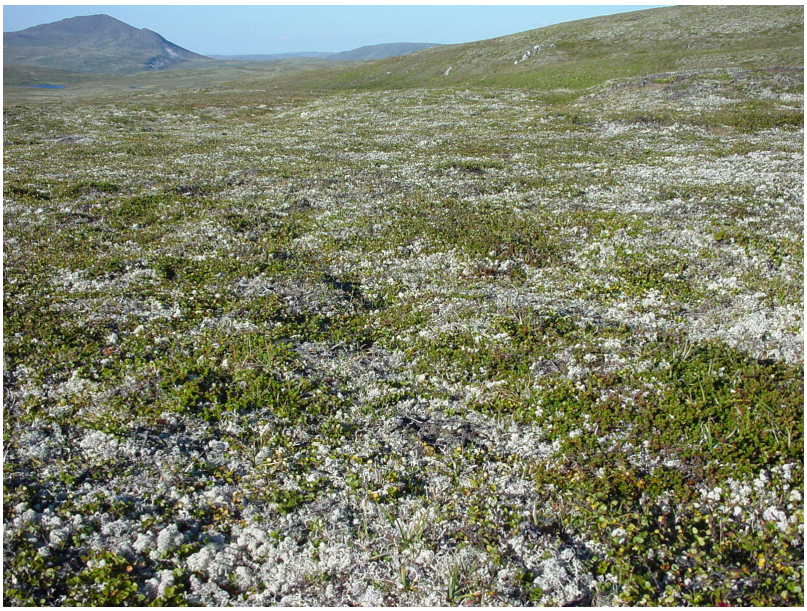
NORGES LANDBRUKSHØGSKOLE
Institutt for kartfag

Reinbeite kartlegging på subpiksel nivå

Subpixel reindeer pasturage mapping



Hovedfagsoppgave

av

Lars Ole U. Engebakken

Institutt for kartfag
Norges landbrukshøgskole
Postboks 5034, 1432 ÅS

Hovedfagsoppgave 2003:2

Forord

Denne hovedoppgaven markerer avslutningen på mitt 5- årige sivilingeniørstudium ved Norges Landbrukshøgskole, Institutt for Kartfag.

Jeg vil takke Leif Kastdalen, prosjekt leder for SatNat prosjektet for at jeg fikk lov til å skrive for de og hjelp underveis. Professor Øystein B. Dick ved Institutt for Kartfag for faglig hjelp underveis.

Jeg vil samtidig rette en stor takk til datakonsulent Gunnar Fjone for programmering av nødvendige skript underveis. Jeg vil også takk Einar Lieng for hjelp og innputt underveis.

For hjelp i slutt fasen vil jeg takke Lars Krangnes for gode tips og råd. Samtidig vil jeg rette en stor takk til klassekamerat Jostein Bjerkan som med sine språkkunnskaper loste meg gjennom utformingen av et "Abstract".

Tilsutt men ikke minst vil jeg takke klassa for fem "røtte" trivelige år, med mye fest og moro.

Ås, NLH 2.juni 2.juni 2003

Lars Ole U. Engebakken

Bilde på forsiden er tatt
av undertegnede på Forelhogna
august 2002

Sammendrag

Denne hovedoppgaven ser på muligheten for å kartlegge reinens tilgang på lavbeite ved hjelp av en sub-piksel tilnærming av Landsat 7 data.

Bakgrunnen for hovedoppgaven er å se på muligheten for å effektivisere dagens metoder for reinbeite kartlegging. I dag foregår det meste av kartleggingen manuelt ved at alle registreringer skjer ved feltarbeid. For å erstatte den manuelle innsamlingen over store områder finnes det ikke så mange andre alternativer enn å se på bruken av satellittbilder.

For at en kartlegging ut fra bare et satellittbilde skal bli bra, er en avhengig av å finne en metode som klarer å identifisere de pikslene som inneholder lav i satellittbildet. For å samle inn treningsdata er det ønskelig å se på egnetheten til flybilder. Disse treningsdataene skal brukes til å bygge regresjonsmodeller som kan prediker forekomster av lav i et Landsat 7 bilde.

I tillegg til å se på om det er mulig å predikere lav ut fra piksel verdiene i et Landsat 7 bilde er det ønskelig å se på om andre faktorer kan være med på å påvirke hvor det er mulig å detektere lav.

Flybildene som skulle brukes som trenningsdata i byggingen av regresjonsmodellene ble først klassifisert og tolket i ERDAS Imagine 8.6 (heretter omtalt som Erdas). For at vi skulle kunne foreta en subpiksel klassifisering måte jeg finne andelen lav som lå innenfor et Landsat 7 piksel. Dette uttrekket ble gjort ArcView og oppsummeringen av andelen ble gjort ved hjelp av en serie skript som foretok ulike filbehandlings opperasjoner. Filene som kom ut av dette kunne brukes til å bygge regresjons modeller ved hjelp av et regresjonstre i programmet Cubist. Disse regresjonsmodellene kun til slutt brukes til å predikere lav på et Landsat 7 bilde sammen med tilleggsdataene høydemodell, vegetasjonsgeografiske soner og varmeindeks.

Resultatet av undersøkelsene viser at en god prediksjon av lavmengden på et Landsat 7 bilde, er avhengig av gode regresjons modeller. Disse modellene må kunne skille mellom og ta hensyn til mange forskjellige kombinasjoner av data, som kan påvise hvor en finner lav. For å få til gode modeller er en igjen avhengig av å ha gode treningsdata. Til anskaffelse av treningsdata er flybilder en egnet kilde, men for at de skal være egnete er en avhengig av å ha foretatt en sikker og god tolkning av bildene før bruk. En feiltolkning i flybildene vil forplante seg til en feiltolkning i satellittbildene.

Konklusjonen er at flybilder kan være egnet til bruk i innsamlingen av treningsdata for tolkning av Landsat 7 bilder, men at det kreves gode bakkesannheter for å etablere en kunnskap om hvordan bildene skal tolkes. Regresjonsmodellene som skal tolke Landsat 7 data må også forbedres slik de unngår å predikere lav der det ikke er lav.

Abstract

This thesis is looking into the possibilities to map reindeer accessibility to lichen by a sub-pixel approximation and Landsat 7 data.

The background for this thesis is to look for possibilities to increase the efficiency of the method for reindeer pasturage mapping. Today most of the mapping is done manually with field registration. To replace the manual collecting for large areas, the use of satellite images is the preferred alternative.

To be able to perform good mapping from a satellite image, it is crucial to find a method that manage to identify the pixels containing lichen. For collection of training data it is desirable to look for suitability in air photos as well. This training data is the starting point for building the regression model to predict the occurrence of lichen in a Landsat 7 image.

In addition to predict occurrence of lichen from the pixel value in a Landsat 7 image, it is of interest to evaluate if other factors (e.g. heights, vegetation and heat) may affect detection of lichen.

The air photos, used as training data, were at first classified and interpreted using ERDAS Imagine. To carry out the sub-pixel classification, the amount of lichen inside each Landsat 7 pixel had to be found. This was done using ArcView, while the amount of lichen was found using different operational scripts. The results of these computations were used to build the regression models by a regression tree in Cubist. To predict lichen in the Landsat 7 image, these regression models were used along with auxiliary data such as an elevation model, geographical vegetation zones and heat index.

The result from the evaluation shows that a good regression model is crucial to get a quality prediction for lichen in a Landsat 7 image. These models have to be separated according to the variety of data combinations that can prove findings of lichen. To be able to get the models needed, good training data is necessary. A suitable source in getting these training data is air photos, but these need to be fully interpreted first. Misinterpretation of air photo will lead to further errors in the satellite images.

The conclusion is that it is a possibility to use Landsat 7 images for mapping lichen, but the methods still need further testing and improvements before they can be used for operational applications.

Innhold

Forord 1
Sammendrag 2
Abstract 3
Innhold 4
1 Innledning 6
1.1 Hypotese 7
1.1.1 Spesifisering 7
2 Bakgrunn 8
2.1 Treningsdata 8
2.1.1 Innsamlingsmetoder 8
2.2 Klassifisering 9
2.2.1 Styrt klassifisering 9
2.2.2 Ikke styrt klassifisering 9
2.3 Subpiksel klassifisering 10
2.3.1 3 metoder for subpiksel klassifisering 11
2.4 Regresjonstre og Beslutningstre 12
2.4.1 Algoritme 13
3 Valg av strategi for predikering av lav 15
4 Datagrunnlag 16
4.1 Satellittbilder 16
4.1.1 Landsat 7 ETM+ 16
4.1.2 SPOT 5 17
4.2 Flybilder 17
4.2.1 DuncanTech MS3100 17
4.2.2 Kodak DCS-760 17
4.2.3 IR-ortofoto 17
4.3 Andre data 17
4.3.1 Høydemodell 17
4.3.2 Vegetasjonsgeografiske soner 17
4.3.3 Varmeindeks 18
5 Studieområde 19
6 Metoder 21
6.1 Feltarbeid 21
6.2 Geometrisk korreksjon 22
6.3 Klassifisering/Identifisering 22
6.4 Bildeanalyse 23
6.4.1 Forberedning 23
6.4.2 ArcView 23
6.4.3 Analyse 24
7 Resultater 26
7.1 Tilrettelegging 26
7.1.1 Feltarbeid 26
7.1.2 Geometrisk korreksjon 26
7.1.3 Klassifisering/Identifisering 26
7.2 Analyse 27
7.2.1 Regresjons modeller 27
7.2.1.1 Resultat for regresjons modellene 27

7.2.1.2 Resultat for testdata 28
Sammenligning predikert og reell lavmengde 28
7.3 Bildeanalyse 31
7.3.1 Duncan 31
7.3.2 Kodak 32
7.3.3 IR-bilde 33
8 Vurdering av metodene/Diskusjon 34
8.1 Geometrisk korreksjon 34
8.1.1 Duncan 34
8.1.2 Kodak 35
8.2 Klassifisering 35
8.2.1 Valg av klasser 35
8.2.1.1 Skygger 37
8.3 Vurdering av metoden 37
8.3.1 Regresjons modellene 37
8.3.1.1 Testdata 37
8.3.2 Bilde analyse 38
8.3.3 Forbedringer i metoden 40
9 Konklusjon 41
10 Litteratur: 42
Vedlegg 44
Vedlegg A 44
Vedlegg B 46
Vedlegg C 47
Vedlegg D 50
Vedlegg E 51
Vedlegg F 52

1 Innledning

Denne hovedoppgaven er knyttet opp mot SatNat prosjektet [10] , [12], som er et samarbeidsprosjekt mellom Direktoratet for Naturforvaltning (DN) og Norsk Romsenter (NRS). Dette prosjektet går på å fremme bruken av satellittdata i forvaltningen av naturverdier, blant annet til bruk i innsamlingen av vegetasjons data for større områder. I tillegg til bruk av satellitt data har det blitt gjort forsøk på bruken av små formats kamera i fly for innsamling av bakkedata. I denne oppgaven er fokuset rettet mot kartlegging av villreinbeite, med Hardangervidda som testfelt.

På Hardangervidda har beiteområdene vært utsatt for stor beitebelastning over lengre tid, noe som har resultert i overbeite og at store deler av lavmassen er i ferd med å forsvine. En grunn til at laven har fått så store skader er at antallet beitedyr er for stort i forhold til beitegrunnlaget. I slike situasjoner må bestanden av beitedyr reguleres slik at presset på beitegrunnlaget ikke blir for stort.

I dagens samfunn blir det stadig færre områder som ikke er påvirket av menneskets utbredelse. Mennesker ferdes nå over alt, og foretar utbygginger på stadig nye steder. I et område som Hardangervidda merkes denne ferdselen gjennom turist trafikk, hytte og vei utbygging. Den økte utbyggingen fører mange steder til at reinen får avskåret sine gamle trekkruter, noe som fører til at de ikke får byttet fra for eksempel sommer til vinter beite. Begrenses reines mulighet til å trekke til andre områder, kan presset på et allerede hardt presset område øke yterligere. Øker beitet, reduseres også vegetasjonens mulighet til å reprodusere seg, noe som vil føre til at beitegrunnlaget forsvinner.

For at det skal være mulig å styre en slik forvaltning er man avhengig av å skaffe tilveie data om hvordan tilstanden til vegetasjonen er, antallet dyr i området og andre faktorer som kan påvirke dyrenes situasjon. I denne sammenhengen blir det fort store mengder data som må skaffes til veie.

