



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2022, 30 stp
BIOVIT

Effekten av flyhøyde på avlingsprediksjon i hvete ved bruk av UAV-bilder

The effects of flight altitude on yield prediction in
wheat using UAV imagery

Fredrik Klaseie
Plantevitenskap

Sammendrag

Effekten av flyhøyde på avlingsprediksjon i hvete ved bruk av UAV-bilder

Med klimakrise og krig i Europa står landbruket i sentrum av begivenhetene. Det krever effektivisering av matproduksjonen samtidig som det krever en grønn omstilling. En del av løsningen på dette kan være utnyttelse av bildeanalyser for avlingsestimering. Dette kan gjøre både bonden og forskeren mer effektiv og bærekraftig.

I denne oppgaven ble det fløyet en UAV med påmontert multispektralt kamera med 6 forskjellige sensorer. Sensorene registrerte refleksjonsbåndene blå, grønn, rød, rededge og nærinfrarød (NIR) i tillegg til RGB. Denne ble fløyet over et forsøksfelt med forskjellige sorter med vårhvete på 4 forskjellige flyhøyder; 20, 50, 70 og 100 meter. Flygingen ble gjentatt 3 forskjellige dager med 4 dagers mellomrom i siste del av kornfyllingen. Feltet var også gjødslet med 2 forskjellige gjødselmengder (7,5 kg N og 15kg N). I tillegg ble den fløyet over et skifte med vårhvete rundt aksskyting som ble stilt til disposisjon av en bonde.

Bildene ble sammenstilt til ortomosaikker i Pix4D. Data fra ortomosaikkene ble ekstrahert i QGIS, og basert på dataene ble det regnet ut 28 forskjellige vegetasjonsindekser som er brukt i forskjellig grad og for forskjellige formål, både i forskning og i landbruk. Avlingen i forsøksfelt ble registrert med forsøksresker, mens avlingen i anvendt landbruk ble registrert med avlingsregistrator montert i treskeren. Indeksene ble også brukt som input i maskinlæringsmodellen LASSO for avlingsprediksjon. Det ble gjort korrelasjonsanalyser for å finne korrelasjonen mellom indeksene fra hver flyhøyde og dato, og den registrerte avlingen fra både forsøksfelt og i anvendt landbruk. Det samme ble gjort med resultatene fra LASSO. Utslagene i NDVI (normalized difference vegetation index) ved forskjellig gjødselmengde ble også undersøkt.

Det ble registrert korrelasjonskoeffisienter helt fra 0,005 til 0,73 i dette forsøket. Indeksene varierte i prestasjon både mellom datoene for innsamling, og innad i datoene. Det var også indekser som var sensitive for endring i flyhøyde.

Abstract

The effects of flight altitude on yield prediction in wheat using UAV imagery

Climate change and war in Europe are greatly impacting modern agriculture. This demands a green change and making food production more efficient. An important part of the solution can be utilizing image analysis, which can give us more accurate yield estimates. This has the potential to make both farmers and scientists more efficient and sustainable.

For collecting necessary research data, we used an UAV equipped with a six-sensor, multispectral camera. The sensors are capable of recording and registering data for these spectral bands: blue, red, green, rededge and near-infrared (NIR). The drone was flown over a test field containing varieties of spring wheat, at 4 different altitudes: 20, 50, 70 and 100 meters. To collect sufficient data this was repeated on 3 separate days, with a 4-day interval between them, during the last period of the grain filling. The field had also been fertilized with 2 different amounts of fertilizer (7.5kg N and 15kg N).

Additional data was collected from another field, containing spring wheat at time of ear formation, put to our disposal by a willing farmer.

The images were converted to orthomosaics in the mapping software Pix4D. Data from the orthomosaics were extracted in QGIS and were used to calculate 28 different vegetation indices. The yield of the test field was registered with a test-dedicated combine harvester, while the yield for the applied agriculture was recorded by a yield registrar mounted inside the combine harvester itself. The indices were also used as input to the the Least Absolute Shrinkage and Selection Operator (LASSO) regression model for prediction. Correlation analysis was applied to find the applicable correlation between the indices for each recorded altitude and date, and between the registered yield for both the test field and applied agriculture. The same analysis was done for the data from LASSO, and the results from NDVI (Normal Difference Vegetation Index) were also examined.

Registered correlation coefficients for this examination ranged from 0,005 to a whole 0,73. The indices varied both between the collection dates and for data collected during the same day. Some indices were even sensitive to adjustment of altitudes of images recorded.

Forord

Jeg ønsker å takke min hovedveileder Sahameh Shafiee spesielt for veiledningen hun har gitt meg gjennom hele arbeidet med oppgaven.

Jeg vil også takke Tomasz Mroz og Morten Lillemo for god veiledning underveis.

Jeg vil takke Lars Gunnar Flatvad for hjelp med korrekturlesing, og Lars Vidar Lien for hjelp med oversettelser.

Til slutt må jeg takke Oliwia Matusiak for å holde ut med meg i den perioden oppgaven er skrevet.

Innhold

Sammendrag	1
Abstract	2
Forord	3
Innhold	4
Figurliste.....	6
Tabelliste	7
1 Innledning.....	8
2 Teori	10
2.1 Fenotyping	10
2.2 Spektral refleksjon	10
2.3 Vegetasjonsindekser	12
2.4 Maskinl�ring	12
2.5 Flyh�yde	12
3 Metode.....	13
3.1 Fors�ksfeltet p� �s, Vollebekk	13
3.1.1 Plantematerialet	13
3.1.2 Feltdesign	14
3.2 Anvendt landbruk	15
3.3 Datainnsamling med UAV	16
3.3.1 Sammenstilte bilder fra PhenoCrop	17
3.4 Dataanalyse.....	18
3.4.1 Indekser	18
3.4.2 NDVI.....	19
3.4.3 SCCCI	19
3.4.4 MTCI.....	19
3.5 Ekstrahering av data i QGIS	20
3.5.1 Fors�ksfelt.....	20
3.5.2 Anvendt landbruk	24
3.6 Maskinl�ring	27
4 Resultater.....	28
4.1 Avlingsdata.....	28
4.2 Korrelasjon mellom NDVI fra forskjellige flyh�yder	30
4.3 Korrelasjon mellom forskjellige flyh�yder og avling	32

4.4	Indekser	33
4.5	Maskinl�ring	36
4.5.1	Prediksjonene	36
4.5.2	Indekser brukt i LASSO	40
4.6	Anvendt landbruk	43
5	Diskusjon.....	46
5.1	Indekser	46
5.1.1	NDVI.....	47
5.1.2	NDVI med forskjellige gj�dselmengder	47
5.1.3	SCCCI	48
5.1.4	MTCI.....	49
5.2	LASSO.....	49
5.2.1	Datagrunnlag	51
5.3	Forslag til videre arbeid	52
5.3.1	Optimalisering av N-indeks	52
5.3.2	Sammenligning av maskinl�ringsmodeller og -innstillinger.....	52
6	Konklusjon	53
7	Referanser.....	54
8	Vedlegg	56
8.1	Kode for LASSO	56

Figurliste

<i>Figur 1</i> Illustrasjon av det elektromagnetiske spektrum [11].....	10
<i>Figur 2</i> Illustrasjon av hvordan planter reflekterer og absorberer lys [12].....	11
<i>Figur 3</i> Feltkart og ortomosaikk over Robotfeltet.	14
<i>Figur 4</i> Ortomosaikk av skiftet.	15
<i>Figur 5</i> Jordsmonnet på skiftet varierer fra siltig mellomsand og siltig finsand» (punkt 1) til grusholdig siltig sand, sandig silt og silt (punkt 2).....	16
<i>Figur 6</i> Rekkefølge for å laste inn filer til QGIS.	20
<i>Figur 7</i> Skjermbilde QGIS.	21
<i>Figur 8</i> Skjermbilde QGIS.	21
<i>Figur 9</i> Skjermbilde QGIS.	21
<i>Figur 10</i> Skjermbilde QGIS.	22
<i>Figur 11</i> Skjermbilde QGIS.	22
<i>Figur 12</i> Skjermbilde QGIS.	23
<i>Figur 13</i> Skjermbilde QGIS.	23
<i>Figur 14</i> Skjermbilde QGIS.	24
<i>Figur 15</i> Skjermbilde QGIS.	25
<i>Figur 16</i> Skjermbilde QGIS.	26
<i>Figur 17</i> Skjermbilde QGIS.	26
<i>Figur 18</i> Fordeling av avlingsmengde i forsøksfelt.	28
<i>Figur 19</i> Fordeling av avlingsmengde i anvendt landbruk etter rensing av data.	29
<i>Figur 20</i> Korrelasjon mellom NDVI beregnet fra forskjellige flyhøyder 28.juli 2022.	30
<i>Figur 21</i> Korrelasjon mellom NDVI beregnet fra forskjellige flyhøyder 01.08.2022.	31
<i>Figur 22</i> Korrelasjon mellom NDVI beregnet fra forskjellige flyhøyder 05.08.2022.	31
<i>Figur 23</i> Korrelasjon mellom NDVI og registrert avling for hver flyhøyde skilt på gjødselmengde 05.august 2022. Korrelasjonskoeffisientene oppgitt i bilde viser den totale korrelasjonskoeffisienten mellom NDVI og avling uavhengig av gjødslingsmengde. Korrelasjonskoeffisientene for de respektive gjødselmengdene er oppgitt i Tabell 3.	33
<i>Figur 24</i> Avlingsprediksjon mot faktisk avling basert på bilder fra forskjellige flyhøyde 28.07.2022.	37
<i>Figur 25</i> Avlingsprediksjon mot faktisk avling basert på bilder fra forskjellige flyhøyde 01.08.	38
<i>Figur 26</i> Avlingsprediksjon mot faktisk avling basert på bilder fra forskjellige flyhøyde 05.08.	39
<i>Figur 27</i> Indekser bruk i LASSO 28.07.2022.	40
<i>Figur 28</i> Indekser bruk i LASSO 01.08.2022.	41
<i>Figur 29</i> Indekser bruk i LASSO 05.08.2022.....	42
<i>Figur 30</i> Korrelasjon mellom predikert avling og registrert avling i anvendt landbruk.....	44
<i>Figur 31</i> Indekser bruk i LASSO for anvendt landbruk.	45
<i>Figur 32</i> Vedlegg 1	56
<i>Figur 33</i> Vedlegg 2	57
<i>Figur 34</i> Vedlegg 3	58

Tabelliste

<i>Tabell 1 Oversikt over sortene som ble brukt i feltforsøket.....</i>	<i>13</i>
<i>Tabell 2 Utregning av vegetasjonsindekser [18-20].</i>	<i>18</i>
<i>Tabell 3 Korrelasjonskoeffisienter mellom NDVI og registrert avling for hver flyhøyde skilt på gjødselmengde..</i>	<i>32</i>
<i>Tabell 4 De 15 mest nøyaktige indeksene for hver dato og hver flyhøyde.....</i>	<i>34</i>
<i>Tabell 5 Korrelasjonskoeffisient mellom hver enkelt indeks og registrert avling for hver høyde og hver dato.....</i>	<i>35</i>
<i>Tabell 6 Korrelasjonskoeffisienter for indekser sammenlignet med registrert avlingsmengde. Sortert etter sterkest korrelasjon.....</i>	<i>43</i>

1 Innledning

Med en stadig økende befolkning, krig i Europa og store klimaforandringer er vi avhengig av et effektivt og bærekraftig landbruk. En større andel av verdens befolkning sulter for hvert år [1]. Konsekvenser av klimaendringene har til felles at de endrer forutsetningene for landbruk, og i de fleste tilfeller gjør de det vanskeligere. Sortsutviklere får en utfordring når de må utvikle sorter for ekstremvær i alle retninger. I tillegg må disse sortene utvikles raskere enn før, for å få de ferdig før vi er tatt igjen av klimaforandringene, og forutsetningene igjen er endret.

Samtidig er selvforsyningsgraden i Norge 45,9% [2]. 3% av det norske landarealet er matjord, og kun 30% av dette egner seg for dyrking av matkorn. Det dyrkes dyrefôr på 90% av norsk matjord. Det betyr at vi bare dyrker matkorn på 1/3 av jorda som egner seg for det.

Dette øker behovet både for effektivt anvendt landbruk, og for sortsutvikling.

Avlingsestimering med UAV (Unmanned Aerial Vehicle) kan være en del av løsningen i begge tilfeller. Bruk av vegetasjonsindekser basert på bilder tatt med UAV for å estimere avlinger kan forkorte tiden det tar fra man begynner utviklingen av en ny sort, til den er ferdig utviklet [3]. De senere årene er også forskjellige modeller for maskinlæring tatt i bruk, med mål om å komme med mer nøyaktige avlingsprediksjoner.

Også anvendt landbruk kan bli både mer bærekraftig og mer effektivt ved bruk av vegetasjonsindekser og/eller maskinlæring for avlingsestimering. Variabel tildeling av gjødsel gjør det mulig å effektivisere gjødslingene. En ting er at slikt utstyr fører til mindre overlapp grunnet seksjonsavstengninger. En annen er at man får muligheten til å tildele riktig mengde gjødsel på riktig sted på skiftet. Dette kan både gjøre at man unngår legde om man gjødsler for mye der det er god vekst, og det kan gjøre at de dårligere områdene på skiftet blir bedre. Et av kravene for å oppnå matkvalitet er proteininnhold [4]. Proteininnholdet henger godt sammen med N-gjødsling, og riktig gjødsling ved aksskyting gir økt proteininnhold [5].

I denne oppgaven blir forskjellige indekser sammenlignet med avlingsmengde for å se hvor godt de korrelerer. Dette vil bli gjort både i et forsøksfelt og på et skifte dyrket på vanlig vis etter norske normer, som ble stilt til disposisjon av en gårdbruker (i denne oppgaven kalt anvendt landbruk). Det blir sett på forskjellige faktorer som kan tenkes å påvirke prestasjonen til indeksene og maskinlæringsmodellen.

I forsøksfelt er flyhøyder i «lavere» sjikt vanligst. I anvendt landbruk kan man dra nytte av å kunne fly noe høyere. Forsøksfelt er oftest plassert bort fra jordekanter og uten hindringer. I anvendt landbruk er det derimot vanligere med potensielle hindringer som høyspentmaster, lyktestolper, åkerholmer med trær og skog i åkerkanten. Ved å fly høyere vil man lettere unngå disse hindringene. I tillegg vil man dekke større områder per bilde, som gjør at man kan spare tid. For et stort skifte vil man også ha utfordringer med batteritid ved lavere flyhøyder.

Spørsmålene som forsøkes besvart i denne oppgaven er følgende:

- Hvordan påvirker *flyhøyde*:
 1. Forskjellige målinger med vegetasjonsindekser?
 2. Korrelasjonen mellom vegetasjonsindeksene og avlingsmengde?
 3. Og prediksjonsmodellens nøyaktighet?
- Hvordan påvirker *vekststadium*:
 1. Nøyaktigheten i prediksjonsmodeller i forsøksfelt?
 2. Nøyaktigheten i vegetasjonsindeksene i forsøksfelt?
- Hvordan påvirker *gjødselmengde* korrelasjonen mellom NDVI og avlingsmengde i forsøksfelt?
- Er prediksjonsmodellen mer nøyaktig enn vegetasjonsindeksene alene?

2 Teori

I dette kapittelet presenteres noe av teorien som ligger til grunn for forsøkene som er gjort i denne oppgaven.

2.1 Fenotyping

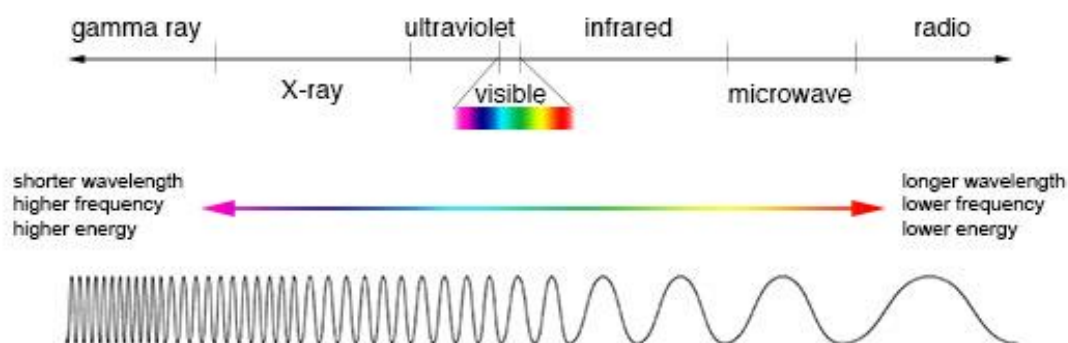
Fenotyping av planter er en kvantitativ beskrivelse av plantens anatomiske, ontogenetiske fysiologiske og biokjemiske egenskaper» [6].

Målet for sortsutvikling i Norge er at bonden skal ha tilgang til sorter som er tilpasset norske behov. Siden vi har korte vekstsesonger med kalde og lange vintere, og varierende lysforhold er det få land med like behov som Norge [7]. En god sort har god resistens mot plantesykdommer og gir høye og stabile avlinger [7].

Klimakrisen gjør det enda viktigere enn før å begrense tiden det tar å utvikle en sort [8]. For å korte ned denne tiden kan verktøy som involverer bildeanalyse være et nyttig verktøy.

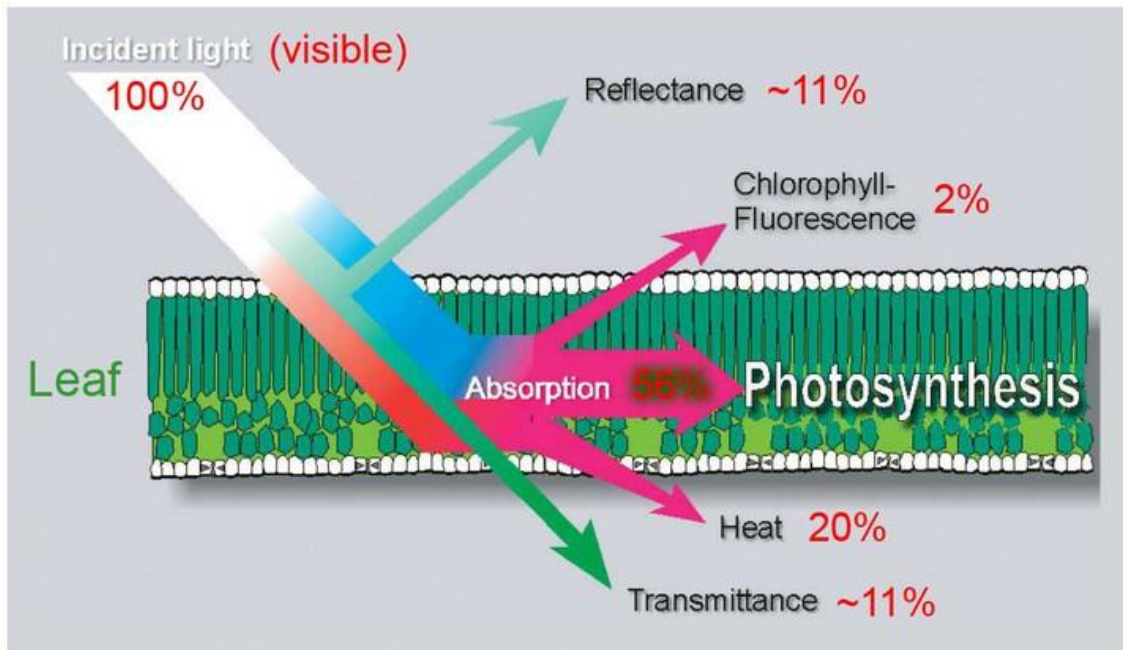
2.2 Spektral refleksjon

Avhengig av materialets fysiske og kjemiske tilstand absorberes og reflekteres solens elektromagnetiske stråling på forskjellige måter (Figur 1). Materialets refleksjonsevne vil også variere med den elektromagnetiske energien. Variasjonene i reflektert lys gjør det mulig å identifisere forskjellige materialer eller egenskaper i jordoverflaten [9]. Bølgelengden på lyset bestemmer fargen mennesker oppfatter. Mennesker kan kun oppfatte bølgelengder mellom cirka 400 og 750 nanometer [10].



