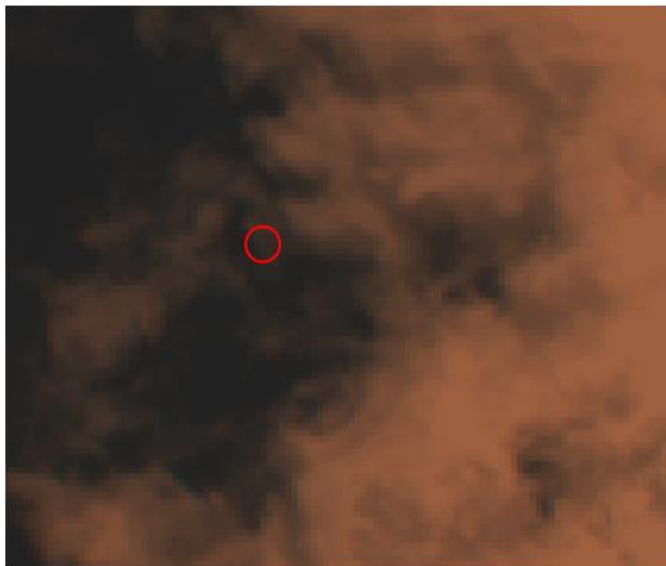
Samme området er fremstilt fra henholdsvis 3. juli 2019 og 5. juli 2019.



Figur 13 Et skogsområde med og uten skydekke, hentet fra Sentinel-2-bilde hhv. 3. og 5. juli 2019.

Høyere verdier i Bånd 2, 3 og 4 er å forvente da fargen på en sky er tilnærmet hvit, og skydekket reflekterer store deler av sollyset.

Av Tabell 5 ser man at med økende bølgelengde avtar verdiene. Dette var forventet, da det brukes lengre bølgelengder (i Sentinel-1 for eksempel, bølgelengde ca. 5,5cm) til å observere jorda gjennom blant annet skydekke (4 ESA, 2023).



Figur 14 Tynt skydekke over skogsområde, hentet fra S2-bilde.

Tabell 6 Verdier for alle 6 bånd i datasettet for en skydekket piksel markert i Figur 14.

Bånd	Verdi – skydekke
2	2276
3	2190
4	2029
8A	3177
11	1935
12	1654

I Figur 14 er det markert et område med båndverdiene i Tabell 6. Da dette skydekket er såpass tynt er det mulig at det ikke er fanget opp ved 'fjerning' av skyer i satellittbildene ved interpolasjon. Langs kantene av skydekket er typiske verdier i RGB-båndene mellom 2000 og 4000. I dette tilfellet reflekterer underlaget mer enn skydekket, og det vil dermed ikke avta med økt bølgelengde i like høy grad som for et tett skydekke. På samme måte som med skyer vil snø forstyrre bildet av skogen da hvite overflater reflekterer spesielt lavere bølgelengder i mye høyere grad enn mørkere vegetasjonsunderlag. Det kan være vanskelig å skille skyer og snø fra hverandre i satellittbilder (Rittger et al., 2020), men ofte vil et snødekke være noe som synes over flere uker, der skyer kontinuerlig er i bevegelse og endring. Terskelverdien settes til 2000 for rødt, grønt og blått bånd, og verdier for alle bånd der ett eller flere av RGB-båndene overskrider terskelverdien fjernes. Det gjøres så en lineær interpolasjon over ukene i S2-tidsserien.

3.7 Jevne ut antall observasjoner per treslag

Random Forest har et svakt punkt ujevne fordelinger av antall observasjoner per klasse (Belgiu & Drăguț, 2016). En skjevfordeling der Dunbjørk, Furu og Gran dominerer i så stor grad som i norsk skog, vil dermed føre til at feilen i treslaget med mange observasjoner vil være nær null. For de resterende klassene vil feilen være veldig høy. Selv om det er signifikant forskjell mellom to treslag, så vil i mange tilfeller alle observasjoner klassifiseres som treslaget med flest observasjoner dersom antallsforskjellen er stor. Derfor må antallet prøveflater per treslag jevnes ut før RF-klassifiseringen gjøres. Antallsskjevheten kommer av mangel på prøveflater for mange av de mindre vanlige treslagene. 50 flater plukkes derfor tilfeldig ut blant furu-, gran- og dunbjørkobservasjoner for å få et godt grunnlag for klassifisering uten at det utkonkurrerer de andre treslagene.

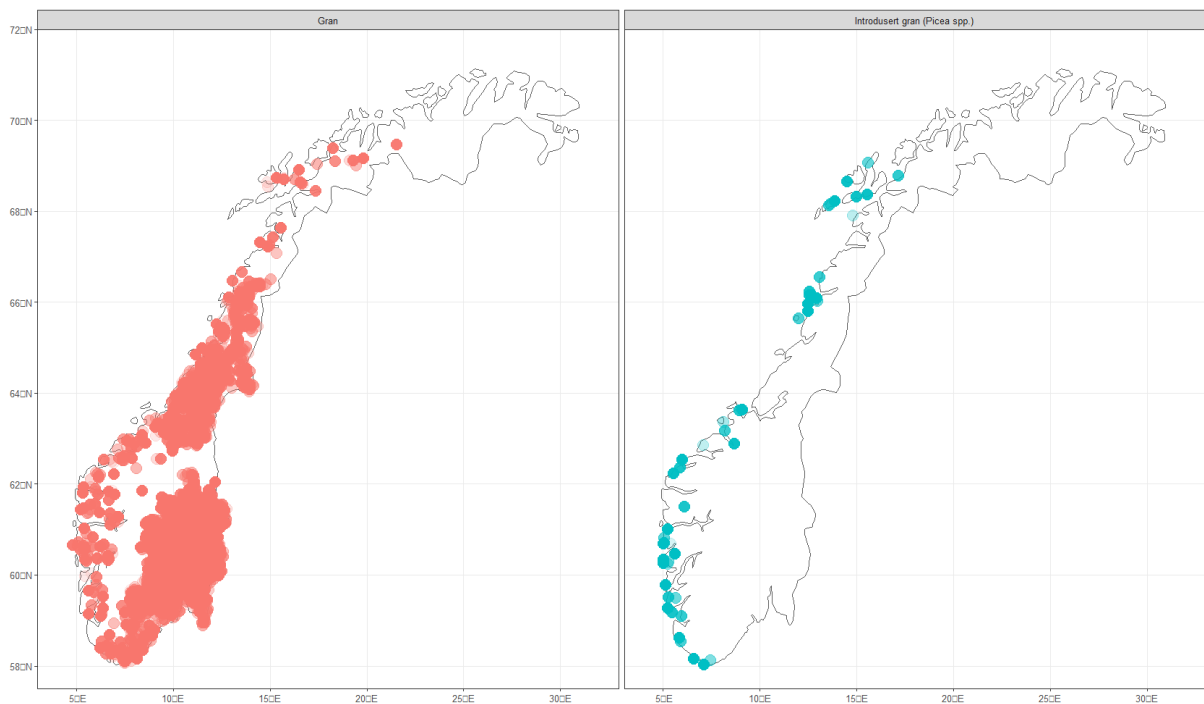
3.8 Klassifisering av treslag

For å finne ut hvilke treslag det er håp om å kunne skille fra resten av treslagene, gjøres en RF-klassifisering for alle treslagene. Basert på dette velges eik og gråor ut blant løvtrærne, og en klassifisering av ulike granarter for videre undersøkelse. Forvirringsmatrisen vises i resultatdelen (i kapittel 4.1) i Tabell 9.

Vedlegg 2 inneholder en tabell med feilklassifiseringsandel ($1 - UA$) fra forvirringsmatrisene for ett og ett treslag klassifisert mot hverandre. Her er antall dunbjørkobservasjoner redusert til 50. Vedlegg 2 har blitt benyttet som supplement for utvelgelse av treslagene undersøkt i oppgaven.

3.8.1 Granklassifisering

4 granarter er representert i datasettet fra Landsskogstakseringen: Gran, introdusert gran, edelgranarter og barlind. Barlind dominerer ikke med over 75% på noen prøveflater, og det er derfor ingen brukbare observasjoner av dette treslaget. Edelgranartene har tre volumdominerende prøveflater i datasettet, og det kan derfor være vanskelig å klassifisere. Edelgranartene er også introduserte granarter (Sunding et al., 2022), og dermed samles granklassene i to; Gran og introdusert gran, der introdusert gran inneholder edelgranartklassen i tillegg til andre introduserte granarter.



Figur 15 Oversiktskart over plassering av prøveflater med dominans av gran og av introdusert gran.

Som vist i Figur 15 vokser de to ulike granartene på ganske ulike steder. Gran vokser over alt utenom de høyeste fjellområdene og de nordligste delene av Norge. Introdusert gran vokser bare langs vestkysten. Selv i kystsonene der det befinner seg gran, ligger introdusert gran i større grad helt mot kystlinja. Ved å snevre inn området for uthenting av granobservasjoner, vil modellen ta for seg forskjeller på treslagene i samme område. Da hindrer man at klassifiseringen skjer nesten utelukkende på bakgrunn av variabler som beskriver plassering i landet (klimatiske variabler). Dette vil også gjøre at modellen kun vil gjelde for områdene brukt i klassifiseringen. Derfor må området avgrenses nok til at forskjellene på granartene i samme område kommer tydelig frem, men ikke i så stor grad at modellen er ubrukelig i klassifisering utover prøveflatene fra Landsskogstakseringen.

For å finne avgrensingsområdet benyttes oversikten i Tabell 7 over avstanden til kyst fra observasjonene av prøveflater der introdusert gran dominerer. Prøveflate A35165 ligger lengst unna kysten av disse, og er observasjonen det tas utgangspunkt i når områdeavgrensningene settes. I og med at det kun er tre observasjoner lengre enn 20 km unna kysten, vil en terskelverdi på 50 km antas å være tilstrekkelig for å få med de fleste større forekomstene av introdusert gran utover Landsskogstakseringsflatene.

Tabell 7 Prøveflater med dominerende introdusert gran og avstanden disse ligger fra kystlinja.

	FlateID	Treslag	Avstand til kyst [m]
1	A04023	Introdusert gran (Picea spp.)	990
2	A06105	Introdusert gran (Picea spp.)	6119
3	A06145	Introdusert gran (Picea spp.)	9055
4	A10204	Introdusert gran (Picea spp.)	141
5	A10224	Introdusert gran (Picea spp.)	200
6	A13059	Introdusert gran (Picea spp.)	1140
7	A14185	Introdusert gran (Picea spp.)	18010
8	A17108	Introdusert gran (Picea spp.)	100
9	A21149	Introdusert gran (Picea spp.)	539
10	A21199	Introdusert gran (Picea spp.)	1432
11	A22114	Introdusert gran (Picea spp.)	671
12	A22161	Introdusert gran (Picea spp.)	1581
13	A25246	Introdusert gran (Picea spp.)	200
14	A26032	Introdusert gran (Picea spp.)	19338
15	A26163	Introdusert gran (Picea spp.)	10735
16	A29048	Edelgranarter (Abies spp.)	200
17	A29156	Introdusert gran (Picea spp.)	5334
18	A33252	Edelgranarter (Abies spp.)	1421
19	A34139	Edelgranarter (Abies spp.)	23044
20	A35165	Introdusert gran (Picea spp.)	46907
21	A42134	Introdusert gran (Picea spp.)	141
22	A52023	Introdusert gran (Picea spp.)	3600
23	B01096	Introdusert gran (Picea spp.)	10102
24	B51025	Introdusert gran (Picea spp.)	28507
25	D57099	Introdusert gran (Picea spp.)	0
26	D58171	Introdusert gran (Picea spp.)	0
27	D62028	Introdusert gran (Picea spp.)	781
28	D62092	Introdusert gran (Picea spp.)	200
29	D62098	Introdusert gran (Picea spp.)	1118
30	D62129	Introdusert gran (Picea spp.)	100
31	D66154	Introdusert gran (Picea spp.)	1421
32	D88221	Introdusert gran (Picea spp.)	1562
33	E27127	Introdusert gran (Picea spp.)	1140

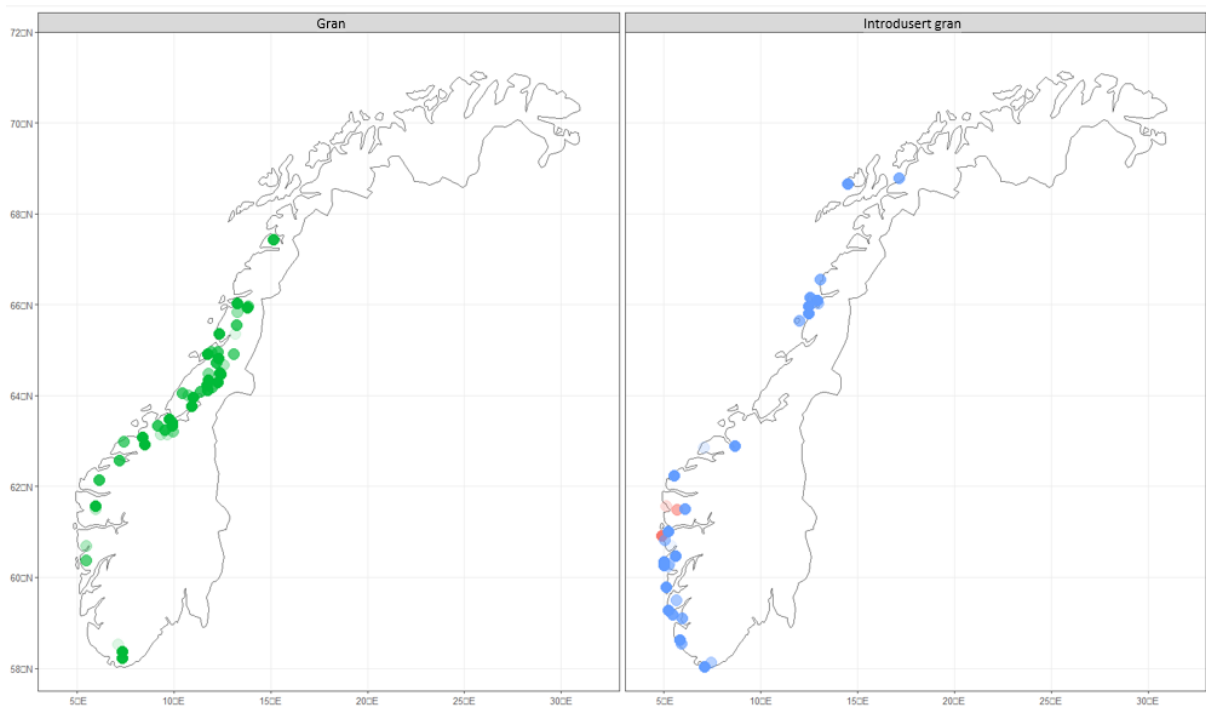
I tillegg til områdebegrensninger fra kystlinja filtreres Østlandet (øst for 8°E og sør for 61°N) og den delen av Norge som ligger lengst øst (øst for 20°E) ut. Det totale avgrensede området er markert i turkis i Figur 16.

Med disse avgrensningene hentes det ut 50 tilfeldig utplukkede granobservasjoner som benyttes i klassifiseringen. Fordelingen av prøveflatene som benyttet i RF-klassifiseringen er fremstilt i Figur 17.

På figuren er granobservasjonene markert i grønt på kartet til venstre. Edelgranartene er markert i rødt på kartet til høyre, og andre introduserte granarter er markert med blått på kartet til høyre.



Figur 16 Visualisering av det avgrensede området for granmodellen.



Figur 17 Observasjoner plukket ut for RF-klassifisering av granarter.