Slik det er i dag må mye av dette arbeidet gjøres manuelt ved feltarbeid. Et arbeid som er både tid- og kostnadskrevende. Det skyldes at det må gjøres målinger og analyser over store områder for å få gode nok data til å kunne si noe om den generelle situasjonen til et området og lokale variasjoner. For å begrense bruken av feltarbeidet til innsamling av data, men samtidig sikre tilgangen til store mengder referansedata har det blitt sett på muligheten for å samle inn referansedata ved hjelp av flybilder. Bruken av flyfoto gjør det mulig å dekke større områder med mindre feltarbeid. Men selv om kostnadene reduseres ved redusert feltarbeid, vil kostnadene bli relativt høye ved innkjøp av flyfoto. Siden det er begrenset hvor stort område et flybilde dekker, trengs det mange bilder for å dekke et større område. Det er i denne sammenhengen man i dag har begynt å se på muligheten for å bruke satellittbilder i kartleggingen av store områder. Satellittbilder har blitt et reelt alternativ i dag siden

kostnadene ved innkjøp av slik opptak har blitt sterkt redusert, samt at tilgjengeligheten øker og oppløsningen blir bedre.
Om satellittbilder er egnet til å erstatte flyfoto vil være avhengig av brukerens krave til nøyaktighet og oppløsning på bildene. Det finnes i dag satellittbilder med en oppløsning på ned til 0.68 meter. Som i vårt tilfelle med Hardangervidda har det blitt brukt to Landsat 7 bilder for å få dekt hele området.
En utfordring som ligger i å bruke satellittdata er at et piksel vil kunne dekke relativt store områder i forhold til et flybildepiksel. Noe som øker muligheten for at satellittbildepikselen vil kunne dekke en flate bestående av forskjellige vegetasjonstyper. Hardangervidda er en slik heterogen flate som er sammen satt av mange forskjellige vegetasjonstyper flettet inn i hverandre. Dette gjør at det må finnes en metode som kan skille ut og beregne andelen av en vegetasjonstype innenfor et område tilsvarende et satellittbildepiksel.

Ved bruke av satellittbilder er man avhengig av å ha referansedata som kan bekrefte det som ses i satellittbildet. I stedet for å benytte seg av feltarbeid, kan det brukes enkelte flybilder eller ortofoto med muligheter for en oppløsning ned til noen cm. Flybilder blir i dag brukt til mange andre formål slik at tilgjengeligheten stadig øker. Dette gjør at det er mulig å slippe å gå til anskaffelse av helt nye flybilder, men i stedet benytte allerede eksisterende bilder. I tillegg til at gamle flybilder blir gjort tilgjengelig på digital form blir også mange gamle kart overfør til den digitale verden. Dette gjør det mulig å trekke inn andre data for å se om de kan lette tolkningen av satellittbildene og om de kan forklare varianser i bildet.

1.1 Hypotese

Problemstillingen for denne oppgaven kan i korthet oppsummeres som:

1. Reinens tilbud av lavbeite kan bli kartlagt ved bruk av en sub-piksel tilnærming av Landsat 7 data.
2. Flyfoto er egnet som treningsdata for en sub-piksel kartlegging.

1.1.1 Spesifisering

For hypotese 2 er det valgt å ta med en liten tilleggs test: Det er valgt å teste tre forskjellige flybilde typer DuncanTech, Kodak og skannede infrarøde (IR) bilder, for å se om noen av de er bedre enn andre til bruk som grunnlag.

I tillegg til å se på ulike kameraer har det blitt valgt å se på noen flere forhold som kan påvirke pikselverdien i bildet. De dataene det er valgt å se på om kan ha noe og si for hvilken pikselverdi vi får er høyden over havet (terrengmodell), vegetasjonssoner og varmekoeffisient.

2 Bakgrunn

2.1 Treningsdata

For at det skal være mulig å foreta en god kartlegging er en avhengig å kjenne til hvordan forholdene virkelig er der kartleggingen skal finne sted. Det trengs derfor en form for fasit eller treningsdata når andre og mer effektive metoder å kartlegge på skal vurderes. For at denne fasiten skal være best mulig må den bestå av store mengder data som representerer de mulige hendelsene innenfor området. Lokaliseringen til dataene må også være bestemt presist slik at de blir koblet mot riktige data i testgrunnlaget. Tanken bak det å endre innsamlingen av treningsdata er ønsket om å finne en effektiv metode å kartlegge et område på, uten å måte gjøre "manuell" innsamling for hele området. Innsamlingen av trenings data kan skje på flere forskjellige måter, og det er et felt hvor det stadig prøves ut nye og mer effektive metoder. Det finnes per i dag ingen metode som det er mulig å omtale som best egnet for alle tilfeller, slik at det i hvert enkelt tilfelle må vurderes hvilken metode som passer best.

2.1.1 Innsamlingsmetoder

Den tradisjonelle måten å samle inn treningsdata er ved at innsamlingen gjøres ved feltarbeid, mens et alternativ til denne "gode gamle" måten kan være bruk av bilder. Fordelen med bruk av bilder er at det er mulig å effektivisere innsamlingen av treningsdata ved at tidsbruken på feltarbeid reduseres og samtidig som at det er mulig å samle inn treningsdata fra et større område.

En måte å gjøre dette på er å merke opp en eller flere ruter i terrenget som kartlegges i detalj. Hvis rutene stedfestes kan data som blir funnet i terrenget kobles med bildedata fra samme sted. En form av dette benyttet Gaare og Tømmervik [6] seg av i overvåkning av lavbeite i Finnmark. De merket opp flere felter i terrenget hvor de foretok en detaljert registrering. I rutene registrerte de blant annet mengden av de forskjellige vegetasjonstypene slik at de kunne regne ut prosentandelen lav i ruta. Ved hjelp av denne prosentandelen kunne de finne ut hvor stor andelen lav som måtte til for å påvirke pikselverdiene i satellittbilde. I tillegg til rene feltobservasjoner ble det tatt bilder med vanlige kamera for analyse. Disse fotografiene ble klassifisert på en tilsvarende måtte som fly og satellittbilder. Målet med dette var å oppnå vegetasjonsinformasjon på samme måte som ved feltarbeid og en subjektiv vurdering av hva som ses.

Dietz og Steinlein [2] har også gjort en vurdering på om bilder er egnet til å klassifisere planter. De fotograferte en serie med ruter bestående av varierende plante dekning. Disse bildene ble så analysert. Resultatet av bildetolkningen ble sammenlignet med målinger gjort med et "Leaf Area Meter". Resultatet fra disse analysene viste at fotografier kunne brukes til fortrinnsvis å kartlegge planter som følger bakken.

Det er ikke bare åpne områder, som fjell og jorder, som kan kartlegges ved bruk av fly- og satellittbilder, skogsområder kan også kartlegges. I USA har Huang, C. m.fl. [7] sett på bruken av digitale ortofoto som treningsdata for statistiske modeller til bruk i analyse av Landsat 7 bilder for kartlegging av skogtetthet. Resultatet av disse undersøkelsen konkluderte med at ortofoto var egnet som referansekilde for en modellbygging.
Nordberg og Hemström [14] har gjort et forsøk der de så på muligheten for å registrere spor etter menneskelig aktivitet i form av veier og stier i fjellet ved hjelp av satellittbilder. I forsøket ble flybilder brukt som referansedata for kontroll av satellittbildene.

I stede for kun å se på det visuelle er det også mulig å se på de spektrale egenskapene til det som er å finne på bakken. For å kunne gjenkjenne hva som finnes på bakken må refleksjonsevnen og refleksjonen fra forskjellige vegetasjonstyper og andre objekter på bakken kjennes.

2.2 Klassifisering

Klassifisering er i hovedsak delt i to typer: styrt og ikke styrt klassifisering. Hvilken metode en velger å bruke vil være avhengig av hva som fins av treningsdata.

2.2.1 Styrt klassifisering

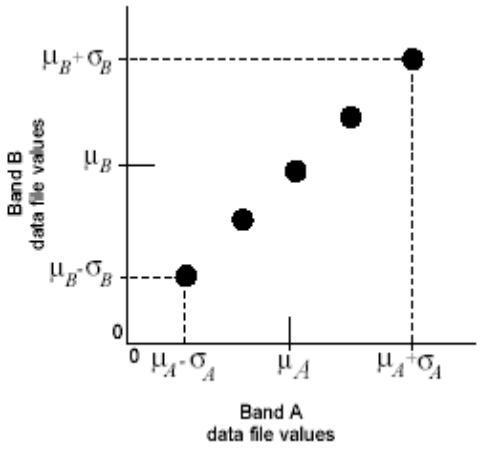
I en styrt klassifisering er det operatøren som setter betingelsene for hvordan inndelingen av klassene skal skje. Betingelsen for en klasse kan defineres ved å trekke opp et polygon rundt et område av interesse i et bilde, et slik område kan være et vann. Siden området inneholder piksler med en spektralsignatur vil disse signaturene være med på å danne referansegrunnlaget til den klassen som polygonet representere. Når en har fått samlet inn treningsdata for de ønskede klassene, kan selve klassifiseringen kjøres. I klassifiseringen vil hvert enkelt piksel bli testet for om det hører til en klasse eller en annen. Denne testingen skjer ved hjelp av en tilordningsmetode, som for eksempel”Maximum Likelihood” eller ”Minimum Distance” [15].

2.2.2 Ikke styrt klassifisering

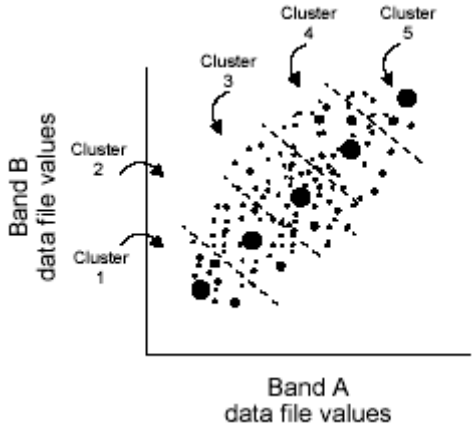
Ved en ikke styrt klassifisering overlater operatøren store deler av arbeidet med inndelingen i klasser til programmet. Det en trenger å gjøre er å gi inn hvor mange klasser det er ønskelig at bildet skal deles inn i, så står en valgt algoritmen for utvelgelsen av de tilhørende clustrene. Et eksempel på en slik algoritme er ISODATA (The Iterative Self-Organizing Data Analysis Technique) [5],[15].

ISODATA algoritmen bruker den spektrale avstanden for å tilordne pikslene til et cluster. For å finne ut hvor clustrene skal plasseres beregnes det en serie med gjennomsnittsverdier for spektraleverdiene som dekker hele datasettet. De beregnede clustersentrene brukes som start verdier i den første iterasjonen (Figur 2.1). Ved hver iterering beregnes det nye clustersentre med utgangspunkt i verdiene fra forrige iterasjon (Figur 2.2). Iterasjonsprosessen gjentas helt

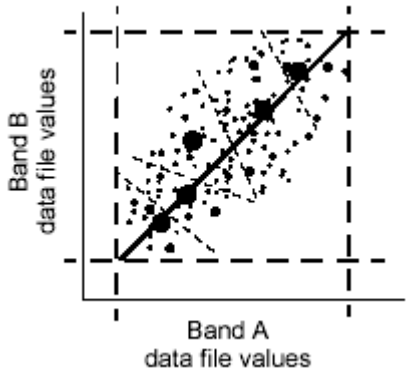
til en maksimal forbedringsgrad mellom to iterasjonene er nådd, eller et på forhånd bestemt antall iterasjoner (Figur 2.3).



Figur 2.1 Gjennomsnittlige spektralverdier ved start



Figur 2.2 Clustersentrum etter en iterasjon



Figur 2.3 Clusetersentrumene etter endt iterasjon

Det er viktig at det på forhånd bestemmes en grense for hvor mange iterasjoner som skal kjøres ellers så kan algoritmen risikere å blir gående i det uendelige. Dette gjelder selv ved få clustre. En grunn til at det er viktig å sette en grense for hvor mange iterasjoner som tillates, er at alle pikslene må kontrolleres mot alle clustrene ved hver nye iterasjon, for å finne ut hvilket cluster de ligger nærmest. Denne kontrollen er med på å sikre enn best mulig oppdeling av datasettet. Etter at clustringen er ferdig bør clustrene vurdering for å se om de kan inneholde ”for” få punkter til at de vil gi noen ny eller nødvendig informasjon. Men det kan også ha skjedd at to clustre har blitt liggende så nær hverandre at de ikke vil gi noe ny informasjon, men i stede fungerer som en unødig oppdeling [15].

2.3 Subpiksel klassifisering

Subpiksel klassifisering er en analyseform som har blitt mer og mer utbredt i de siste årene, men så langt har denne metoden vært lite brukt i Norge. Det er ingen fast metoder å forta en subpiksel klassifisering på. Det finnes derfor flere forskjellige metoder å foreta en slik klassifisering på.