Figur 1 Illustrasjon av det elektromagnetiske spektrum [11].

Når det gjelder planter absorberer en frisk plante mye elektromagnetisk energi i den synlige delen av lysspekteret. Klorofyll absorberer mye blått og rødt lys. Mens det reflekterer mye grønt lys (Figur 2). Derfor oppfatter mennesker en frisk plante som grønn. En frisk plante reflekterer også mye nær-infrarødt (NIR), med bølgelengder mellom 700 og 1 300 nanometer. Dette er bølgelengder som mennesker ikke oppfatter [9].



Figur 2 Illustrasjon av hvordan planter reflekterer og absorberer lys [12].

2.3 Vegetasjonsindekser

Vegetasjonsindekser regnes ut basert på to eller flere bølgelengder for å analysere spesifikke karakteristikk ved planten. Bølgelengdene ekstraheres vanligvis fra bilder tatt med UAV eller satellitter. Indeksene kan brukes for å bestemme gjødselmengde, eller se tegn til svakheter i åkeren, som sykdom, tørke eller vannmetning. I anvendt landbruk brukes indekser allerede i stor grad for å bestemme gjødselmengde i programmer som CropSat (MSAVI2) og AtFarm (bl.a. NDVI) [13, 14].

2.4 Maskinlæring

I oppgaven blir en maskinlæringsmodell brukt for å estimere avlingsmengden basert på dataen som er ekstrahert fra bildene tatt med UAV. Resultatene vil så sammenlignes med den registrerte avlingsmengden og med relevante indekser. Dette kan fortelle oss noe om potensialet i å bruke maskinlæring for avlingsprediksjon.

For de fleste maskinlæringsalgoritmer er det viktig at alle faktorer er i samme format. Rådataene blir konvertert til et standardisert format ved å bruke lineær transformasjon, som tilpasser alle faktorer til én og samme skala. Dette kalles datanormalisering. Metoden for datanormalisering er ikke standardisert. Derfor er det opp til brukeren å velge hvordan normaliseringen skal gjøres. I denne oppgaven brukes Z-score som datapreprosesseringsmetode. Etter dette er gjennomsnittet av variablene 0 og verdiene varierer fra -1 til 1.

2.5 Flyhøyde

Både for delen med maskinlæring og delen med indekser flys det i 4 forskjellige høyder; 20, 50, 70 og 100 meter, for å se hvordan det påvirker resultatene. Ved lavere høyder vil man få høyere oppløsning på bildene, mens ved høyere høyder kan man forvente lavere oppløsning. Høyere oppløsning kan gjøre estimatene både mer og mindre nøyaktige. Med ømfintlige indekser kan for høy oppløsning gi forstyrrelser mens lavere oppløsning vil gjøre det lettere for indeksene å oppfatte det store bildet [15]. I tillegg vil den relative farten til dronen, i forhold til vekstene, være høyere jo nærmere sensoren er planten, selv om farten er den samme [15].

3 Metode

I dette kapittelet vil metodene og forutsetningene for forsøkene og analysene presenteres. Først vil stedene for både forsøksfeltet og for anvendt landbruk bli presentert, så vil fremgangsmåten for selve forsøkene blir presentert.

3.1 Forsøksfeltet på Ås, Vollebekk

Forsøkene i oppgaven er delt i to: én del i forsøksfelt og én del i vårhveteåker som er gjort tilgjengelig av bonden.

3.1.1 Plantematerialet

En kolleksjon av 24 vårhvetesorter og -linjer var samlet. Alle med en relevans til det norske markedet de siste 50 årene. Kolleksjonen dekker alle de viktigste sortene i norsk landbruk siden 1972 frem til nå og foredlingsfremgangen som er oppnådd i denne perioden av norske og europeiske foredlere er representert. 23 av 24 sorter er utviklet av norske (Graminor/IPK) eller svenske (Lantmännen/Weibull) foredlere. Den siste sorten (Arabella) er en polsk sort, som dyrkes til fôrhvete grunnet sin høye avling og gode tilpasning til norske forhold (Tabell 1).

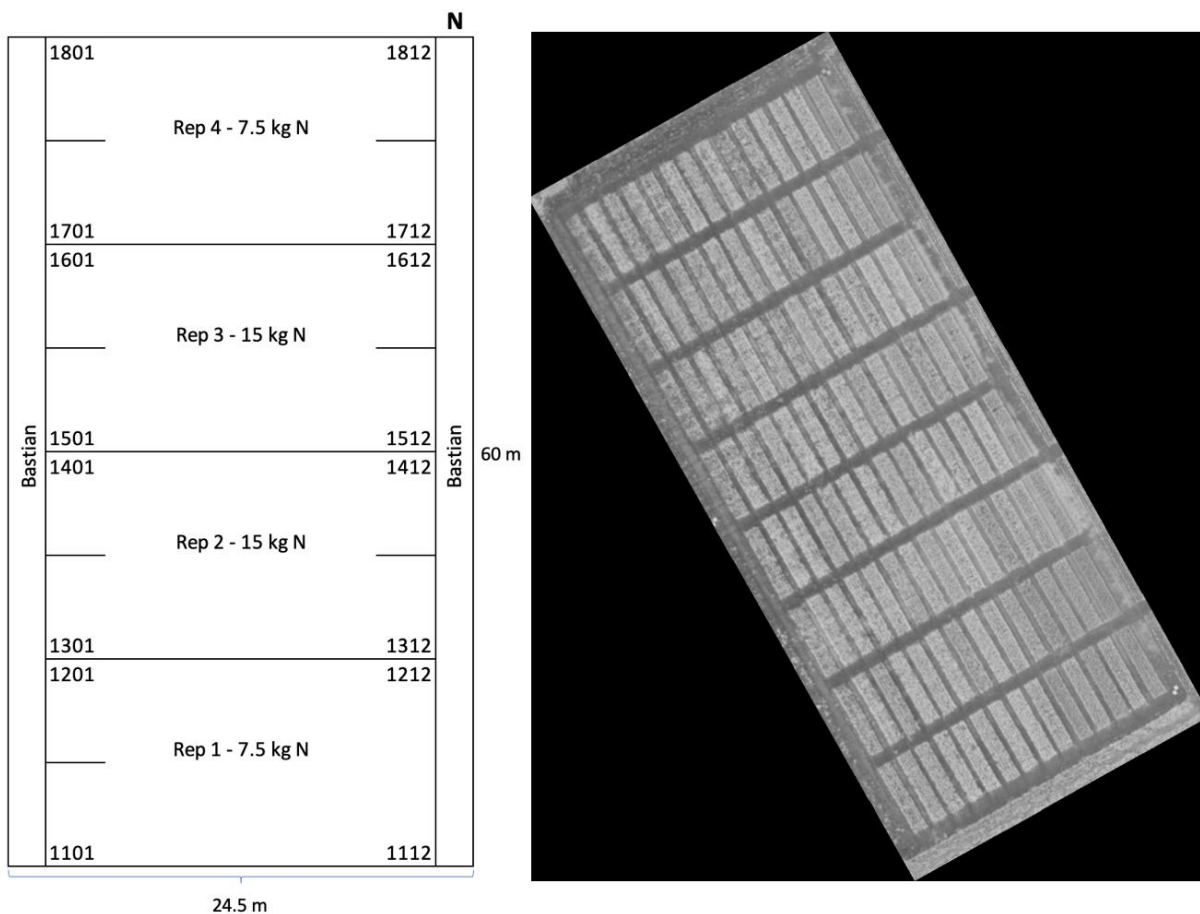
Tabell 1 Oversikt over sortene som ble brukt i feltforsøket.

Sort	Opprinnelse/Selskap	Frigitt i
Runar	Norge/IPK	1972
Reno	Norge/IPK	1975
Tjalve	Sverige/Weibull	1987
Bastian	Norge/IPK	1989
Polkka	Sverige/Lantmännen	1992
Avle	Sverige/Lantmännen	1996
Zebra	Sverige/Lantmännen	2001
Bjarne	Norge/Graminor	2002
Demonstran	Norge/Graminor	2008
Krabat	Norge/Graminor	2010
Mirakel	Norge/Graminor	2012
Rabagast	Norge/Graminor	2013
Arabella	Polen/Danko	2014
Seniorita	Norge/Graminor	2014
Willy	Norge/Graminor	2016
Caress	Sverige/Lantmännen	2017
Zombi	Norge/Graminor	2018
Eleven	Sverige/Lantmännen	2019
Felgen	Sverige/Lantmännen	2019
Betong	Norge/Graminor	2019
Alarm	Norge/Graminor	2019
Festus	Norge/Graminor	2021
Malvolio	Sverige/Lantmännen	2022
Helmi	Norge/Graminor	2022

Kolleksjonen ble valgt på grunn av sin høye fenotypiske variasjon i avling og andre egenskaper [16]. Aksskyting skjedde mellom 24. og 29 juni og gulmodning skjedde mellom 12.08 og 24.08. Det ble ikke registrert vekststadier i forsøket, men vi vet på bakgrunn av dette at de tre datainnsamlingene skjedde i den siste halvdel av kornfyllinga.

3.1.2 Feltdesign

Forsøksfeltet var plassert på Vollebekk Forsøksgård, Ås, Sørøst-Norge (59,66E, 10,75E). Eksperimentet var arrangert som et «randomized incomplete block split plot design», med to replikasjoner av gjødsling og replikasjoner av hver sort i hver replikasjon av gjødsling. Det gir totalt 96 forsøksruter (Figur 3). Hver rad innenfor hver replikasjon inneholdt 6 ruter. Feltet ble sådd og gjødslet 28.april 2022 med 24,9 g såkorn per kvadratmeter, og med NPK gjødsel (YaraMila 22-3-10) i to forskjellige mengder; 7,5 kg N per dekar og 15 kg N per dekar. Hver forsøksrute var 1,5 x 7,5 meter. Feltet ble behandlet etter vanlig praksis i norsk landbruk og holdt sykdoms- og ugrasfritt ved sprøyting med fungicider. Det ble høstet etter modning sent i august for å registrere oppnådde avlinger. Forsøksfeltet ble høstet med forsøktresker for fremskaffing av avlingsdata.



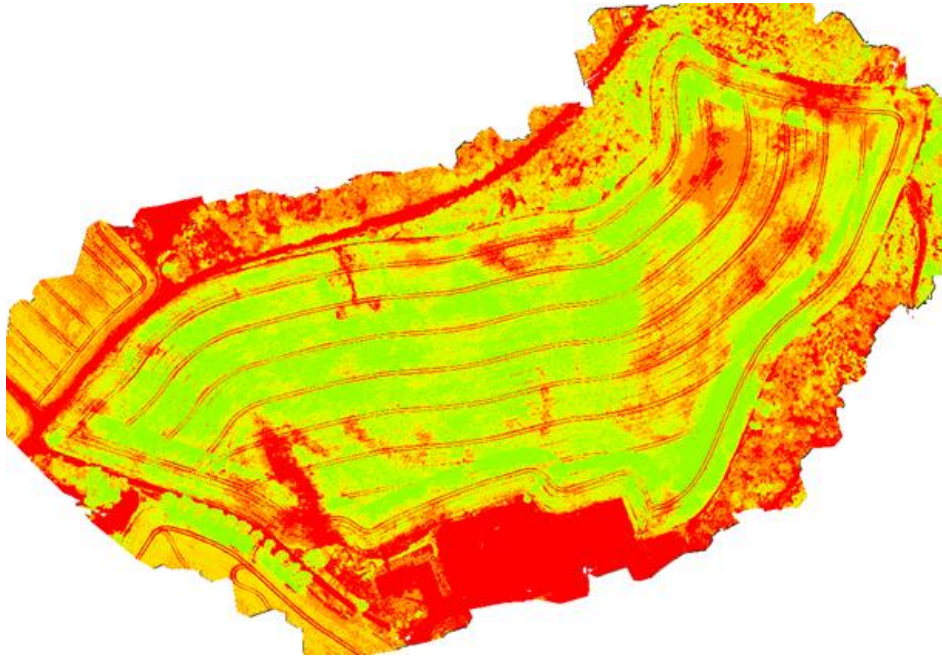
Figur 3 Feltkart og ortomosaikk over Robotfeltet.

3.2 Anvendt landbruk

Grunnet forskjellige feil med dataene fra vekstsesongen 2022 ble det brukt data fra 2021.

Også i 2021 var det problemer med kvaliteten på dataene. Derfor er det kun brukt data fra én bonde.

Skiftet er plassert i Hof, Holmestrand, Sørøst-Norge ($59^{\circ}31'54.82''\text{N}$, $10^{\circ}5'58,99''\text{E}$). Skiftet ble dyrket av eier Knut Lindseth og var 75 dekar stort. Figur 4 viser en ortomosaikk av skiftet.



Figur 4 Ortomosaikk av skiftet.

Jordarten på skiftet går ifølge NIBIO Kilden fra «Siltig mellomsand og siltig finsand» på den vestre delen av jordet (punkt 1 på Figur 5), og over i «grusholdig siltig sand, sandig silt og silt» på den østre delen av jordet (punkt 2 på Figur 5). Kornsorten var Mirakel og det ble sådd 22.04.2021. Avlingen ble høstet med en tresker av typen Claas Tucano 430 med avlingsregistrator fra Claas. Fra avlingsregistratoren ble det hentet ut en shapefil, som senere bearbejdes videre i QGIS (se kapittelet om QGIS - anvendt landbruk).



Figur 5 Jordsmonnet på skiftet varierer fra siltig mellomsand og siltig finsand» (punkt 1) til grusholdig siltig sand, sandig silt og silt (punkt 2).

3.3 Datainnsamling med UAV

Bildene som brukes i oppgaven er tatt med en UAV (unmanned aerial vehicle). Det ble fløyet med en DJI phantom 4 med P4 Multispectral sensor. Dette er et kamera med 6 sensorer. En sensor er en vanlig RGB-sensor, som oppfatter synlig lys. De fem andre oppfatter hver sin av refleksjonene blå, grønn, rød, nær-infrarød og rededge. Brennvidden på kameraet er 5,74 mm, sensorstørrelsen er 4,87 mm x 3,96 mm og bildestørrelsen er 1600x1300 piksler.

DJI GS Pro appen ble brukt for å sette et flymønster [17]. Bildene ble tatt fra forskjellige flyhøyder med 80% overlapp i lengderetning, og 85% overlapp sidelengs. Maksimal hastighet ble satt til 1 m/s. og maksimalt antall bilder ble satt til 1 per sekund. Bildene ble tatt med nadir-vinkel.

For forsøksfeltet blir det brukt data fra tre forskjellige datoer; 28. juli 2022, 01. august 2022 og 05. august 2022. For hver av disse datoene ble det tatt bilder på 20, 50, 70 og 100 meters høyde. Denne oppgaven vil undersøke hvordan de forskjellige tidspunktene og høydene vil påvirke resultatet for prediksjon av avling.

I anvendt landbruk ble datainnsamlingen gjort 26.juni 2022. Dette var rundt aksskyting.

3.3.1 Sammenstilte bilder fra PhenoCrop

RedEdge kalibreringsmodellen ble brukt for å konvertere rådataene fra pikslene til radianverdier. Den kompenserer for svarte området, sensitiviteten for sensoren, forsterkninger fra sensoren, eksponeringsinnstillinger og mørke kanter i bildene. RedEdge-kalibreringen konverterer rådataene til absolutte spektrale radiansverdier. $W/m^2/sr/nm$. Så konverterer en overføringsfunksjon radians til refleksjon for hvert bånd. Alle parametere som blir brukt i modellen kan bli lest fra XMP metadataen i TIFF filen, lagret av RedEdge-kameraet. Bilder av kalibreringspanel (med Albedo-verdier av 0,58, 0,59, 0,60, 0,59 og 0,56, for henholdsvis blå, grønn, rød, RedEdge og NIR) ble tatt rett før og etter flyvning.

Prosessering av UAV-bilder inkludert geometrisk korreksjon, sammenstilling til mosaikker og radiometrisk kalibrasjon ble gjort i Pix4D software med romlig oppløsning 1,32 cm/piksel.

3.4 Dataanalyse

3.4.1 Indekser

I denne oppgaven brukes flere indekser som er brukt for å fortelle noe om plantens helse eller vekstpotensial. Utregningen av de forskjellige vegetasjonsindeksene er vist i Tabell 2.

Det vil også undersøkes hvordan utviklingsstadium for plantene, og flyhøyde for bildetaging påvirker resultatet.

NDVI brukes i mange sammenhenger, både av forskere og i landbruket. Oppgaven ser på hvordan NDVI påvirkes av flyhøyde og forskjellige gjødselmengder. Det er flydd på tre forskjellige dager og oppgaven undersøker hvordan vekststadium påvirker korrelasjonen mellom NDVI og avling.

Tabell 2 Utregning av vegetasjonsindekser [18-20].

Indeks	Navn på indeks	Utregning av indeks	Referanse
NDVI	Normalized Difference Vegetation Index	$(NIR - R\delta d) / (NIR + R\delta d)$	[18]
DVI	Difference Vegetation Index	$NIR - R\delta d$	[18]
GDVI	Green Difference Vegetation Index	$NIR - Gr\delta nn$	[18]
MTCI	MERIS Terrestrial Chlorophyll Index	$(NIR - RedEdge) / (RedEdge - R\delta d)$	[19]
EXG	Excess Green Vegetation Index	$2Gr\delta nn - R\delta d - Bl\delta$	[19]
EXGR	Excess Green minus Excess Red	$ExG - 1,4R\delta d - Gr\delta nn$	[20]
RDVI	Renormalized Difference Vegetation Index	$(NIR - R\delta d) / \sqrt{NIR + R\delta d}$	[18]
TDVI	Transformed Difference Vegetation Index	$1,5 (NIR - R\delta d) / \sqrt{NIR^2 + R\delta d + 0,5}$	[18]
GNDVI	Green Normalized Difference Vegetation Index	$(NIR - Gr\delta nn) / (NIR + Gr\delta nn)$	[18]
NDRE	Normalized Difference RedEdge	$(NIR - RedEdge) / (NIR + RedEdge)$	[18]
SCCI	Simplified Canopy Chlorophyll Content Index	$NDRE/NDVI$	[18]
EVI	Enhanced Vegetation Index	$2,5 * (NIR - R\delta d) / (NIR + 6R\delta d - 7,5Bl\delta + 1)$	[18]
TVI	Triangular Vegetation Index	$0,5 [120 (NIR - Gr\delta nn)] - 200 (R\delta d - Gr\delta nn)$	[18]
VARI	Visible Atmospherically Resistant Index	$(Gr\delta nn - R\delta d) / (Gr\delta nn + R\delta d - Bl\delta)$	[18]
GARI	Green Atmospherically Resistant Index	$NIR - Gr\delta nn - (1,7 (Bl\delta - R\delta d)) / (NIR + Gr\delta nn - (1,7 (Bl\delta - R\delta d)))$	[18]
GCI	Green Chlorophyll Index	$(NIR / Gr\delta nn) - 1$	[18]
GLI	Green Leaf Index	$(Gr\delta nn - R\delta d - Bl\delta) / (2Gr\delta nn + R\delta d + Bl\delta)$	[18]
NLI	Non-Linear Index	$(NIR^2 - R\delta d) / (NIR^2 + R\delta d)$	[18]
MNLI	Modified Non-Linear Index	$(NIR^2 - R\delta d) * (1+0,5) / (NIR^2 + R\delta d + 0,5)$	[18]
SAVI	Soil-Adjusted Vegetation Index	$1,5 * (NIR - R\delta d) / (NIR + R\delta d + 0,5)$	[18]
GSAVI	Green Soil-Adjusted Vegetation Index	$1,5 * (NIR - Gr\delta nn) / (NIR + Gr\delta nn + 0,5)$	[18]
OSAVI	Optimized Soil-Adjusted Vegetation Index	$(NIR - R\delta d) / (NIR + R\delta d + 0,16)$	[18]
GOSAVI	Green Optimized Soil-Adjusted Vegetation Index	$(NIR - Gr\delta nn) / (NIR + Gr\delta nn + 0,16)$	[18]
MSAVI2	Modified Soil-Adjusted Vegetation Index 2	$(2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - R\delta d)}) / 2$	[18]
MSR	Modified Simple Ratio	$(NIR / R\delta d) - 1 / \sqrt{NIR/R\delta d + 1}$	[18]
GRVI	Green Ratio Vegetation Index	$NIR / Gr\delta nn$	[18]
WDRVI	Wide Dynamic Range Vegetation Index	$(0,1 NIR - R\delta d) / (0,1 NIR + R\delta d)$	[18]
SR	Simple Ratio	$NIR / R\delta d$	[18]

3.4.2 NDVI

NDVI (Normalized Difference Vegetation Index) brukes både i anvendt landbruk og i forskning [18]. Indeksen baserer seg på forholdet mellom det nær-infrarøde båndet og det røde båndet (Tabell 2) og er ment for å estimere biomasseutviklingen i veksten [18]. Forsøk har vist at NDVI blir mindre nøyaktig etter hvert som biomassen øker [18]. Dette kan være fordi den er følsom for reflektert grønnfarge fra klorofyll [21].