3.8.2 Eikedeteksjon

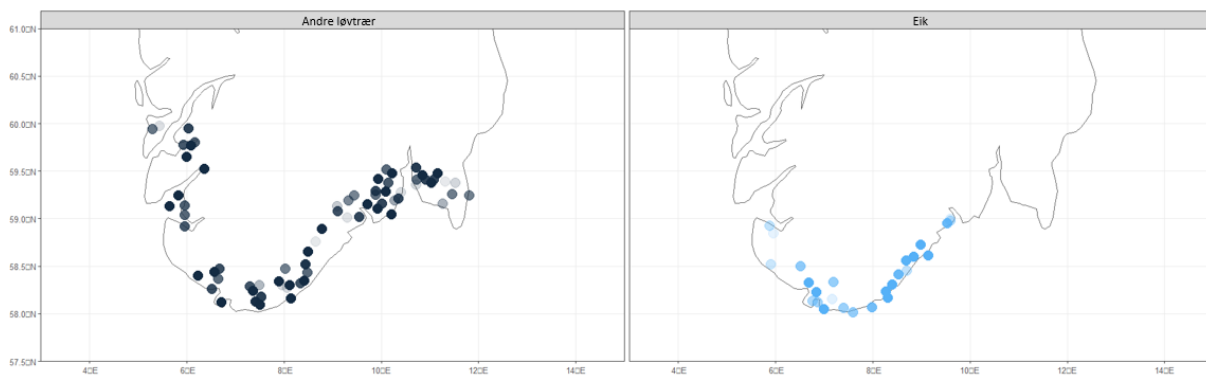
På samme måte som med introdusert gran, kan også eikedeteksjon avgrenses til et mindre geografisk område i landet. Dominerende eikeflater befinner seg langs kysten sør for 60°N.

Avstand til kyst er avgrensa til 32 km da 30 512 meter er observasjonen lengst unna kysten av prøveflatene. Prøveflaten nest lengst unna kysten ligger 22 811 meter unna, og det er kun tre observasjoner lengre unna enn 13 km. 32 km antas derfor tilstrekkelig som kystavgrensning.



Figur 18 Visualisering av det avgrensede området for eikemodellene.

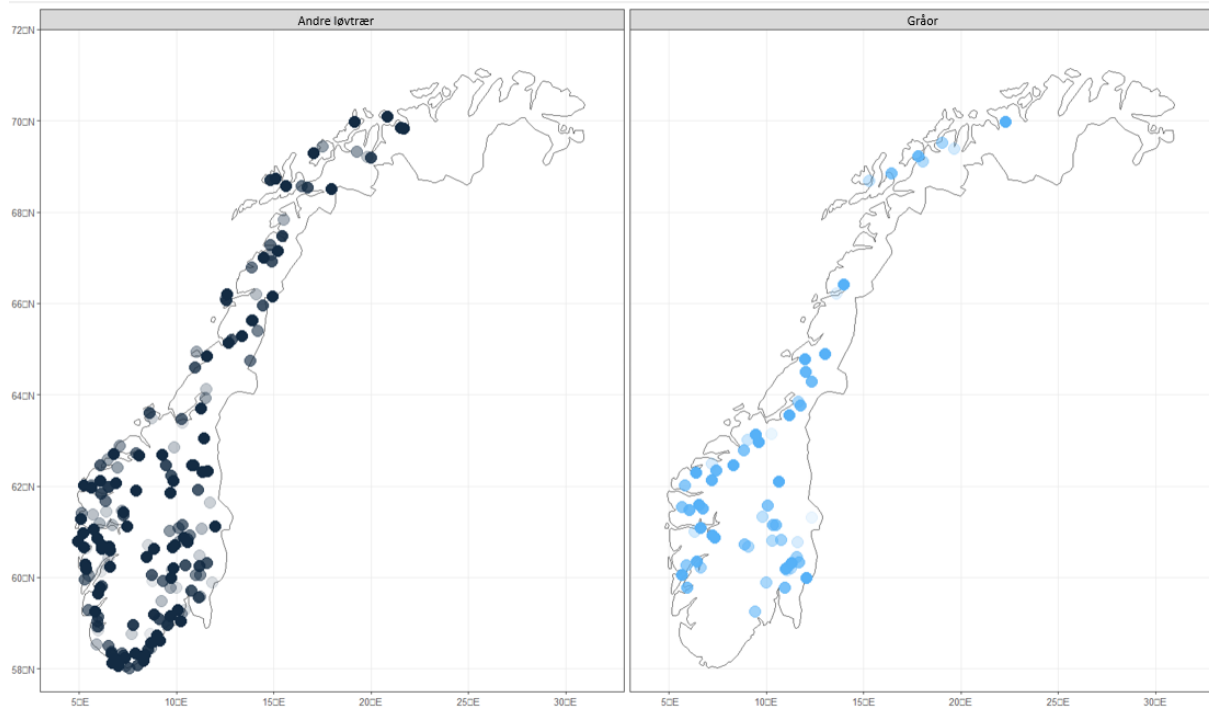
50 dunbjørkobservasjoner plukkes ut innen det avgrensede området vist i Figur 18, og sammen med resten av løvtreslagene innen området samles disse til én klasse. Figur 19 viser fordelingen av prøveflatene for Eikeobservasjonene på høyre side og den samlede klassen for andre løvtrær på venstre side.



Figur 19 Plassering av Eikeflater og andre løvtræflater til bruk i RF-klassifisering.

3.8.3 Gråordeteksjon

I motsetning til eik og introdusert gran befinner gråor seg i hele landet. Derfor benyttes ingen områdeavgrensninger når dunbjørkobservasjonene plukkes ut til klassifisering. I denne klassifiseringen inneholder *annet løv* 50 tilfeldig utplukkede dunbjørkobservasjoner, samt alle andre prøveflater der et annet løvtreslag dominerer volumet av trærne på prøveflaten. Dette utgjør til sammen 187 flater. Med 62 gråordominante flater har modellen observasjonene som brukes i RF-klassifiseringen. Plasseringen av disse er visualisert i Figur 20. Her er de blå observasjonene til høyre gråorobservasjonene, og de svarte på kartet til venstre er resten av løvtreobservasjonene med redusert antall dunbjørkobservasjoner.



Figur 20 Plassering av gråorflater og andre, delvis utplukkede løvtræflater til bruk i RF-klassifisering.

3.8.4 Forskjellen mellom løvtreslag

For å se nærmere på hvilke løvtreslag det er mulig å skille, gjøres en RF-klassifisering av alle kombinasjoner av to treslag. Her er resten av treslagene filtrert ut og det er plukket ut 50 tilfeldig utplukkede dunbjørkflater uten geografiske avgrensninger. Vedlegg 2 inneholder alle klassifiseringsfeilene for alle modellene, for hvert av de to treslagene i hver RF-klassifisering med minst to observasjoner. To observasjoner er minimum for at et treslag skal kunne være både i bootstrap-datasettet og blant OOB-observasjon som testes på beslutningstreet. Av tabellen i vedlegget ser man at det er mange treslag som har for få observasjoner til å ha et grunnlag for å trene modellen til å gjenkjenne treslaget.

Bakgrunnen for mulige skiller mellom klasser kan nærmere undersøkes ved at treslagene samles i grupper med treslag brukt i en ny RF-klassifisering. Dette gjøres for å avgjøre om snevrere inndeling i restklassen kan ha noe for seg, og for å se nærmere på grunnlaget for differensiering mellom treslagene på et overordnet vis.

Basert på forvirringsmatrisene fra klassifisering av alle treslag, og klassifiseringene mellom hvert enkelt løvtreslag settes en ny inndeling opp. Her er gråor og eik plasseres i hver sin klasse. Dette fordi eik og gråor er blant de to treslagene andre treslag oftest feilklassifiseres som (lav PA). I tillegg er dette treslag som er forsøkt å isolere fra andre løvtreslag, samt at de godt adskilt fra hverandre i modell II) og IV) i vedlegg 2. Resten av løvtreslagene fordeles på disse to klassene ut ifra hvilke treslag de i flest tilfeller feilklassifiseres som. I denne RF-klassifiseringen ser de nye samlede treslagsklassene ut som følger:

- Klasse 1: Gråor, Alm, Hassel, Hegg, Lavlandsbjørk, Osp, Platanlønn, Selje og Annet lauv
- Klasse 2: Eik, Dunbjørk, Rogn, Kristtorn, Lind, Svartor, Villeple, Ask, Bøk og Lerk

3.9 Utjevning for variasjon i antallsreduerte treslag og kryssvalidering

Når det kun velges ut en liten del av observasjonene for et treslag for å jevne ut antallet per klasse, vil resultatene bli veldig ulike hver gang klassifiseringen gjøres med nye data for et nytt utvalg observasjoner. For å jevne ut dette deles granobservasjonene og dunbjørkobservasjonene tilfeldig inn i grupper med rundt 40 observasjoner i hver gruppe. For granklassifiseringen deles observasjonene inn i 10 deler. For dunbjørk deles observasjonene inn i 60 grupper i gråorklassifiseringen og 3 grupper for eikedeteksjon. Gruppeantallet er så forskjellig da avgrensningene for eik filtrerer vekk flesteparten av dunbjørkobservasjonene i datasettet, mens gråorklassifiseringen ikke inneholder noen avgrensninger. Resultatet blir gjennomsnittet av nøyaktighetene fra hver iterasjon med RF-klassifisering.

For den beste klassifiseringen blant iterasjonene på de resterende dataene i en form for kryssvalidering. I illustrasjonen fra kapittel 2.6 om kryssvalidering ville eksempelvis antall iterasjoner vært 10 i granmodellen, og introduserte granarter ville vært en del av treningsdatasettet i alle 10 iterasjonene. I og med at bare én klasse er representert i testdatasettet skal bare én klasse ideelt sett klassifiseres som seg selv i denne kryssvalideringen. Ytterligere validering av modellens nøyaktighet gjøres så med et helt uavhengige testdatasett. Dette er likevel bare for å validere den klassen med redusert antall observasjoner, og ikke for å differensiere mellom to klasser i den uavhengige testen. PA og UA for klassen som bare er representert i treningsdatasettet er dermed uvisst etter denne kryssvalideringen. Det plottes videre UA for antallsdominant treslagsklasse for modellens beste iterasjon mot en UA fra de resterende usette testdataene mot hverandre. Dette vil fortelle oss i hvor stor grad økt nøyaktighet gjør modellen overtilpasset.

4 Resultater

4.1 Treslagsutvelgelsen

Treslagsutvelgelsen er gjort på bakgrunn av feilklassifiseringsandelen ($1 - UA$) for hvert treslag i en klassifisering med alle treslagene representert i datasettet. En verdi på 1 tilsier at alle observasjonene av treslaget er klassifisert som noe annet enn de er i realiteten. På samme måte vil en verdi på 0 tilsie at alle observasjonene av treslaget er klassifisert riktig.

Datasettet med interpolerte verdier over outliere, og med reduserte antall dunbjørk-, gran- og furuobservasjoner klassifisert i de fire modellene vises i Tabell 8. Treslagene for videre klassifiseringer gjort i oppgaven er markert i grønt.

Tabell 8 Klassifiseringsfeil for hvert treslag for alle fire modellene i random forest.

	Modell I)	Modell II)	Modell III)	Modell IV)
Gran	0.5208333	0.5000000	0.4791667	0.4375000
Introdusert gran	0.5172414	0.4137931	0.5172414	0.4482759
Edelgranarter	1.0000000	1.0000000	1.0000000	1.0000000
Furu	0.3333333	0.3750000	0.3750000	0.3541667
Contortafuru	1.0000000	1.0000000	1.0000000	1.0000000
Lerk	1.0000000	1.0000000	1.0000000	1.0000000
Annet bar	1.0000000	1.0000000	1.0000000	1.0000000
Dunbjørk	0.5319149	0.4893617	0.5319149	0.4893617
Lavlandsbjørk	0.9565217	1.0000000	1.0000000	1.0000000
Osp	0.8148148	0.7777778	0.8148148	0.8148148
Eik	0.9230769	0.1538462	0.9230769	0.1538462
Bøk	1.0000000	1.0000000	1.0000000	1.0000000
Ask	1.0000000	1.0000000	1.0000000	1.0000000
Alm	1.0000000	1.0000000	1.0000000	1.0000000
Lind	1.0000000	1.0000000	1.0000000	1.0000000
Platanlønn	1.0000000	1.0000000	1.0000000	1.0000000
Gråor	0.2096774	0.2258065	0.2741935	0.1935484
Svartor	1.0000000	1.0000000	1.0000000	1.0000000
Selje	1.0000000	1.0000000	1.0000000	1.0000000
Rogn	0.9230769	0.9615385	1.0000000	0.9615385
Hegg	1.0000000	1.0000000	1.0000000	1.0000000
Hassel	1.0000000	1.0000000	1.0000000	1.0000000
Villeple	1.0000000	1.0000000	1.0000000	1.0000000
Annet lauv	1.0000000	1.0000000	1.0000000	1.0000000
Kristtorn	1.0000000	1.0000000	1.0000000	1.0000000

De to klassene med størst grad av riktig klassifiserte observasjoner er eik og gråor. I tillegg er både gran og introdusert gran skilt ut til en viss grad fra resten av treslagene.

Forvirringsmatrise for alle treslagene brukt i inndelingen av de to samlede treslagsklassene i modell IV) finnes i Tabell 9, og diagonalen med riktig klassifiserte observasjoner er markert.

Tabell 9 Forvirringsmatrise fra RF-klassifisering av alle treslag med modell IV), og tilhørende feilklassifiseringsandel.

	1	2	3	10	11	20	29	30	31	32	40	41	42	43	44	49	50	51	52	53	54	55	57	59	70	Feil
1	27	3	0	5	0	0	0	8	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0.4375000
2	5	16	0	3	0	0	1	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4482759
3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000000
10	8	1	0	23	0	0	0	8	0	1	1	0	0	0	0	0	3	0	0	1	0	0	0	0	0	0.5000000
11	3	2	0	2	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000000
20	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000000
29	0	1	0	2	0	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8333333
30	1	0	0	2	0	0	0	32	0	0	0	0	0	0	0	0	10	0	0	1	0	0	0	0	0	0.3043478
31	1	0	0	6	0	0	0	0	0	2	4	0	0	0	0	0	9	0	0	1	0	0	0	0	0	1.0000000
32	2	0	0	6	0	0	0	4	0	4	2	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0.8518519
40	0	1	0	2	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1153846
41	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1.0000000
42	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0	1	0	0	0	0	0	1.0000000
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1.0000000
44	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000000
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1.0000000
50	3	1	0	4	0	0	0	5	1	0	1	0	0	0	0	0	47	0	0	0	0	0	0	0	0	0.2419355
51	0	0	0	1	0	0	0	0	0	0	4	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1.0000000
52	3	0	0	1	0	0	0	3	0	0	0	0	0	0	0	0	7	0	0	2	0	0	0	0	0	1.0000000
53	1	2	0	2	0	0	1	9	0	1	1	0	0	0	0	0	7	0	1	1	0	0	0	0	0	0.9615385
54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	1.0000000
55	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	1	0	0	0	0	0	1.0000000
57	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000000
59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1.0000000
70	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000000

Treslagene er representert med tall som er oversatt under:

Gran

- 1 – Gran
- 2 – Introdusert gran
- 3 – Edelgranarter

Furu

- 10 – Furu
- 11 – Contortafuru
- 20 – Lerk
- 29 – Annet bar

Løvtrær

- 30 – Dunbjørk
- 31 – Lavlandsbjørk
- 32 – Osp
- 40 – Eik
- 41 – Bøk
- 42 – Ask
- 43 – Alm
- 44 – Lind
- 49 – Platanlønn
- 50 – Gråor
- 51 – Svartor
- 52 – Selje
- 53 – Rogn
- 54 – Hegg
- 55 – Hassel
- 57 – Villeple
- 59 – Annet lauv
- 70 – Kristtorn

4.2 Gran

Tabell 10 til Tabell 13 viser RF-modellene for granklassifiseringene med geografiske avgrensninger. I tillegg er UA (User's accuracy), PA (Producer's accuracy) og OA (Overall accuracy) regnet ut for forvirringsmatrisene.