Tanken bak subpiksel klassifisering er ønsket om å foretar en bildeanalyse der det trekkes ut informasjon fra pikslene som har en høyere detaljeringsgrad enn oppløsningen til selve pikslene. For at det skal være mulig å trekke ut slik informasjon er en avhengig av å vite hva det letes etter. For å få en oversikt over dette trengs det en "fasit" som sier en hvordan den søkte informasjonen oppfører seg under ulike forhold eller hvilke egenskaper de kan ha. Et eksempel på dette er spektralsignaturen til en plante. Denne "fasiten" omtales også ofte som treningsdata.

Felles for de fleste av metodene for subpiksel klassifisering er at pikslene enten tilordnes en type eller klasse, eller at det blir beregnet en andel for hvor mye det er av forekomsten i et piksel eller hvor stor sannsynlighet det er for å finne en forekomst i et piksel.

2.3.1 3 metoder for subpiksel klassifisering

Som det er nevnt tidligere finnes det flere måter å foreta en subpiksel klassifisering på. Tre måter å foreta en subpiksel klassifisering som er knyttet opp mot bruk av ERDAS Imagine 8.6 (heretter omtalt som Erdas)er:

1) Denne metoden er nærmere beskrevet i Erdas [4]. Resultatet av denne metoden er et bilde som er klassifisert ut fra hvor sannsynlig det er for at et bestemt piksel tilhørerer en bestemt klasse. For å komme fram til dette må det først lages et grid som tilsvarer pikslene i det lav oppløselige bilde. Dette gridet kan så legges over et høyoppløselig bilde. En kan så velge ut enkelte celler fra gridet som det er ønskelig å foreta en enda bedre spesifisering av. Deretter legges et punkt grid inne i de utvalgte cellene, for så og tilordne punktene til område av interesse eller ikke. Etter at alle punktene er bestemt kan analysen kjøres og til slutt få ut det tidligere nevnte "sannsynlighetsbildet".

2) "Imagine Subpixel Classifier" er en tilleggsmodul tilgjengelig for Erdas.
Denne metoden går på analyse av den spektrale signaturen i piksler. Pikslene blir klassifisert etter hvor mye de inneholder av en forekomst. Metoden går dypere i vurderingen av de forskjellige pikslene ved at den er i stand til å skille ut en forekomst av interesse selv om den søkte forekomsten ikke er enerådende innenfor et piksel, i et multispektralt bilde. Det er ved denne metoden mulig å finne piksler som inneholder ned til 20% av arealet med den søkte forekomsten. For at spektral signaturen til den søkte forekomsten skal bli funnet, trengs det et eller flere pikseler som inneholder den søkte signaturen, eller referanse signaturen til det søkte forekomsten (MOI-materia of intrest). Programmet gir beskjed om funn når avstanden mellom referanse signaturen og residualspektrum er innenfor en toleranse grense. Mens tradisjonell klassifisering kombinerer alle signaturene i treningssettet får "Imagine Subpixel Classifier" bestemt kun en forekomst av gangen. Dette gjør at ved klassifisering av flere forekomster [1] må det kjøres flere runder for å få klassifisert alle.

3) Denne metoden bygger på mye av det samme som i metode 1), men i stede for å estimerer sannsynlighet for at et piksel hører til en søkt forekomst, finner vi her hvor mye det finnes av en forekomst. Etter at et høyoppløselige bildet har blitt klassifisert, vil det være lurt å ta en tur ut i terrenget for å se om klassene i bildet stemmer med virkeligheten. Hvis det er bra nok kan det foretas en opptelling som finner andelen høyoppløselige piksler av en klasse innenfor området til et lavoppløselig piksel. Denne andelen kan sammen med for eksempel piksel verdiene fra forskjellige bånd i et satellittbilde brukes til bygge regresjonsmodeller for predikering av nye forekomster. For å klare å bygge slike modeller kan det brukes et regresjonstre. Ved å se på flere piksler er det mulig finne ut hva som må til for at en områdetype skal kunne gjøre seg gjeldene i pikselverdiene til et lavoppløselig bilde.

2.4 Regresjonstre og Beslutningstre

"Beslutningstre og beslutningsregler er typiske "data-mining" tekniker som hører til en klasse av metoder som gir utdata i form av logiske modeller". [9] (fritt oversatt)

Regresjonstre er statistisk metode som blir mye benyttet i analyse av data i databaser.

Oppbygningen av et regresjonstre og beslutningstre skjer på samme måte. Det som skiller et regresjonstre fra et beslutningstre er resultatet i løvnodene[1]. Et beslutningstre gir tilhørigheten til en klasse i løvnodene mens i et regresjonstre får en ut en numerisk verdi [16]. For å komme fram til verdien i løvnoden i et regresjonstre blir det dannet en regresjonsmodell som prøver å predikere verdi for punktet ut fra gitte regler. Valget av attributter underveis avgjør hvordan regresjonsmodellen blir. All splitting av dataene blir gjort slik at den samlede feilen for to delsett etter en splitt, er betydelig mindre enn feilen før splitting.

Regresjonstre algoritmen produserer regel baserte modeller for prediksjon av kontinuerlige variable, basert på treningsdata. Hvert enkelt regelsett definerer betingelsene for danningen av en mulitpel lineær regresjonsmodell. Et regresjonstre kan gjøre rede for den ikke lineære sammenhenger mellom predikert og reell verdi, og tillater både kontinuerlige og diskrete variable som inndata.

Kvaliteten på et konstruert regresjonstre kan måles ved en gjennomsnittsfeil R for et tre, T (hentet fra Yang, L. m.fl.[18]):

$$R(T) = \frac{1}{N}\sum_{i=1}^{n}\left| y_i - g\left(\overset{\rho}{x}_i\right)\right|$$

Der funksjonen $g\left(\overset{\rho}{x}_i\right)$ representerer regresjonsplanet gjennom eksempel settet, og N er antallet punkter som er brukt til å få etablere treet.

[1] Løvnode – den ytterste noden i et data tre, den siste noden en kommer til ved å følge stien fra roten og helt ut.

I alle tilfeller hvor det brukes og lages beslutnings- eller regresjonstre skjer oppbygningen ved hjelp av styrte metoder der et sett med ”input/output” data brukes til å bygge modellene [9]. Løsningene finnes ved å foreta et ”fra topp til bunn” søk i deler av søke rommet [9]. Resultatet blir i de fleste tilfellene et enkelt tre, men resultat treet trenger nødvendigvis ikke være det enkleste treet.

Et beslutningstre består av noder hvor attributtene blir testet, og greinene ut fra nodene tilsvarer de mulige utfallene av en attributt test i en node. Sagt på en annen måte så er hver grein en klassifikasjonsreglen. Når alle eksemplene i en node er klassifisert får vi en sti til en løvnode som representerer en klassifikasjonsregel. Det kritiske valget ved denne klassifikasjonen er valget av attributt i noden.

Et tre med en flervalgsdeling har en enkel representasjonsform, men samtidig er denne formen med på å legge en begrensning i modellens mulighet til å uttrykke sammenhengen mellom dataene som er blitt lagt inn i modellen. En hver begrensning i treets representasjon fører signifikant til begrensning i funksjonaliteten og tilnærmingsstyrken til modellen som er bygd ut fra treet. Det vil si at modellen ikke klarer å modellere virkeligheten like godt som det er ønskelig om eller mulig.

2.4.1 Algoritme

Algoritmen som ligger til grunn for et beslutningstre er algoritmene C4.5 som igjen er en utvidelse av Quinlan’s ID3-algoritme [9]. Beskrivelsen av denne algoritmen passer også for mye av det som ligger tilgrunn for byggingen av regresjonstre i Cubist. For å få et lite hint om hva som skjer under byggingen av et beslutnings- eller regresjonstre er denne beskrivelsen tatt med.
I ID3 baserer valget av attributter seg på antagelser om at kompleksiteten til et beslutningstre er sterkt knyttet til mengden av informasjon det er mulig å få ut av verdien til en attributt. Dette gjør at valget havner på den atributten som gir mest informasjon ut fra sin verdi. I C4.5 favoriseres attributter med lav klasse entropi[2], det vil si at klassen har liten variasjon/varians. Et eksempel på dette er når flesteparten av forekomstene i en attributt tilhører en klasse av klassifiserte data. Videre velger algoritmen den atributten som viser den størst grad av forskjell mellom klassene lokalt.

Det viktigste som skjer i C4.5 algoritmen er generering av beslutningstre basert på trenings data. Dette danner et tre med to mulige node typer. 1) Beslutningsnode: som spesifiserer tester som kjøres på en attributt verdi med en gren og subtre som mulig utfall. 2) Løvnode: Indikerer hvilken klasse som verdien hører til.

[2] Entropi – variasjon, i form av for eksempel varians, konfidensintervall, maksimal feil.
Brukes i den kartografiske verden om varias.

Testene som brukes for å bygge modell blir valgt slik at treet skal bli mest mulig kompakt. Dette gjør at en ikke kan kjøre en ny test med nye kombinasjoner når et utfall først er funnet. For at en søkeprosess ikke skal kunne gå i det uendelige, blir ikke andre mulige løsninger undersøkt når en løsning er funnet. Dette er en av de negative tingene med denne algoritmen. Den kan fort bli kostnadskrevende ved at den krever mye maskin resurser ved bygging av en modell.

3 Valg av strategi for predikering av lav

En av problemstillingen i denne oppgaven er å se på om flybilder er egnet som en kilde for innsamling av treningsdata. Det er derfor naturlig å velge dette som kilde for trenningsdataene. Samtidig er det valgt å bruke tre forskjellige typer flybilder, for å se på om flere typer flybilder er egnet til bruk som kilde for treningsdataene ved en analyse. For at forsøket skulle ha blitt best mulig burde jeg hatt ytterligere et sett med treningsdata fra feltundersøkelser som med sikkerhet kunne si hva jeg var i stand til å se i flybildene. Men siden det ikke fantes slike data måtte jeg klare meg uten. De eneste dataene som var tilgjengelig var noen grov kontrollerte bilder fra noen utvalgte DuncanTech (heretter bare kalt Duncan) bilder på Hardangervidda. Mangel på treningsdata vil kunne medføre en stor feilkilde hvis ikke bildene blir tolket korrekt. Det er derfor i dette tilfellet viktig å få en best mulig visuell vurdering av bildene. Valget av flybilder som treningsdata baserer seg å på som nevnt i innledning at flybilder blir stadig mer brukt i andre sammenhenger og får med det en økt tilgjengelighet. Samtidig er det ønske om å finne en metode som kan redusere bruken av feltarbeid.
Som nevnt ovenfor fantes det ikke trenningsdata for alle bilden slik at skal vi få klassifisert bildene må vi benytte oss av en ikke-styrt klassifisering, også omtalt som clustring, for å få trekt ut de ønskede pikslene i flybildene.

For subpiksel klassifiseringen av Landsat 7 bildet har jeg valgt å benytte meg av metode 3 omtalt i kapittel 2.3.1. Valget av å bruke regresjonstre kommer av at det var ønskelig å finne en modell som kunne predikere andelen lav, for et stort område ut fra treningsdata fra mindre områder. Det var ønskelig å se på flere faktorer som kunne ha noe å si for hva som ble oppdaget ved en analyse av bildet. Kan kombinasjonen av forskjellige piksel verdier i et bilde fortelle oss noe om hva som finnes av vegetasjon i et området dekt av et piksel? I tillegg til bare å benytte seg av pikselverdiene i en tolkning kan det være ønskelig å trekke inn flere kilder for å se om det er flere faktorer som kan påvirke hvor det er mulig å finne lav (i dette tilfellet). I denne oppgaven har det blitt valgt å se på om høyde over havet, vegetasjonsgeografiske soner og solinnstrålingen kan ha noe si for hvor det er mulig å predikere lav.

4 Datagrunnlag

4.1 Satellittbilder

4.1.1 Landsat 7 ETM+

Landsat 7 bildet som blir benyttet i denne oppgaven består av to Landsat 7 scener tatt opp den 30.juni 1999. De to scenene er 198_18 og 199_18.

Disse to scenene har gjennomgått en radiometrisk korreksjon og belysnings korreksjon ved bruk av et tilleggsverktøy i Erdas kalt Atcor. De har så blitt satt sammen til et bilde kalt Landsat-7-atcor (heretter bare omtalt som Landsat 7 bildet).
I tillegg til har bildet blitt geometrisk korrigert og resamplet til 25 meters piksler. Dette arbeidet har blitt gjort av Einar Lieng knyttet til SatNat prosjektet.