3.4.3 SCCC

SCCCI (Simplified Canopy Chlorophyll Content Index) er basert på forholdet mellom NDRE og NDVI (Tabell 2). NDRE er basert på NIR og RedEdge. Dette kan gjøre at indeksen fanger opp flere parametere enn NDVI. Ved å sette sammen flere indekser kan man få et mer nøyaktig estimat av klorofyll og plantens biomasse [22]. SCCC viste seg som den mest nøyaktige indeksen i flere vekststadier i et forsøk i mais [18].

3.4.4 MTCI

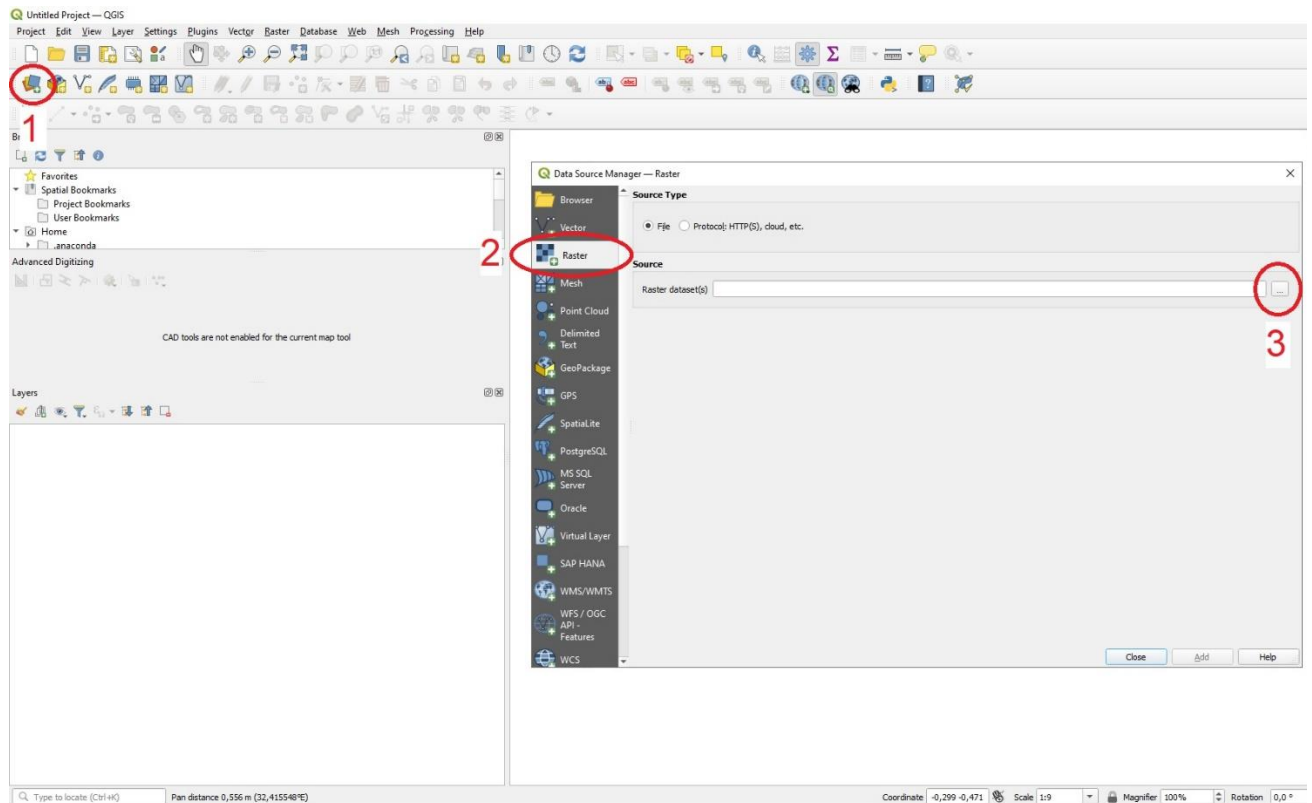
MTCI (Meris Terrestrial Chlorophyll Index) er basert på nær-infrarødt, RedEdge og rødt. MTCI estimerer klorofyll-innhold og skal være mindre sårbar for endringer i plantedekke gjennom sesongen [21]. Det kan tenkes at dette gjør at nøyaktigheten i estimatene vil avta mindre gjennom sesongen med MTCI enn med NDVI.

3.5 Ekstrahering av data i QGIS

QGIS ble brukt for å ekstrahere dataene fra bildene som ble sammenstilt i Pix4D. I denne delen vil det bli gjort rede for arbeidet gjennomført i QGIS. Versjon 3.26.2 av QGIS for Windows ble brukt i denne oppgaven. Dersom man bruker en senere versjon kan det forekomme endringer i programmet.

3.5.1 Forsøksfelt

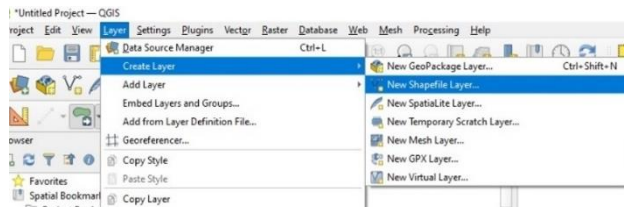
Etter å ha åpnet et nytt prosjekt må man laste opp bildene. Dette gjør man med å følge rekkefølgen i Figur 6 og finner frem til mappa med de ønskede filene i filutforskeren på datamaskinen. Filene som skal lastes inn skal være i TIFF-format (tagged image file). Disse filene kalles raster.



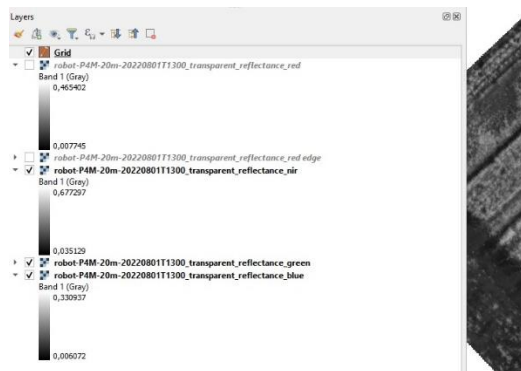
Figur 6 Rekkefølge for å laste inn filer til QGIS.

Når raster-lagene er lagt inn lages et grid. I denne oppgaven er det brukt følgende fremgangsmåte for dette:

Først lages en Shapefil, som vist i Figur 7. Da får man opp et nytt kartlag med navnet man har gitt i menyen som kommer opp etter man har fulgt retningslinjene i Figur 7. Da vil menyen til høyre se ut som i Figur 8.



Figur 7 Skjerm bilde QGIS.



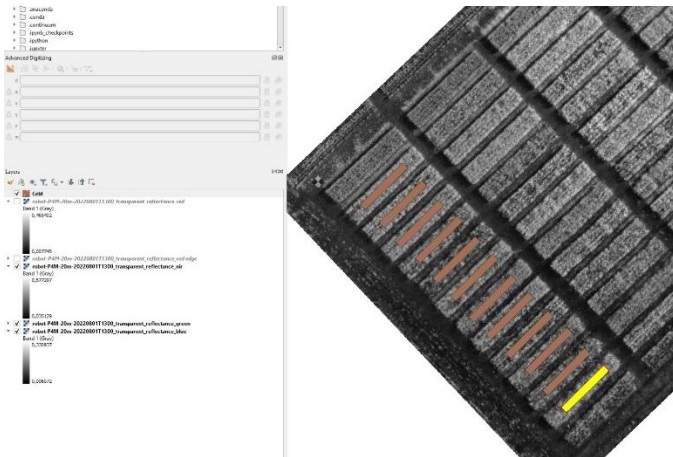
Figur 8 Skjerm bilde QGIS.

Deretter lages det en polygon. Først velges symbolet som er merket med 1 i Figur 9. Da får man opp tegneverktøyet som gjør det mulig å tegne et rektangel inne i forsøksrutene.



Figur 9 Skjerm bilde QGIS.

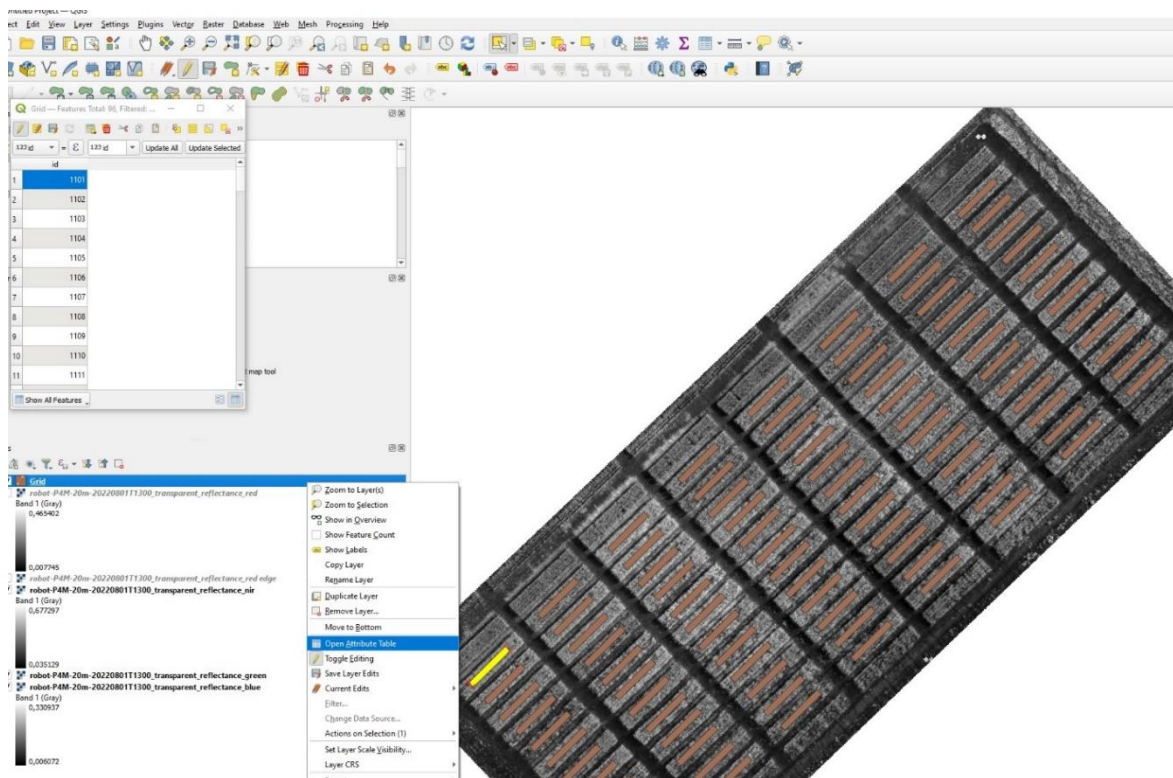
Polygonet tegnes inn manuelt i den første ruten. Det er viktig at rektangelet ikke er så stort at det tar med kantene av forsøksruta som vist i Figur 10. Etter å ha tegnet inn polygonet i en rute kopierer man polygonet ved å markere det med verktøyet markert med 2 i Figur 9. Da skifter det markerte polygonet farge (vist i Figur 10). Da kopierer og limer man inn med **CTRL+C** og **CTRL+V**. Det nye polygonet blir limt inn oppå det forrige og man må da flytte det med verktøyet markert med 3 i Figur 9. Med det valgt kan det nye polygonet dras til neste forsøksrute. Det er viktig å plassere ut polygonene i kronologisk rekkefølge for en lettest mulig jobb med å navngi rutene senere.



Figur 10 Skjerm bilde QGIS.

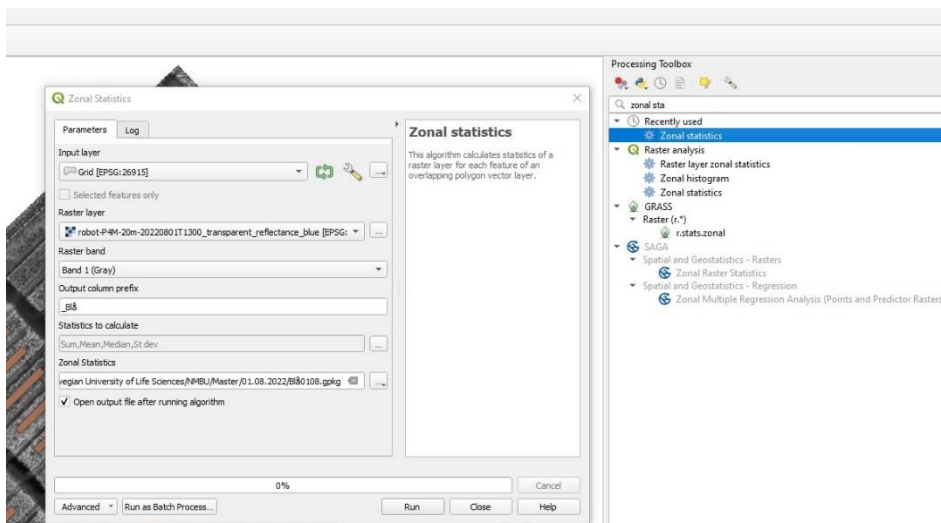
Når en rad er ferdig, kan den kopiere og limes inn på de resterende radene på feltet. Deretter justeres de polygonene som ikke er midt i forsøksruta enkeltvis. Begge disse handlingene gjøres med samme framgangsmåte som tidligere forklart for klipping og liming.

Polygonene må navngis så de samsvarer med feltkartet (Figur 3Figur 6). Da åpner man attributt-tabell i nedtrekksmenyen som vist i Figur 11. Da kommer vinduet som er vist til venstre i Figur 11. I denne menyen skriver man inn navn i kolonnen *id*. Navnet er nummeret på forsøksruta.



Figur 11 Skjerm bilde QGIS.

Dataen ekstraheres ved å velge *Zonal Statistics*. I nedtrekksmenyen *Processing* i menylinja i toppen åpnes *Toolbox*. Da kommer det opp en meny til høyre hvor man søker opp *Zonal Statistics* og velger det (vist i Figur 12). Da kommer vinduet vist i Figur 12 opp. Her brukes griddet man har laget som *Input layer*, og en fil for en av fargebåndene som *Raster band*. I *Statistics to calculate* velges hvilken statistikk man ønsker. I denne oppgaven ekstraherte vi sum, gjennomsnitt, median og standardavvik. Fila lagres på ønsket lokasjon.



Figur 12 Skjerm bilde QGIS.

Deretter åpnes attributt-tabellen igjen (vist i Figur 11). Da får man opp all statistikken man har ekstrahert i vinduet vist i Figur 13. Ved å markere hele tabellen kan den kopieres og limes inn i Excel.

fid	id	_Blåsum	_Blåmean	_Blåmedian	_Blåstdev
1	1	1101 487,2860440807...	0,027060923201...	0,025634335353...	0,008541998608...
2	2	1102 431,2922003669...	0,023942056198...	0,022614801302...	0,007685678860...
3	3	1103 485,8279438875...	0,026973957242...	0,025135247036...	0,009347578107...
4	4	1104 453,1489680362...	0,025174942668...	0,022942778654...	0,009020959750...
5	5	1105 496,6076968740...	0,027570935869...	0,025902648456...	0,008203771492...
6	6	1106 465,1246776310...	0,025817311147...	0,023531839251...	0,008497147827...
7	7	1107 458,3491223221...	0,025444050312...	0,023903987370...	0,008009513452...
8	8	1108 532,6986260255...	0,029594368112...	0,027215288951...	0,010622779247...
9	9	1109 452,6979915862...	0,025131737721...	0,023840744048...	0,007629627876...
10	10	1110 477,1737391930...	0,026489049583...	0,024517595767...	0,008268557512...
11	11	1111 492,4607366584...	0,027354370752...	0,025902764871...	0,009004570728...
12	12	1112 561,3986676260...	0,031173228253...	0,028993751853...	0,011307084479...
13	13	1201 469,4490116974...	0,026061678326...	0,025026960298...	0,008638591045...
14	14	1202 510,9748815987...	0,028392225459...	0,027063077315...	0,009400224047...

Figur 13 Skjerm bilde QGIS.

Fremgangsmåten gjentas for hver farge og hver flyhøyde. Gridet lagres og brukes igjen. Det er viktig å sjekke for hver høyde og dato at gridet fortsatt passer. Noen justeringer kan være nødvendig.

3.5.2 Anvendt landbruk

Fremgangsmåten for å ekstrahere data for anvendt landbruk er nesten den samme som i forsøksfelt. Det som skiller disse er at avlingene i forsøksrutene er registrert per rute, mens man får mange punkt med avlingsregistrering per hektar fra avlingsregistratoren i treskeren.

Man laster inn dataene på samme måte som for anvendt landbruk, men her lager man ikke et grid. Avlingsdataen lastes ned som en shapefil og punktene for avlingsregistrering ligger i denne. Deretter må man rense fila for å fjerne punkt som gir feil data. Dette gjøres med å slette punkter som ligger dobbelt eller på annen måte ikke ser ut til å følge mønsteret til de andre punktene. For å slette punktene bruker man markeringsverktøyet for å markere punkter og sletter de med *delete selected*-knappen (markert i Figur 14). I denne oppgaven blir hele vendeteigen fjernet for å gjøre arbeidet enklere og samtidig være sikker på å få data av god kvalitet.



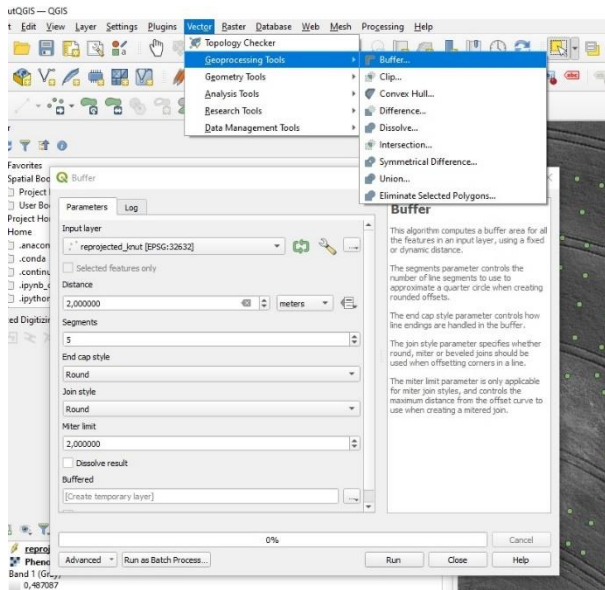
Figur 14 Skjerm bilde QGIS.