Tabell 10 Forvirringsmatrise for RF-klassifisering av granarter med modell I), og tilhørende nøyaktighetsmål.

Modell I)		Treslag fra klassifisering		
Faktiske treslag		Gran	Introdusert gran	UA
		Gran	39	9
	Introdusert gran	18	14	44 %
PA		68 %	61 %	
OA				66 %

Tabell 11 Forvirringsmatrise for RF-klassifisering av granarter med modell II), og tilhørende nøyaktighetsmål.

Modell II)		Treslag fra klassifisering		
Faktiske treslag		Gran	Introdusert gran	UA
		Gran	39	9
	Introdusert gran	10	22	69 %
PA		80 %	71 %	
OA				76 %

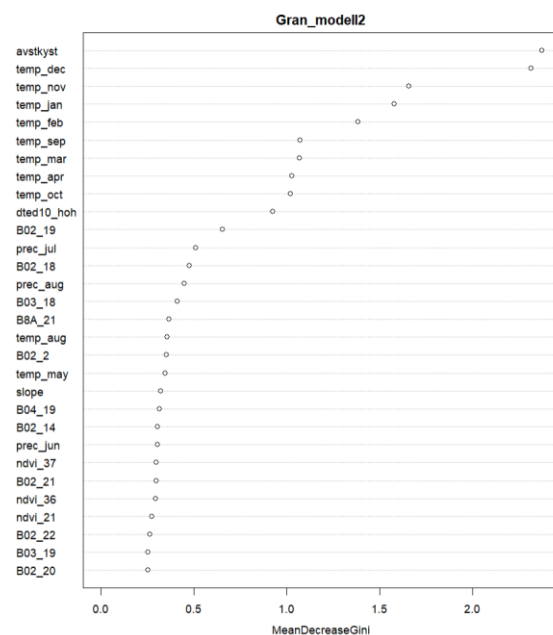
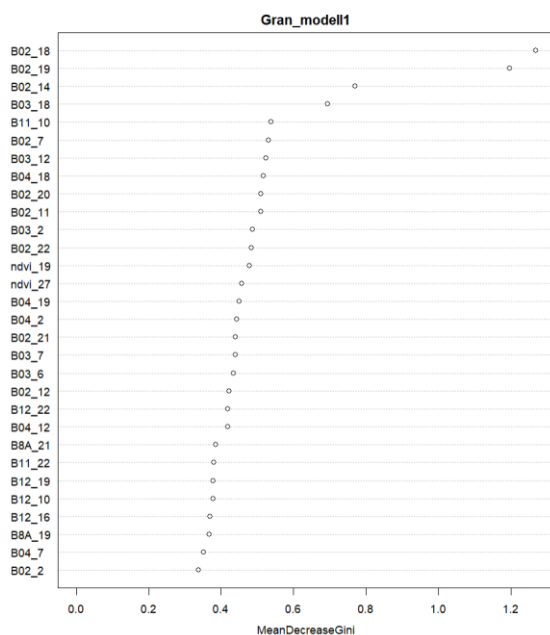
Tabell 12 Forvirringsmatrise for RF-klassifisering av granarter med modell III), og tilhørende nøyaktighetsmål.

Modell III)		Treslag fra klassifisering		
Faktiske treslag		Gran	Introdusert gran	UA
		Gran	39	9
	Introdusert gran	16	16	50 %
PA		71 %	64 %	
OA				69 %

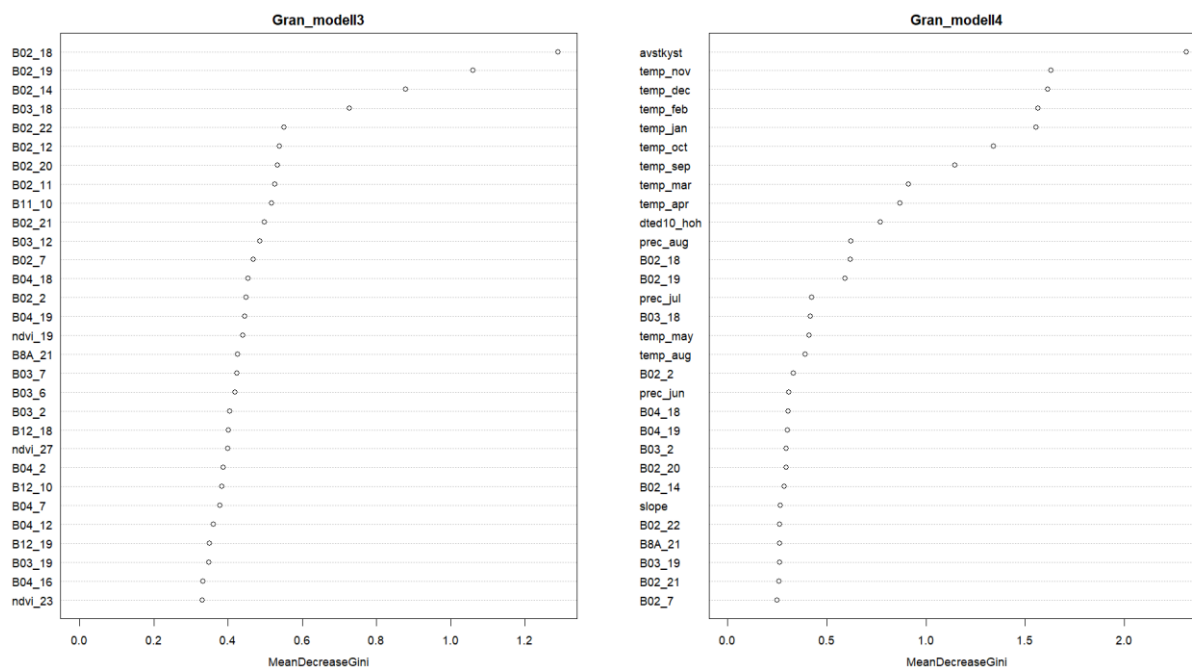
Tabell 13 Forvirringsmatrise for RF-klassifisering av granarter med modell IV), og tilhørende nøyaktighetsmål.

Modell IV)		Treslagsklasse fra modellen		
Faktiske treslag		Gran	Introdusert gran	UA
		Gran	40	8
	Introdusert gran	9	23	72 %
PA		82 %	74 %	
OA				79 %

Videre er variabelviktigheten plottet for hver av modellene. Variabler med høyere MeanDecreaseGini har større betydning i klassiferingen, som vist for alle modellene i Figur 21 og Figur 22.



Figur 21 Variable importance plots for hhv. modell I) og II) for RF-klassifisering av granarter. Merk ulik målestokk på MeanDecreaseGini.



Figur 22 Variable importance plots for hhv. modell III) og IV) for RF-klassifisering av granarter Merk ulike målestokk på MeanDecreaseGini.

For modell I) og III) er det tydelig at det blå båndet (B02) fra de seinere ukene i tidsserien har mest å si for klassifiseringen. For modell II) og IV) er det de klimatiske variablene og spesielt kystavstand og temperaturer i vinterhalvåret som har størst innvirkning.

Gjennomsnittlige nøyaktigheter og tilhørende standardavvik til granmodell IV) for de 10 iterasjonene med ulikt utvalg granobservasjoner finnes i Tabell 14.

Tabell 14 Gjennomsnittlige nøyaktigheter for alle iterasjoner av granmodell IV) og tilhørende standardavvik.

	UA	sd	PA	sd
Gran	83 %	6,1 %	81 %	3,1 %
Introdusert gran	72 %	6,3 %	77 %	4,6 %
OA	79 %	3,0 %		

Forvirringsmatrisen og tilhørende nøyaktigheter for den beste iterasjonen for modell IV) ser man i Tabell 15.

Tabell 15 Forvirringsmatriser og tilhørende nøyaktigheter for den beste iterasjonen av granklassifiseringene.

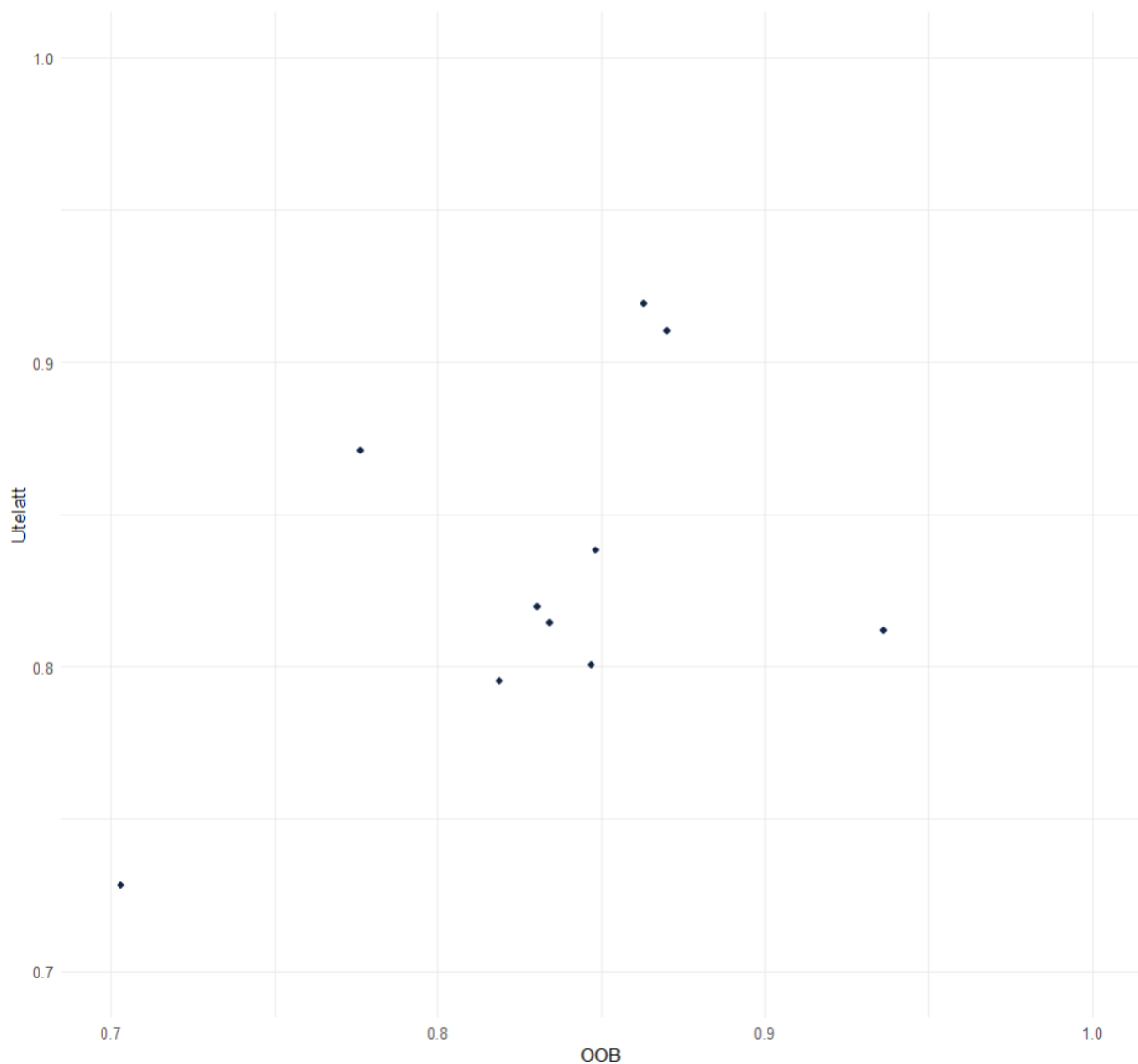
Modell IV)	Treslagsklasse fra modellen			
Faktiske treslag		Gran	Introdusert gran	UA
	Gran	37	8	82 %
	Introdusert gran	6	26	81 %
PA		86 %	76 %	
OA				82 %

De granobservasjonene som er utelatt i den beste modellen blant iterasjonene, er klassifisert med samme modell fra Tabell 15, og forvirringsmatrise med tilhørende UA vises i Tabell 16. UA for gran er her 80 %, og nesten lik den i Tabell 15.

Tabell 16 Forvirringsmatrise og tilhørende nøyaktighet for granobservasjoner i testdataene med modell IV) fra den beste iterasjonen i granklassifisering.

Modell IV)		Treslag fra klassifisering		
Faktisk treslag		Gran	Introdusert gran	UA
Gran	311	80		80 %

Videre er UA for OOB-klassifiseringsmodellen i den beste iterasjonen, og for modellen brukt på de utelatte observasjonene plottet i Figur 23.



Figur 23 Plot for UA for OOB-modellen og for modellen påført de utelatte granobservasjonene.

4.3 Eik

På samme måte som for gran er forvirringsmatrisene for eikemodell I) til IV) og tilhørende nøyaktighetsmål regnet ut i Tabell 17 til Tabell 20.

Tabell 17 Forvirringsmatrise fra RF-klassifisering for deteksjon av eik med modell I), og tilhørende nøyaktighetsmål.

Modell I)		Treslag fra klassifisering		
Faktiske treslag		Annet løv	Eik	UA
		Annet løv	70	3
	Eik	21	5	19 %
PA		77 %	63 %	
OA				76 %

Tabell 18 Forvirringsmatrise fra RF-klassifisering for deteksjon av eik med modell II), og tilhørende nøyaktighetsmål.

Modell II)		Treslag fra klassifisering		
Faktiske treslag		Annet løv	Eik	UA
		Annet løv	67	6
	Eik	12	14	54 %
PA		85 %	70 %	
OA				82 %

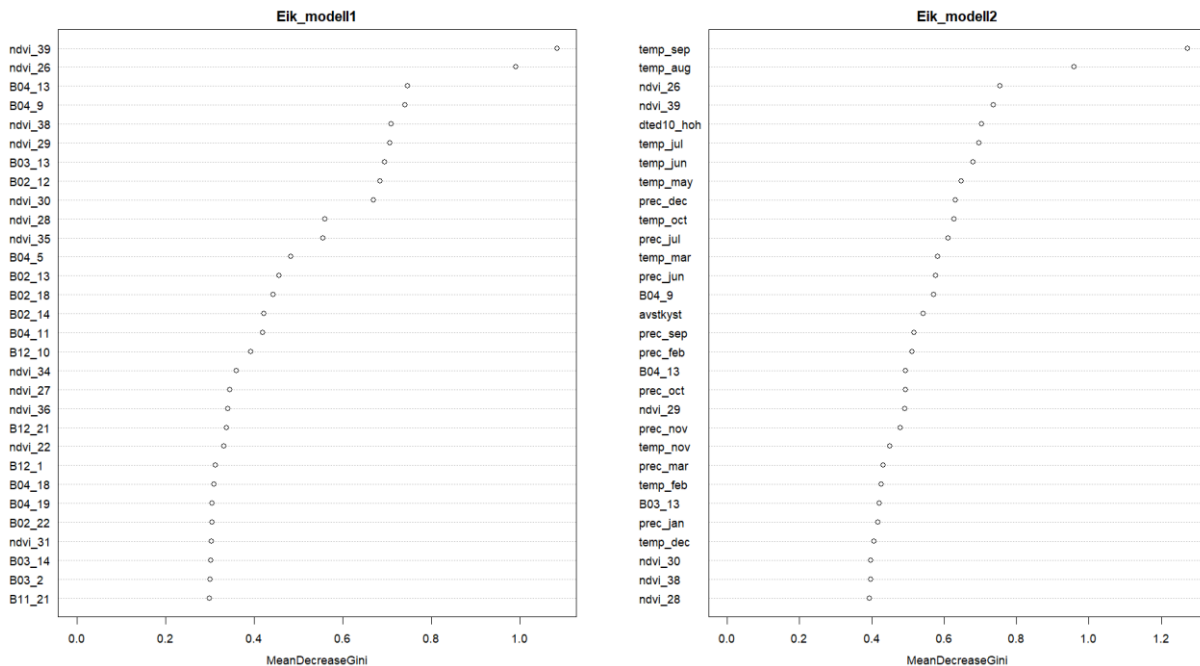
Tabell 19 Forvirringsmatrise fra RF-klassifisering for deteksjon av eik med modell III), og tilhørende nøyaktighetsmål.