Tabell 4.1 Egenskaper og oppløsning til Landsat ETM+ [15]

Bånd	**Spektral oppløsning (μm)**	**Oppløsning (m) v/nadir**	**Egenskap**
1	0,450 - 0,515	30 x 30	Har god evne til gjennomtrenging av vann, men er også til støtte ved analyse, egner seg også til støtte for analyse av jord og vegetasjon
2	0,525 - 0,605	30 x 30	Registrerer den grønne refleksjonen fra frisk vegetasjon.
3	0,630 - 0,690	30 x 30	Registrer rødt klorofyll fra grønn vegetasjon. Viktig i skille mellom vegetasjon.
4	0,750 - 0,900	30 x 30	Infrarødt bånd, registrerer biomassen i en scene. ”Ser” kontraster jord/avling og land/vann
5	1,55 - 1,75	30 x 30	Følsom for vannmengden i planter, brukes ved kartlegging av tørke. Skjelner mellom skyer, snø og is.
6	10,40 – 12,50	60 x 60	Måler størrelsen på den infrarøde radiant fluxen utsendt fra jorda. Topografiske forhold i fjell områder
7	2,08 - 2,35	30 x 30	Skjelner mellom geologiske stein formasjoner.
8 (pan)	0,52 - 0,90	15 x 15	

I forhold til de tidligere Landsat TM satellittene har ETM+ mulighet til å gjøre opptak i et ekstra bånd, et pankromatisk bånd med oppløsning 15 x 15 m. Dette båndet har ikke blitt benyttet i denne oppgaven.

4.1.2 SPOT 5

En pankromatisk scene fra SPOT5 har i denne oppgaven blitt brukt som referanse kilde til oppretting av Duncan og Kodak bilder. SPOT5 scenen som ble benyttet i denne oppgaven var 073_226 tatt opp 3/9-02 og hadde 2,5 meters oppløsning.

4.2 Flybilder

4.2.1 DuncanTech MS3100

DunkcanTech [3], [11] er et digitalt kamera med mulighet til å ta bilder i det infrarøde spektret. Kameraet består av 3 CCD bilde sensorer og har en oppløsning på 1024 x 1280 bildeelement, tilsvarende 1,3-megapiksel. På bakken vil denne oppløsningen tilsvare 20-40 cm og gi bilder som gir 300-500m brede stripper, avhengig av flyhøyden. Ved for eksempel en fly høyde på 1000 meter vil oppløsningen være 27 cm. Kameraet har 4 spektrale bånd det kan registrere i, der av 3 bånd i det synlige og 1 i det infrarøde lyset.
Duncan bilder som har blitt brukt i denne oppgaven er hentet fra flystriper tatt opp 22 og 23 august 2001, med en gjennomsnittlig oppløsning på ca 20 cm.

4.2.2 Kodak DCS-760

Kodak DCS-760 [11] (heretter bare omtalt som Kodak) er et vanlig digitalt proff-kamera med RGB farger som ble plassert i et fly ved fotografering. Kameraet har en 6,1-megapiksels CCD brikke, med 3032 x 2008 bildeelement.
Kodak-bildene som ble brukt i denne oppgaven har en oppløsning på omkring 20 cm på bakken og ble tatt opp 23 august 2001.

4.2.3 IR-ortofoto

Dette er analoge infrarøde flybilder som Statens Kartverk har skannet inn på digital form og opprettet, slik at de nå foreligger som digitale infrarøde ortofoto med en meters oppløsning. Det som her omtales som et IR-bilde er i realiteten flere bilder som har blitt satt sammen til sammenhengende bilder slik at de dekker et mye større område enn Kodak og Duncan bildene.

4.3 Andre data

4.3.1 Høydemodell

For uttrekk av høyde data ble det brukt et utsnitt av Statens Kartverks nyutviklede høydemodell basert på høydekurver, vann og terrenglinjer fra N50 kartdata. Denne Høydemodellen har en oppløsning på 25m.

4.3.2 Vegetasjonsgeografiske soner

"Vegetasjonssone er et begrep som i de fleste sammenhenger brukes for enheter som mer eller mindre sammenhengende går rundt jordkloden og som viser den horisontale variasjonen

fra sør til nord. I Vegetasjonsatlaset og her brukes vegetasjonssone også om høgdebetingete enheter, eller den vertikal variasjonen. " (Moen,[13])

De vegetasjonsgeografiske sonene over Hardangervidda, som ble benyttet har blitt lagd til av Einar Lieng. Dette kartet er basert på Statens Kartverks høydemodell (kap.4.3.1) som Lieng har tilpasset Moens vegetasjonsatlas med enkelt forandringer. Det har blant annet blitt lagt inn at tregrensen kryper høyere til fjells i sør-sørvest vendte skråninger som følger av at de blir sterkere eksponert for sol lys.
Dette kartet ligger lagret på digital form med en oppløsning på 25m.

Vegetasjonsgeografiske soner er kontinuerlige dataverdier som forteller om hvilke betingelser vegetasjonen har for å klare seg. 0-1 nemoral (N) (finnes helt sør i Norge), 1-2 boreonemoral (BN), 2-3 sørboreal (SB), 3-4 mellomboreal (MB), 4-5 nordboreal (NB), 5-6 lågalpin/sørarktisk (LA/SA), 6-7 mellomalpin (MA), 7-8 høgalpin (HA).

4.3.3 Varmeindeks

Det siste data laget som ble benyttet var et rasterlag med 25 m oppløsning som sier noe om hvor sterk solinnstrålingen er på et punkt. Varmeindeksen beregnes ut fra helningsgraden og solinnstrålningen.
Varme indeksen kan beregnes ved formelen [13]:

Varmeindeks = cos (eksposisjonen i º - 225) x tan (helningsgraden i º)

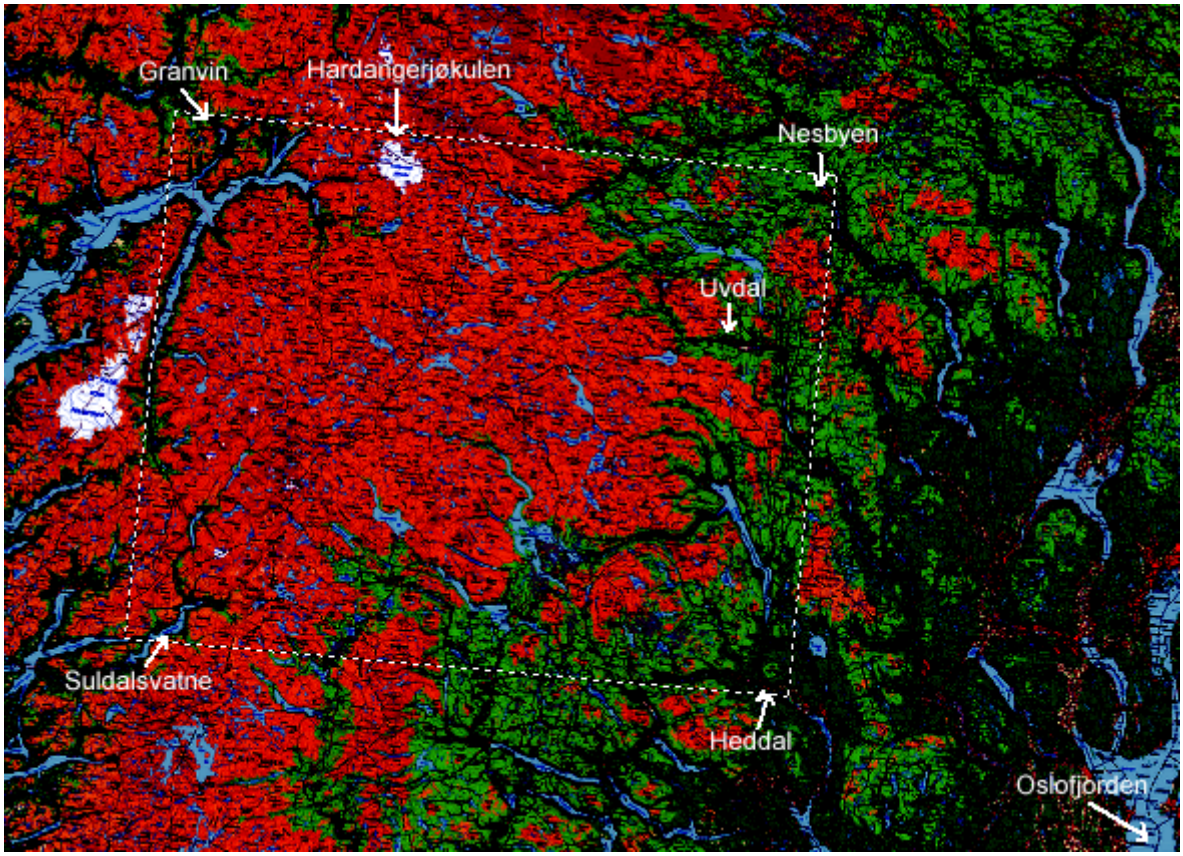
der eksposisjonen i, tilsvarer kompasretningen.
Jo høyere verdi varmeindeksen får, dess mer solinnstråling.

5 Studieområde

Studieområdet i denne oppgaven er Hardangervidd. Hele området dekkes og avgrenses av det sammen satte Landsat 7 bildet (kap.4.1.1) med hjørnekoordinater (WGS84, UTM sone 32):

Øvre venstre hjørne	X:	363025.000	Y:	6719900.000
Øvre høyre hjørne	X:	508050.000	Y:	6719900.000
Nedre venstre hjørne	X:	363025.000	Y:	6605125.000
Nedre høyre hjørne	X:	508050.000	Y:	6605125.000

Referert til stedsnavn tilsvare dette et område fra omkring Granvin i nordvest, Nesbyen i nordøst, Suldalsvatnet i sørvest og Heddal i sørøst.



Figur 5.1 Kartutsnittet viser beliggenheten til Landsat utsnittet. [Kart utsnitt hentet fra Statens Kartverks N250]

Bildet dekker et område som går fra 0 til 1900 meter over havet. men majoriteten av arealet ligger på ca 1100-1300 meter over havet. Dette gjør at høyfjellsvegetasjonen dominerer og består blant annet av lav, lyng og busker. I den østlige delen av bildet ligger det større områder med skog. Deler av området er for øvrig småkupert.

Innenfor området som dekkes av Landsat 7 bildet, finnes det et stort utvalg av flybilder. Det store utvalget av flybildet gjør at de fleste naturtypene kan dekkes i en analysen. Av disse er

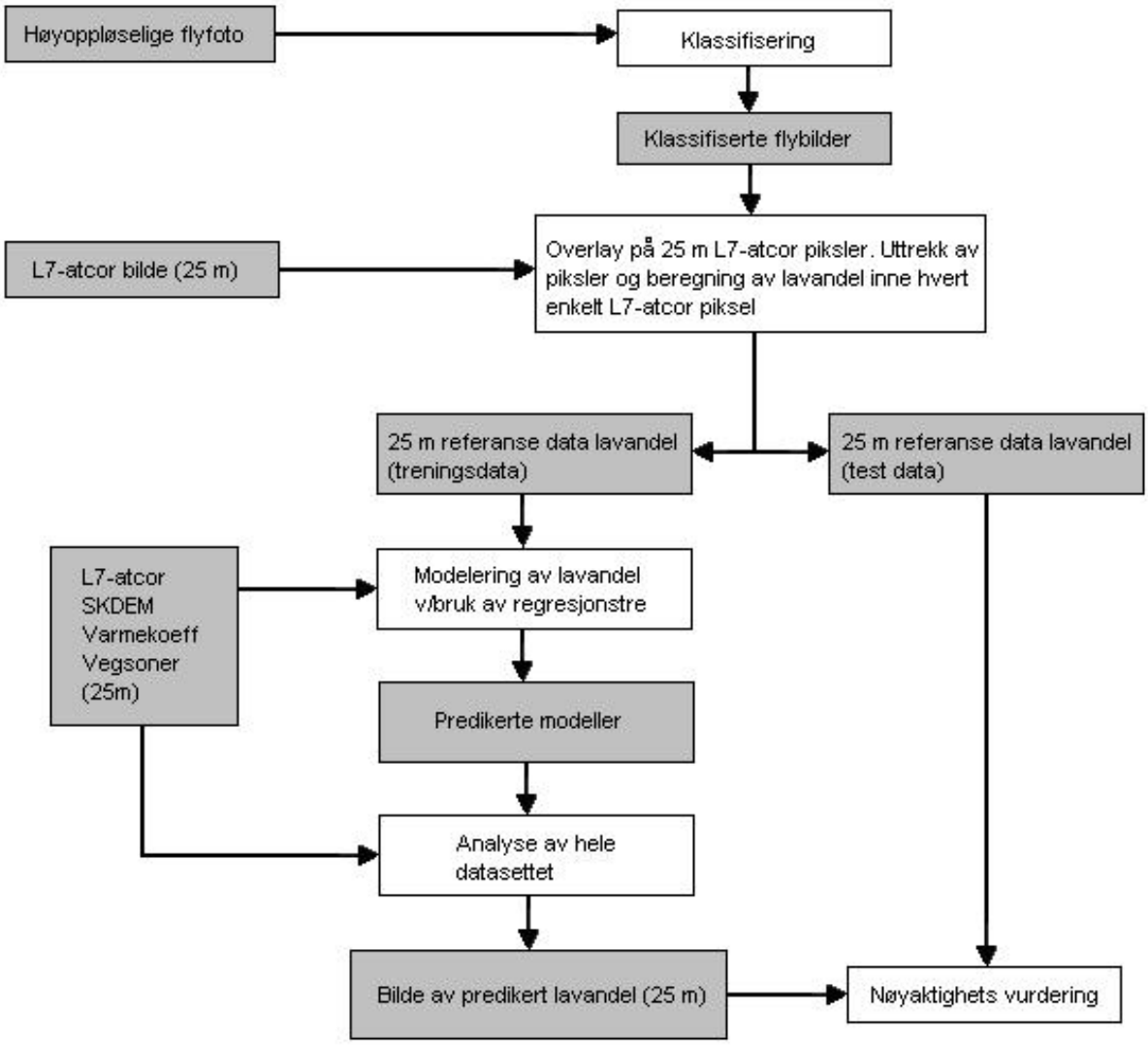
det gjort et utvalg på noen. Bruk av alle de utvalgte bildene ble for tidkrevende, slik det også måtte gjøres et utvalg av disse igjen, for at jeg skulle rekke å få kjørt tilstrekkelig analyser. Flybildene måtte da velges slik at de representerte et størst mulig område, samtidig som det skulle være mulig å identifisere de fleste vegetasjonstypene.