For denne oppgaven ser avlingskartet ut som i Figur 15 når det er ferdig renset. Disse punktene blir utgangspunktet for hvor man skal ekstrahere data fra. Ut fra hvert punkt blir det laget en buffer med en radius på 4 meter.



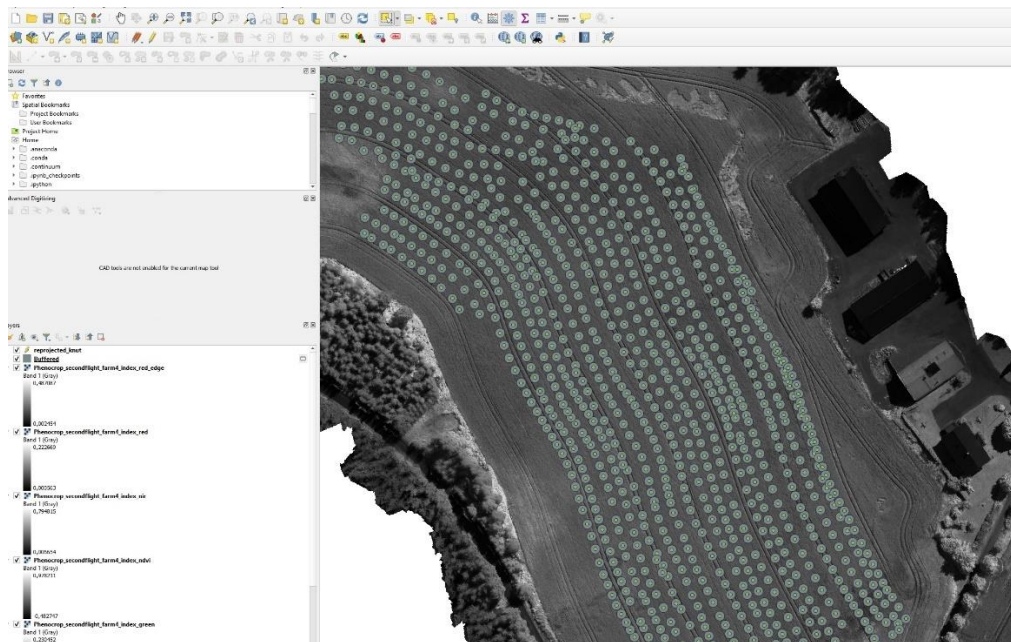
Figur 15 Skjerm bilde QGIS.

Å lage en buffer gjør man ved å velge *vector* i menyen øverst i QGIS. I nedtrekksmenyen velger man *Geoprocessing Tools -> Buffer* (Figur 16). Da får man opp vinduet som er vist i Figur 16. Her velger man vektoren med avlingskartet som input, og velger ønsket størrelse på bufferen. Deretter trykker man på *run* nederst i vinduet.



Figur 16 Skjerm bilde QGIS.

Når bufferen er laget vil det se ut som på Figur 17. Herfra følges samme fremgangsmåte som i forsøksfeltet; man kjører *zonal statistics* for hver farge. I anvendt landbruk ble det kun fløyet i én høyde.



Figur 17 Skjerm bilde QGIS.

3.6 Maskinl ring

For all maskinl ring og utregning av indekser i oppgaven brukes Python 3.10. For bruk av Python blir Jupyter Notebook 4.6.8 brukt. Det meste gj res p  en PC med AMD Ryzen 5 5600X Prosessor, mens noe gj res p  en MacBook Pro, 2019 med 1,4 GHz firekjerners Intel Core i5 prosessor. For LASSO brukes pakken Sckit-learn. Dette er en «open source»-pakke som inneholder forskjellige verkt y for prediktive dataanalyser [23]. Som input i all dataanalyse ble median brukt, for   v re mindre s rbar for ekstremverdier enn man ville v rt med gjennomsnitt.

I denne oppgaven brukes LASSO regresjonsmodell. Metoden skal begrense feilmarginen i prediksjoner. Dette gj r den ved   straffe h ye verdier ved   vektlegge disse mindre, og p  den m ten blir modellen mindre sensitiv for ekstremverdier. Modellen f r f rst 70% av alle dataene, med b de indekser, og «fasit» i form av registrert avling. Deretter f r den kun indeksene, uten «fasit» for de siste 30%. M let er at modellen skal finne en kombinasjon basert p  de f rste 70%, slik at den kan predikere en avlingsmengde s  n yaktig som mulig. Hvor hardt man straffer ekstremverdiene bestemmes av niv et p  *lambda*. Verdien p  lambda settes et sted mellom 0 og ∞ . Jo h yere lambda er, jo mer blir ekstremverdier nedjustert. Verdiene som er lavere enn 0 blir utelatt fra modellen [3]. I denne oppgaven er lambda satt til 0,001.

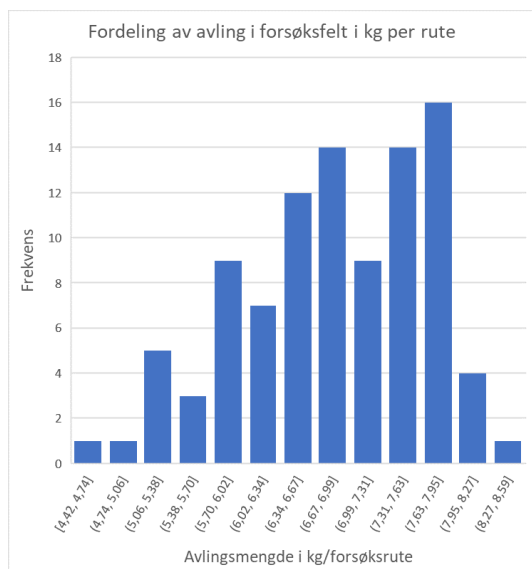
For   vurdere kvaliteten for metoden i denne oppgaven brukes korrelasjonskoeffisienten.

4 Resultater

I dette kapittelet skal resultatene av de presenterte metodene presenteres.

4.1 Avlingsdata

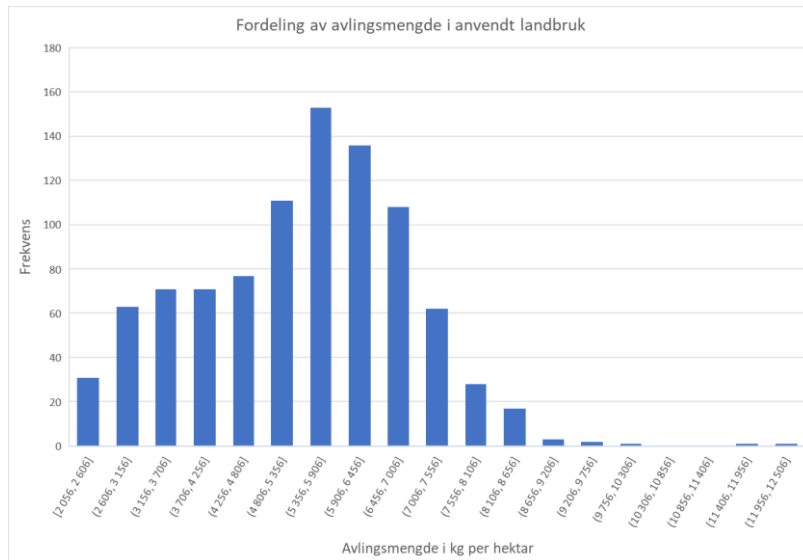
Avlingene i forsøksfeltet varierte fra 4,42 til 8,59 kg per forsøksrute (Figur 18).



Figur 18 Fordeling av avlingsmengde i forsøksfelt.

Dataen fra avlingsregistratoren i treskeren inneholdt mange punkt med som var registrert som 0 kg. Disse ble fjernet. I tillegg ble alle registreringspunkt på vendeteigen (ca 30m rundt hele skiftet) fjernet.

Registreringer på under 200 kg per dekar ble fjernet og det samme ble registreringer over 1 200 kg per dekar. Etter rensingen av data var avlingsmengden fordelt som i Figur 19.

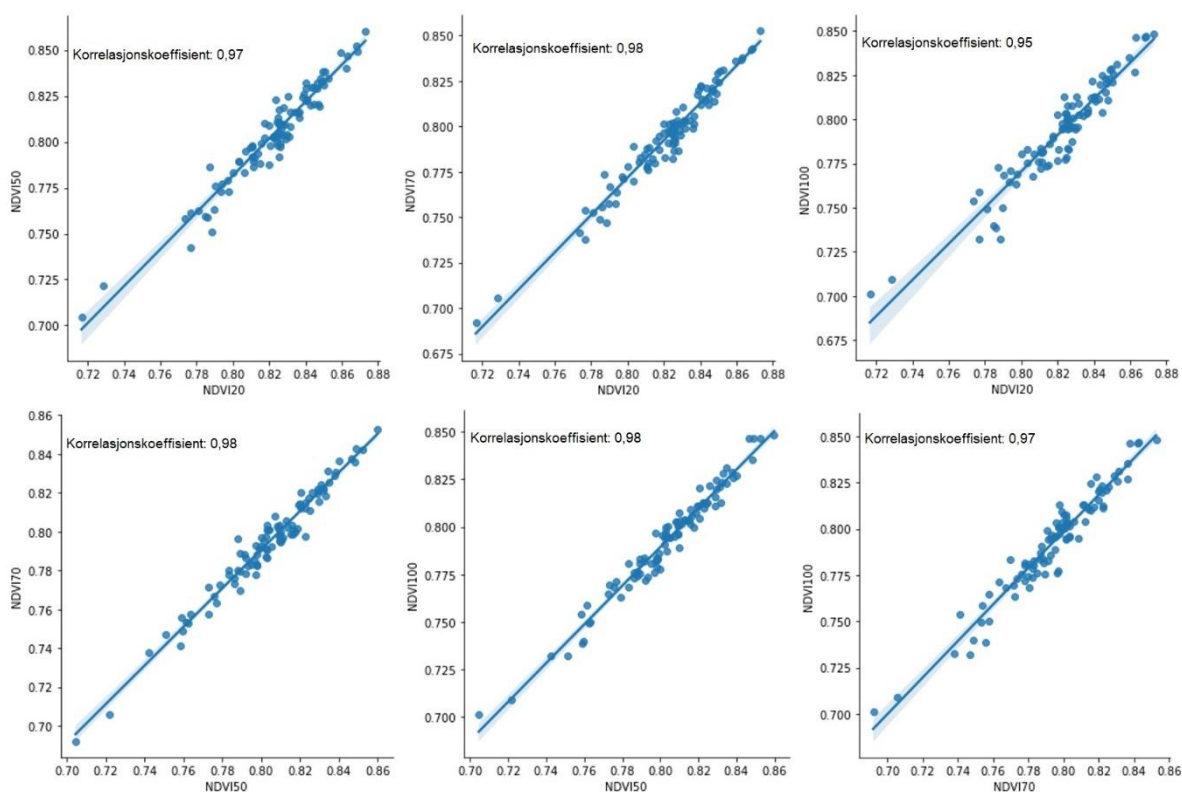


Figur 19 Fordeling av avlingsmengde i anvendt landbruk etter rensing av data.

4.2 Korrelasjon mellom NDVI fra forskjellige flyhøyder

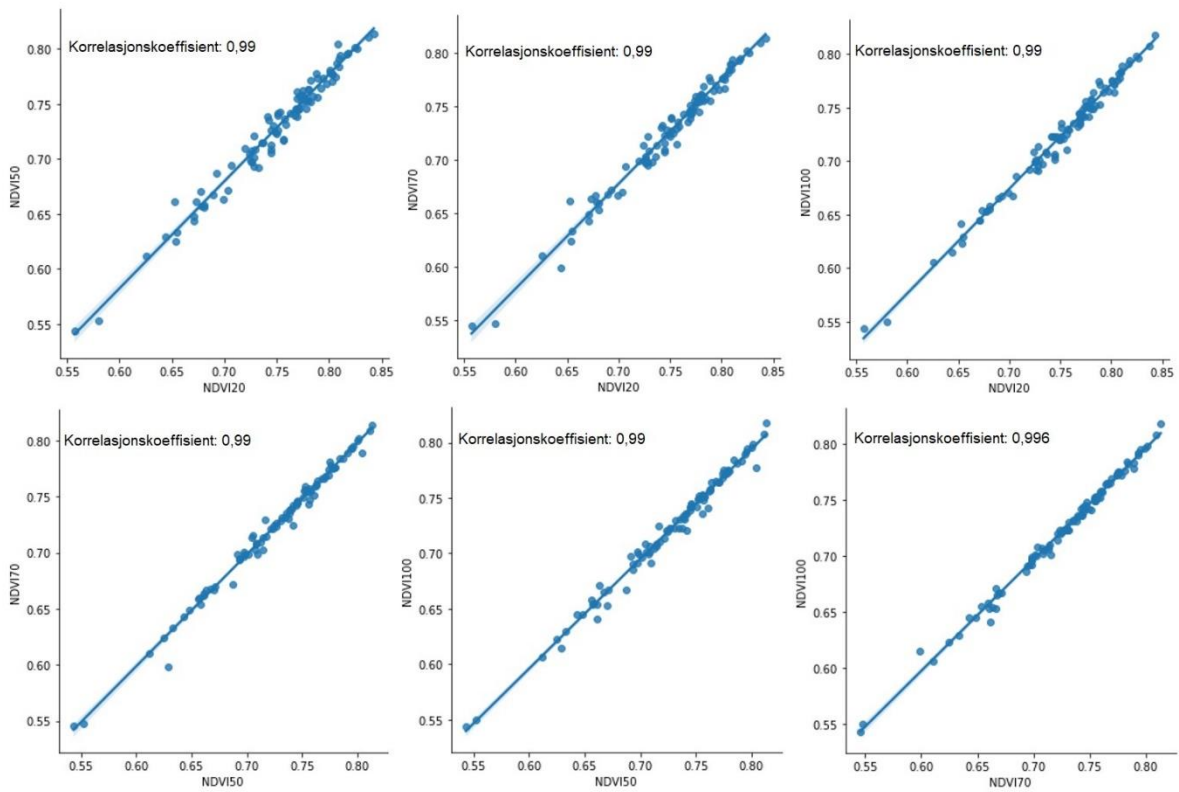
Korrelasjonen mellom NDVI fra de forskjellige høydene kan si noe om hvordan høyden for flyging påvirker kvaliteten på resultatene. Resultatet viser at NDVI generelt varierer lite fra høyde til høyde på samme dag.

Figur 20 viser at korrelasjonen er noe svakere den tidligste dagen for flygning (28. juli), men korrelasjonen er generelt sterk. Den 28. juli er den svakeste korrelasjonskoeffisienten mellom 20 og 100 meter (0,95), mens den er sterkest mellom 20 og 70, 50 og 70, og 50 og 100 (0,98).

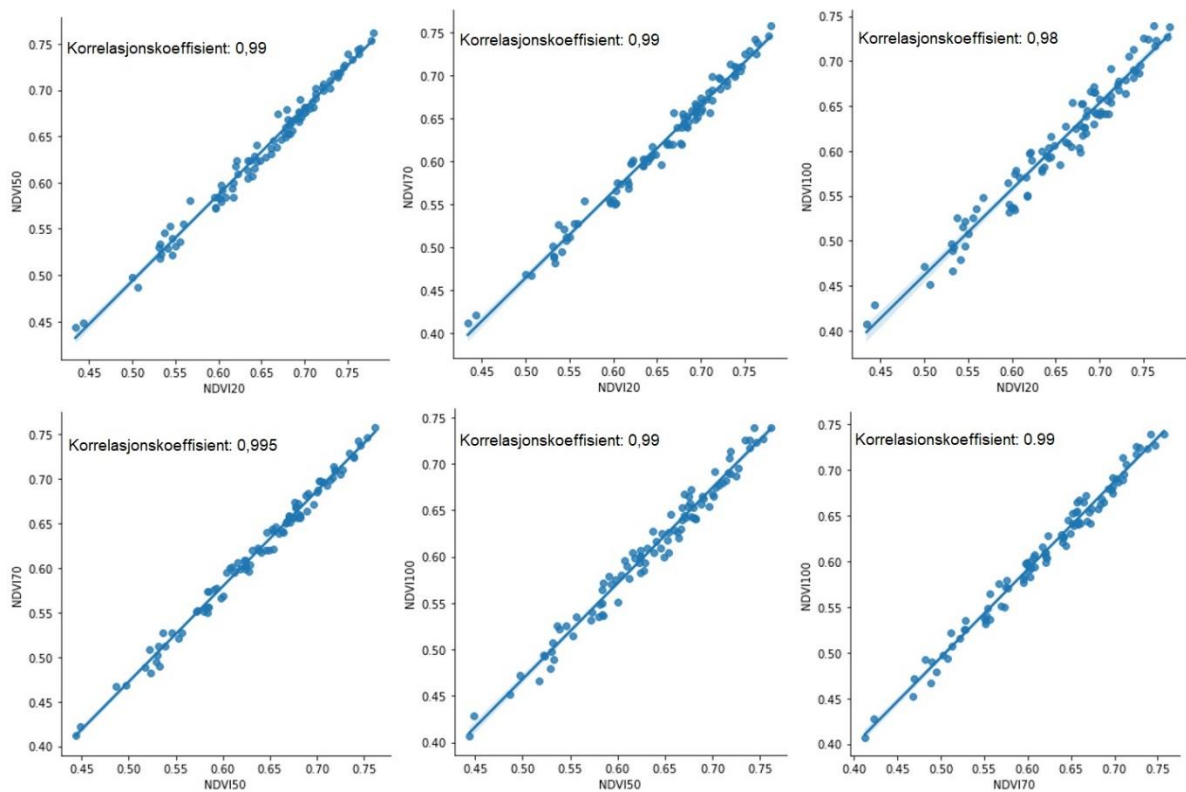


Figur 20 Korrelasjon mellom NDVI beregnet fra forskjellige flyhøyder 28.juli 2022.

Figur 21 og Figur 22 viser forskjellen ved de forskjellige høydene for henholdsvis 1. og 5. august. For 1. august er den svakeste korrelasjonskoeffisienten 0,99 mens for 5. august er den svakeste 0,98.



Figur 21 Korrelasjon mellom NDVI beregnet fra forskjellige flyhøyder 01.08.2022.



Figur 22 Korrelasjon mellom NDVI beregnet fra forskjellige flyhøyder 05.08.2022.