Modell III)		Treslag fra klassifisering		
Faktiske treslag		Annet løv	Eik	UA
		Annet løv	69	4
	Eik	20	6	23 %
PA		78 %	60 %	
OA				76 %

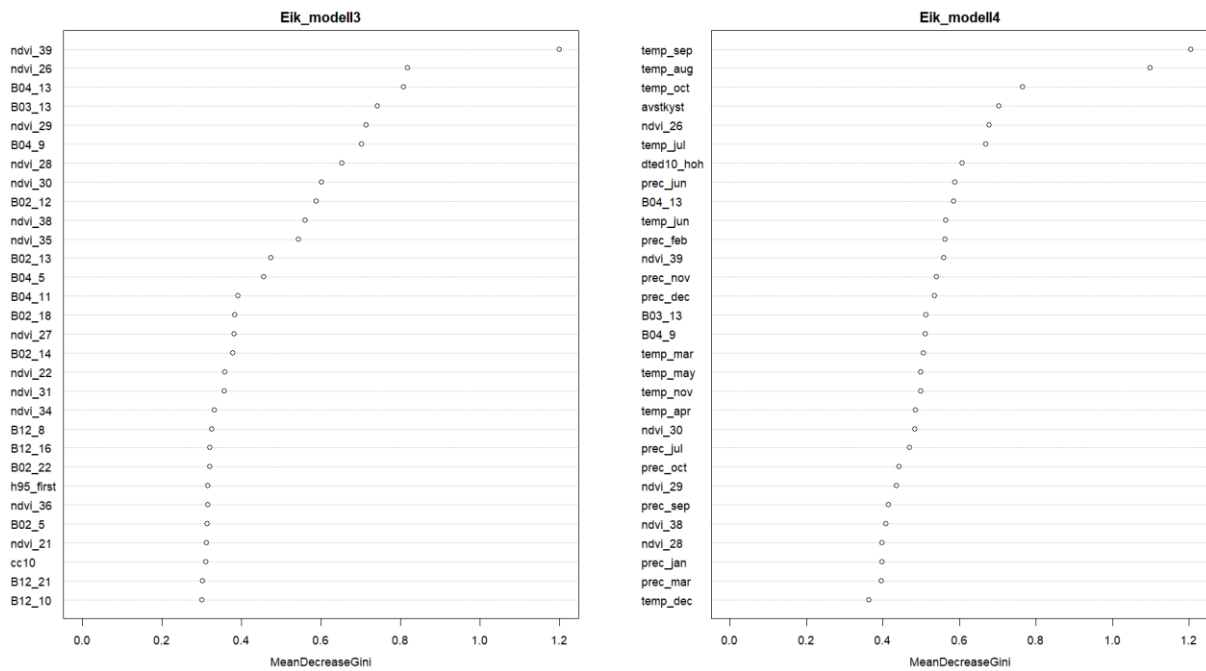
Tabell 20 Forvirringsmatrise fra RF-klassifisering for deteksjon av eik med modell IV), og tilhørende nøyaktighetsmål.

Modell IV)		Treslag fra klassifisering		
Faktiske treslag		Annet løv	Eik	UA
		Annet løv	68	5
	Eik	12	14	54 %
PA		85 %	74 %	
OA				83 %

Viktighetsplot for de 30 viktigste variablene i RF-klassiferingen av eik blant løvtrær er å finne i Figur 24 og Figur 25.



Figur 24 Variable importance plots for hhv. modell I) og II) for eikedeteksjon med RF-klassifisering. Merk ulike målestokker på MeanDecreaseGini.



Figur 25 Variable importance plots for hhv. modell III) og IV) for eikedeteksjon med RF-klassifisering. Merk ulik målestokk på MeanDecreaseGini.

Variablene av størst viktighet er her i størst grad NDVI, og temperaturer der modellene inkluderer det.

Gjennomsnittlige nøyaktigheter og tilhørende standardavvik til eikemodell IV) for de 3 iterasjonene med ulikt utvalg dunbjørkobservasjoner finnes i Tabell 21.

Tabell 21 Gjennomsnittlige nøyaktigheter for alle iterasjoner av eikemodell IV) og tilhørende standardavvik.

	UA	sd	PA	sd
Annet løv	94 %	3,2 %	83 %	1,3 %
Eik	36 %	7,9 %	69 %	12,0 %
OA	81 %	2,6 %		

4.4 Gråor

På samme måte som for gran og eik er forvirringsmatrisene for modell I) til IV) og tilhørende nøyaktighetsmål regnet ut i Tabell 22 til Tabell 25.

Tabell 22 Forvirringsmatrise fra RF-klassifisering for deteksjon av gråor med modell I), og tilhørende nøyaktighetsmål.

Modell I)		Treslag fra klassifisering		
Faktiske treslag		Annet løv	Gråor	UA
	Annet løv	181	6	97 %
	Gråor	55	7	11 %
PA		77 %	54 %	
OA				76 %

Tabell 23 Forvirringsmatrise fra RF-klassifisering for deteksjon av gråor med modell II), og tilhørende nøyaktighetsmål.

Modell II)		Treslag fra klassifisering		
Faktiske treslag		Annet løv	Gråor	UA
	Annet løv	181	6	97 %
	Gråor	43	19	31 %
PA		81 %	76 %	
OA				80 %

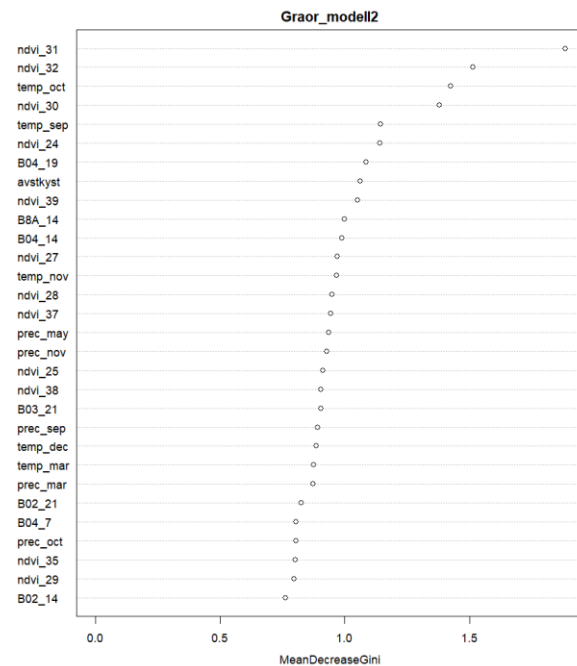
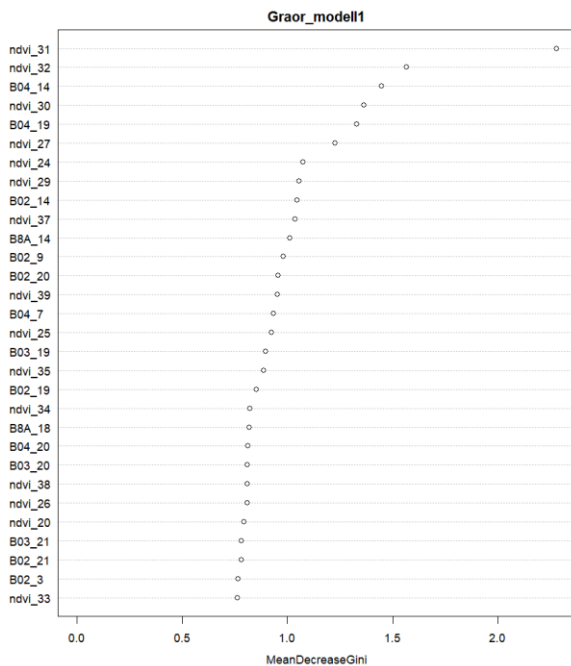
Tabell 24 Forvirringsmatrise fra RF-klassifisering for deteksjon av gråor med modell III), og tilhørende nøyaktighetsmål.

Modell III)		Treslag fra klassifisering		
Faktiske treslag		Annet løv	Gråor	UA
	Annet løv	181	6	97 %
	Gråor	56	6	10 %
PA		76 %	50 %	
OA				75 %

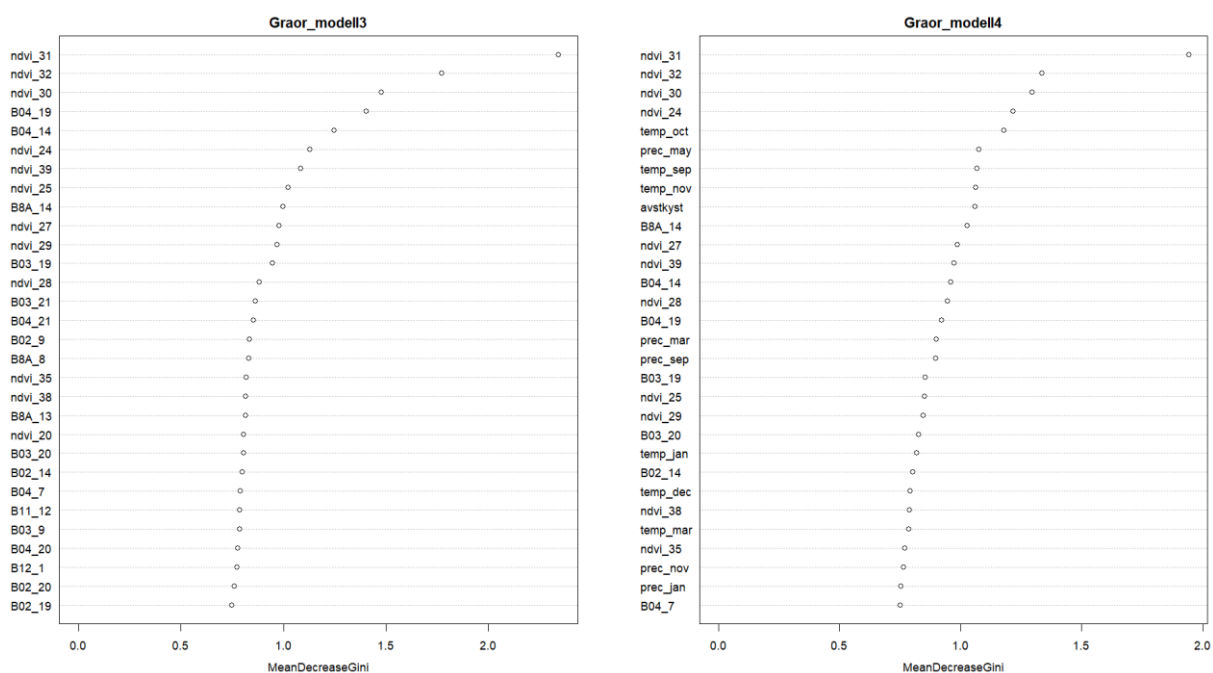
Tabell 25 Forvirringsmatrise fra RF-klassifisering for deteksjon av gråor med modell IV), og tilhørende nøyaktighetsmål.

Modell IV)		Treslag fra klassifisering		
Faktiske treslag		Annet løv	Gråor	UA
	Annet løv	180	7	96 %
	Gråor	42	20	32 %
PA		81 %	74 %	
OA				80 %

Viktighetsplot for de 30 viktigste variablene i RF-klassiferingen av gråor blant løvtrær er å finne i Figur 26 og Figur 27.



Figur 26 Variable importance plots for hhv. modell I) og II) for gråordeteksjon med RF-klassifisering. Merk ulike målestokker på MeanDecreaseGini.



Figur 27 Variable importance plots for hhv. modell III) og IV) for gråordeteksjon med RF-klassifisering. Merk ulik målestokk på MeanDecreaseGini.

NDVI-verdiene er viktigst i alle modellene, men for modell II) og IV) blir også spesielt høsttemperaturer viktige i klassifiseringene.

Gjennomsnittlige nøyaktigheter og tilhørende standardavvik til gråormmodell IV) for de 60 iterasjonene med ulikt utvalg dunbjørkobserasjoner finnes i Tabell 26.

Tabell 26 Gjennomsnittlige nøyaktigheter for alle iterasjoner av gråormmodell IV) og tilhørende standardavvik.

	UA	sd	PA	sd
Annet løv	90 %	1,6 %	76 %	1,1 %
Gråor	40 %	3,3 %	68 %	3,4 %
OA	65 %	2,1 %		

4.5 Fordeling i to løvtreklasser

Tabell 27 viser forvirringsmatrisen tilhørende RF-klassifisering gjort med de to klassene som er konstruert basert på likheten til andre løvtreslag. Å plukke ut 100 tilfeldige observasjoner fra klasse 1 med samme klassifiseringsmodell gir forvirringsmatrisen i Tabell 28.

Her er modell IV) benyttet da denne er totalt best i klassifisering for både gråor og eik.

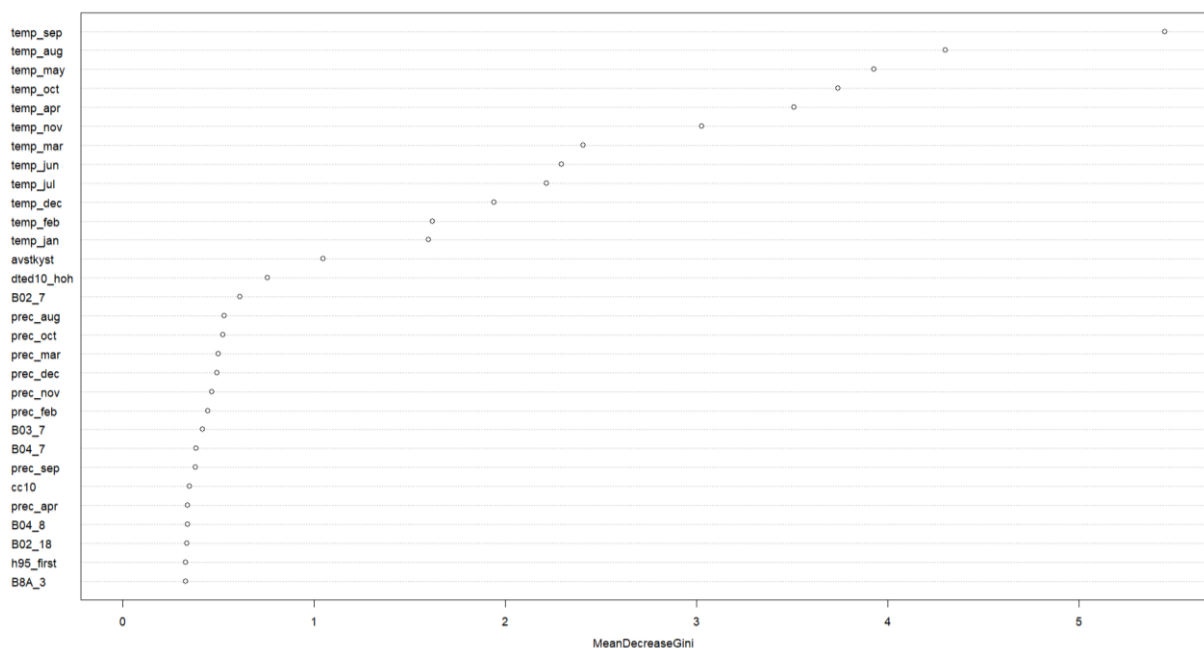
Tabell 27 Forvirringsmatrise for RF-klassifisering av ny inndeling i to klasser med modell IV).