Duncan bildene som er brukt i denne oppgavene er plukket ut av flystripper som går på kryss og tvers av Hardangervidda. Stripene ligger i nordsør- og i østvest retning, med flygning i henholdsvis begge retninger.

IR bildene fra Kartverket dekker et sammenhengene område i Uvdal.

6 Metoder

Gangen i metode delen kan i korthet beskrives ved et flytdiagram:



Figur 6.1 Gangen i prosessen. Grå bokser er datalag eller modeller. Prosedyrer i hvite bokser.

6.1 Feltarbeid

I august 2002 ble det gjennomført feltundersøkelse på noen utvalgte Duncan bilder på Hardangervidda av med-student Ole Marin Aanonsen. Dette arbeidet ble brukt som støtte i bilde analysene. Arbeidet ble gjort ved at det ble tatt med utskrifter av Duncan bilder ut i felt for å identifisere hva som virkelig fantes av vegetasjon på stedene og som vi kunne se igjen i bildene. Bakgrunnen for dette var å foreta en sikrere identifikasjon av tilsvarende områder i andre Duncan bilder ved klassifisering av bildene. De områdene som det ble lagt vekt på å identifisere var de områdene med lav og hvordan de kunne skilles fra andre områder, da i hovedsak de med jord og stein. I områden med lav ble det gjort grove registret av hva slags lav som fantes og hvor tykt lavlaget var. Det ble også gjort et grovt overslag hvor stor andel

laven utgjorde i området. I enkelte utvalgte Duncan bilder ble det også gjort GPS-målinger med tanke på georefereringen av bildene til seinere bruk.

6.2 Geometrisk korreksjon

Skal det være mulig å jobbe med flere forskjellige bilder i samme prosjekt, og benytte de om hverandre må de være referert til samme koordinatsystem.

Duncan og Kodak bildene hadde fått tilordnet et sett koordinater i form av koordinatene til projeksjonssentret i det bildet ble tatt. Men disse hadde ikke en bedre nøyaktighet enn ± 50 meter. Noe som ikke er bra nok til vårt formål. For å at bilde skal bli nøyaktige nok trenges det flere og mer nøyaktige koordinater. For å få dette til trenges det en nøyaktig kilde, valget falt da på å et SPOT5 bilde med 2,5 meters oppløsning.

Opprettingen skjer i Erdas ved hjelp av ”Geometric correction” verktøyet. Det bildet som skal rettes opp legges i en viewer, mens SPOT5 bildet som det skal rettes opp mot legges inn i en annen viewer. I opprettingen av bildene i denne oppgaven er det valgt å bruke 5-6 referansepunkter i hvert enkelt flybilde. Disse punktene må plasseres på detaljer som er lett å identifisere i begge bildene. 4 av punktene ble forsøkt plassert ut mot kanten av bildene og 1-2 ble plassert i midten av bildet. Modellen som ble valgt for den geometriske korreksjonen var en 1.ordens polynom funksjon.

Siden SPOT-bildet ikke var stort nok til å dekke hele området hvor det fantes flybilder, måtte andre kilder også benyttes. Derfor ble Kartverkets skannede IR-ortofoto brukt til å rette opp enkelte av Kodak bildene. Nøyaktigheten på IR ortofoto skulle tilfredsstille våre krav. I tillegg ble to bilder rettet opp ved hjelp av GPS-målte punkter. Disse punktene var merket på papirkopier av bildene slik at det var mulig å identifisere og plassere de.

6.3 Klassifisering/Identifisering

For å få skilt ut de områdene med lav, som jeg var på utkikk etter, var det ønskelige å foreta en klassifisering av flybildene. Det finnes, som tidligere nevnt, flere klassifiserings metoder som kan være egnet til dette formålet. En styrt klassifisering vil nok ha vært å foretrekke, men ved bruk av denne metoden er en avhengig av å ha sikre treningsdata å gå ut i fra. På grunn av at jeg måtte bruke flybilder som det var gjort lite feltarbeid på, ble det valgt å bruke en ikke-styrtklassifisering eller clusting i Erdas. For å være sikker på at all laven skulle bli skilt ut til egne klasser, måtte det kjøres flere runder med forskjellige antall clustre. Startet med få clustre for så å øke antallet. Det ble gjort forsøk med flere clustre enn strengt tatt nødvendig for å se om det gav noe bedre resultat.

Etter at bildene var blitt clustret måtte de forskjellige områdene identifiseres visuelt. Duncan bildene ble tolket med støtte fra dataene samlet inn av Aanonsen. Noe av denne kunnskapen ble også benyttet som hjelp i tolkningen av IR bildene. I tolkningen av Kodak bildene måtte jeg stole på at det jeg så i bildene var korrekt. Ettersom det bare var områder med lav som jeg

var på utkikk etter, var det kun disse områdene som ble identifisert og klassifisert til seinere bruk.

6.4 Bildeanalyse

6.4.1 Forberedning

Etter at klassifiseringen av flybildene var gjort måtte bildene tilpasses videre bruk i ArcView. For at jeg skulle kunne hente ut informasjon om antallet flybildepiksler med lav som lå innenfor området til et Landsat piksel, måtte jeg først definere et rutenett med rutestørrelse tilsvarende et Landsat 7-piksel. Rutenettet kunne etterpå legges ovet flybildet. Deretter var det mulig å foreta en opptelling av antallet lavpiksler, og beregne prosentandelen lav innenfor et Landsat piksel. Prosentandelen av ble senere brukt i regresjonstreet.

Rutenettene ble lagd i Erdas ved å lage utsnitt av Landsat 7 bildet tilsvarende størrelsen på flybildene. Bilde utsnittene ble så eksportere ut som shape-filer, som leses av ArcView. For at et klassifiserte flybildet skal kunne kobles mot rutenette i ArcView blir det eksportert ut som ei GRID-fil, ei polygonfil med piksler tilsvarende pikslene i flybildet.

Det ble i Erdas eksportert ut ascii-filer bestående av de forskjellige lagene med informasjons som vi ønsket å se om kunne ha en forklarende effekt på gjengivelsen av lav i Landsat pikslene. Disse lagene besto av bånd 1-7 fra Landsat 7, høydemodell, vegetasjonsgeografiske soner og varmeindeks.
For at programmet CubistSatNat skulle klare å lese de lagene som skulle tolkes, måtte de eksporteres til et rådata format der dataene ble lagret piksel for piksel.

6.4.2 ArcView

Alle skriptene i det følgende arbeidet er laget av Gunnar Fjone.
Etter at flybildene var blitt eksportert ut fra Erdas kunne de hentes inn i ArcView. I ArcView ble det ved hjelp av et skript (Vedlegg D) gjort et uttrekk av alle pikslene i bildet, det il si at alle pikselverdiene innenfor ei rute i rutenette ble skrevet ut til ei tekstfil. Antallet filer ble nå så stort og uhåndterlig at det måtte foretas enn oppsummering av antall piksler for hver enkelt klasse. Til denne oppsummeringen ble det benyttet et skript (Vedlegg A) som telte opp hvor mange piksler som hører til hver enkelt klasse, innefor ei rute, og skrev antallet til ei fil for hele bildet. De rutene hvor det ikke var overlapp mellom flybildet og rutenettet ble utelatt siden de ikke er av interesse, det samme ble ruter hvor det ikke var full overlapp mellom rutenett og flybilde. Denne filen kunne så redigeres med et skript (Vedlegg B) som først summerte antall piksler innefor ei rute, for så å plukke ut og summere antall piksler i klassene med lav. Disse verdiene ble til slutt brukt til å regne ut hvor mange prosent lav det var innenfor ei rute tilsvarende et Landsat piksel. Det er denne prosent andelen som skal benyttes til slutt i regresjonstreet.

For at jeg skulle kunne koble filen med prosentandel lav for et Landsat piksel og ascii-filen med de ønskede test verdiene, måtte jeg trekke ut ei fil med koordinat og rute ID fra rutenettet i ArcView. Dette ble gjort ved hjelp av et skript i ArcView (Vedlegg E).

Ettersom det ble fjernet mange ruter fra rutenettet som ble laget i starten, måtte det også fjernes piksler fra filen med de tilsvarende testverdiene. Dette ble gjort med et skript (Vedlegg C) som sammenlignet de forskjellige filene med koordinater og ID. Punkter som ikke fantes i alle filene ble fjernet. En kolonne med "ProsentLav" ble lagt til i slutt fila, slik at vi fikk en fil med alle de ønskede test attributtene. Denne fila ble til slutt skrevet ut som en DATA-fil som kunne analysere i Cubist. Det er denne filen Cubist bruker i byggingen av regresjonsmodellene.

6.4.3 Analyse

For at jeg skulle få lagd regresjonsmodeller for datasettet ble programmet Cubist benyttet. Dette programmet krever at dataene som skal analyseres er lagret i filer med navn *.data og at det fines en egen fil med navn på attributtene som skal ligge til grunn for modelleringen. Denne filen må ha ending .names. og gi informasjon om hvilken attributt som er test verdien som det er ønskelig å bygge modellen mot og hvilke attributter som det er ønskelig å se om har noen påvirkning på denne verdien.
Attributtene i denne filen er som følger:

Attributt	**Type**	**Kommentar/beskrivelse**
ID:	label.	ID er ikke med i modellen
F1B1:	continuous.	L7-atcor. band 1 (Landsat bildet)
F1B2:	continuous.	band 2
F1B3:	continuous.	band 3
F1B4:	continuous.	band 4
F1B5:	continuous.	band 5
F1B6:	continuous.	band 6
F1B7:	continuous.	band 7
F2B1:	continuous.	skdem - terrengmodellen
F3B1:	continuous.	varmekoeff – sol innstråling
F4B1:	continuous.	vegsoner - vegetasjonssoner
LavAndel:	continuous.	prosent andel lav som vi ønsker å teste på

Ved kjøring av Cubist er det forutsatt at navn-fila og data-fila ligger i samme mappe og med samme fornavn. Dette gjør at jeg bare trenger å hente opp data-fila, så henter programmet resten av de tilgjengelige filene selv.

Det er flere muligheter for hvilke kriterier som skal ligge til grunnlag for byggingen av modellen. Jeg valgte å la programmet selv velge kriterier som modellene skal bygges ut fra.

Fra denne kjøringen får jeg ut en serie med forskjellige regresjonsmodeller som er bygd på forskjellige kriterier og kombinasjoner av datasettet. Disse regresjonsmodellene kan videre brukes til å teste hele Landsat bildet og de andre data lagene som vi ønsker å se på om påvirker hvor vi kan påvise lav. Denne analysen av datasettet gjøres i programmet CubistSatNat laget av Gunnar Fjone. Dataene leses inn sammen med modellen for analyse. Resultatet av denne analysen kommer ut som bilde med pikselverdier tilsvarende antall prosent lav det er innenfor området til et piksel på 25x25 meter.
Under prosessen i CubistSatNat blir det plukket ut 200 piksler som kan brukes som test verdier i en ”Cross-reference” i Cubist. Disse verdiene kan brukes til å se om de predikerte verdiene i Cubist stemmer med de som finnes i prosent bildet.

Sidene BIP-filen som kommer ut fra CubistSatNat ikke er noen bildefil men ei fil bestående av rådata må den importers i Erdas til Img-fil. Etter at det nye bildet er blitt hentet inn i Erdas kan de vurderes for å se hvor gode modellen fra Cubist har klart å modellere lav andelen i forhold hvordan det virkelig er.