4.3 Korrelasjon mellom forskjellige flyhøyder og avling

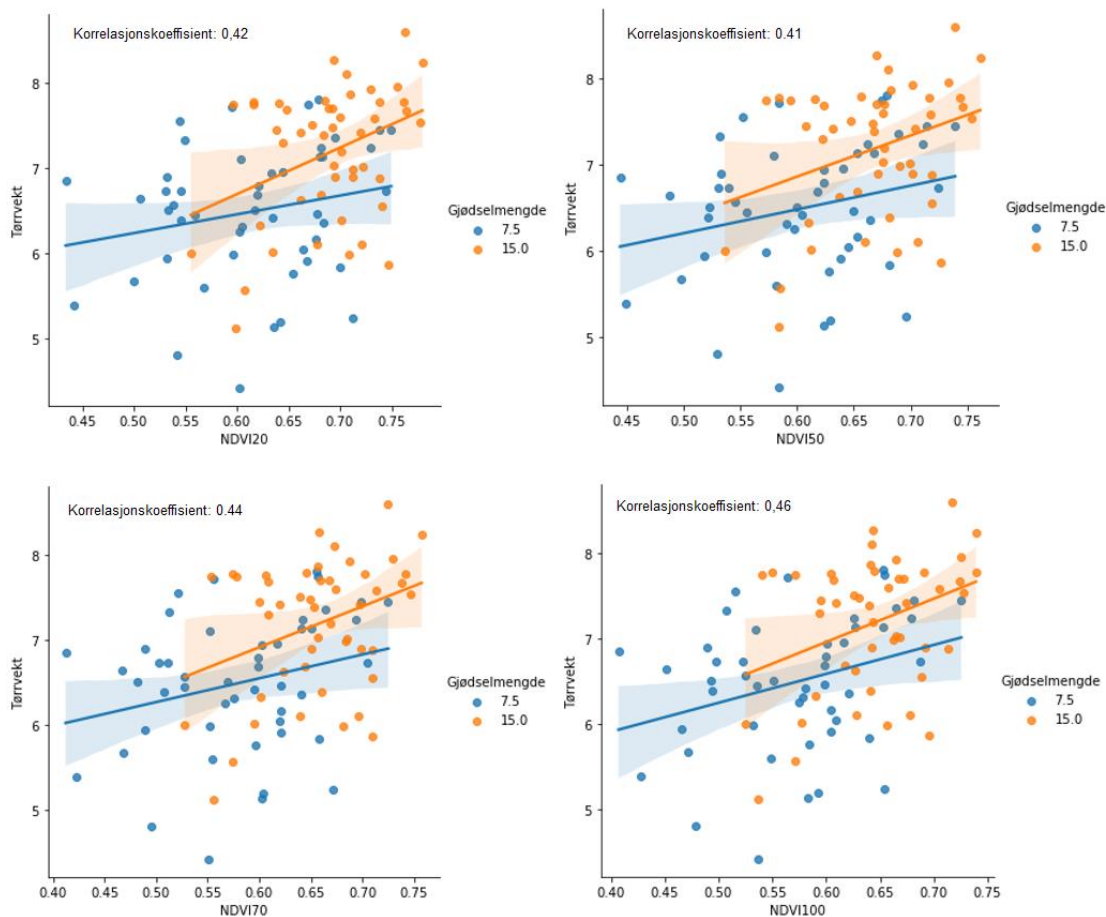
NDVI ble sammenlignet med registrert avlingsmengde. Tabell 3 viser resultatene for de tre forskjellige datoene for flyvning. Disse viser at det ikke er signifikante forskjeller på nøyaktigheten ved de forskjellige høydene. Det er svakest korrelasjon mellom NDVI og registrert avlingsmengde 28. juli (0,13-0,17). Korrelasjonen er sterkere 01. august (0,23-0,24). Datoen med sterkest korrelasjon mellom NDVI og avlingsmengde er 05. august (0,41-0,46).

Tabell 3 Korrelasjonskoeffisienter mellom NDVI og registrert avling for hver flyhøyde skilt på gjødselmengde.

Dato	Høyde	Korrelasjonskoeffisient		
		7,5 kg N	15 kg N	Uavh. av gj.mengde
28.jul	20m	0,005	-0,03	0,17
	50m	0,06	-0,11	0,16
	70m	0,02	-0,07	0,13
	100m	0,05	-0,08	0,17
01.aug	20m	0,12	-0,1	0,23
	50m	0,17	-0,14	0,23
	70m	0,16	-0,12	0,24
	100m	0,16	-0,13	0,23
05.aug	20m	0,21	0,36	0,42
	50m	0,25	0,32	0,41
	70m	0,27	0,34	0,44
	100m	0,31	0,35	0,46

Det var gjødslet med to forskjellige gjødselmengder i forsøksfeltet. Den ene halvdel fikk 7,5 kg N per dekar, mens den andre fikk 15 kg N per dekar. For 28. juli og 1. august er korrelasjonen mellom NDVI og avling positiv for lav gjødselmengde, mens den er negativ for høy gjødselmengde. For 05. august er korrelasjonen positiv for begge. Figur 23 viser korrelasjonen mellom NDVI med de forskjellige gjødslingsmengdene.

Korrelasjonskoeffisienten som er mellom NDVI og avling uavhengig av gjødslingsmengde. Korrelasjonskoeffisientene for de respektive gjødselmengdene er oppgitt i Tabell 3.



Figur 23 Korrelasjon mellom NDVI og registrert avling for hver flyhøyde skilt på gjødselmengde 05.august 2022. Korrelasjonskoeffisientene oppgitt i bilde viser den totale korrelasjonskoeffisienten mellom NDVI og avling uavhengig av gjødslingsmengde. Korrelasjonskoeffisientene for de respektive gjødselmengdene er oppgitt i Tabell 3.

4.4 Indekser

28 forskjellige indekser ble kalkulert basert på dataene ekstrahert i QGIS. Alle ble kalkulert for alle datoene for datainnsamling og det ble utført en korrelasjonsanalyse i Python for å se hvilke indekser som korrelerer best med registrert avling. Tabell 4 viser de 15 indeksene som korrelerer sterkest med registrert avlingsmengde for hver flyhøyde hver dato. Tabellen involverer også de enkelte fargebåndene som indeksene er basert på. Indeksene samsvarer generelt bedre med avlingsmengde 05. august, som var den seneste datoen for innsamling av data. Dette var også tilfellet ved avlingspredikering med maskinlæring. For 20, 50 og 70 meters flyhøyde korrelerte indeksen GDVI best med avlingsmengde (korrelasjonskoeffisient 0,66, 0,55 og 0,59). Ved 100 meter korrelerer GDVI dårligere og er den tolvte sterkeste (korrelasjonskoeffisient 0,48). GSAVI er jevnere, og korrelasjonskoeffisienten varierer bare fra 0,52 til 0,64 for alle høydene.

Tabell 4 De 15 mest nøyaktige indeksene for hver dato og hver flyhøyde.

Indekser som korrelerer sterkest med registrert avlingsmengde - 28.07.2022							
20m		50m		70m		100m	
Indeks	Korr.koff Indeks og avling	Indeks	Korr.koff Indeks og avling	Indeks	Korr.koff Indeks og avling	Indeks	Korr.koff Indeks og avling
MTCI_CI	0.25	RedEdge	-0.28	SCCI	0.30	RedEdge	-0.47
NDRE	0.24	GARI	-0.28	RedEdge	-0.29	NDVI	-0.47
SCCI	0.19	Red	-0.26	NDVI	-0.29	Green	-0.43
GCI	0.19	Green	-0.25	MTCI_CI	0.28	GARI	-0.43
GRVI	0.19	EXG	-0.24	Red	-0.25	SCCI	0.43
GNDVI	0.18	MTCI_CI	0.24	Green	-0.24	Red	-0.41
MSR	0.18	SCCI	0.24	NDRE	0.24	EXG	-0.32
WDRVI	0.18	NDRE	0.22	GARI	-0.22	Blue	-0.31
SR	0.18	VARI	0.21	Blue	-0.18	NIR	-0.31
GSAVI	0.17	NDVI	0.16	VARI	0.17	GDVI	-0.26
OSAVI	0.17	MSR	0.16	NIR	-0.15	DVI	-0.25
GOSAVI	0.17	WDRVI	0.16	EXG	-0.14	TVI	-0.24
RDVI	0.16	SR	0.16	MSR	0.14	MSAVI2	-0.23
NLI	0.16	NIR	-0.13	WDRVI	0.14	MTCI_CI	0.22
MNLI	0.16	GNDVI	0.13	SR	0.14	NDRE	0.22
Indekser som korrelerer sterkest med registrert avlingsmengde - 01.08.2022							
20m		50m		70m		100m	
Indeks	Korr.koff Indeks og avling	Indeks	Korr.koff Indeks og avling	Indeks	Korr.koff Indeks og avling	Indeks	Korr.koff Indeks og avling
SCCI	0.30	SCCI	0.30	MTCI_CI	0.29	GARI	-0.33
MTCI_CI	0.29	MTCI_CI	0.29	SCCI	0.29	Red	-0.30
NDRE	0.28	NDRE	0.27	NDRE	0.28	MTCI_CI	0.29
Red	-0.25	Red	-0.26	Red	-0.26	SCCI	0.29
NDVI	0.23	GARI	-0.26	GARI	-0.25	Green	-0.27
GNDVI	0.23	NDVI	0.23	NDVI	0.24	NDRE	0.27
WDRVI	0.23	VARI	0.22	VARI	0.23	RedEdge	-0.26
Green	-0.22	WDRVI	0.22	WDRVI	0.23	VARI	0.24
GCI	0.22	Green	-0.21	GNDVI	0.22	NDVI	0.23
MSR	0.22	GNDVI	0.21	MSR	0.22	WDRVI	0.23
GRVI	0.22	MSR	0.21	SR	0.22	MSR	0.22
SR	0.22	SR	0.21	NLI	0.21	SR	0.22
GARI	-0.20	OSAVI	0.20	OSAVI	0.21	GNDVI	0.21
NLI	0.20	GCI	0.19	Green	-0.20	GCI	0.19
OSAVI	0.20	NLI	0.19	GCI	0.20	GRVI	0.19
Indekser som korrelerer sterkest med registrert avlingsmengde - 05.08.2022							
20m		50m		70m		100m	
Indeks	Korr.koff Indeks og avling	Indeks	Korr.koff Indeks og avling	Indeks	Korr.koff Indeks og avling	Indeks	Korr.koff Indeks og avling
GDVI	0.66	GDVI	0.55	GDVI	0.59	GOSAVI	0.53
NIR	0.65	GSAVI	0.54	GSAVI	0.58	MTCI_CI	0.53
DVI	0.64	DVI	0.53	DVI	0.57	GSAVI	0.52
GSAVI	0.64	MSAVI2	0.53	MSAVI2	0.57	NDRE	0.52
MSAVI2	0.64	NIR	0.52	TDVI	0.56	EVI	-0.51
TDVI	0.63	TDVI	0.52	TVI	0.55	SCCI	0.51
TVI	0.62	TVI	0.52	EVI	-0.54	MNLI	0.50
SAVI	0.60	EVI	-0.51	NIR	0.54	SAVI	0.49
GOSAVI	0.60	SAVI	0.51	SAVI	0.54	RDVI	0.49
EVI	-0.59	GOSAVI	0.51	GOSAVI	0.54	NLI	0.49
MNLI	0.58	MNLI	0.50	NIR	0.54	OSAVI	0.49
RDVI	0.58	RDVI	0.50	RDVI	0.53	GDVI	0.48
NLI	0.56	NLI	0.49	NLI	0.52	TDVI	0.48
OSAVI	0.54	OSAVI	0.48	OSAVI	0.50	MSR	0.48
RedEdge	0.50	NDRE	0.42	MTCI_CI	0.48	WDRVI	0.48

For 28. juli og 01. august er korrelasjonen mellom indeksene og avlingen generelt svakere. I tillegg er det i større grad de samme indeksene som gjør det best. Tabell 5 viser korrelasjonskoeffisienten for alle indeksene sammenlignet med avlingsmengde. For 28. juli ser man generelt lave korrelasjonskoeffisienter. Der er også mange indekser som varierer fra høyde til høyde innenfor samme dato.

Tabell 5 Korrelasjonskoeffisient mellom hver enkelt indeks og registrert avling for hver høyde og hver dato.

	28.jul.22				01.aug.22				05.aug.22			
	20m	50m	70m	100m	20m	50m	70m	100m	20m	50m	70m	100m
Blue	0.07	-0.11	-0.18	-0.31	-0.06	-0.04	-0.05	-0.11	0.34	0.11	0.01	-0.07
Green	-0.12	-0.25	-0.24	-0.43	-0.22	-0.21	-0.20	-0.27	0.25	0.07	-0.02	-0.12
Red	-0.15	-0.26	-0.25	-0.41	-0.25	-0.26	-0.26	-0.30	-0.03	-0.18	-0.23	-0.30
RedEdge	-0.05	-0.28	-0.29	-0.47	-0.14	-0.17	-0.16	-0.26	0.50	0.31	0.22	0.08
NIR	0.14	-0.13	-0.15	-0.31	0.11	0.07	0.09	-0.01	0.65	0.52	0.54	0.37
NDVI	-0.05	0.16	-0.29	-0.47	0.23	0.23	0.24	0.23	0.42	0.41	0.44	0.46
DVI	0.15	-0.09	-0.11	-0.25	0.16	0.13	0.15	0.07	0.64	0.53	0.57	0.47
GDVI	0.15	-0.09	-0.12	-0.26	0.16	0.12	0.13	0.06	0.66	0.55	0.59	0.48
MTCI_CI	0.25	0.24	0.28	0.22	0.29	0.29	0.29	0.29	0.42	0.40	0.48	0.53
EXG	-0.14	-0.24	-0.14	-0.32	-0.16	-0.11	-0.08	-0.17	0.44	0.35	0.33	0.19
EXGR	-0.05	-0.12	0.02	-0.06	0.05	0.11	0.13	0.09	0.40	0.38	0.40	0.35
RDVI	0.16	-0.03	-0.05	-0.15	0.19	0.18	0.19	0.14	0.58	0.50	0.53	0.49
TDVI	0.15	-0.07	-0.09	-0.21	0.17	0.15	0.17	0.11	0.63	0.52	0.56	0.48
GNDVI	0.18	0.13	0.11	0.16	0.23	0.21	0.22	0.21	0.40	0.38	0.43	0.47
NDRE	0.24	0.22	0.24	0.22	0.28	0.27	0.28	0.27	0.42	0.42	0.48	0.52
SCCI	0.19	0.24	0.30	0.43	0.30	0.30	0.29	0.29	0.37	0.34	0.46	0.51
EVI	-0.15	0.03	0.04	0.12	-0.19	-0.17	-0.18	-0.14	-0.59	-0.51	-0.54	-0.51
TVI	0.15	-0.09	-0.11	-0.24	0.16	0.14	0.15	0.08	0.62	0.52	0.55	0.46
VARI	0.15	0.21	0.17	0.19	0.19	0.22	0.23	0.24	0.37	0.39	0.41	0.39
GARI	0.02	-0.28	-0.22	-0.43	-0.20	-0.26	-0.25	-0.33	-0.01	-0.19	-0.21	-0.30
GCI	0.19	0.13	0.12	0.16	0.22	0.19	0.20	0.19	0.40	0.38	0.42	0.46
GLI	-0.07	-0.04	-0.01	-0.08	0.07	0.12	0.14	0.14	0.35	0.38	0.40	0.37
NLI	0.16	0.06	0.04	-0.01	0.20	0.19	0.21	0.17	0.56	0.49	0.52	0.49
MNLI	0.16	-0.05	-0.08	-0.19	0.19	0.16	0.17	0.12	0.58	0.50	0.54	0.50
SAVI	0.16	-0.04	-0.06	-0.17	0.18	0.17	0.19	0.14	0.60	0.51	0.54	0.49
GSAVI	0.17	-0.05	-0.07	-0.18	0.18	0.16	0.17	0.11	0.64	0.54	0.58	0.52
OSAVI	0.17	0.02	-0.00	-0.06	0.20	0.20	0.21	0.18	0.54	0.48	0.50	0.49
GOSAVI	0.17	0.01	-0.02	-0.08	0.20	0.19	0.20	0.16	0.60	0.51	0.54	0.53
MSAVI2	0.15	-0.08	-0.10	-0.23	0.16	0.14	0.15	0.08	0.64	0.53	0.57	0.46
MSR	0.18	0.16	0.14	0.17	0.22	0.21	0.22	0.22	0.42	0.42	0.45	0.48
GRVI	0.19	0.13	0.12	0.16	0.22	0.19	0.20	0.19	0.40	0.38	0.42	0.46
WDRVI	0.18	0.16	0.14	0.17	0.23	0.22	0.23	0.23	0.42	0.42	0.45	0.48
SR	0.18	0.16	0.14	0.17	0.22	0.21	0.22	0.22	0.42	0.42	0.45	0.48

Den 01. august har generelt noe sterkere korrelasjon mellom avling og indekser. Indeksen med sterkst korrelasjon er GARI, som har en negativ korrelasjonskoeffisient på -0,33. Også for de to andre datoene er korrelasjonskoeffisienten negativ for GARI, med unntak ved 20 m den 28.juli (0,02).

4.5 Maskinlæring

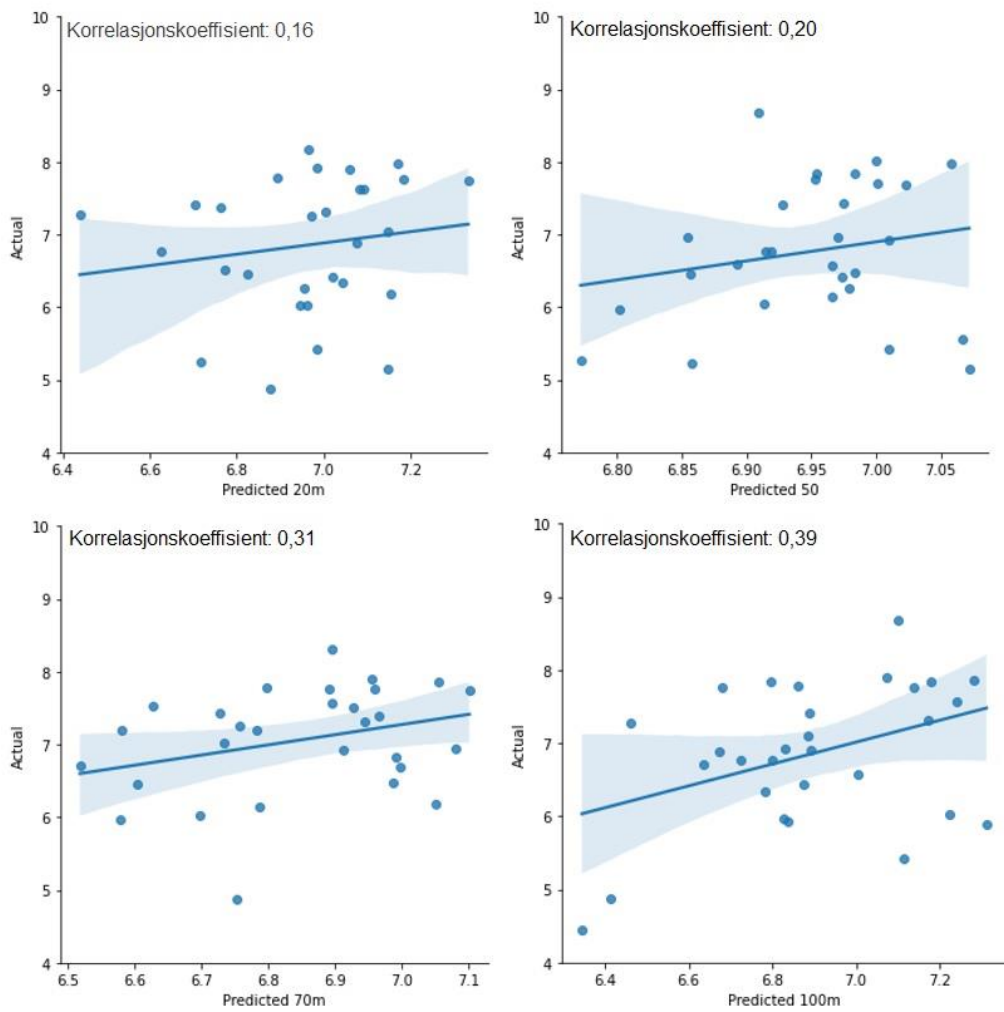
Data fra alle de forskjellige datoene og høydene ble brukt for å lage en avlingsprediksjon med LASSO regresjon i Python.