Modell IV)		Treslag fra klassifisering		
Faktiske treslag		1	2	UA
	1	195	6	97 %
2	13	31	70 %	
PA		94 %	84 %	
OA				92 %

Tabell 28 Forvirringsmatrise for RF-klassifisering av ny inndeling med reduserte antall observasjoner i klasse 1 med modell IV).

Modell IV)		Treslag fra klassifisering		
Faktiske treslag		1	2	UA
	1	94	6	94 %
2	7	37	84 %	
PA		93 %	86 %	
OA				91 %

I begge tilfeller er temperaturer, deretter kystavstand og deretter nedbør de viktigste variablene i klassiferingen. For modell I) er det 94 % (UA₂ = 6 %) feilklassifisering av klasse 2, og 77 % (UA₂ = 23 %) feilklassifisert for modell III) med ytterligere redusert antall observasjoner i klasse 1, som i Tabell 28. Variabelviktighetsplottet for modell IV) fra Tabell 27 vises i Figur 28.



Figur 28 Variable importance plot for RF-klassifisering med de to nyinndelte klassene.

De klimatiske variablene, og spesielt temperatur, er avgjørende i denne klassiferingen.

Gjennomsnittlige nøyaktigheter og tilhørende standardavvik til modell IV) for de 60 iterasjonene med ulikt utvalg dunbjørkobservasjoner finnes i Tabell 29.

Tabell 29 Gjennomsnittlige nøyaktigheter for alle iterasjoner av modell IV) for den nye klasseinndelingen og tilhørende standardavvik. Ytterligere reduksjon av observasjonsantall i klasse 1 er ikke foretatt.

	UA	sd	PA	sd
1	98 %	0,5 %	93 %	0,6 %
2	68 %	3,0 %	86 %	2,7 %
OA	92 %	0,7 %		

5 Diskusjon

5.1 Datagrunnlaget

Dataene fra Sentinel-2 har 20 meters bakkeoppløsning. Det betyr at det er store områder med samlet verdi for pikslene i de representerte båndene. Det kan være store variasjoner innad i hver piksel, men dette vil ikke synes i S2-bildene. På den ene siden gjør dette det vanskelig å finne homogene områder å trene en modell på, i tillegg til at det vil kunne være verdifull informasjon som ikke vises. På den andre side blir det mulig å gjøre storskala analyser. Dersom man finner en god modell for klassifisering av treslag med denne oppløsningen, vil det kunne brukes over alt hvor satellittene passerer over, og andre kriteriene som settes for bruk av modellen er oppfylt.

Som forklart i kapittel 2.3 om hvordan Landsskogstakseringsflatene og S2-tidsseriene ser ut over hverandre, ser man tydelig at det ikke er en naturlig overlapp. I verste fall vil fire piksler være brukt i utregningen av reflektert verdi. Det vil da være store deler av pikselområdet til S2-datasettet som ikke tilhører observasjonene innenfor prøveflaten den representerer i Landsskogstakseringsdataene. Det er derfor umulig å være sikker på at treslaget man benytter i analysene er dominant i like stor grad i satellitt-dataene, som i feltobservasjonene på prøveflaten. Dette kan utgjøre en mulig feilkilde for tidsseriene. Likevel er det sannsynligvis store likheter i treslagssammensetningen i det omliggende området, og det dominerende treslaget innad i hver prøveflate vil i mange tilfeller bre seg utover flateavgrensningene. For å se noen treslag oftere vokser blandet enn andre lages Tabell 30. Av tabellen er det vanskelig å se et tydelige mønster i hvilke treslag som er lettere å skille basert på graden av treslagets dominans for prøveflatene. Man kan derfor tenke seg at for datasettet brukt i analysene, er det trygt å anta omtrent lik grad av homogenitet i prøveflatene for alle treslag.

Tabell 30 Antall prøveflater per treslag for ulike dekningsgrader.

	75% - 85%	85% - 95%	95% <
Alm	1	0	1
Annet bar	6	0	0
Annet lauv	0	0	1
Ask	4	0	1
Bøk	1	0	1
Contortafuru	6	2	3
Dunbjørk	1887	359	252
Edelgranarter	1	1	1
Eik	13	7	8
Furu	1759	701	458
Gran	1094	665	525
Gråor	28	21	14
Hassel	3	0	1
Hegg	2	1	0
Introdusert gran	17	10	3
Kristtorn	1	0	0
Lavlandsbjørk	9	0	9
Lerk	2	0	0
Lind	1	0	1
Osp	8	11	10
Platanlønn	0	1	0
Rogn	20	6	5
Selje	7	7	6
Svartor	2	2	3
Villeple	1	0	0

5.2 Metode

5.2.1 Interpolasjon

Skogen i satellittbilde vil i de fleste tilfellene i datasettet være delvis tilslørt av skyer noen steder i landet. Interpolasjonen som er gjort vil produsere syntetiske verdier tilnærmet de som er forventet å finne under skydekket. Dette betyr at refleksjonsverdiene i disse interpolerte områdene ikke er reelle, og dermed kan være avvikende fra de sanne reflekterte verdiene. I tillegg til den geografiske interpolasjonen for skydekket i S2-bildene, er det etter datasettet i denne oppgaven ble mottatt, gjort en ytterligere lineær interpolasjon for verdier som er mistenkt oversett i denne skydeteksjonen. Denne interpolasjonen er gjort i tidsaspektet av datasettet, og vil kunne gi riktige verdier for geografisk plassering, men vil mulig gi et mindre korrekt bilde av endring over tid. Interpolasjon er i utgangspunktet ikke ønskelig da det forstyrrer målte data. Likevel er disse interpolasjonene i dette tilfellet nødvendige for å få et datasett som kun viser jordoverflaten så det er mulig å benytte modelleringsverktøyet random forest.

5.2.2 Observasjoner – antall og fordeling

Det største hinderet i klassifiseringen er observasjonsantallet. Dette gjelder både mangel på tilstrekkelig observasjonsantall for mange av treslagene, men også skjevfordelingen av observasjonsantall mellom treslagene. Dette gir store utslag på forvirringsmatrisene, som viser at en økt variasjon i observasjonsgrunnlaget for ett treslag gjør det vanskeligere for modellene å skille treslagene fra hverandre. Denne variasjonen finnes både innad i treslag og mellom treslag som er samlet i én klasse. Gråor- og eikedeteksjon er spesielt utsatt for dette, der *Annet løv*-klassen inneholder 18 ulike treslag.

5.2.3 Variabelutvelgelse og korrelasjoner

Romlig autokorrelasjon ville kunne forstyrre klassifiseringen av treslagene (Comber et al., 2012; Fassnacht et al., 2016), men i og med at Landsskogstakseringsflatene er plassert minst 3 km unna hverandre (Viken, 2021), kan man se bort ifra korrelasjon mellom flatene for laserdata og S2-tidsserier. For de klimatiske variablene vil geografisk plassering ha betydning. Med tettere prøveflater ville dette kunne forstyrre klassifiseringen (Fassnacht et al., 2016), men med den jevne fordelingen fra Landsskogstakseringen favoriseres ikke klynger med observasjoner i klassifiseringen. Likevel er det ingen garanti for at det er balanse i geografisk fordelingen av feilklassifiseringer, som ikke nødvendigvis kan oppdages i forvirringsmatrisene med tilhørende nøyaktigheter (Comber et al., 2012).

Det som i større grad vil kunne være problematisk i klassifiseringsanalysen er korrelasjon mellom variablene (Janitza et al., 2018). I denne oppgaven er omtrent alle variablene sterkt korrelert med minst én annen variabel. Tidsserieverdiene for hvert av de representerte båndene i Sentinel-2 vil for alle treslagene øke jevnt i ukene mot midten av tidsrommet, og vil kunne vises som korrelasjon mellom båndene. Figur 9 viser denne korrelasjonen mellom bånd med bølgelengder nær hverandre på det elektromagnetiske spekteret. Figuren viser også at ukene er så tett plassert i tidsaspektet at det er en korrelasjon mellom nærliggende uker. Dette gjør at resultatet kunne sett annerledes ut dersom variablene brukt i modellen var gjort en mer kritisk vurdering av. Dersom det for eksempel var regnet en snittverdi per måned for hvert bånd i S2-tidsseriene og for hvert kvartal for temperaturer og nedbør, ville muligens resultatene sett annerledes ut. Klassifiseringene vil også kunne forbedres ytterligere ved å fjerne variablene fra klassifiseringen som ikke har en relevans i klassifiseringsprosessen, da det er mulig at en del av variablene har en negativ påvirkning på

klassifisering (Janitza et al., 2018). Forbedringer med variabelutvalg kunne eksempelvis også vært undersøkt i en PCA (Principal Component Analysis), eller ved hjelp av dyplæring.

5.2.4 Kyssvalidering og utjevning for observasjonsreduksjon

For å sikre korrektheten i en modell gjøres gjerne en kryssvalidering for klassifisering. Innebygd i random forest ligger en *Out-of-Bag (OOB)* validering som skal i hindre skjevhet og feil i modellen til en viss grad på samme måte som en kryssvalidering vil gjøre. Inndelingen i OOB er likevel ikke helt objektiv når det antas at observasjonene er helt uavhengige av hverandre, mens de i realiteten sjelden er dette (Cánovas-García et al., 2017). I RFs OOB-validering testes likevel modellen på observasjoner som ikke er brukt for å bygge hvert beslutningstre i modellen, og på denne måten motvirker man overtilpasning (Bhatia, 2021). Ytterligere valideringer med et uavhengig testdatasett utelatt fra modellen vil kunne hjelpe med et enda sikrere resultat, (Fassnacht et al., 2016) men med et mindre antall observasjoner vil dette være u hensiktsmessig. Derfor er ikke dette gjort i denne oppgaven. Det er heller testet på uavhengige testdata for en av klassene der observasjonsantallet er betydelig høyere for ett treslag.

Det er ikke gjort en tradisjonell kryssvalidering for resultatene av RF-klassifiseringene, men heller en type validering som tar høyde for variasjonene i observasjonene som er redusert i antall, for å se forskjell på treslagene i sin helhet. Det er stor variasjon innad i hvert av treslagene for mange av variablene. Ved å gjøre en utjevning der alle observasjoner av det reduserte antallet prøveflater tar del i én iterasjon, der gjennomsnittet av nøyaktighetene regnes ut for alle iterasjonene, vil dette jevnes ut til en viss grad. På denne måten vil variasjonen være bevart, men antallet vil være jevnet ut for hver klassifisering. Likevel vil kombinasjonen av utvalgte observasjoner variere hver gang klassifiseringskoden kjøres.

Videre er enda en validering gjort for den beste granmodelliterasjonen. Dette er alternativet som er brukt for et utelatt testdatasett fra opprinnelige data. På samme måte som et uavhengig testdatasett er ingen av observasjonene benyttet i bygging av modellen. Gran er likevel den eneste klassen som er representert, og det blir ikke gjort en fullstendig test av modellens evne til å klassifisere introdusert gran. Likevel ble 80% av granobservasjonene i testdatasettet klassifiseres rett med modellen fra den beste iterasjonen. Valideringen viser (som er plottet i Figur 23) at forholdet mellom UA for gran i OOB-modellen og denne modellen påført testdataene er tilfeldig spredt i plottet. Dette betyr at økt nøyaktighet i OOB-modellen ikke er på grunn av overtilpasning, da modellen er overførbart til usette testdata. En negativ korrelasjon ville her indikert det motsatte.

5.3 Random forest som klassifiseringsverktøy

Med tidsseriedata er det mange variabler per observasjon, og antall observasjoner for noen av treslagene representert i dataene er ikke så høyt. Random forest er robust i møte med både overtilpasning og stort antall variabler i forhold til observasjoner (Breidenbach, Waser, et al., 2020). Dette gjør at modellen egner seg godt til klassifiseringene i oppgaven. Det var likevel store utfordringer knyttet opp mot ujevn fordeling av observasjoner per treslag, og analyse av variabelviktighet i modellene (Janitza et al., 2018).

95,4% av alle observasjonene i datasettet brukt i analysen er fordelt på de tre mest utbredte treslagene i Norge: Gran (*Picea abies*), furu (*Pinus sylvestris*) og vanlig bjørk (*Betula pubescens*) (Breidenbach, Granhus, et al., 2020). Dette gjorde at å skille ut flere treslag enn det som allerede er klassifisert i SR16 – Skogressurskart, uten tilpasning av observasjonsantallet, var tilnærmet umulig å få noen resultater på. Ved å redusere antall observasjoner mister man en variasjon, og man risikerer

å klassifisere basert på et lite representativt utvalg observasjoner. Likevel er dette nødvendig for å få noe ut av klassifiseringene gjort med random forest. Det jevnes også ut til en viss grad gjennom å benytte gjennomsnittlige nøyaktigheter for RF-klassifiseringene med ulike utvalg av overrepresenterte treslag.

En enkel RF-klassifisering godtar ikke inputobservasjoner uten verdier for alle variabler. Dette betyr at vurderingen av variabelers påvirkning på klassifiseringen, må settes opp mot antall frafallende observasjoner med manglende verdier. I og med at usikkerheten rundt om disse variablene har noe innvirkning på resten av variablene, kan heller ikke de manglende verdiene beregnes. 'dtw' reduserte i dette tilfellet observasjonsantallet betraktelig, og ble fjernet fra inputvariablene. Manglende verdier for denne variabelen er ikke på grunn av mangelfulle feltobservasjonene, men det er ikke alle områder der det eksisterer vannmengder i bakken under prøveflaten.

I interpolasjonen av outliere som er fjernet fra tidsserien, er det imidlertid avgjørende å fylle ut manglende verdier for å kunne bruke random forest som verktøy i klassifiseringen. Det finnes alternativer til lineær interpolasjon der manglende verdier erstattes med alternative verdier trent på resten av datasettet (Tang & Ishwaran, 2017). Alternative fremgangsmetoder kunne her være undersøkt for å fylle ut manglende verdier i tidsseriene. I et tilfelle der nabovertiene er så tett korrelerte, vil dette likevel være et mindre viktig fokusområde.

5.4 Vurdering av resultater

5.4.1 Gran

Laservariablene i modell III) ser ut til å ha liten påvirkning på klassifiseringen. Det er en liten forbedring i klassifiseringen fra modell I) til III), men det er ingen stor påvirkning synlig hverken alene eller i kombinasjon med klimatiske variabler. Forbedringene kan her skyldes tilfeldigheter når forskjellene er såpass små.

Modell II) og IV) gav de beste klassifiseringene. Figur 21 gir et inntrykk av at temperaturer er mindre viktige enn avstanden til kyst i klassifiseringa, men i og med at de er så tett korrelerte, vil viktigheten av temperatur være spredt over flere variabler som er relativt like (Janitza et al., 2018).

Variabelviktighetsplottene kan derfor gi et feilaktig inntrykk av den reelle virkningen av variabelenes påvirkning på klassifiseringene. Fra variablene i modell I) er i stor grad båndene for synlig lys viktigst for klassifiseringen. I tillegg er noen NDVI-variabler med, men dette kan skyldes at de er regnet ut delvis basert på bånd 4. 'Prec'-variablene beskriver nedbørsmengde, og kan som sett på Figur 9 ha sammenheng med både temperaturer og avstand til kystlinja, som begge deler er viktige variabler for klassifiseringen. Nedbør er likevel ikke like essensielt i klassifiseringen av grantyper.