For å se hvor gode regresjonsmodellene var, ble det i Cubist brukt et sett med data fra områder som ikke var brukt til selve modell byggingen. Disse dataene ble så kjørt sammen med de eksisterende modellene.

7 Resultater

7.1 Tilrettelegging

7.1.1 Feltarbeid

Aanonsen besøkte 18 Duncan bilder på Hardangervidda.

7.1.2 Geometrisk korreksjon

Landsat 7 bildene som ligger til grunn for bildet over Hardangervidda var rettet opp på forhånd. Denne opprettingen hadde skjedd ved hjelp av N50 data. Nøyaktigheten på denne opprettingen tilsvarte en RMS på 10-15 meter. I tillegg til unøyaktigheten i selve opprettingen kommer unøyaktigheten i N50 dataene. Unøyaktigheten i N50 er ±10 meter for nyere kartblad og ±50 meter for noe eldre digitaliserte kartblad.

Alle de benyttede Duncan og Kodak bildene ble rettet opp til WGS84, UTM sone 32.
Duncan og Kodak bildene ble rettet opp ved hjelp av en 1.ordens polynom funksjon. Det ble valgt 5-6 punkter i hvert bilde til bruk i opprettingen. De bildene som ble korrigert er listetopp i vedlegg 6. Det ble rettet opp i alt 50 bilder; 16 Kodak og 34 Duncan.

For at opprettingen skulle bli nøyaktig nok ble det valgt å sette som krav at RMS verdien ikke skulle overstige 7. Noe som tilsvarer 1- 2,8 meter i terrenget, avhengig av oppløsningen til flybildene. De fleste punktene i Duncan bildene fikk en RMS under 6, som ved en oppløsning på 20 cm vil tilsvare 1,20 meter i terrenget.
Kodak bildene i serien fra 493 til 530 ble rettet opp mot Kartverket skannede IR bilder. I enkelte av disse bilden var det vanskelig å få rettet opp alle punktene med god nok RMS. I enkelte av bildene fikk jeg punkter med en RMS opp mot 100. Selv med flere forsøk var det nesten umulig å få ned feilen. De resterende bildene ble rettet opp mot SPOT5-bildet. For disse gikk RMS verdien greit under den valgte grensen på 7.

Siden IR bildene kommer ferdig opprettet fra Statens Kartverk antas det at de har en god nøyaktighet. Slik at det ikke har blitt foretatt noen korreksjon av disse. De manglet kun kartprojeksjon slik at den ble satt til WGS84, UTM sone 32.

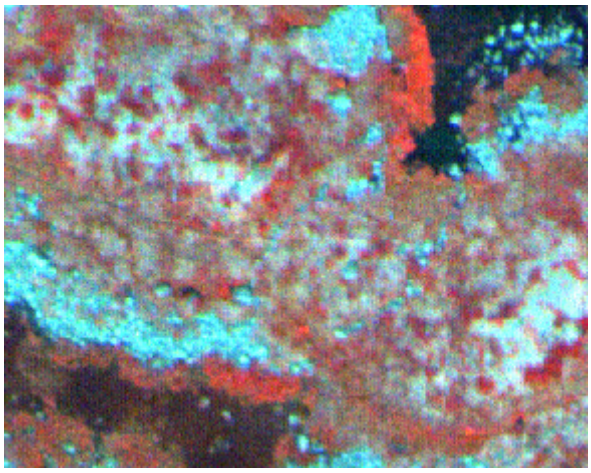
7.1.3 Klassifisering/Identifisering

Alle de bilden som ble geometrisk opprettet ble også klassifisert, slik at vi fikk i alt 64 klassifiserte bilder; 34 Duncan, 16 Kodak og 14 IR bilder.

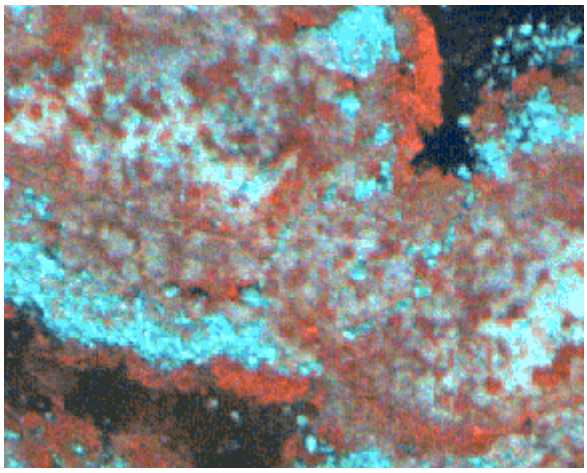
Etter en del prøving og feiling endte jeg tilslutt opp med 70 klasser for alle tre bilde typene. Selv om det i enkelte tilfeller kunne virke som det ikke var nok, ble ikke effekten av å øke antallet noe bedre. Det ble derfor valgt å beholde dette antallet.

Med 70 klasser fikk jeg detaljerte bilder med en detaljeringsgrad som lå veldig nære de uklassifiserte bildene. Av disse 70 klassene var 4-6 som kunne tolkes som lav. Et eksempel på denne detaljeringen er synlig ved å sammenligne Figur 7.1 som viser et original bilde og Figur 7.2 som viser det klassifiserte bildet.

Ble det færre enn 70 klasser, fikk jeg flere tilfeller hvor laven havnet i samme klasse som stein og jord. Men i enkelte bilder var det nesten umulig å skille laven fra områder med stein og jord selv med 70 klasser.



Figur 7.1 Duncan orginal, aug22-5962954, bånd 1,2,4



Figur 7.2 Duncan 70 klasser aug22-5962954_70.img

7.2 Analyse

7.2.1 Regresjons modeller

Det ble i byggingen av de tre forskjellige regresjons modeller benyttet 13 Duncan bilder, 2 Kodak bilder og 1 IR bilde. For testing av modellene ble det brukt 4 Duncan bilder, 1 Kodak bilde og 1 IR bilde. Hvilke bilder som er benyttet er merket i vedlegg 6. Bilder som er brukt til modellbyggingen er merket M og testbildene er merket T.
Testbilder (omtales å som testdata) er bilder som har blitt brukt til å se om regresjonsmodellene passer på andre data enn de som ble brukt til å bygge modellene.

7.2.1.1 Resultat for regresjons modellene

Kolonnen ”Antall Landast piksler” i Tabell 7.1 og Tabell 7.2 sier noe om hvor mange områder tilsvarende et Landsat 7 piksel, som er hentet fra en type flybilder, brukt til å bygge regresjons modellene eller i testdatasettet.
”Gj.snitt feil” er det samme som gjennomsnitts feil, denne verdien sier noe om hvor mye det kan forvente at de predikerte verdiene kan avvike fra de reelle verdiene. Den relative feilen sier noe om forholdet mellom den gjennomsnittlige feilen og feilen vi vil få ved alltid å estimere gjennomsnittsverdien. Denne verdien skal være under 1 for at regresjonsmodellene skal være brukelige.

”Korrelasjonenkoeffisienten” sier noe om forholdet mellom de reelle attributt verdiene i en forekomst og den predikerte verdien fra modellen. Jo nærmere korrelasjonen er 1, jo bedre samsvar er det mellom atributtene og den prediekerte verdien.

Tabell 7.1 Statistisk resultat fra modellbyggingen

Bilder	Gj.snitt feil	Relativ feil	Korrelasjon	Antall Landsat piksler
Duncan	2,441	0,551	0,71	1296
Kodak	5,904	0,45	0,84	2384
Skannede IR	7,175	0,37	0,86	28440

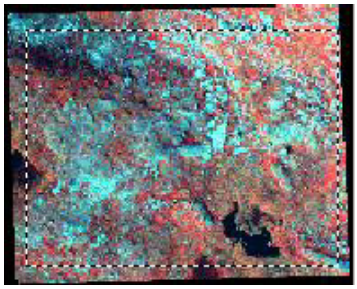
7.2.1.2 Resultat for testdata

Tabell 7.2 Statistiske verdier fra test av regresjonsmodellen

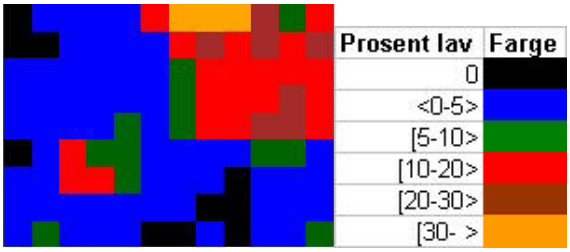
Bilder	Gj.snitt feil	Relativ feil	Korrelasjon	Antall Landsat piksler
Duncan	6,833	1,09	0,01	354
Kodak	11,842	0,86	0,42	1583
Skannede IR	18,979	1,37	0,37	4166

Sammenligning predikert og reell lavmengde

Ved en sammenligning av predikerte og reelle lavmengder ser det ut til at regresjonsmodellene generelt har en tendens til å predikere litt for mye lav. Figur 7.4 viser et utsnitt fra Landsat 7 bilde med beregnet lav mengde fordelt i grupper (å omtalt som reell mengde). Figur 7.5 viser avviket mellom predkiert lav mengde og reell mengde. Negativ mengde viser at regresjonsmodellene har predkiert for mye lav. Stipla linje i Figur 7.3 viser avgrensning for området hvor det er beregnet lav.



Figur 7.3 Flybilde over området
Klassifisert bilde geo41883.img



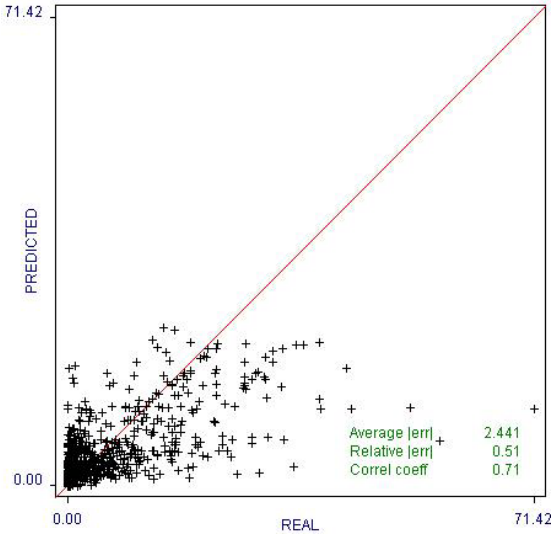
Figur 7.4 Reell lav mengde for bildet



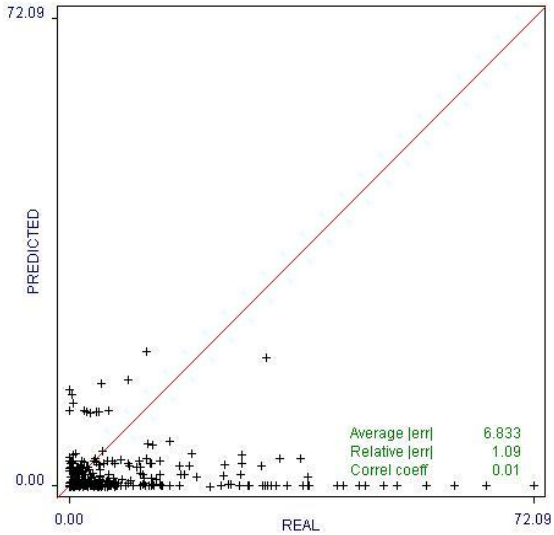
Figur 7.5 Avvik mellom reell og predikert i bildet

I Cubist er det etter modellbygging mulighet for å få ut plot som viser sammenhengen mellom de reelle og predikerte verdiene. Det tilsvarende er mulig for treningsdata som er testet på modellene. Figur 7.6, Figur 7.8 og Figur 7.10 viser reell mot predikert verdi for modellene og Figur 7.7, Figur 7.9 og Figur 7.11 forholdet for test verdiene.

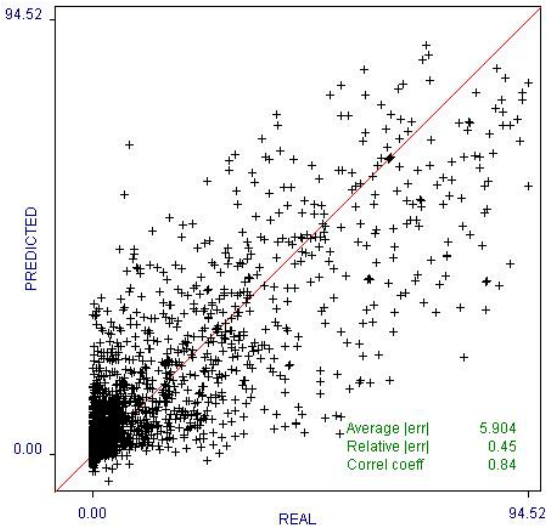
For modelldataene øker spredningen langs linjen jo flere punkter som er med i modellen. Generelt for alle tre modellene kommer det fram at modellene predikerer for mye lav når den reelle lav mengden er liten. Men ved økt lavmengde, øker underpredikeringen, slik at modellene predikerer for lite lav ved store reelle mengder. For test dataene kommer det fram at modellene ikke passer så godt for disse dataene. En ser her at Duncan og Kodak modellene stort sett bare klarer å predikere lite eller ingen lav, mens IR modellene overpredikerer lav mengden.