4.5.1 Prediksjonene

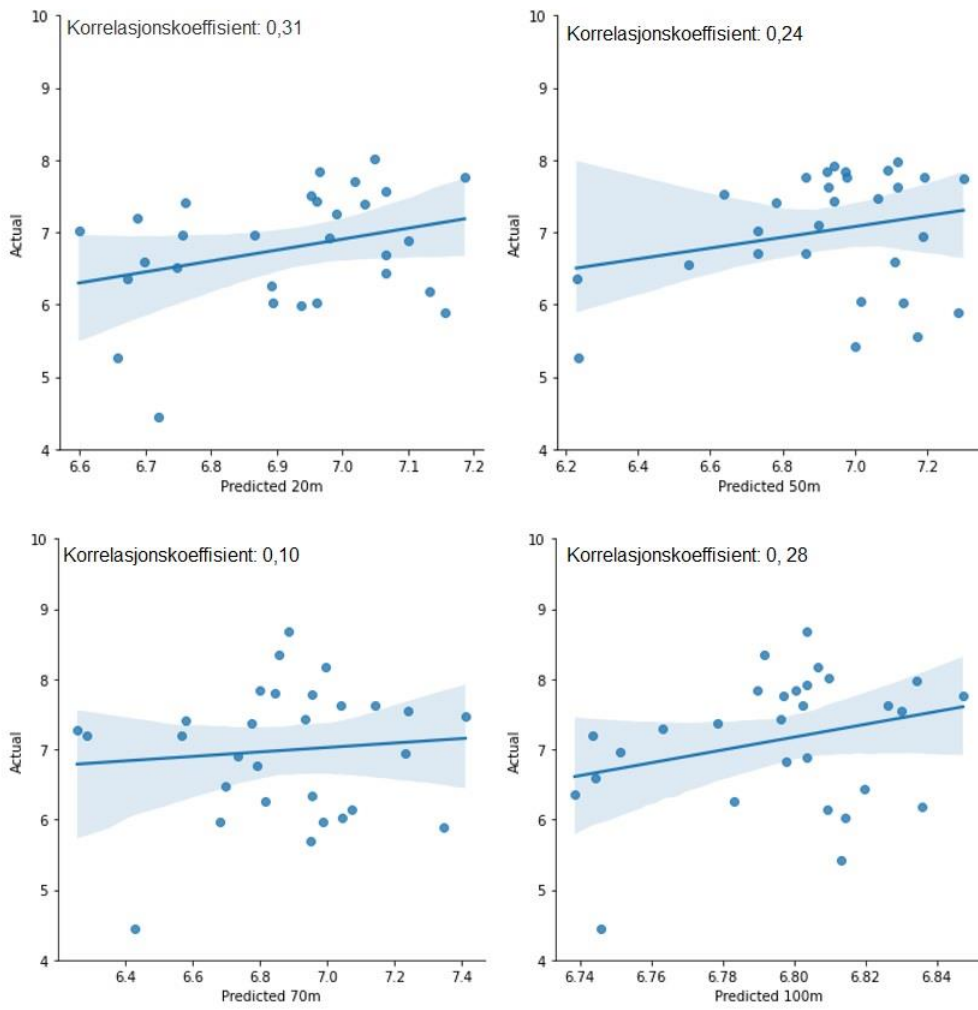
Resultatene viser ganske stor forskjell på korrelasjonen mellom prediksjon og registrert avling fra dato til dato. Den første datoen for flyvning ga lite nøyaktige prediksjoner (Figur 24). Her var det også størst forskjell mellom de forskjellige høydene (0,16-0,39) og korrelasjonen ble sterkere hvor hver økning av flyhøyde.

For den midtre datoen var forskjellen mellom høydene mindre, selv om korrelasjonskoeffisienten for 50 meters høyde 01.august var 0,10. For de andre høydene varierte det fra 0,24 til 0,31 og utviklingen i korrelasjonskoeffisienten var ikke parallell med økningen av høyde (Figur 25).

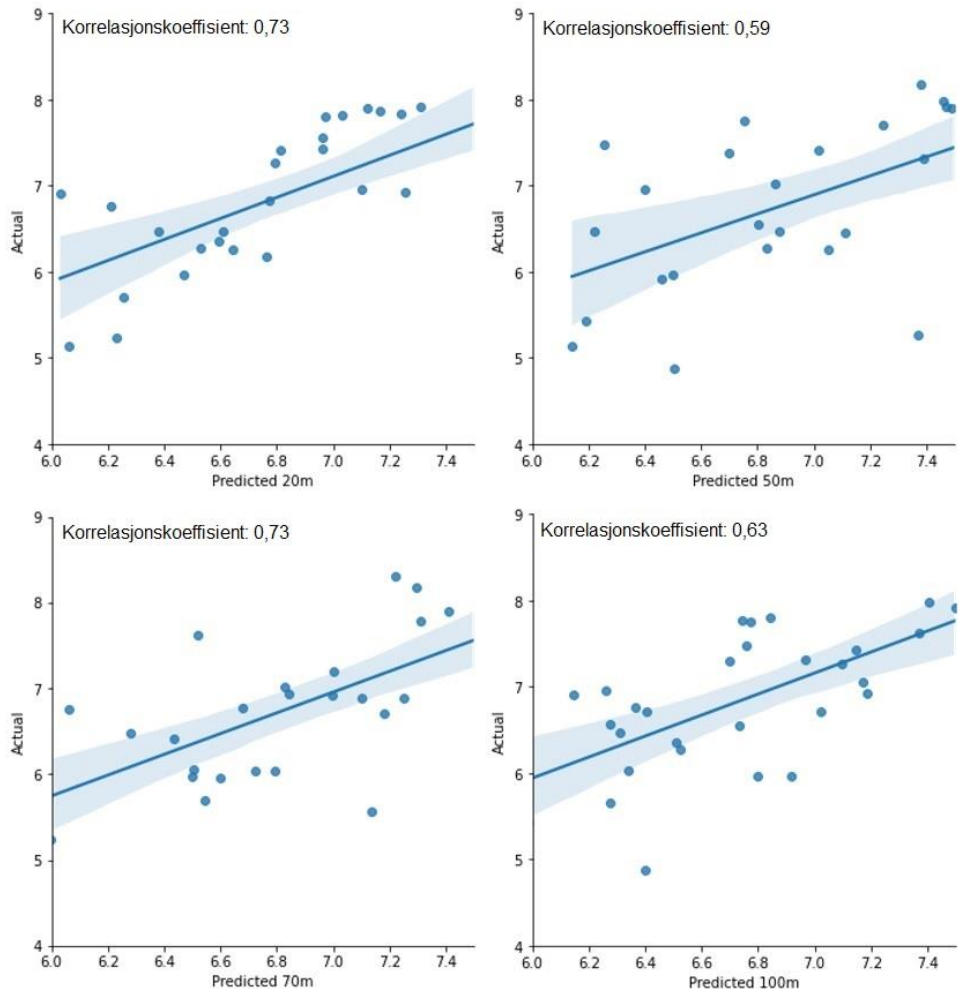
For den seneste datoen er korrelasjonen mellom den registrerte avlingsmengden og den predikerte avlingen sterkere enn for de andre datoene (0,59-0,73). Her er heller ikke utviklingen i korrelasjon parallell med økningen av flyhøyde (Figur 26).



Figur 24 Avlingsprediksjon mot faktisk avling basert på bilder fra forskjellige flyhøyde 28.07.2022.



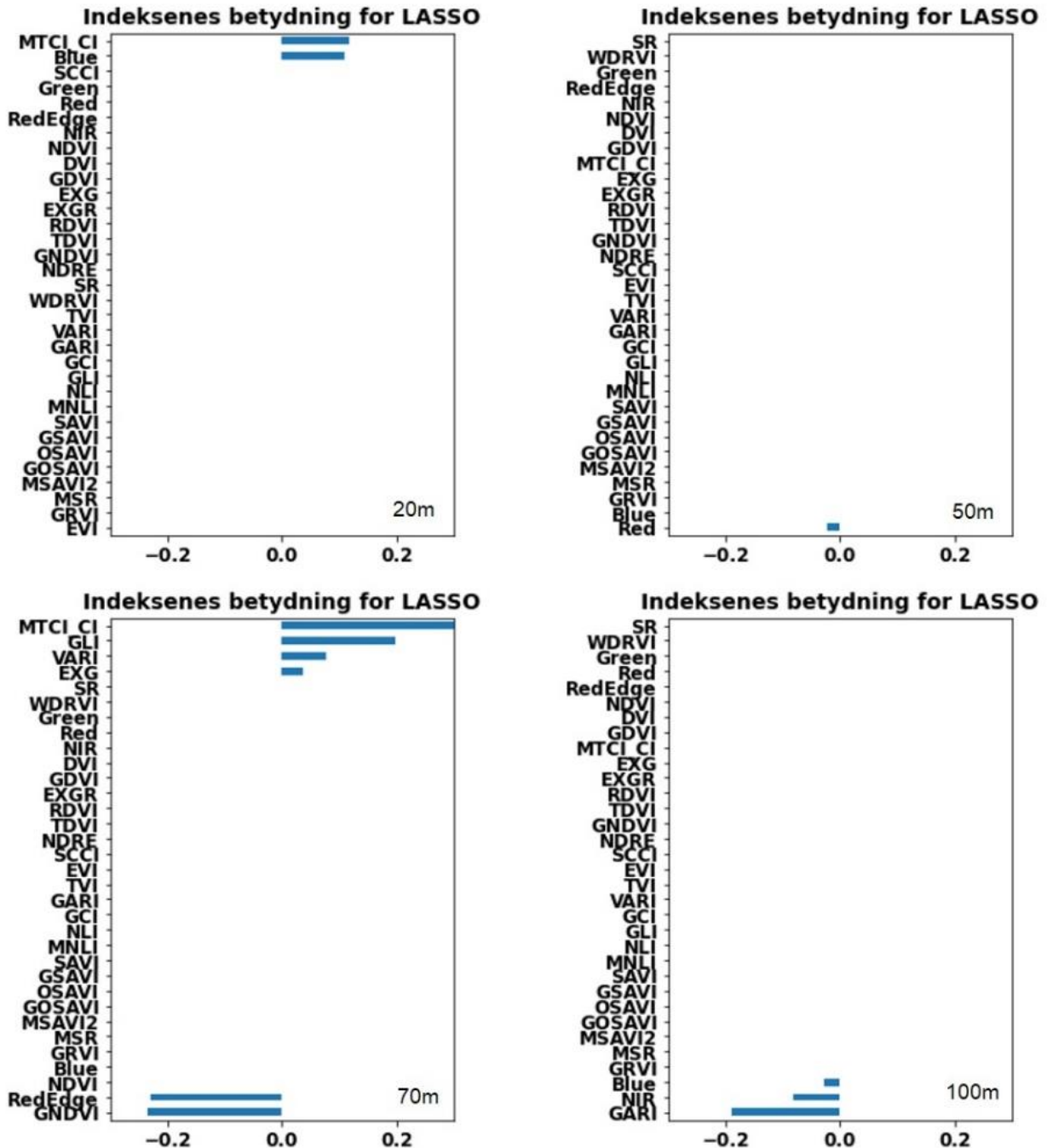
Figur 25 Avlingsprediksjon mot faktisk avling basert på bilder fra forskjellige flyhøyde 01.08.



Figur 26 Avlingsprediksjon mot faktisk avling basert på bilder fra forskjellige flyhøyde 05.08.

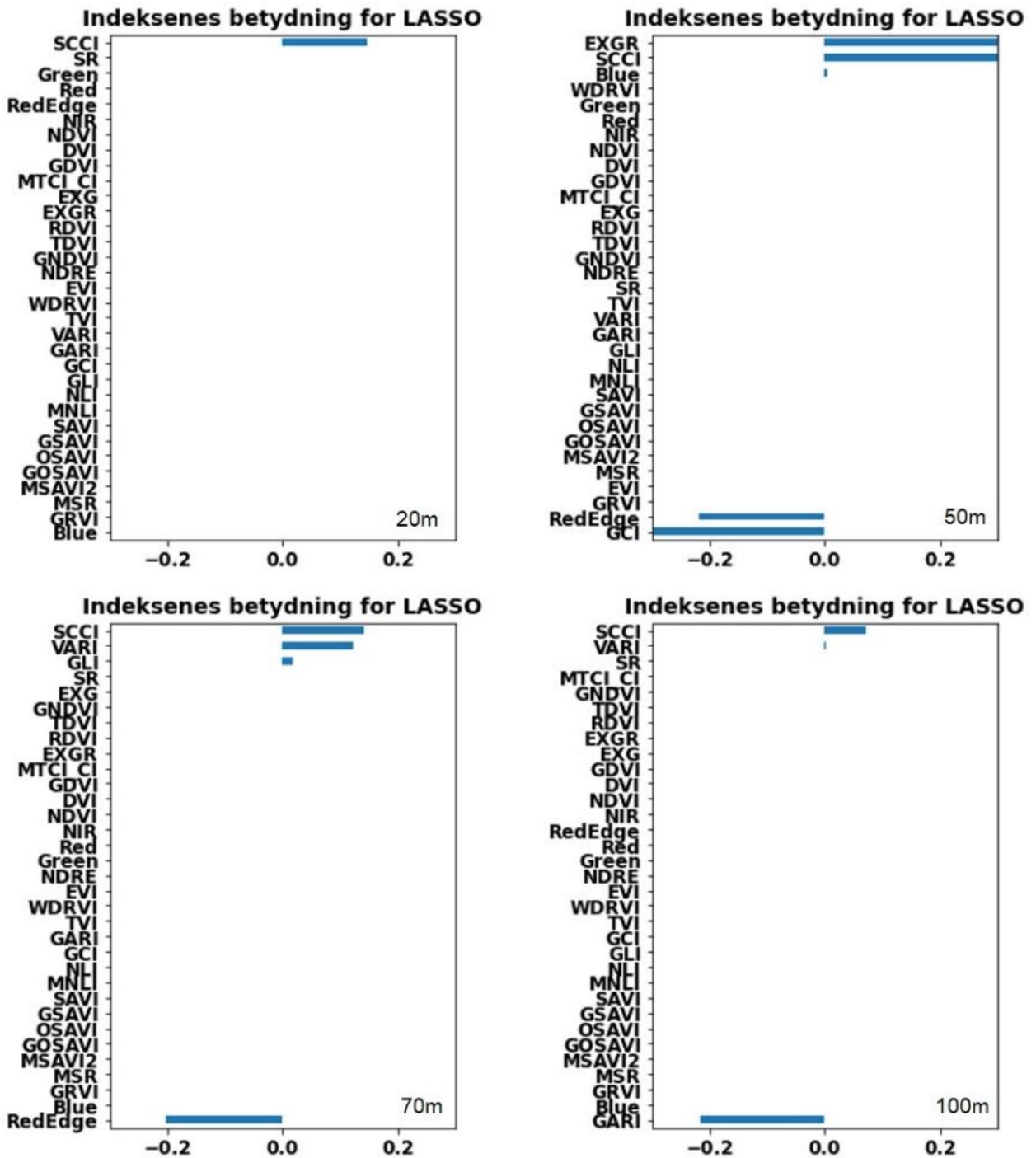
4.5.2 Indekser brukt i LASSO

Indeksene LASSO tok i bruk for å komme frem til sitt resultat ved de forskjellige tidspunktene og for de forskjellige høydene er vist i Figur 27, Figur 28, Figur 29.



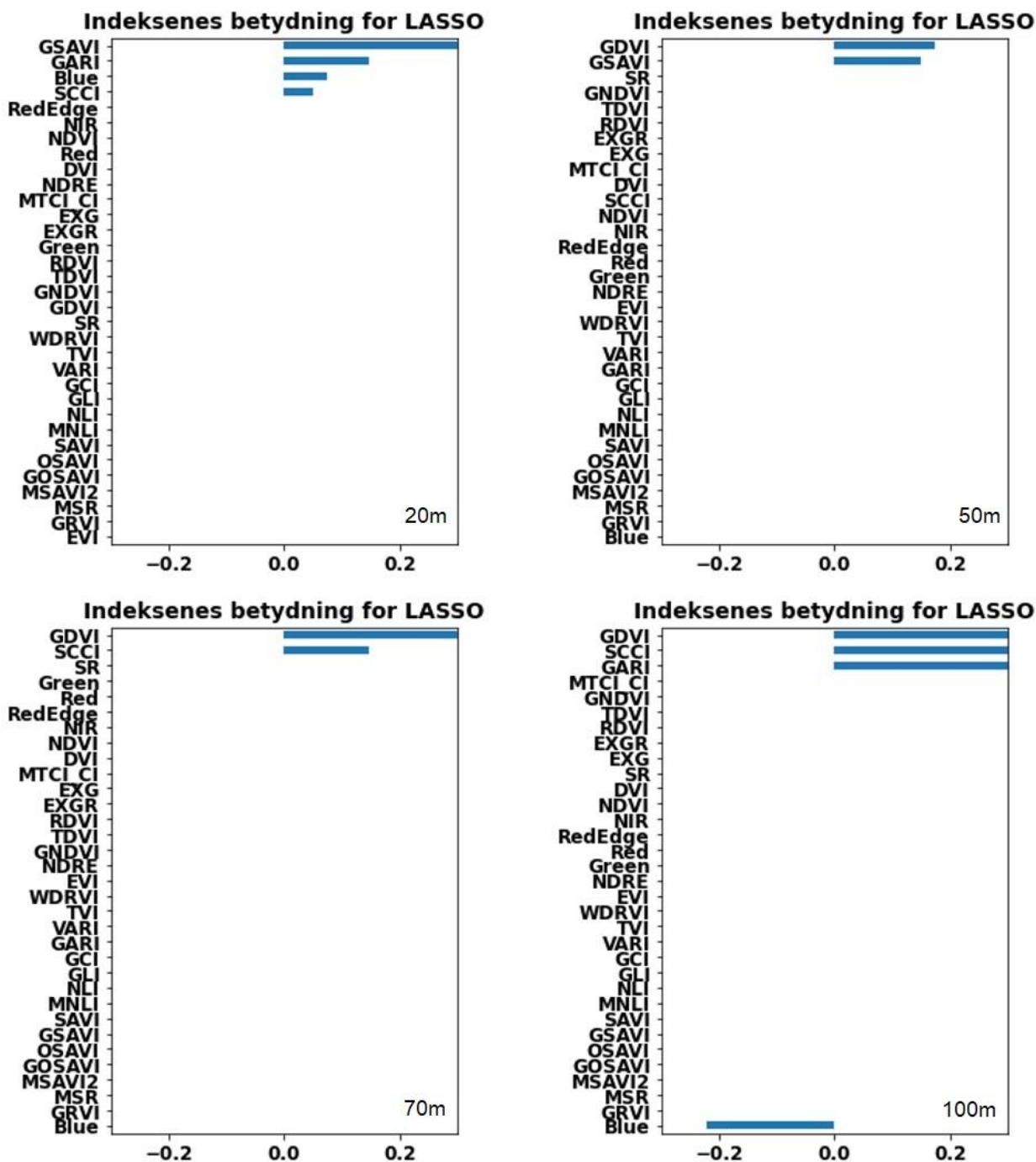
Figur 27 Indekser bruk i LASSO 28.07.2022.

For den første dagen for datainnsamling var MTCI viktig for LASSO for 20 og 70 meter. De to andre høydene var annerledes. For 50 meter ble bare Rød brukt og for 100 meter var GARI viktigst, mens NIR veide noe mindre, mens blå enda mindre.



Figur 28 Indekser bruk i LASSO 01.08.2022.

For den andre dagen for datainnsamling var SCCCI viktig for alle de fire høydene. For 50 meters høyde var EXGR enda viktigere enn SCCCI. RedEdge er også tungt vektet negativt for både 70 meter og 50 meter.



Figur 29 Indekser bruk i LASSO 05.08.2022

Den siste dagen for datainnsamling er GDVI den tyngst vektete indeksen for de tre høyeste høydene. For de to høyeste er SCCI nest tyngst vektet. SCCI er også noe vektet ved 20 meter, men ikke ved 50 meter. Blå er vektet negativt ved 100m.

Det ble brukt få indekser for hver høyde og dato, og antall indekser i modellen varierte fra 1 til 6.