For den beste modellen i granklassifiseringen (Modell IV)) var *avstand til kyst (avstkyst)* den viktigste variabelen i plottet i Figur 22. Granobservasjoner er plukket ut fra området som er avgrenset 50km fra kysten. Dette betyr at i teorien kan alle observasjonene av vanlig gran befinne seg 50km unna kysten, og modellen vil kunne benytte kun avstand til kyst for å få en relativt presis klassifisering. Dette later ikke til å være tilfellet i utplukkede granobservasjoner, selv om det er en tydelig forskjell på gjennomsnittlige kystavstand på prøveflatene i de to granklassene.

De introduserte granartene består i svært stor grad av Sitkagran, som på 1950-tallet ble plantet langs vestkysten av Norge på grunn av behovet for skogreising (Nordbø, 1954). Sitkagran ble tatt i bruk her

da vanlig gran har større problemer med å vokse nær kysten i vindutsatte områder og med salt sjøsprøyt (Frøystad, 1961). Dette indikerer at avstanden til kysten har betydning for evnen de ulike granartene har til å vokse nær kysten. Her er derfor *avstkyst*-variabelen en god indikator på forskjellen mellom klassene og ikke bare et resultat av redusert antall granobservasjoner. Dette støttes også av resultatene i Tabell 14, der variasjonen i alle granobservasjoner innen det avgrensede området er jevnet ut med gjennomsnittet for hver iterasjon. Her er det små forskjeller i de gjennomsnittlige nøyaktighetsmålene, og nøyaktighetene fra 50 tilfeldige utplukkede granobservasjoner fra Tabell 13.

Modellen fra den beste iterasjonen fra utjevningsskoden for observasjonsreduksjonene for gran, ser ut til å være god for de resterende granobservasjonene som er utelatt i denne iterasjonen. Dette gjør at det er en mulighet å benytte den beste iterasjonen for prediksjoner utover treningsdatasettet med større sikkerhet. Likevel sier ikke dette noe om hvor god modellen er til å klassifisere introdusert gran utover de dataene brukt for å lage modellen. Ved å påføre denne modellen til områder utenfor prøveflatene i Landsskogstakseringen, vil derimot ytterligere undersøkelse av modellenes nøyaktighet kunne gjøres ved feltobservasjoner av områder der introdusert gran antas å vokse.

5.4.2 Eik

I RF-klassifiseringen for klassene Eik og andre løvtreslag klarer ikke noen av modellene å skille ut flere enn rundt halvparten av eikeobservasjonene fra andre løvtreslag. Eik var treslaget det var lettest å skille i klassifiseringen med alle prøveflatene fra datasettet med mer enn 75% dominans. Dette kan komme av at dunbjørkobservasjonene som var plukket ut i modellen for alle treslag befant seg på helt andre steder enn bare langs sørkysten. Det vil også være mange færre observasjoner av andre løvtreslag med avgrensningene i eikeklassifiseringene, enn for klassifiseringen av alle treslagene uten geografiske avgrensninger. Dette vil si at variabler som temperatur, høyde over havet og avstand til kyst kan ha vært veldig forskjellige fra de utplukkede observasjonene i eik-klassifiseringen. Det vil da ha vært lettere å skille ut eikeflater utelukkende basert på variabler som avhenger av plasseringen i landet da alle observasjoner var inkludert i klassifiseringen. Derfor er det viktig å avgrense området der dunbjørkobservasjonene plukkes ut før klassifiseringen gjøres.

Det er stor forskjell i kvaliteten på modellene for eikedeteksjon. Det er en veldig liten forbedring fra modell I) til III), og på samme måte fra modell II) til IV). Dette kan tyde på at laservariablene har liten evne til selv å skille ut eiketrær. Forskjellen fra modell I) til modell II) er derimot stor, og plottet over variabelviktighet viser at spesielt temperaturer og kystavstanden er viktig i klassifiseringen.

Ut ifra OA kan det se ut som modellen gjør en relativt god jobb med klassifiseringen. Likevel er ikke modellen bedre enn det svakeste ledd. I dette tilfellet er dette UA for eik. Selv om det meste av klassen *Annet løv* er riktig klassifisert, er det ingen verdi i dette dersom eik også antas å høre til samme klasse. Et gjennomgående problem med løvtreklassifisering er skjevfordelingen av antall prøveflater per treslag. Den beste løsningen på denne utfordringen ville vært å ha flere observasjoner av eik og eventuelt andre løvtreslag med færre dominante prøveflater. Dette er likevel noe som vil kreve feltarbeid utover det som allerede gjøres i Landsskogstakseringen, og som ikke nødvendigvis er mulig i så stor grad dersom treslagsbestanden er liten totalt sett. Likevel vil utjevningen for reduksjon i observasjonsantall av dunbjørk hjelpe på troverdigheten i nøyaktighetene for eikedeteksjon.

Resultatene fra utjevninger er for eikedeteksjon en del dårligere enn den som er gjort med et tilfeldig utvalg observasjoner. Med så høyt standardavvik som vist i Tabell 21, er det flere iterasjoner som gir veldig forskjellige nøyaktighetsresultater. Dette kan tyde på at eik ikke er en spesielt forskjellig fra alle løvtreslag, men heller forskjellig fra de observasjonene som tilfeldigvis er plukket ut i den beste iterasjonen.

5.4.3 Gråor

De fleste av gråorobservasjonene klassifiseres i alle modeller som *Annet løv*. For gråordeteksjon er det stor forskjell i antall observasjoner, da det ikke er noen geografisk avgrensning som begrenser antallet observasjoner i modellen. I kombinasjon med stor variasjon mellom treslagene, vil det være vanskelig å skille ut gråor.

I alle modellene er NDVI-variablene viktigst, og dette skiller seg tydelig fra de andre klassifiseringene. De klimatiske variablene har ikke like stor innvirkning når gråordominante prøveflater finnes over hele landet. Spredningen av gråorprøveflater gjør at det ikke er hensiktsmessig å avgrense modellområdet. Dette medfører en økt skjevfordeling blant observasjonsantallet for de to klassene. Det er vanskelig å vite hvor godt RF skiller mellom gråor og andre løvtrær når variasjonen i *annet løv*-klassen ikke er garantert bevart ved reduksjon av det totale antallet av løvtreobservasjoner. Antallsfordelingen gjør at nesten all gråor klassifiseres som *annet løv*, og det er vanskelig å se for seg at gråor vil kunne skilles ut uten at den resterende løvtreslagsklassen deles inn i flere kategorier i klassifiseringen.

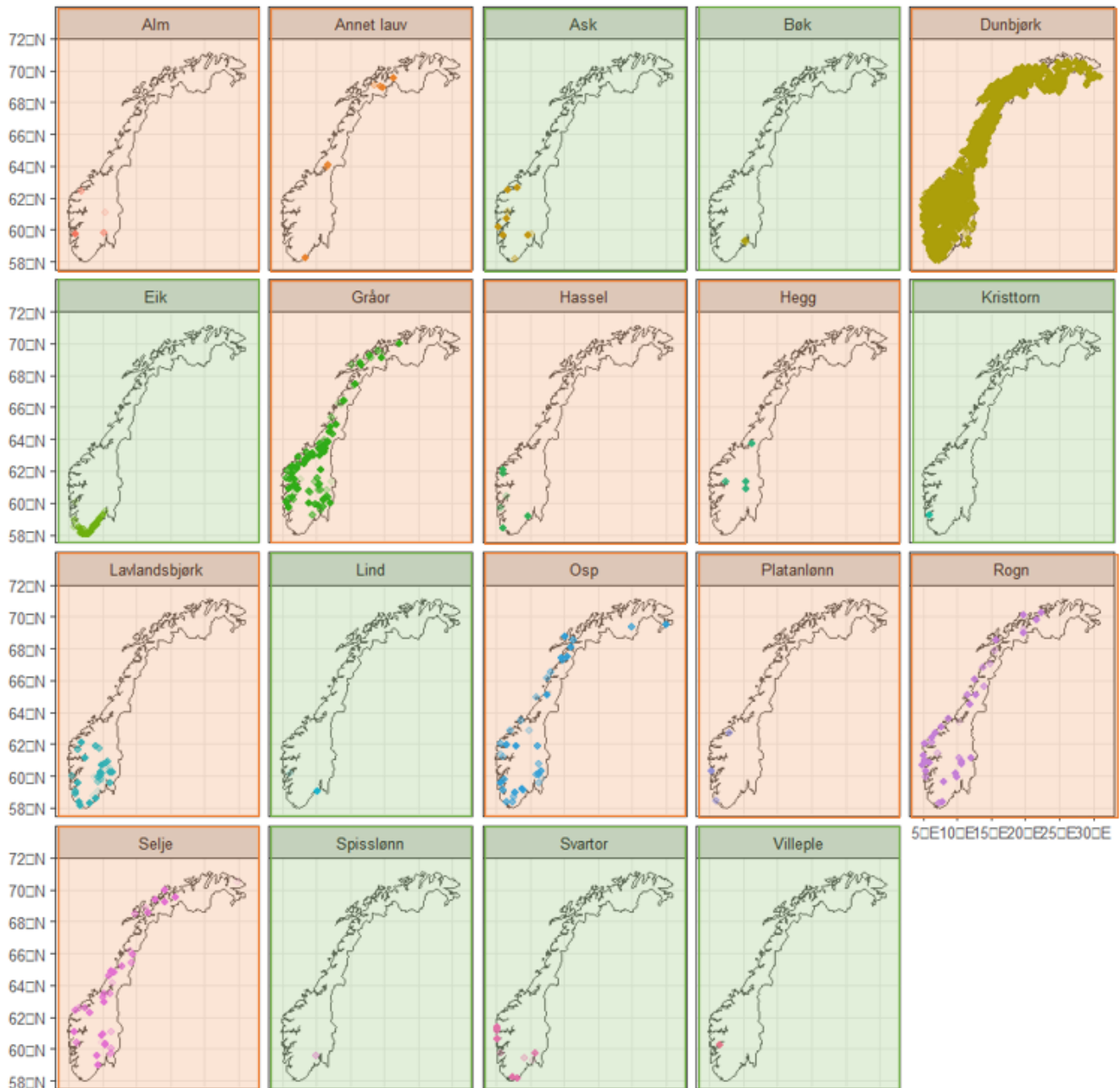
For de utjevnedte observasjonene med gjennomsnitt av iterasjonene er standardavviket spesielt lavt for *annet løv*. Dette er fordi det er veldig få av observasjonene herfra som klassifiseres feil i alle iterasjoner, da både antall observasjoner og variasjon i variabler innad i klassen er mye høyere. UA og tilhørende standardavvik for gråor påvirkes igjen negativt av dette.

5.4.4 Klassifisering med fordelte løvtrær på 2 klasser

Bakgrunnen for ønsket om å gjøre en slik inndeling av løvtreslagene i en ny modell, bygger på behovet for å se nærmere på variasjonene i restklassen for løvtrærne i eik- og gråorklassifiseringene. Ved å prøve å isolere ett treslag fra løvtrærne, vil den totale variasjonen for restklassen og antallsskjevheten i observasjonene forstyrre en mulig klassifisering av treslag. Ved å se om det går an å dele inn i underkategorier av løvtreslag, vil det kunne være andre muligheter for videre å skille enkelttreslag ut fra underkategoriene.

I fordelingen av løvtreslagene på to klasser var feilklassifiseringene i forvirringsmatrise i Tabell 9 avgjørende. De to nye treslagskategoriene går med en viss nøyaktighet an å skille fra hverandre. Likevel er det grunner til dette som gjør det vanskelig å stole på resultatene. Nesten utelukkende klimatiske variabler utgjør klassifiseringsgrunnlaget. Dette er også tydelig i vedlegg 2, der forskjellene mellom eik og gråor er betydelig enklere å skille for modell II) og IV) enn for modell I) og III).

For modell I) ble 96 % av alle observasjoner i klasse 2 feilklassifisert som klasse 1. Dette betyr at S2-tidsseriene ikke har stor betydning når de to klassene skal skilles fra hverandre. Laservariablene økte graden av riktig klassifiseringer og andel feilklassifiseringer var her 77 %. Det er likevel ingen håp for klassifiseringen før de klimatiske variablene tilføres. Dette kan bety at inndelingen av klasse 1 og 2 har sammenheng med plasseringen av prøveflatene i stor grad. I Figur 29 synes en tydelig trend, der treslagene i klasse 2 (markert i grønt) befinner seg langs kysten i sørdelen av landet, og det er større geografisk spredning i prøveflatene blant treslagene i klasse 1 (markert i rødt).



Figur 29 Fordeling av prøveflater blant løvtreslag. Klasse 1 markert i rødt og klasse 2 markert i grønt.

Det at forskjellene på de to klassene i stor grad er basert på plasseringen av prøveflatene, gjør at det ikke er sikkert at modellen er i stand til å se forskjeller på treslagene som befinner seg på samme sted. Heller er modellen trent til å bruke klimatiske variabler til å forstå hvilket område observasjonene befinner seg i. Dermed er det begrenset hvor mye informasjon som kan hentes fra denne klassifiseringen. Det vil dermed være vanskelig å benytte denne inndelingen til videre isolasjon av andre treslag innenfor klasse 1 og 2.

5.5 Forslag til videre arbeid

Innen klassifisering av treslag er det mange muligheter til videre forskning. Maskinlæringsmetoden benyttet i denne oppgaven er bare én av mange. I tillegg er det en voksende mulighet til å benytte kunstig intelligens i forbedringen av modeller og for å se andre sammenhenger og muligheter i klassifiseringene.

Også ved tilpasning av RF-modellen med endring av parametere og variabelutvelgelse, vil resultatene for klassifiseringene i denne oppgaven kunne forbedres. Det er også utallige muligheter blant andre treslag og ulike grupperinger blant dem, som observert i undersøkelsene gjort i vedlegg 2.

I tillegg kan det finnes potensiale i satellittbåndene fra Sentinel-2 som ikke er representert i datasettet i oppgaven. Ved i tillegg å benytte oppløsning på 10 meter for B02, B03, B04 og B08 vil muligens også reflekterte verdier fra prøveflatene være mer representative for feltobservert treslagsfordeling.

6 Konklusjon

Forskjeller på granarter er mulig å se med multispektrale satellittbilder fra Sentinel-2 i kombinasjon med klimatiske variabler i ganske stor grad med maskinlæringsverktøyet random forest. I den beste RF-klassifiseringene gjort med modell IV) var det mulig å skille ut introduserte granarter med en OA på 82%, og UA for introdusert gran på 81%. Resultatene later til å være overførbare til usette granområder i stor grad, men er ikke testet for usette introduserte granområder i denne oppgaven. Klimatiske variabler er avgjørende i denne klassifiseringen.

Eikedeteksjon med modellene i denne oppgaven er mindre lovende. Selv for den beste modellen med alle tidsserievariabler fra S2-bilder, laserdata og klimatiske variabler ble ikke UA høyere enn 36%, selv med en OA 81% for modell IV) etter utvalgsutjevning. De viktigste variablene for eikedeteksjon var i stor grad temperaturer for hver måned av året. Skillet mellom eik og andre løvtreslag med RF-klassifisering kan med en løsning på skjevfordelingen av observasjonsantall mellom klasser muligens gi bedre resultater.

Gråordeteksjon blant andre løvtreslag er vanskelig i Norge da variasjonen blant andre løvtrær er stor og vanskelig å beholde dersom en utjevning av observasjonsantall skal gjøres. Dette gjør at isolasjon av gråor basert på satellittbilder, laserdata og klimatiske variabler er vanskelig når ikke random forest har flere enn to klasser å fordele treslagene på. Modell IV) gav best resultater med en kombinasjon av alle variabler, og NDVI var det viktigste i klassifiseringen. Modellen gav likevel ikke mer enn 40% riktig klassifisering av gråorobservasjoner med utjevnete resultater for utvalg av dunbjørkflater.