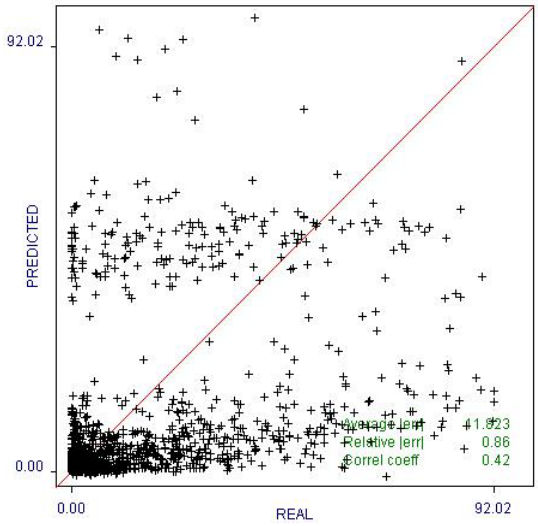
Figur 7.6 Kryssreferanse modelldata Duncan



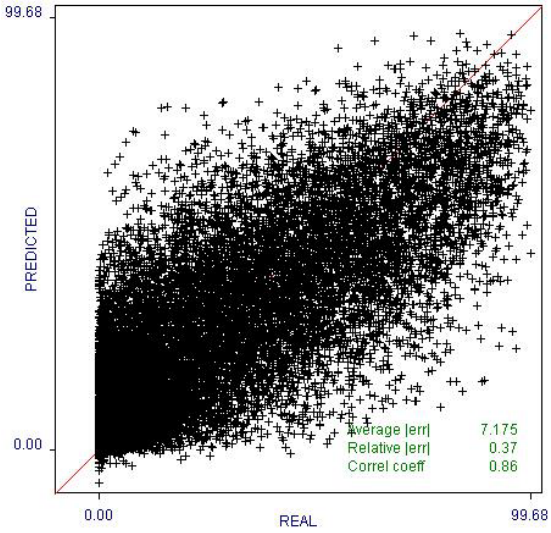
Figur 7.7 Kryssreferanse testdata Duncan



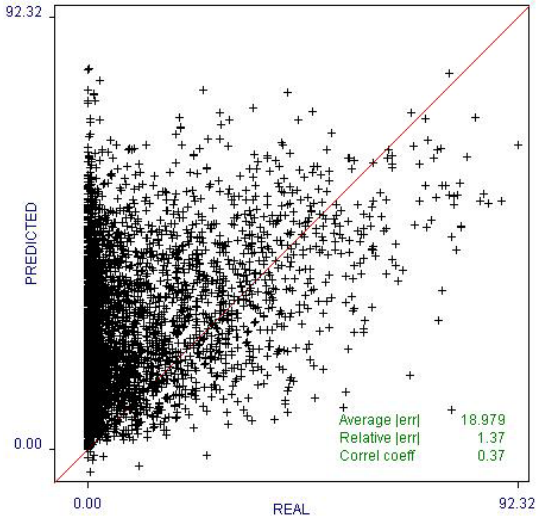
Figur 7.8 Kryssreferanse modelldata Kodak



Figur 7.9 Kryssreferanse testdata Kodak



Figur 7.10 Kryssreferanse modelldata IR



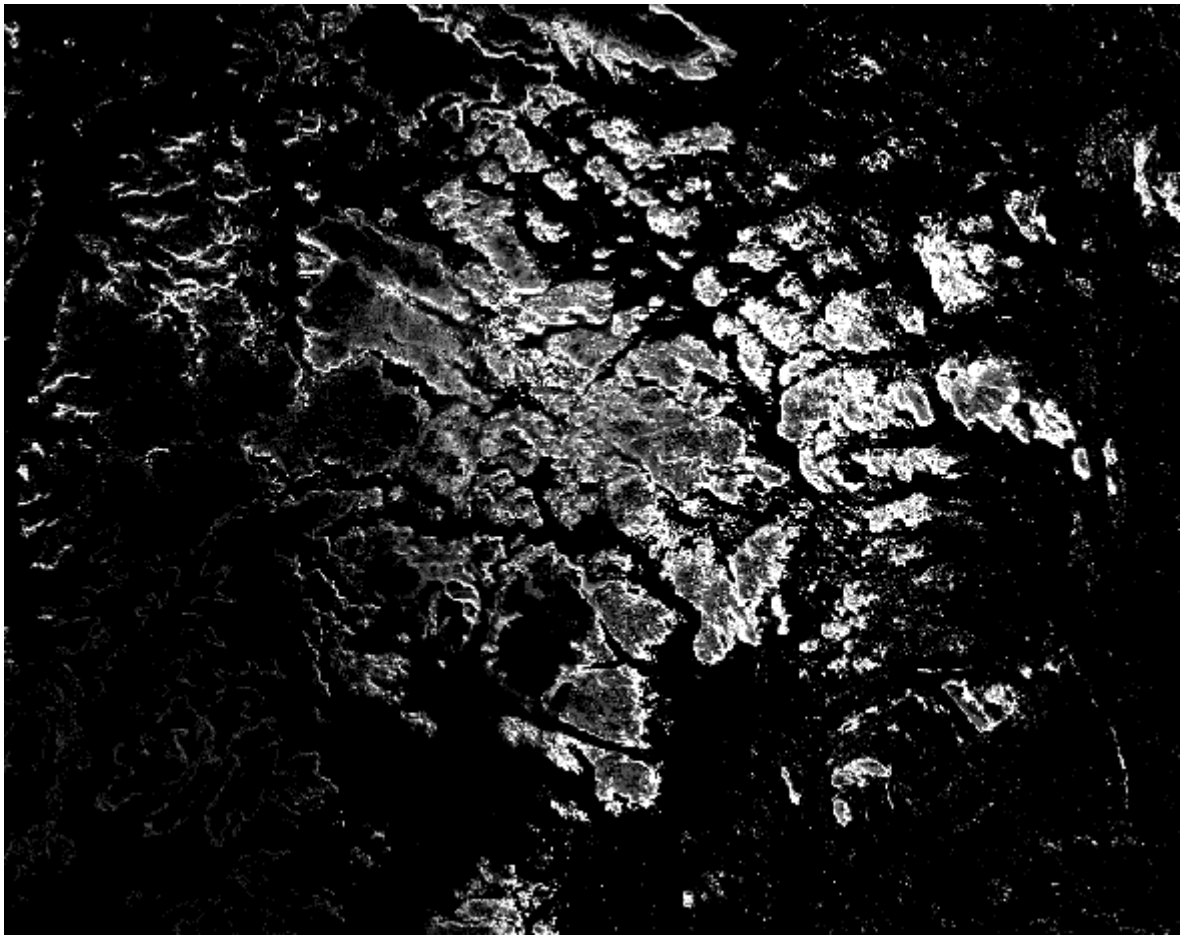
Figur 7.11 Kryssreferanse testdata IR

7.3 Bildeanalyse

Bildene som står under er et resultat av at Landsat 7 bildet og de tre andre data lagene har blitt tolket ved hjelp av regresjonsmodellene i CubistSatNat. De viser alle hvor mye lav som skal finnes i et piksel ut fra regresjonsmodellenes tolkning.

Felles for alle tre bildene er at de er i svart-hvitt. Jo lysere et piksel er jo høyere andel lav er det i pikslet, det vil si at svart tilsvarer null prosent lav.

7.3.1 Duncan



Figur 7.12 Predikert lavandel ut fra Duncan bilder

Resultatbilde Figur 7.12 er konstruert ut fra regresjonsmodeller bygd på analyse av Duncan bildene. Fra ”Image Info” i Erdas finner jeg at middelverdiene for pikslene er 1,285 og standardavviket 4,077. Minste verdi er 0 og høyeste er 47.

7.3.2 Kodak



Figur 7.13 Predikert lavandel ut fra Kodak bilder

Resultatbilde Figur 7.13 er konstruert ut fra regresjonsmodeller bygd på analyse av Kodak bildene. Middelverdiene for pikslene er 18,065 og standardavviket 31,995. Minste verdi er 0 og høyeste er 255.

7.3.3 IR-bilde



Figur 7.14 Predikert lavandel ut fra IR-bildet

Resultatbilde Figur 7.14 konstruert ut fra regresjonsmodeller bygd på analyse av Kodak bildene. Middelverdiene for pikslene er 23,908, standardavviket 44,453. Minste verdi er 0 og høyeste er 255.

8 Vurdering av metodene/Diskusjon

8.1 Geometrisk korreksjon

Ved en geometrisk oppretting kan en aldri forvente å få en nøyaktighet som er bedre enn nøyaktigheten til kilden referansepunktene hentes fra. Dette gjør at ved bruk av satellittbilder eller flybilder er en avhengig av at den som har korrigert referansebildet, har gjort en bra jobb. Samtidlig vil nøyaktigheten også påvirkes av hvor god oppløsning bildene har.

Modellen som ble valgt i opprettingen av bildene er som nevnt i kapittel 7.1.2 en 1.ordens polynom funksjon. Den ble antatt god nok til å tilfredsstille kravene til opprettingen i denne oppgaven sammen med 5-6 kontroll punkter. Selv om det ikke trengs mer enn 3 punkter i en 1.ordens polynom funksjon [5] er flere punkter å foretrekke. Dette for å få overbestemmelse. Ved bruk av overbestemmelse er det lettere å kontroller om de valgte referansepunktene er korrekt plassert eller ikke.

Ved overlay mellom flybilder og satellittbilder er det viktig at den geometriske nøyaktigheten til bildene er god nok til at informasjon kan overføres mellom to bilder og samtidig havne på samme sted. I vår sammenheng gjelder dette når vi skal overføre informasjon fra flybilder til Landsat bildet. Laven som identifiseres i flybildene må kobles mot riktig piksel i Landsat 7 bildet for at jeg skal klare å identifisere laven i andre piksler, hvor jeg ikke har treningsdata. Samtidig vil en forskyvning av plasseringen kunne føre til at jeg får beregnet feil lavmengde i et området. Feil her gjør at jeg får feil treningsdata i byggingen av regresjonsmodellene, som igjen gjør at prediksjonen av lav andre steder i bildet blir feil.
Forskyves et flybilde en halv Landsat piksel i forhold til Landsat bildet får vi en feil klassifisering på 50% i forhold til hva det skulle ha vært. Dette gjør at det kan bli store feil i den predikerte lavmengden. For oss vil dette ha noe å si ved at lav andelen som blir koblet opp mot et punkt (Landsat piksel) blir feil i forhold til den reelle mengden, og dermed vil modellene predikere lav på feil grunn lag.

”Nøyaktigheten på Landsat bildet som er blitt benyttet kan ikke forventes er noe bedre enn 20-30 meter. Det samme gjelder høyde modellen som er et utsnitt av Kartverkets digitale terrengmodell.” [personlig meddelelse fra Rune Ødegård , HiG]

8.1.1 Duncan

Disse bildene ble i all hovedsak rettet opp mot SPOT5 bildet.
Siden de fleste bilden måtte rettes opp mot SPOT bildet ble det vanskelig å tilfredstille det opprinelige kravet om at referanse punktene skulle ha en nøyaktighet på bedre enn 1 meter ettersom oppløsningen til SPOT5 er på 2,5 meter. Siden Duncan bildene har en oppløsning på ca 20cm vil en RMS på 5-6 piksler tilsvare ca. 100 – 120 cm. Noe som må sies å være bra for vår bruk.

8.1.2 Kodak

Hva de store RMS feilene for enkelte av Kodak bildene kan skyldes er vanskelig å si, men en feilkilde kan være store høyde variasjoner i terrenget som dekkes av bildene. Dette gjør at den 1.ordens polynom funksjon ikke klarer å fange opp de store variasjonene i terrenget innenfor område. En mulighet å fange opp dette på kan være å bruke flere referansepunkter, eller å bruke en høyere ordens polynom funksjon som tar hensyn til flere forhold, men på grunn av tidsnød ble ikke dette prøvd ut noe nærmere . En annen kilde som må tas med i vurderingen er hvor god opprettingen av de skannede IR bildene til Kartverket er. Men i utgangspunktet skulle en kunne anta at disse bildene var i orden slik at den kan utelates. Opprettingen av de resterende Kodak bildene mot SPOT5 så som tidligere nevnt, ut til å ha gått bra. Disse bildene fikk en nøyaktighet tilsvarende den som ble oppnådd for Duncan bildene. Men ved kontroll er det mulig å se at i områder med kupert terreng vil opprettingen bli litt for dårlig. Hvis referanse punktene som er valgt ligger oppe på høydedrag oppstår det litt større avvik nede i søkk og daler. Disse virkningene skyldes at linja som polynom funksjonen representere tar lite hensyn til store høyde variasjon, samt at det har litt å si hvor punktene plasseres.