4.6 Anvendt landbruk

Alle indeksene ble også regnet ut fra data fra anvendt landbruk ekstrahert i QGIS. Tabell 6 viser korrelasjonskoeffisienten mellom registrert avling og indeksene. Alle indeksene har en negativ korrelasjon til registrert avling utenom EVI (0,39) og VARI (0,18).

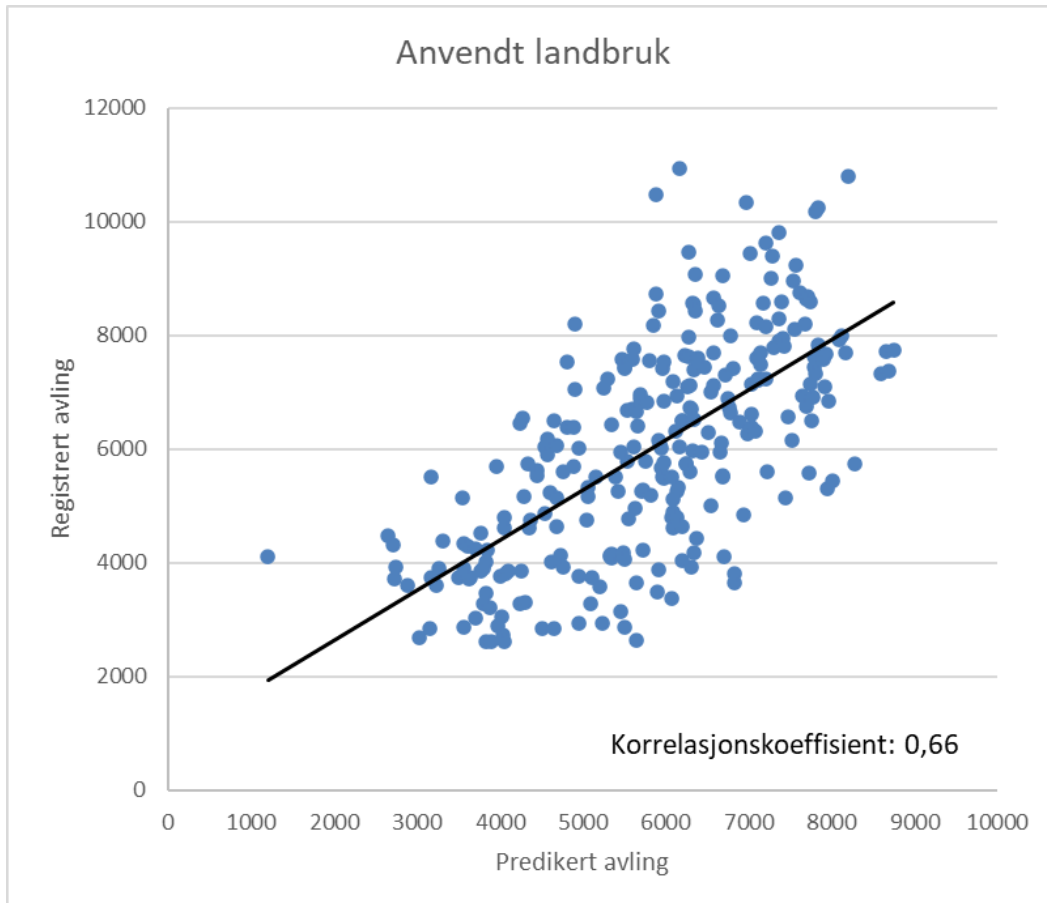
Den sterkeste korrelasjonen finner man for blå. Blå alene korrelerer altså sterkere med avling i dette tilfellet enn alle indeksene. Indeksen med sterkest korrelasjon er GARI (-0,52). Det er en sterkere korrelasjon for GARI i anvendt landbruk enn i forsøksfelt hvor den sterkeste målte korrelasjon var ved 100 meters flyhøyde den 28. juli (Tabell 5).

Den mest nøyaktige indeksen er mindre nøyaktig enn prediksjonen utført med maskinlæring (0,52, 0,66).

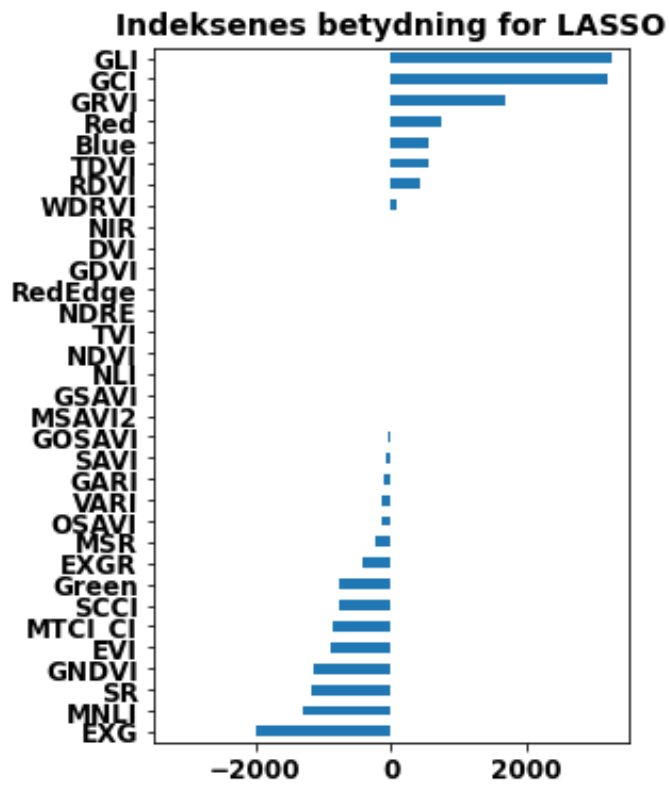
Tabell 6 Korrelasjonskoeffisienter for indekser sammenlignet med registrert avlingsmengde. Sortert etter sterkest korrelasjon.

Indekser	Tørrvekt
Blue	0.70
GARI	-0.52
EVI	0.39
NIR	-0.38
GDVI	-0.38
DVI	-0.37
TVI	-0.37
TDVI	-0.36
MNLI	-0.36
GSAVI	-0.36
MSAVI2	-0.36
RDVI	-0.34
SAVI	-0.34
GOSAVI	-0.33
RedEdge	-0.32
OSAVI	-0.31
GCI	-0.30
GRVI	-0.30
MTCI_CI	-0.28
GLI	-0.27
GNDVI	-0.26
NDRE	-0.26
SCCI	-0.26
NLI	-0.26
MSR	-0.21
WDRVI	-0.21
SR	-0.21
NDVI	-0.20
EXG	-0.20
VARI	0.18
EXGR	-0.13
Red	-0.06
Green	-0.05

Ved prediksjon i anvendt landbruk brukte vi også Lasso Regresjon i Python, men vi trente modellen med registrert avling fra avlingsregistrator på treskeren. Den predikerte avlingsmengden sammenlignet med den registrerte avlingen gir en korrelasjonskoeffisient på 0,66 (Figur 30). Denne korrelasjonen er sterkere enn ved prediksjon i forsøksfelt, men det er også et langt større datagrunnlag fra anvendt landbruk enn fra forsøksfelt.



Figur 30 Korrelasjon mellom predikert avling og registrert avling i anvendt landbruk.



Figur 31 Indeks bruk i LASSO for anvendt landbruk.

I anvendt landbruk ble GLI vektet tungt i LASSO. Det samme ble GCI – nesten like tungt. 23 indekser er gitt vekt i LASSO-modellens prediksjon (Figur 31).

5 Diskusjon

I dette kapitlet vil alt som er presentert i denne oppgaven diskuteres. Resultatene presentert i forrige kapittel vil diskuteres opp imot hverandre og mulige svakheter eller styrker i materialet eller metoden for oppgaven. Noen steder vil det også gjøres sammenligninger med andre forsøk gjort av andre.

5.1 Indekser

Den mest nøyaktige indeksen i anvendt landbruk var GARI. Denne korrelerte negativt med avlingsmengde med en korrelasjonskoeffisient på $-0,52$. Den neste var EVI på $0,39$. Dette er svakere enn de mest nøyaktige indeksene i forsøksfeltet for den siste datoen for flygning, men sterkere enn for de to første datoene.

GARI hadde den tredje sterkeste korrelasjonen med registrert avling når den var regnet ut fra 100 meter for den første datoen ($0,43$). For 20 meter samme dato var korrelasjonskoeffisienten bare $0,02$. For de andre datoene presterte den også bedre ved høyere flyhøyde. Dette tyder på at GARI er mest nøyaktig noe tidligere i sesongen og ved høyere flyvninger. Grunnen til dette er uviss, men det kan bety at GARI er mer nøyaktig med lav oppløsning.

Det kan generelt se ut som nøyaktigheten på resultatene fra LASSO og indeksene i grove trekk utvikler seg i takt med hverandre når det gjelder vekststadium. Både indeksene og LASSO presterer betydelig bedre på den siste datoen for datainnsamling. Det er sannsynlig at en betydelig andel av bladene har gulnet på dette punktet. Det er interessant at dette gjør både indeksene og maskinlæringsmodellen mer treffsikre. NDVI er unntaket når det gjelder dette, og presterer klart best den første datoen. Den første datoen er NDVI mest nøyaktig på 100 meters høyde. Dette er det samme som GARI var i anvendt landbruk. GARI har også den samme utviklingen hvor den presterer best den første dagen i forsøksfeltet.

5.1.1 NDVI

Et av målene for oppgaven var å se hvordan flyhøyde påvirket resultatet ved bruk av UAV for avlingsprediksjon. NDVI ble brukt som indeks grunnet dens popularitet både i anvendt landbruk og i forskning. Det ble observert små forskjeller mellom indeksen basert på de forskjellige høydene. Det var få dager, over relativt kort tid (8 dager). For å bedre kunne se utviklingen i NDVI burde vi hatt bilder fra tidligere i sesongen, og kanskje også med mindre tid mellom hver dag for datainnsamling.

5.1.2 NDVI med forskjellige gjødselmengder

Det var betydelig forskjell på korrelasjon mellom NDVI og avling ved høy og lav gjødselmengde. Det var også her stor forskjell mellom de forskjellige dagene for datainnsamling. For de to første dagene var korrelasjonen mellom registrert avling og NDVI negativ ved 15 kg N, mens den var positiv for 7,5 kg N. Det er også en viss forskjell på styrken i korrelasjonen, men denne forskjellen er av mindre størrelse. Forskjellene mellom flyhøydene er heller ikke stor, men korrelasjonen er generelt svakest for 20 m og sterkest for 50 m for de to første dagene, som stemmer med tidligere forsøk hvor 60 meter har vist seg som det mest nøyaktige [15]. Et forsøk på unge oljepalmer i Malaysia viste at den mest nøyaktige beregningen av NDVI var fra en flyhøyde på 60m, med 80m som nest mest nøyaktig og 20m som minst nøyaktig [15]. Et annet forsøk med mål om å finne de best mulige innstillingene for bruk av UAV i steds-spesifikk plantedyrking fløy også på 20, 60 og 80 meter. De kom også frem til at 60m var den høyden som ga mest nøyaktig beregning av NDVI [15, 24].

For den tredje dagen er det positiv korrelasjon for både lav og høy gjødselmengde. Ved lav gjødselmengde blir korrelasjonen sterkere i takt med økningen av flyhøyden (0,21-0,25-0,27-0,31). Ved høy gjødselmengde er det mindre variasjoner, og de følger ikke økningen i flyhøyde, som tyder på at de små variasjonene her kan være tilfeldige. På 100 meter er korrelasjonen ganske lik for de to gjødselmengdene (0,31 og 0,35). Det kan se ut som at NDVI er mer sårbar for høyde ved lave gjødselmengder. Dette kan henge sammen med at den er sårbar for forstyrrelser, som det kan komme mer av ved mindre plantedekke. Ved høy gjødselmengde er det naturlig å anta at det er tettere plantedekke, og her er den sterkeste korrelasjonen for både den siste og den midterste datoen sterkest ved flygning på 20 meters høyde.

Bakgrunnen for å velge NDVI som en «representativ» indeks var basert på dens popularitet. I forsøket viser den seg å være relativt jevn de to siste målingene. I motsetning til i andre forsøk

ser vi ikke nedgang i nøyaktigheten senere i sesongen hos NDVI i dette forsøket.

Nøyaktigheten går derimot noe opp de to siste dagene. Enkelte andre indekser har likevel en større økning på de siste dagene og kommer ut som mer nøyaktige enn NDVI. Alle datoene for datainnsamling er sent i vekstsesong, og det ville vært interessant å se utviklingen fra et tidligere stadium. Guan et.al så på korrelasjonen mellom NDVI med gjødselmengde og avling i ris og hvete [25]. De observerte at NDVI korrelerte ganske sterkt med avling ved tidlig reproduktivt stadium. De så også at NDVI var følsom for forskjellige gjødselmengder, som kan gjøre den aktuell å bruke for grunnlag for tildeling av gjødsel basert på N [25].

5.1.3 SCCCI

SCCCI har i andre forsøk vist seg som nøyaktig. I dette forsøket er SCCCI blant de mest nøyaktige indeksene for de to første dag for datainnsamling. For den første dagen var det generelt svake korrelasjoner. SCCCI er klart sterkest korrelert med registrert avling ved 100 meters høyde (0,43). Det er og generelt noe sterkere korrelasjoner for 100 meter enn for de andre høydene. SCCCI er også sterkest blant alle indekser og fargeområder for 70 meter, men korrelasjonen er svakere enn for 100 (0,30).

For den andre dagen for datainnsamling er SCCCI mest nøyaktig blant alle fargebånd og indekser for de to laveste høydene. Korrelasjonskoeffisienten er jevn mellom alle høydene (0,29-0,30). Det vil si at resultatene i dette forsøket tyder på at SCCCI blir mindre påvirket av flyhøyde i senere vekststadium. Også for SCCCI er det generelt høyere korrelasjon mellom indeksen og registrert avling for den siste datoen for datainnsamling (0,37-0,51), men den sterkeste korrelasjonen er for 100 meters høyde. Generelt viser SCCCI seg som en god indeks for de to første datoene mens det er flere indekser i alle høyder som er mer nøyaktige for den siste datoen. Det er interessant om det er SCCCI som underpresterer eller om de andre indeksene overpresterer på den siste datoen. Siden vi ikke har nøyaktige vekststadier i dette forsøket kan det også tenkes at den siste dagen for datainnsamling er på et så sent stadium at det påvirker resultatene negativt for SCCCI. At SCCCI utmerker seg som en nøyaktig indeks stemmer også med forsøk i mais fra Malaysia [18].

5.1.4 MTCI

Variasjonen i korrelasjonen mellom MTCI og avling er mindre enn for mange andre indekser mellom de forskjellige datoene og høydene (0,22-0,53).

Det er noe svakere korrelasjon den første dagen enn den andre, mens det, som for de fleste, er sterkest den siste dagen. Dette er det samme som for mange andre indekser, og også det samme som for LASSO. Den ser også ut til å bli relativt lite påvirket av flyhøyde.

Sammenlignet med de andre indeksene er MTCI altså relativt jevn mellom datoene for innsamling som også innebærer at den presterer bedre enn mange tidligere i sesongen. Det er tidligere gjort funn som tyder på at MTCI egner seg godt for å prediksjoner tidlig i sesongen, hvor det også er gjort prediksjoner tidligere enn i denne oppgaven [3]. Dette underbygger trenden i denne oppgavens resultater. MTCI har også vist seg å korrelere sterkere med avling enn NDVI i forsøk på høsthvete i Kina [26]. Dette viste også at MTCI fungerte best ved aksskyting, mens NDVI fungerte best i kornfyllingsfasen. Vi antar at datainnsamlingen i denne oppgaven skjedde i siste del av kornfyllingsfasen, og det stemmer godt med at NDVI presterer bedre på den første datoen for datainnsamling.

5.2 LASSO

LASSO-regresjon ble valgt for maskinlæringsdelen i oppgaven. Modellen har gjort det bra i tidligere forsøk og ble ansett som en reell modell for å sammenligne med vegetasjonsindeksene. Avlingsprediksjonen fra LASSO varierte stort fra dato til dato, men også fra høyde til høyde. For den tidligste datoen var korrelasjonen generelt noe svak, men den ble sterkere jo høyere man fløy. Dette kan tyde på at modellen så tidlig er sårbar for bilder med forstyrrelser, og at den på dette stadiet fungerer bedre med større avstand, og mindre detaljer. For de to senere datoene følger ikke korrelasjonen flyhøyden på samme måte. Det er noe variasjon, men man ser ikke et tydelig mønster og det er mulig at denne variasjonen er tilfeldig. Den siste datoen for flyging ga den klart mest nøyaktige prediksjonen. For både 20 og 100 meter var korrelasjonen 0,73. Dette er den sterkeste korrelasjonen i hele oppgaven og kan bety at LASSO får de beste resultatene ved senere stadier. Det tyder også på at den kan oppnå gode resultater både ved høye og lave flyhøyder. Som nevnt tidligere er trolig en del av bladmassen gul ved siste dato for datainnsamling.

LASSO ga ganske gode avlingsprediksjoner i anvendt landbruk. Korrelasjonskoeffisient på 0,66 er en av de sterkeste målt i hele oppgaven. Det var bare den siste datoen i forsøksfelt som hadde nøyaktighet i dette landskapet (0,59-0,73). Det er et langt større datagrunnlag for

anvendt landbruk enn i forsøksfeltene. Selv om LASSO straffer ekstremverdier vil ekstremverdier påvirke resultatet mer med et lite datagrunnlag enn med et stort. Funnene i denne oppgaven tyder på at LASSO gir bedre treffsikkerhet enn indeksene, men også her trengs det flere forsøk, helst med større datagrunnlag.

At maskinlæringsmodellen presterer ganske jevnt, men ikke bedre enn indeksene alene stemmer med et forsøk gjort i Japan. De sammenlignet to maskinlæringsmodeller med vegetasjonsindekser og konkluderte at maskinlæringsmodellene ikke klarte å utkonkurrere indeksene [27].

Indeksene LASSO brukte for å komme frem til sine resultater varierte. Det var et sprik mellom forsøksfelt og anvendt landbruk; både i hvilke indekser, og i antall indekser modellen la til grunn. I anvendt landbruk ble det brukt betydelig flere indekser (23) enn i forsøksfeltet (1-6). Dette kan skyldes at det er et langt større datagrunnlag for anvendt landbruk. Et større datagrunnlag kan tenkes å gjøre det lettere for modellen å finne bredere sammensatte sammenhenger.

NDVI ble ikke brukt som grunnlag i modellen for noen av datoene eller høydene. Dette skiller seg fra tidligere forsøk hvor NDVI har utmerket seg som den indeksen LASSO foretrekker [3]. I denne oppgaven er den første datoen for datoinnsamling 28. juli, mens den seneste datoen for innsamling i dette forsøket var 19. juli. I det samme forsøket gjorde NDVI det dårlig 02 juli [3]. Siden vi i denne oppgaven ser at NDVI er mest nøyaktig på den første datoen kan dette tyde på at NDVI er mer nøyaktig litt ut i sesongen, men taper seg igjen når det blir sent.

For de tre forskjellige datoene for datainnsamling i forsøksfelt var det litt forskjell på hvilke indekser som ble brukt. Variasjonen kan forklares med den tidligere omtalte forskjellen i vekststadier. Indeksene som ble brukt stemte ganske godt med indeksene som isolert korrelerte best med registrert avling.