Ved å skille løvtreslagene i to klasser finnes ingen overbevisende forskjell utover klimatiske variabler som beskriver geografisk plassering, for noen av treslagsklassifiseringer gjort i denne oppgaven. Med manglende kvantitativt aspekt for mange av treslagene, er også en ytterligere splittelse i snevrere treslagskategorier vanskeligere.

Generelt sett er deteksjon av enkelttreslag vanskelig uten en inndeling i undergrupper for treslag, dette gjelder også blant løvtreslag. Med ytterligere tilpasning av RF-parametere eller en form for dyplæring har imidlertid alle modellene potensiale for å bedres, der granklassifisering med mindre variasjon blant observasjoner har best utgangspunkt for en god klassifisering utover oppgavens presentert feltobservasjoner.

7 Referanser

- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.*, *114*, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Berrar, D. (2018). Cross-validation. In (pp. 8). 2-12-1-S3-70 Ookayama, Meguro-ku, Tokyo 152-8550, Japan: Data Science Laboratory, Tokyo Institute of Technology.
- Bhatia, N. (2021). What is Out of Bag (OOB) score in Random Forest? - Towards Data Science. *Medium*. <https://towardsdatascience.com/what-is-out-of-bag-oob-score-in-random-forest-a7fa23d710>
- Breidenbach, J., Granhus, A., Hysten, G., Eriksen, R., & Astrup, R. (2020). A century of National Forest Inventory in Norway – informing past, present, and future decisions. *For. Ecosyst.*, *7*(1), 1–19. <https://doi.org/10.1186/s40663-020-00261-0>
- Breidenbach, J., Waser, L. T., Debella-Gilo, M., Schumacher, J., Rahlf, J., Hauglin, M., Puliti, S., & Astrup, R. (2020). National mapping and estimation of forest area by dominant tree species using Sentinel-2 data. *Can. J. For. Res.* <https://doi.org/https://doi.org/10.1139/cjfr-2020-0170>
- Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. <https://doi.org/10.1023/a:1010933404324>
- Cánovas-García, F., Alonso-Sarría, F., Gomariz-Castillo, F., & Oñate-Valdivieso, F. (2017). Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. *Comput. Geosci.*, *103*, 1–11. <https://doi.org/10.1016/j.cageo.2017.02.012>
- Comber, A., Fisher, P., Brunsdon, C., & Khmag, A. (2012). Spatial analysis of remote sensing image classification accuracy. *Remote Sens. Environ.*, *127*, 237–246. <https://doi.org/10.1016/j.rse.2012.09.005>
- ESA. (2023, 2023/04/20/). *Radiometric - Resolutions - Sentinel-2 MSI - User Guides - Sentinel Online - Sentinel Online*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/Sentinel-2-msi/resolutions/radiometric>
- ESA. (2023, 2023/04/14/). *SAR Instrument - Sentinel-1 SAR Technical Guide - Sentinel Online - Sentinel Online*. <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-1-sar/sar-instrument>
- ESA. (2023). *Sentinel-2 - Satellite Description - Sentinel Online - Sentinel Online*. In.
- ESA. (2023, 2023/04/13/). *Spatial - Resolutions - Sentinel-2 MSI - User Guides - Sentinel Online - Sentinel Online*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial>
- Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.*, *186*, 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>
- Frøystad, B. (1961). Leplanting på fastmark og på myr i verharde strøk. In.
- Grabska, E., Hostert, P., Pflugmacher, D., & Ostapowicz, K. (2019). Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sens.*, *11*(10), 1197. <https://doi.org/10.3390/rs11101197>
- Grindeland, J. M. (2022). bjørk. In: Store norske leksikon.
- Hauglin, M. Skript for uthenting av S2-tidsserier. In.
- Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.*, *32*(1), 1–6. <https://doi.org/10.1007/s11676-020-01155-1>
- Janitzka, S., Celik, E., & Boulesteix, A.-L. (2018). A computationally fast variable importance test for random forests for high-dimensional data. *Adv. Data Anal. Classif.*, *12*(4), 885–915. <https://doi.org/10.1007/s11634-016-0276-4>
- Kartverket. (2022). Høydedata og dybdedata. In.
- Lisein, J., Michez, A., Claessens, H., & Lejeune, P. (2015). Discrimination of Deciduous Tree Species from Time Series of Unmanned Aerial System Imagery. *PLoS One*, *10*(11), e0141006. <https://doi.org/10.1371/journal.pone.0141006>
- Ljumović, M., & Klar, M. (2015). Estimating expected error rates of random forest classifiers: A comparison of cross-validation and bootstrap. In *2015 4th Mediterranean Conference on Embedded Computing (MECO)* (pp. 212–215). IEEE. <https://doi.org/10.1109/meco.2015.7181905>
- Matthew W., M. (2011). Bias of the Random Forest Out-of-Bag (OOB) Error for Certain Input Parameters. *Open Journal of Statistics*, *2011*. <https://doi.org/10.4236/ojs.2011.13024>
- Mueller-Dombois, D. (1964). Effect of Depth to Water Table on Height Growth of Tree Seedlings in a Greenhouse. *For. Sci.*, *10*(3), 306–316. <https://doi.org/10.1093/forestscience/10.3.306>
- Nordbø, R. (1954). Landbruksdepartementet st. meld. nr. 44 (1954) om innstilling ifra skogkommissjonen av 1951 om skogreising vestafjells. In: Stortinget.
- Rittger, K., Raleigh, M. S., Dozier, J., Hill, A. F., Lutz, J. A., & Painter, T. H. (2020). Canopy Adjustment and Improved Cloud Detection for Remotely Sensed Snow Cover Mapping. *Water Resour. Res.*, *56*(6), e2019WR024914. <https://doi.org/10.1029/2019wr024914>
- Story, M., & Congalton, R. G. (2008). Accuracy Assessment: A User's Perspective. In (pp. 3): American Society for Photogrammetry and Remote Sensing.
- Sunding, P., Foslie, M., Aune, E. I., & Mæhlum, L. (2022). edelgran – Store norske leksikon. In.
- Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *10*(6), 363–377. <https://doi.org/10.1002/sam.11348>
- van Aardt, J. A. N., & Wynne, R. H. (2007). Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. In: Taylor & Francis.

Viken, K. O. (2021). *Landsskogtakseringens feltinstruks – 2021*. NIBIO. <https://nibio.brage.unit.no/nibio-xmliui/handle/11250/2826859>

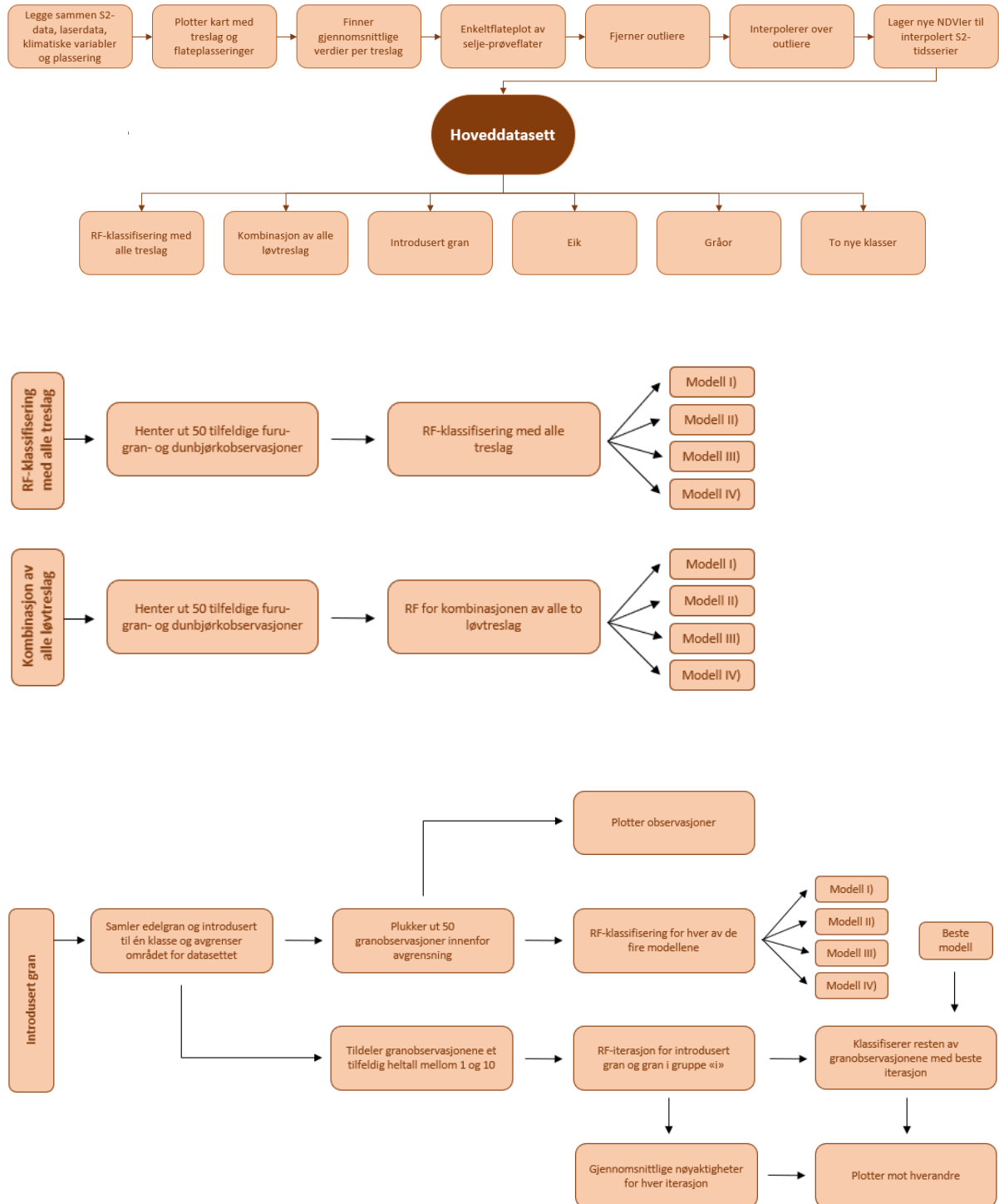
Zhang, F., & Yang, X. (2020). Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sens. Environ.*, 251, 112105. <https://doi.org/10.1016/j.rse.2020.112105>

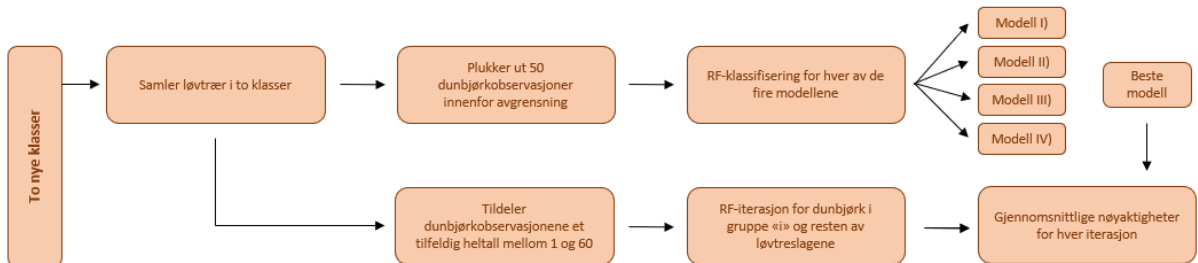
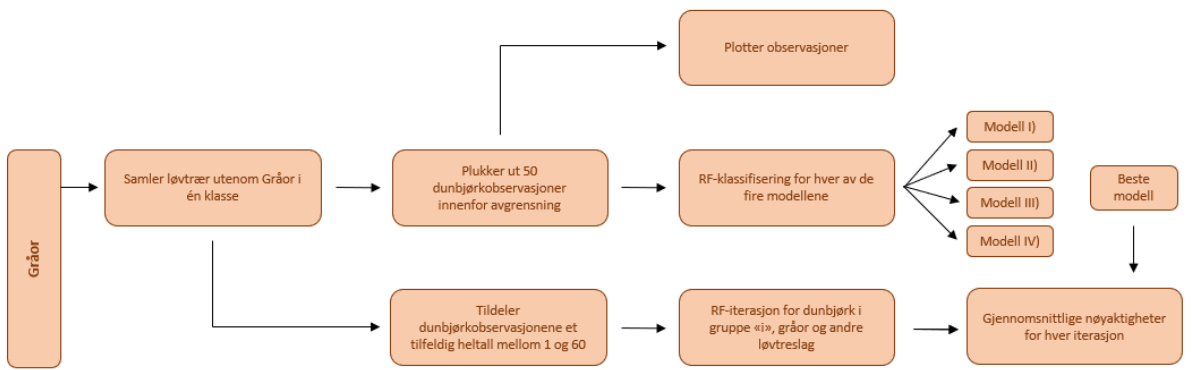
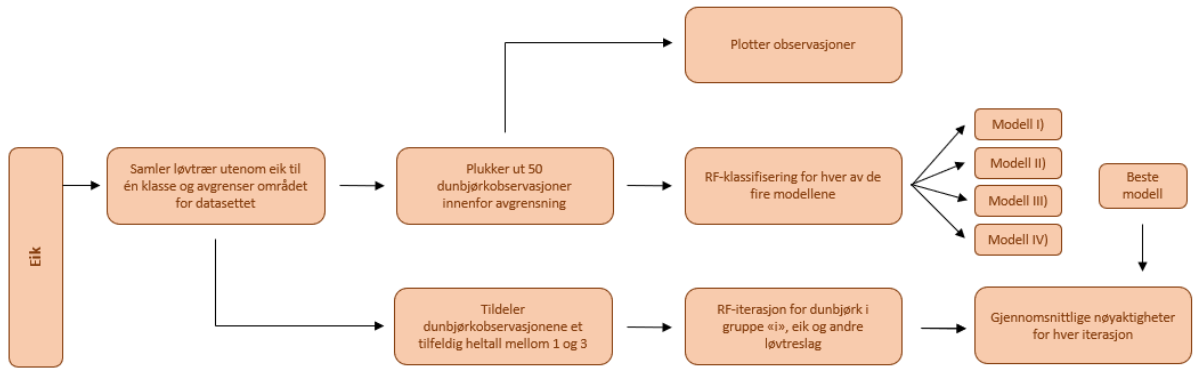
Vedlegg 1

R-skriptet brukt i oppgaven ligger i Github-lenken under:

<https://github.com/hjonland/Masterskript.git>

Flytdiagram for koden





Vedlegg 2

RF klassifiseringsfeil for hver løvtreslag mot hverandre

Alle kombinasjoner av to løvtreslag. NA-verdier i feil-kolonnene er filtrert bort.

30 – Dunbjørk	50 – Gråor
31 – Lavlandsbjørk	51 – Svartor
32 – Osp	52 – Selje
40 – Eik	53 – Rogn
41 – Bøk	54 – Hegg
42 – Ask	55 – Hassel
43 – Alm	57 – Villeple
44 – Lind	59 – Annet lauv
49 – Platanlønn	70 – Kristtorn

	Modell I)		Modell II)		Modell III)		Modell IV)			
	Treslag1	Treslag2	Feil treslag1	Feil treslag2	Feil treslag1	Feil treslag2	Feil treslag1	Feil treslag2		
1	31	40	0.39130435	0.11538462	0.21739130	0.03846154	0.56521739	0.15384615	0.26086957	0.00000000
2	31	32	0.60869565	0.25925926	0.56521739	0.33333333	0.56521739	0.18518519	0.56521739	0.40740741
3	31	42	0.04347826	1.00000000	0.04347826	1.00000000	0.04347826	1.00000000	0.08695652	1.00000000
4	31	50	0.82608696	0.04838710	0.73913043	0.04838710	0.86956522	0.03225806	0.86956522	0.03225806
5	31	53	0.65217391	0.30769231	0.34782609	0.26923077	0.43478261	0.30769231	0.43478261	0.15384615
6	31	43	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
7	31	44	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
8	31	52	0.21739130	0.43750000	0.17391304	0.31250000	0.21739130	0.37500000	0.21739130	0.31250000
9	31	51	0.04347826	0.85714286	0.08695652	0.85714286	0.04347826	0.85714286	0.08695652	0.85714286
10	31	55	0.04347826	1.00000000	0.08695652	1.00000000	0.04347826	1.00000000	0.08695652	1.00000000
11	31	54	0.04347826	1.00000000	0.04347826	1.00000000	0.04347826	1.00000000	0.04347826	1.00000000
12	31	41	0.04347826	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
13	30	31	0.06382979	0.52173913	0.08510638	0.21739130	0.06382979	0.60869565	0.10638298	0.26086957
14	31	40	0.47826087	0.11538462	0.21739130	0.03846154	0.56521739	0.23076923	0.26086957	0.03846154
15	32	40	0.44444444	0.34615385	0.11111111	0.03846154	0.48148148	0.34615385	0.07407407	0.07692308
16	40	42	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	0.80000000	0.00000000	1.00000000
17	40	50	0.61538462	0.04838710	0.07692308	0.03225806	0.46153846	0.01612903	0.07692308	0.04838710
18	40	53	0.23076923	0.38461538	0.03846154	0.19230769	0.26923077	0.23076923	0.03846154	0.07692308
19	40	43	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
20	40	44	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
21	40	52	0.07692308	0.56250000	0.00000000	0.06250000	0.07692308	0.43750000	0.00000000	0.06250000
22	40	51	0.00000000	0.85714286	0.00000000	0.85714286	0.00000000	0.85714286	0.00000000	0.85714286
23	40	55	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
24	40	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
25	40	41	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
26	30	40	0.06382979	0.38461538	0.04255319	0.03846154	0.06382979	0.42307692	0.04255319	0.03846154
27	31	32	0.65217391	0.29629630	0.43478261	0.33333333	0.60869565	0.25925926	0.43478261	0.29629630
28	32	40	0.37037037	0.38461538	0.07407407	0.07692308	0.48148148	0.34615385	0.07407407	0.03846154
29	32	42	0.03703704	0.60000000	0.03703704	0.80000000	0.03703704	0.60000000	0.03703704	0.80000000
30	32	50	0.62962963	0.11290323	0.66666667	0.04838710	0.66666667	0.06451613	0.59259259	0.08064516
31	32	53	0.22222222	0.30769231	0.22222222	0.34615385	0.22222222	0.26923077	0.22222222	0.26923077
32	32	43	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
33	32	44	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
34	32	52	0.07407407	0.31250000	0.07407407	0.31250000	0.07407407	0.31250000	0.11111111	0.37500000
35	32	51	0.03703704	1.00000000	0.03703704	1.00000000	0.03703704	0.85714286	0.03703704	1.00000000
36	32	55	0.03703704	1.00000000	0.00000000	1.00000000	0.03703704	1.00000000	0.00000000	1.00000000
37	32	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
38	32	41	0.03703704	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
39	30	32	0.10638298	0.48148148	0.06382979	0.37037037	0.06382979	0.37037037	0.06382979	0.33333333
40	31	42	0.04347826	1.00000000	0.04347826	1.00000000	0.08695652	1.00000000	0.08695652	1.00000000
41	40	42	0.00000000	0.80000000	0.00000000	0.80000000	0.00000000	0.60000000	0.00000000	0.80000000
42	32	42	0.00000000	0.80000000	0.03703704	0.80000000	0.03703704	0.80000000	0.03703704	0.80000000
43	42	50	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.01612903	1.00000000	0.00000000
44	42	53	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
45	42	43	0.00000000	1.00000000	0.20000000	1.00000000	0.00000000	1.00000000	0.40000000	1.00000000
46	42	44	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
47	42	52	1.00000000	0.06250000	0.80000000	0.06250000	1.00000000	0.06250000	1.00000000	0.06250000
48	42	51	0.80000000	0.14285714	0.80000000	0.14285714	0.80000000	0.28571429	1.00000000	0.14285714
49	42	55	0.40000000	0.50000000	0.60000000	0.50000000	0.60000000	0.50000000	0.80000000	0.50000000
50	42	54	0.20000000	1.00000000	0.40000000	1.00000000	0.40000000	1.00000000	0.40000000	1.00000000
51	41	42	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	0.50000000	0.00000000
52	30	42	0.00000000	1.00000000	0.00000000	0.60000000	0.00000000	1.00000000	0.00000000	0.60000000

53	31	50	0.78260870	0.08064516	0.82608696	0.04838710	0.82608696	0.06451613	0.73913043	0.03225806
54	40	50	0.61538462	0.06451613	0.07692308	0.04838710	0.57692308	0.06451613	0.07692308	0.04838710
55	32	50	0.66666667	0.09677419	0.59259259	0.03225806	0.62962963	0.06451613	0.59259259	0.06451613
56	42	50	1.00000000	0.01612903	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
57	50	53	0.09677419	0.65384615	0.09677419	0.69230769	0.09677419	0.57692308	0.08064516	0.57692308
58	43	50	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
59	44	50	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
60	50	52	0.03225806	1.00000000	0.01612903	1.00000000	0.03225806	1.00000000	0.01612903	0.93750000
61	50	51	0.00000000	1.00000000	0.00000000	0.85714286	0.00000000	1.00000000	0.00000000	1.00000000
62	50	55	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
63	50	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
64	41	50	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
65	30	50	0.31914894	0.16129032	0.27659574	0.09677419	0.27659574	0.14516129	0.27659574	0.11290323
66	31	53	0.52173913	0.34615385	0.34782609	0.19230769	0.34782609	0.30769231	0.39130435	0.15384615
67	40	53	0.19230769	0.38461538	0.03846154	0.15384615	0.26923077	0.15384615	0.03846154	0.11538462
68	32	53	0.25925926	0.34615385	0.25925926	0.34615385	0.22222222	0.26923077	0.22222222	0.26923077
69	42	53	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
70	50	53	0.04838710	0.73076923	0.11290323	0.65384615	0.11290323	0.57692308	0.08064516	0.61538462
71	43	53	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
72	44	53	1.00000000	0.03846154	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
73	52	53	0.87500000	0.11538462	0.81250000	0.15384615	0.68750000	0.11538462	0.75000000	0.07692308
74	51	53	0.85714286	0.03846154	0.85714286	0.03846154	0.57142857	0.03846154	0.57142857	0.07692308
75	53	55	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
76	53	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
77	41	53	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
78	30	53	0.14893617	0.69230769	0.19148936	0.57692308	0.12765957	0.57692308	0.19148936	0.61538462
79	31	43	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
80	40	43	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
81	32	43	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
82	42	43	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
83	43	50	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
84	43	53	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
85	43	44	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000
86	43	52	1.00000000	0.06250000	1.00000000	0.00000000	1.00000000	0.06250000	1.00000000	0.06250000
87	43	51	1.00000000	0.14285714	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
88	43	55	1.00000000	0.25000000	1.00000000	0.25000000	1.00000000	0.50000000	1.00000000	0.25000000
89	43	54	0.50000000	0.66666667	0.00000000	0.66666667	1.00000000	1.00000000	0.50000000	0.33333333
90	41	43	0.50000000	0.00000000	0.50000000	0.00000000	0.50000000	0.00000000	0.50000000	0.00000000
91	30	43	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
92	31	44	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
93	40	44	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
94	32	44	0.03703704	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
95	42	44	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
96	44	50	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
97	44	53	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
98	43	44	1.00000000	1.00000000	0.50000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000
99	44	52	1.00000000	0.06250000	1.00000000	0.00000000	1.00000000	0.06250000	1.00000000	0.00000000
100	44	51	1.00000000	0.14285714	1.00000000	0.14285714	1.00000000	0.00000000	1.00000000	0.00000000
101	44	55	1.00000000	0.25000000	1.00000000	0.25000000	1.00000000	0.25000000	1.00000000	0.25000000
102	44	54	1.00000000	0.33333333	1.00000000	0.00000000	1.00000000	0.33333333	1.00000000	0.66666667
103	41	44	0.50000000	1.00000000	0.50000000	1.00000000	0.50000000	1.00000000	0.50000000	1.00000000
104	30	44	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
105	31	52	0.21739130	0.37500000	0.17391304	0.37500000	0.21739130	0.31250000	0.21739130	0.31250000
106	40	52	0.11538462	0.43750000	0.00000000	0.12500000	0.07692308	0.62500000	0.00000000	0.06250000
107	32	52	0.07407407	0.31250000	0.03703704	0.37500000	0.03703704	0.37500000	0.07407407	0.37500000
108	42	52	1.00000000	0.06250000	1.00000000	0.06250000	1.00000000	0.06250000	1.00000000	0.06250000
109	50	52	0.03225806	1.00000000	0.01612903	1.00000000	0.03225806	1.00000000	0.00000000	0.93750000
110	52	53	1.00000000	0.19230769	0.81250000	0.15384615	0.75000000	0.07692308	0.68750000	0.15384615
111	43	52	1.00000000	0.06250000	1.00000000	0.06250000	1.00000000	0.06250000	1.00000000	0.06250000
112	44	52	1.00000000	0.12500000	1.00000000	0.00000000	1.00000000	0.06250000	1.00000000	0.00000000
113	51	52	0.71428571	0.18750000	0.71428571	0.00000000	0.71428571	0.12500000	0.57142857	0.06250000
114	52	55	0.06250000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
115	52	54	0.06250000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
116	41	52	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
117	30	52	0.00000000	0.75000000	0.00000000	0.87500000	0.00000000	0.93750000	0.00000000	0.75000000
118	31	51	0.04347826	0.85714286	0.08695652	0.85714286	0.04347826	0.85714286	0.08695652	0.85714286
119	40	51	0.00000000	0.85714286	0.00000000	0.85714286	0.00000000	0.85714286	0.00000000	0.85714286
120	32	51	0.03703704	1.00000000	0.03703704	1.00000000	0.00000000	0.03703704	1.00000000	0.00000000
121	42	51	0.80000000	0.28571429	0.80000000	0.14285714	1.00000000	0.14285714	1.00000000	0.14285714
122	50	51	0.00000000	1.00000000	0.01612903	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
123	51	53	0.85714286	0.07692308	0.85714286	0.03846154	0.57142857	0.00000000	0.71428571	0.00000000
124	43	51	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
125	44	51	1.00000000	0.14285714	1.00000000	0.14285714	1.00000000	0.14285714	1.00000000	0.14285714
126	51	52	0.71428571	0.18750000	0.57142857	0.06250000	0.71428571	0.18750000	0.57142857	0.06250000
127	51	55	0.28571429	1.00000000	0.28571429	1.00000000	0.28571429	1.00000000	0.28571429	1.00000000
128	51	54	0.14285714	1.00000000	0.14285714	1.00000000	0.28571429	1.00000000	0.14285714	1.00000000
129	41	51	1.00000000	0.14285714	1.00000000	0.14285714	1.00000000	0.14285714	1.00000000	0.14285714

130	30	51	0.00000000	0.71428571	0.02127660	0.57142857	0.00000000	0.71428571	0.02127660	0.57142857
131	31	55	0.04347826	1.00000000	0.08695652	1.00000000	0.04347826	1.00000000	0.08695652	1.00000000
132	40	55	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
133	32	55	0.03703704	1.00000000	0.00000000	1.00000000	0.03703704	1.00000000	0.00000000	1.00000000
134	42	55	0.40000000	0.50000000	0.80000000	0.50000000	0.60000000	0.50000000	0.40000000	0.50000000
135	50	55	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
136	53	55	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
137	43	55	1.00000000	0.25000000	1.00000000	0.00000000	1.00000000	0.25000000	1.00000000	0.00000000
138	44	55	1.00000000	0.25000000	1.00000000	0.25000000	1.00000000	0.25000000	1.00000000	0.25000000
139	52	55	0.06250000	1.00000000	0.06250000	1.00000000	0.06250000	1.00000000	0.00000000	1.00000000
140	51	55	0.28571429	1.00000000	0.28571429	1.00000000	0.28571429	1.00000000	0.28571429	1.00000000
141	54	55	1.00000000	0.50000000	1.00000000	0.25000000	1.00000000	0.50000000	1.00000000	0.00000000
142	41	55	1.00000000	0.25000000	1.00000000	0.25000000	1.00000000	0.00000000	1.00000000	0.25000000
143	30	55	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
144	31	54	0.04347826	1.00000000	0.04347826	1.00000000	0.04347826	1.00000000	0.04347826	1.00000000
145	40	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
146	32	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
147	42	54	0.40000000	1.00000000	0.40000000	1.00000000	0.00000000	1.00000000	0.20000000	1.00000000
148	50	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
149	53	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
150	43	54	1.00000000	1.00000000	0.50000000	0.66666667	1.00000000	1.00000000	0.50000000	0.66666667
151	44	54	1.00000000	0.33333333	1.00000000	0.33333333	1.00000000	0.66666667	1.00000000	0.00000000
152	52	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
153	51	54	0.28571429	1.00000000	0.14285714	1.00000000	0.28571429	1.00000000	0.14285714	1.00000000
154	54	55	1.00000000	0.50000000	1.00000000	0.00000000	1.00000000	0.50000000	1.00000000	0.00000000
155	41	54	1.00000000	0.66666667	1.00000000	0.66666667	1.00000000	0.66666667	1.00000000	0.66666667
156	30	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
157	31	41	0.04347826	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
158	40	41	0.03846154	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
159	32	41	0.00000000	1.00000000	0.00000000	1.00000000	0.03703704	1.00000000	0.00000000	1.00000000
160	41	42	1.00000000	0.00000000	1.00000000	0.20000000	1.00000000	0.00000000	1.00000000	0.00000000
161	41	50	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
162	41	53	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
163	41	43	0.50000000	0.00000000	0.50000000	0.00000000	0.50000000	0.00000000	0.50000000	0.00000000
164	41	44	0.50000000	1.00000000	0.50000000	1.00000000	0.50000000	1.00000000	0.00000000	1.00000000
165	41	52	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000
166	41	51	1.00000000	0.14285714	1.00000000	0.14285714	1.00000000	0.14285714	1.00000000	0.14285714
167	41	55	1.00000000	0.25000000	1.00000000	0.25000000	1.00000000	0.00000000	1.00000000	0.25000000
168	41	54	1.00000000	0.66666667	1.00000000	0.66666667	1.00000000	0.33333333	1.00000000	0.33333333
169	30	41	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
170	30	31	0.10638298	0.52173913	0.08510638	0.17391304	0.04255319	0.52173913	0.10638298	0.26086957
171	30	40	0.06382979	0.42307692	0.04255319	0.03846154	0.08510638	0.42307692	0.04255319	0.03846154
172	30	32	0.08510638	0.44444444	0.08510638	0.40740741	0.08510638	0.33333333	0.08510638	0.33333333
173	30	42	0.00000000	1.00000000	0.00000000	0.60000000	0.00000000	1.00000000	0.00000000	0.60000000
174	30	50	0.25531915	0.14516129	0.29787234	0.11290323	0.29787234	0.14516129	0.27659574	0.12903226
175	30	53	0.14893617	0.69230769	0.21276596	0.69230769	0.10638298	0.69230769	0.21276596	0.61538462
176	30	43	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
177	30	44	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
178	30	52	0.00000000	0.87500000	0.00000000	0.93750000	0.02127660	0.87500000	0.00000000	0.87500000
179	30	51	0.00000000	0.85714286	0.02127660	0.57142857	0.00000000	0.71428571	0.02127660	0.71428571
180	30	55	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
181	30	54	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000
182	30	41	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000	0.00000000	1.00000000



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

Postboks 5003
NO-1432 Ås
Norway