5.2.1 Datagrunnlag

Problemer med dataene fra årets vekstsesong gjorde at data fra fjorårets sesong måtte tas i bruk. Fra i fjor fantes det bare data med registrert avlingsmengde fra én gård. Dette gir et lite grunnlag å konkludere på. Sammenligningsgrunnlaget for anvendt landbruk med forsøksfelt ville også blitt bedre med data fra samme vekstsesong. Det ville uansett vært en betydelig forskjell mellom forsøksfelt og anvendt landbruk. De to feltene ligger et stykke fra hverandre geografisk, og selv med ganske små avstander kan man ha ganske store klimatiske forskjeller på Østlandet i Norge. Dette både i form av temperatur, men også i form av lokal nedbør. Det er likevel sannsynlig at de store klimatiske linjene ville vært likere om dataene fra feltene var fra samme vekstsesong.

I anvendt landbruk var korrelasjonen mellom det blå lyset og registrert avling sterkere enn alle indeksene i anvendt landbruk. Korrelasjonskoeffisienten er 0,70 og er også sterkere enn ved LASSO. Friske planter med høyt innhold av klorofyll absorberer mye blått lys [9]. Hvor mye blått som reflekteres kan derfor være en indikator på klorofyllinnhold. Det er verdt å merke seg at både rød og grønn korrelerer svært svakt med registrert avling (-0,06 og -0,05). En plante med høyt klorofyllinnhold reflekterer mye grønnfarge, så dette kan være et tegn på at plantenes bladverk begynner å gulne. Siden vi kun har et skifte som bakgrunn kan vi ikke konkludere med noe ut fra dette funnet, og det hele kan dreie seg om tilfeldigheter. Det er likevel et interessant funn som vil være interessant å undersøke videre.

Det er viktig å være bevisst på at avlingsvariasjonene i forsøksfelt og anvendt landbruk ikke nødvendigvis vil være sammenlignbare. I anvendt landbruk identifiserer man variasjon innad i en sort, mens i forsøksfelt identifiserer man variasjon mellom sorter, med forskjellig gjødselmengde.

5.3 Forslag til videre arbeid

5.3.1 Optimalisering av N-indeks

Det vil være interessant å finne ut mer om de forskjellige indeksenes reaksjon på forskjellig tildeling av N. God kunnskap om dette er viktig for å kunne gjøre gode vurderinger av hvilke indekser som egner seg til å kartlegge variasjon i N-behov i åkeren. For norsk landbruk er variabel tildeling av gjødsel veldig aktuelt for flere og flere. For å kunne bruke funn som råd i anvendt landbruk trengs det forsøk hvor man ser på indeksenes prestasjoner tidligere i sesongen – når det er aktuelt å gjødsle. Flere forsøk i anvendt landbruk vil være svært interessant for den praktiske delen av norsk landbruk.

5.3.2 Sammenligning av maskinlæringsmodeller og -innstillinger

I denne oppgaven har vi kun sett på resultater fra LASSO-regresjon med standard-verdier. Det kan se ut som modellen blir i overkant sårbar ved lite datagrunnlag. Det vil både være interessant å se hvordan LASSO fungerer med andre parametere, for eksempel med en høyere lambda, og et større datagrunnlag med flere gjentak. Det vil også være interessant å se LASSO sammenlignet med andre maskinlæringsmodeller. Mer diskusjon om dette kommer. Jeg må skrive mer om dette i teoridelen først.

6 Konklusjon

Flyhøyde påvirker prestasjonen til indeksene. Det virker som om noen indekser er mer følsom for forstyrrelser og presterer bedre med høyere flyhøyde, mens andre presterer bedre ved lavere høyder. Maskinlæringsmodellen fungerte bedre ved høyere flyhøyder den første datoen for datainnsamling, men virker å være mindre avhengig av høyde ved de to seneste.

Det var store forskjeller på nøyaktigheten i prediksjonene mellom datoene. Det var også forskjellige datoer som presterte godt på forskjellige datoer. Også maskinlæringsmodellen presterte bedre senere i sesongen enn tidligere.

Det var en noe høyere treffsikkerhet for indeksene ved høye gjødselmengder enn ved lave.

Maskinlæringsmodellen var generelt på nivå med eller litt bedre enn vegetasjonsindeksene alene.

7 Referanser

1. Nations, F.a.A.O.o.t.U. *The world is at a critical juncture*. 2021 [cited 2022 15.12]; Available from: <https://www.fao.org/state-of-food-security-nutrition/2021/en/>.
2. Britt Lande, L.J., Thea Nørgaard Breili *Utviklingen i norsk kosthold 2022*. 2022: helsedirektoratet.no.
3. Shafiee, S., et al., *Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery*. *Computers and electronics in agriculture*, 2021. **183**: p. 106036.
4. Graminor. *Hvete*. u.d [cited 2022 15.12]; Available from: https://www.helsedirektoratet.no/rapporter/utviklingen-i-norsk-kosthold/Utviklingen%20i%20norsk%20kosthold%202022%20-%20Kortversjon.pdf/_attachment/inline/b8079b0a-fefe-4627-8e96-bd979c061555:2c2d7308d81f9c591365dacdcf94e6c5d55e5fd1/Utviklingen%20i%20norsk%20kosthold%202022%20-%20Kortversjon.pdf.
5. Pandey, N.L., *Baking quality in wheat : effect of delayed harvest, cultivars, growing conditions and nitrogen fertilization*. 2014, Norwegian University of Life Sciences, Ås.
6. Walter, A., F. Liebisch, and A. Hund, *Plant phenotyping: from bean weighing to image analysis*. *Plant Methods*, 2015. **11**(1): p. 14-14.
7. Graminor. *Om Graminor*.
8. Atlin, G.N., J.E. Cairns, and B. Das, *Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change*. *Glob Food Sec*, 2017. **12**: p. 31-37.
9. University, H.S. *Spectral Reflectance*. u.å; Available from: http://gsp.humboldt.edu/olm/Courses/GSP_216/lessons/reflectance.html.
10. Leksikon, S.N. *Bølgelengde*. 2021 [cited 2022 02.12]; Available from: <https://snl.no/b%C3%B8lgelengde>.
11. NASA. *The Electromagnetic Spectrum 2013* [cited 2022 15.12]; Available from: <https://imagine.gsfc.nasa.gov/science/toolbox/emspectrum1.html>.
12. Malcolm Davidson, M.B., Ismael Moya, José Moreno, Tuomas Laurila, Marc-Phillipe Stoll, John Miller *Photosynthesis from Space - a new vegetation-fluorescence technique*. 2003.
13. Alshihabi, O., K. Piikki, and M. Söderström. *CropSAT – A Decision Support System for Practical Use of Satellite Images in Precision Agriculture*. Cham: Cham: Springer International Publishing.
14. AtFarm. *Biomasse/avlingsovervåking*. 2022 [cited 2022 04.12]; Available from: https://support.at.farm/hc/no/articles/360019194959-Hvordan-kan-jeg-bruke-biomassekart-for-%C3%A5-overv%C3%A5ke-avlingsutviklingen-#h_01FVMQ05RBS8K3BJZSNYSQJMP1.
15. Avtar, R., et al., *Assessing the Influence of UAV Altitude on Extracted Biophysical Parameters of Young Oil Palm*. *Remote sensing (Basel, Switzerland)*, 2020. **12**(18): p. 3030.

16. Mróz, T., J.A. Dieseth, and M. Lillemo, *Historical grain yield genetic gains in Norwegian spring wheat under contrasting fertilization regimes*. *Crop science*, 2022. **62**(3): p. 997-1010.
17. DJI. *DJI GS PRO*. 2022 [cited 2022 12.12]; Available from: <https://www.dji.com/no/ground-station-pro>.
18. Barzin, R., et al., *Use of UAS Multispectral Imagery at Different Physiological Stages for Yield Prediction and Input Resource Optimization in Corn*. *Remote sensing (Basel, Switzerland)*, 2020. **12**(15): p. 2392.
19. Feng, W., et al., *An optimized non-linear vegetation index for estimating leaf area index in winter wheat*. *Precision agriculture*, 2019. **20**(6): p. 1157-1176.
20. Yang, W., et al., *Greenness identification based on HSV decision tree*. *Information processing in agriculture*, 2015. **2**(3-4): p. 149-160.
21. Boyd, D.S., et al., *Phenology of vegetation in Southern England from Envisat MERIS terrestrial chlorophyll index (MTCI) data*. *International journal of remote sensing*, 2011. **32**(23): p. 8421-8447.
22. Sun, Q., et al., *Improving the Retrieval of Crop Canopy Chlorophyll Content Using Vegetation Index Combinations*. *Remote sensing (Basel, Switzerland)*, 2021. **13**(3): p. 470.
23. Pedregosa, F., et al., *Scikit-learn: Machine Learning in Python*. 2012.
24. Mesas-Carrascosa, F.-J., et al., *Assessing Optimal Flight Parameters for Generating Accurate Multispectral Orthomosaics by UAV to Support Site-Specific Crop Management*. *Remote sensing (Basel, Switzerland)*, 2015. **7**(10): p. 12793-12814.
25. Guan, S., et al., *Assessing Correlation of High-Resolution NDVI with Fertilizer Application Level and Yield of Rice and Wheat Crops using Small UAVs*. *Remote sensing (Basel, Switzerland)*, 2019. **11**(2): p. 112.
26. Zhang, S. and L. Liu, *potential of the MERIS Terrestrial Chlorophyll Index for crop yield prediction*. *Remote sensing letters*, 2014. **5**(8): p. 733-742.
27. Zhou, X., et al., *Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches*. *Plant production science*, 2021. **24**(2): p. 137-151.

8 Vedlegg

8.1 Kode for LASSO

```
In [10]: import numpy as np
import pandas as pd

In [13]: data1=pd.read_excel('Knut0-600.xlsx')
data2=pd.read_excel('Knut0-600.xlsx')
Y=pd.DataFrame(data2, columns=['Y1d_Mass_D'])/1000
Y1=pd.DataFrame(data2, columns=['Y1d_Mass_W'])/1000

In [17]: #####MTCI_CI is the calculated MTCI based on median value of plots
#####MTCI is the extracted from pix4D
#####indices definition
DF['DVI']=DF['NIR']-DF['Red']
DF['GDVI']=DF['NIR']-DF['Green']
DF['MTCI_CI']=(DF['NIR']-DF['RedEdge'])/(DF['RedEdge']-DF['Red'])
DF['EXG']=(2*DF['Green'])-DF['Red']-DF['Blue']
DF['EXGR']=(3*DF['Green'])-(2.4*DF['Red'])-DF['Blue']

import math
from math import sqrt
DF['RDVI']=(DF['NIR']-DF['Red'])/np.sqrt(DF[['NIR','Red']].sum(axis=1))
DF['TDVI']=1.5*(DF['NIR']-DF['Red'])/np.sqrt((np.power(DF['NIR'],2)+DF['Red']+0.5))
DF['GNDVI']=(DF['NIR']-DF['Green'])/(DF['NIR']+DF['Green'])
DF['NDRE']=(DF['NIR']-DF['RedEdge'])/(DF['NIR']+DF['RedEdge'])
DF['SCCI']=DF['NDRE']/DF['NDVI']
DF['EVI']=2.5*(DF['NIR']-DF['Red'])/(DF['NIR']-6*(DF['Red'])-(7.5*DF['Blue'])-1)
DF['TVI']=0.5*(120*(DF['NIR']-DF['Green'])-200*(DF['Red']-DF['Green']))
DF['VARI']=(DF['Green']-DF['Red'])/(DF['Green']+DF['Red']-DF['Blue'])
DF['GARI']=(DF['NIR']-DF['Green'])-(1.7*(DF['Blue']-DF['Red']))/(DF['NIR']+DF['Green'])-(1.7*(DF['Blue']-DF['Red']))
DF['GCI']=(DF['NIR'])/DF['Green']-1
DF['GLI']=(DF['Green']-DF['Red']-DF['Blue'])/(2*DF['Green']+DF['Red']+DF['Blue'])
DF['NLI']=(np.power(DF['NIR'],2)-DF['Red'])/(np.power(DF['NIR'],2)+DF['Red'])
DF['MNLI']=(np.power(DF['NIR'],2)-DF['Red'])*1.5/(np.power(DF['NIR'],2)+DF['Red']+0.5)
DF['SAVI']=((DF['NIR']-DF['Red'])*1.5)/(DF['NIR']+DF['Red']+0.5)
DF['GSAVI']=(DF['NIR']-DF['Green'])*1.5/(DF['NIR']+DF['Green']+0.5)
DF['OSAVI']=(DF['NIR']-DF['Red'])/(DF['NIR']+DF['Red']+0.16)
DF['GOSAVI']=(DF['NIR']-DF['Green'])/(DF['NIR']+DF['Green']+0.16)
DF['MSAVI2']=(2*DF['NIR'])+1-np.sqrt(np.power((2*DF['NIR']+1),2)-8*(DF['NIR']-DF['Red'])))2
DF['MSR']=(DF['NIR']/DF['Red'])-(1/np.sqrt(DF['NIR']/DF['Red']))
DF['GRVI']=(DF['NIR']/DF['Green'])
DF['WDRVI']=(0.1*DF['NIR']-DF['Red'])/(0.1*DF['NIR']+DF['Red'])
DF['SR']=(DF['NIR']/DF['Red'])

In [18]: X=pd.DataFrame(DF, columns=['Blue', 'Green', 'Red', 'RedEdge', 'NIR',
'NDVI', 'DVI', 'GDVI', 'MTCI_CI', 'EXG', 'EXGR', 'RDVI', 'TDVI',
'GNDVI', 'NDRE', 'SCCI', 'EVI', 'TVI', 'VARI', 'GARI', 'GCI', 'GLI',
'NLI', 'MNLI', 'SAVI', 'GSAVI', 'OSAVI', 'GOSAVI', 'MSAVI2', 'MSR',
'GRVI', 'WDRVI', 'SR'])
Y=pd.DataFrame(data2, columns=['Y1d_Mass_D'])
Y1=pd.DataFrame(data2, columns=['Y1d_Mass_W'])
X.to_excel("Out.xlsx")

In [19]: from sklearn.preprocessing import StandardScaler
names=X.columns
scaler=StandardScaler()
scaled_df = scaler.fit_transform(X)
scaled_df = pd.DataFrame(scaled_df, columns=names)

from sklearn.model_selection import train_test_split
X_train, X_test,y_train, y_test = train_test_split(
scaled_df, Y1, test_size=0.30)

y_test.to_excel('y_test.xlsx')

In [18]: X=pd.DataFrame(DF, columns=['Blue', 'Green', 'Red', 'RedEdge', 'NIR',
'NDVI', 'DVI', 'GDVI', 'MTCI_CI', 'EXG', 'EXGR', 'RDVI', 'TDVI',
'GNDVI', 'NDRE', 'SCCI', 'EVI', 'TVI', 'VARI', 'GARI', 'GCI', 'GLI',
'NLI', 'MNLI', 'SAVI', 'GSAVI', 'OSAVI', 'GOSAVI', 'MSAVI2', 'MSR',
'GRVI', 'WDRVI', 'SR'])
Y=pd.DataFrame(data2, columns=['Y1d_Mass_D'])
Y1=pd.DataFrame(data2, columns=['Y1d_Mass_W'])
X.to_excel("Out.xlsx")
```

Figur 32 Vedlegg 1

```
In [19]: from sklearn.preprocessing import StandardScaler
names=X.columns
scaler=StandardScaler()
scaled_df = scaler.fit_transform(X)
scaled_df = pd.DataFrame(scaled_df, columns=names)

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
    scaled_df, Y1, test_size=0.30)

y_test.to_excel('y_test.xlsx')
```

In []:

```
In [20]: from sklearn.cross_decomposition import PLSRegression
pls2 = PLSRegression(n_components=3)
pls2.fit(X_train, y_train)

y_train_pred=pls2.predict(X_train)
y_test_pred=pls2.predict(X_test)
# b = y_test.values.ravel()

from matplotlib.pyplot import figure

figure(figsize=(8, 6), dpi=80)
plt.scatter(pls2.predict(X_test), y_test)

plt.xlabel('Predicted grain yield')
plt.ylabel('Measured grain yield')
```

Figur 33 Vedlegg 2

```
In [20]: Text(0, 0.5, 'Measured grain yield')
```

```
In [21]: from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
print('MSE train: %.3f, test: %.3f' % (mean_squared_error(y_train, y_train_pred), mean_squared_error(y_test, y_test_pred)))
print('MAE train: %.3f, test: %.3f' % (mean_absolute_error(y_train, y_train_pred), mean_absolute_error(y_test, y_test_pred)))
from sklearn.metrics import r2_score
print('R^2 train: %.3f, test: %.3f' % (r2_score(y_train, y_train_pred), r2_score(y_test, y_test_pred)))
```

```
MSE train: 2810947.428, test: 2468995.589
MAE train: 1235.648, test: 1216.593
R^2 train: 0.447, test: 0.379
```

```
In [ ]:
```

```
In [22]: Y1=Y1.values.ravel()
```

```
In [29]: ###overfitting solution with LassoCV
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    scaled_df, Y1, test_size=0.30)

from sklearn.linear_model import LassoCV
reg = LassoCV(cv=3).fit(X_train, y_train)
reg.score(X_train, y_train)
coef = pd.Series(reg.coef_, index = X_train.columns)
imp_coef = coef.sort_values()
import matplotlib
matplotlib.rcParams['figure.figsize'] = (4, 6.0)
imp_coef.plot(x='y', kind = "barh", fontsize=12)
plt.rcParams['font.serif']='Times New Roman'
plt.rcParams['font.size']='14'
plt.rcParams['font.weight']='bold'

plt.title("Feature importance using Lasso Model", fontsize=14, fontweight='bold')
plt.fill(True)
plt.grid(False)
coef.to_excel('coef.xlsx')

y_train_pred=reg.predict(X_train)
y_test_pred =reg.predict(X_test)
#y_train_pred=clf_sgd.predict(X_train)
#y_test_pred=clf_sgd.predict(X_test)
#B =df.y.values.ravel()
# f, axs = plt.subplots(1,2,figsize=(15,15))
# plt.subplot(221)
figure(figsize=(8, 6), dpi=80)
plt.scatter(y_test_pred , y_test)
plt.xlabel('Predicted grain yield(ton/hect)', fontsize=12)
plt.ylabel('Measured grain yield (ton/hect)', fontsize=12)
plt.tick_params(axis='y', which='major', labelsize=12, pad=2)
# plt.plot(y_test, y_test, '-',color='red')
# plt.scatter(y_train_pred, y_train_pred - y_train,c='steelblue', marker='o', edgecolor='white', label='Training data')
# plt.scatter(y_test_pred, y_test_pred - y_test, c='limegreen', marker='s', edgecolor='white', label='Test data')
# plt.xlabel('Predicted values (ton/hect)', fontsize=12)
# plt.ylabel('Residuals', fontsize=12)
# plt.tick_params(axis='y', which='major', labelsize=12, pad=2)
# plt.legend(loc='upper left')
# plt.hlines(y=0, xmin=0, xmax=7.5, color='black', lw=5)
# plt.xlim([1,7.5])
plt.show()

from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
print('MSE train: %.3f, test: %.3f' % (mean_squared_error(y_train, y_train_pred), mean_squared_error(y_test, y_test_pred)))
print('MAE train: %.3f, test: %.3f' % (mean_absolute_error(y_train, y_train_pred), mean_absolute_error(y_test, y_test_pred)))
from sklearn.metrics import r2_score
print('R^2 train: %.3f, test: %.3f' % (r2_score(y_train, y_train_pred), r2_score(y_test, y_test_pred)))
print("Best alpha using built-in LassoCV: %f" % reg.alpha_)
#print("Best score using built-in LassoCV: %f" % reg.score(df_X, df_y))
# coef = pd.Series(reg.coef_, index = X_train.columns)
# ###Plotting feature importance using Lasso Model

# imp_coef = coef.sort_values()
# import matplotlib
# matplotlib.rcParams['figure.figsize'] = (5, 100.0)
# imp_coef.plot(kind = "barh")
# plt.title("Feature importance using Lasso Model")
# plt.fill(False)
```

Figur 34 Vedlegg 3



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

Postboks 5003
NO-1432 Ås
Norway