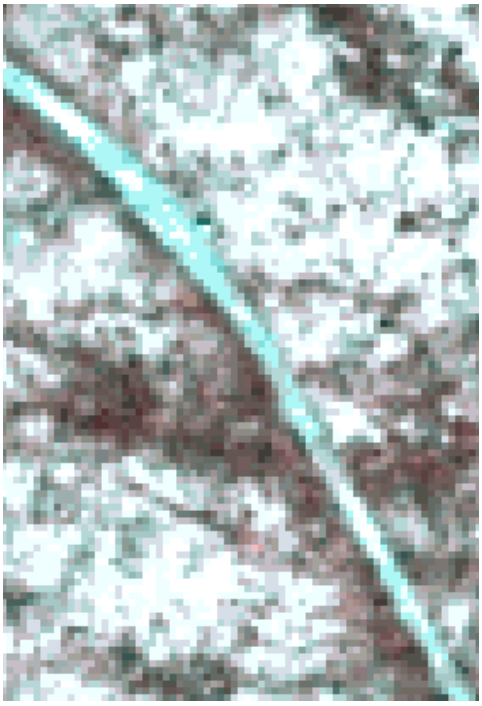
8.2 Klassifisering

8.2.1 Valg av klasser

Som nevnt i kapittel 7.1.3 måtte jeg velge 70 klasser i Duncan bildene for at det skulle være mulig å skille ut alle områdene som kunne være lav. Dette kan skyldes at bildene var blitt uskarpe under fotografering, eller at den spektrale refleksjonen fra de ulike områdene er så lik at de blir registrert som lik i det synlige spektret. Det tilsvarende problemet fikk jeg også i Kodak og IR bildene. I IR bildene var det godt synlig at det ville bli vanskelig å skille lav fra andre områder (Figur 8.1). Dette medførte at i klassifisering kom enkelte ikke-lav områder med i de samme klassene som lav og omvendt (Figur 8.2)



Figur 8.1 Uklassifisert bilde uv_ir_0_2



Figur 8.2 Klassifisert bilde 70kl. uv_ir_0_2

Ettersom det i enkelt områder kan være større områder med feil klassifisering, vil dette være med på å gi feil i regresjonsmodellen i Cubist. En modell som predikere for mye lav vil gjøre at vi får lav i piksler som ikke inneholder lav, eller at det blir antatt for mye lav i et Landsat piksel

Siden IR bildene er skannet fra analogt format kan dette ha påvirke fargene i bildene. Dette er gjelder spesielt hvis det er papirkopier av bildene, som har blitt skannet [17]. Faren med dette er at det innføres nye farge og kontrast variasjoner i bilene. Det er derfor å anbefale og bruke original filmen til skanning. Hvis dette skulle være tilfelle at bildene er skannet fra en farge kopi kan dette forklare noe av problemet med å skille lav og stein i bildene.

I Kodak bildene kom noen av de samme problemene som i IR bildene, nemlig å skille mellom lav, stein og jord. Selv med over 100 klasser var det umulig å se noen forskjell i enkelte områder. Dette gjøre at det blir vanskelig å stole på disse analysene så lenge en ikke har noe feltarbeid å kontrollere mot eller at en besitter gode kunnskaper i hvordan skille de forskjellige områdetypen i bildene. Siden Kodak bildene kun ”registrer” synlig lyset kan nok en del av feilene forklares på bakgrunn av dette. Grunnen til noe av dette ligger i at på avstand kan endel lav og stein typer se veldig like ut, slik at det blir vanskelig å skille de fra hverandre.

Laven forekommer ikke bare på homogene flater. Den kan også finnes som tynne lag opp stein og jord. Dette gjør at hvis laget er for tynt så kan de bli feil tolket ved at den i det synlige lyset blir registrert sammen med stein.

8.2.1.1 Skygger

Et problem som var synlig i flere flybilder var skygger. Dette gjorde at i enkelte bilder ble områder mørkere enn de ellers skulle ha vært. De samme effektene var å se i enkelt områder med hellende terreng og nede i daler. Dette gjør at det er vanskelig å stole blindt på den automatiske klassifiseringa. Det er derfor viktig med visuell kontroll av de klassifiserte dataene for å se om noen områder må flyttes til andre klasser enn de foreslåtte. I Landsat 7 bildet blir ikke disse effektene like synlig ved at de blir redusert ved at de integreres i signalet opp tilsatellitten.

En måte å kunne kontrollere hva de forskjellige områden er, er å se på deres spektralsignatur. Men for at det skal være mulig å finne ut noe ved hjelp av denne metoden er en avhengig av å kjenner signaturen til forskjellige planter og andre materialer. Er signaturen kjent kan en gå inn i en signatur editor (her: i Erdas) å foreta en klassifisering på signatur verdiene.

8.3 Vurdering av metoden

8.3.1 Regresjons modellene

Ut fra verdiene for korrelasjon i Tabell 7.1 ser vi at alle tre modellene har fått en bra tilpasning til referanse dataene. Det vil si at modellene klarer å predikere verdier som er veldig lik de som har blitt brukt til å bygge modellene. Dette gir en forhåpning om at regresjonstre er en godt egnet metode å modellere satellitt- og andre data på. Men ved sammenligning av reelle og predikerte verdiene for lav ser vi at modellene i enkelte tilfeller har en tendens til å underpredikere mengden lav i piksler, som skal inneholde mye lav. Dette kan nok skyldes at det i referanse grunnlaget er for få piksler med store lav mengder, slik at disse blir dårlig representert og dermed faller utenfor når modellene bygges. Dette kommer som følger av at det ikke bygges nok modeller til at alle forekomster av lav kan fanges opp og bli predikert rett. Det vil i stedet bli dannet modeller som går på kompromisser mellom forskjellige faktorer, noe som resulterer i modeller som gir en form for middelverdier.

8.3.1.1 Testdata

For å se hvor god modellene er, er det beste å benytte seg av et sett med testdata som ikke er har blitt brukt i modell byggingen. Som vi ser for disse modellen i Tabell 7.2 så kan det virke som om modellene ikke passer så godt for testdataene siden korrelasjons verdiene faller drastisk. Dette kan skyldes at ikke alle mulige forekomster av for eksempel vegetasjon har vært med i de bildene som ble brukt til modell byggingen, noe som gjør at modellene ikke passer på testdataene. For modellene bygd på data fra Duncan bildene er dette god synlig. Grunnen til at dette kan være at det er benyttet færre piksler i modell grunnlaget for Duncan, 1296 piksler, samt at det er liten variasjon i vegetasjon og terreng type. Sammenligner vi med IR er de regresjonsmodellene bygd på bakgrunn av 28440 piksler, og et mer variert terreng som går fra nede i Uvdal og opp til fjellet.

Studeres plotene for testdataene av de reelle og predikerte verdiene er det synlig at modellene er svært dårlig til å predikere for Duncan bildene. Noe som kan tyde på at de har piksel verdier som avviker noe fra de som modellene ble bygd for.
For Kodak bildet blir det to skikt i plotet. Et hvor det predikeres for lite lav og et hvor det ser ut til at det predikeres for mye lav. Ved nærmere studier av selve regresjonsmodellene og de predikerte verdiene fant jeg ut at det var to modeller som skilte seg ut når det gjaldt å predikere for mye lav.

I testplotet for IR får jeg en litt motsatt effekt av Duncan og Kodak. Her er det i utgangspunktet lite lav, men regresjonsmodellene predikere mye lav i nesten samtlige forekomster. Grunnen til denne overprediksjonen kan være at pikselverdiene og de tilhørende datalagene i området inneholder verdier som er veldig lik de som i modellbyggingen har blitt brukt til å predikere lav, eller at de oppfyller betingelser for de modellene som predikerer mye lav.

Dette viser hvor viktig det er med stort og variert referansegrunnlag for all modellbygging, og spesielt i slike tilfeller som i denne oppgaven hvor det er et så stort og variert område som skal modelleres.

8.3.2 Bilde analyse

Ved sammenligning av de tre resultat bilden kommer det fort fram at de tre settene med regresjonsmodeller ikke klarer å predikere like mengder med lav. Duncan er ganske moderat med genereringen av inntill 47% lav. Noe som skulle kunne være en reell mengde. Modellene fra Kodak overpredikerer ved å generere over 100% lav. I tillegg pressenteres piksel verdier som skulle tilsvare 255% lav. Ved nærmere studie av disse pikslene kom det fram at dette var piksler som hadde fått predikert en negativ lavandel av regresjonsmodellene, men ved kjøring av ”CubistSatNat” har de fått en ekstrem verdi for at de skal skilles fra restene av pikslene. De tilsvarende problemene som i Kodak dukket også opp i bildet med modeller fra IR bildene.

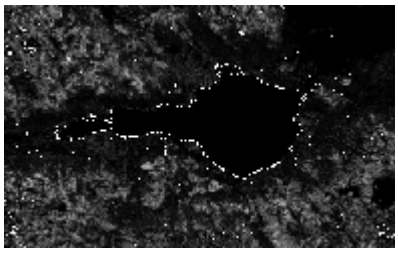
Duncan
Ved nærstudie av Figur 7.12 kan det se ut som om det er bare i sentrale deler av Hardangervidda hvor det har blitt predikert lav. Studeres bildet sammen med en høydemodell kommer det fram at mesteparten av laven har blitt predikert i en bestemt høyde over havet, 1200-1360 meter. Men det finnes som vanlig noen unntak i form av spredte piksler utenfor de sentrale delene av Hardangervidda. I det sør vestlige hjørnet er det striper som antyder omriss av terrengformasjoner som skiller seg ut.
Det samme kan ses ut fra betingelsene i regresjonsmodellene for Duncan bildene. Dette kan tyde på at siden de benyttede Duncan bildene ligger innen for et begrenset høyde intervall, så er det med på å begrense modellenes bruksområde. En annen attributt verdi som ser ut til å ha like stor betydning er vegetasjonskoeffisienten (vegetasjonsgeografiske soner). På samme måte som høydedataene har denne fått et begrenset intervall. Noe som å skylles at flybildene

de er hentet fra ligger i et lite heterogent område, når det gjelder vegetasjonsgeografiske soner.
Ser en på hvor det er blitt predikert lav, ser en at de områdene som har fått de største mengden lav ligger i sør – sørvest vendte. Grunnen til dette kan komme av at det er i disse områdene sola er sterkest og dermed sørger for de beste vekst vilkårene.

Kodak
Figur 7.13 ser ved første øyekast ut til å ikke ha noen uventede mønster, men ved nærmere studier kommer det fram at regresjonsmodellene har predikert lav der det i virkeligheten er snø. Et eksempel er på Hardangerjøkulen. Det ble også oppdaget at skyer har blitt beregnet som lav.

Skannede IR
Resultat bildet for IR Figur 7.14 viser et tydelig skille med en lys og en mørk del. Dette skillet deler bildet i nesten to like store deler. På den venstre delen ligger piksel verdiene stort sett under 50 mens de til høyre ligger over. En grunn til at vi får dette skillet kunne ha vært at Landsat 7 bildet som vi bruker egentlig er satt sammen av to scener. Slik feil kan fort oppstå ved at ikke alle ulikhetene i lys og refleksjon har blitt fjernet, ved sammenslåing av bildene. Men etter å ha sett på et bilde som viser overlappen mellom de to bildene, kommer det fram at skillet går lengre øst enn det skillet som har oppstått i lav bildet. Noe som kan tyde på at dette ikke er noe problem i denne sammenhengen. Skulle det ha vært et problem burde det også ha vært synlig på bildene fra de andre regresjonsmodellene.

Ved nærmere studie av bildet viser det seg at det er mange tilfeller av ekstreme verdier 247 – 255. Dette kan tyde på at regresjonsmodellene ikke har vært bra nok for disse tilfellene. I den østre delen bildet er det enkelte steder en systematikk i disse feile verdiene. Der er mulig å se at mange ekstreme piksler ligger rundt vann og viser konturene i form av hvite piksler med verdi opp mot 255 (Figur 8.3). En grunn til dette kan være stein og jord som ikke passer inn i modellene. I Figur 8.3 er nok dette feilprediksjonen på grunn av at det rundt vannet går en kant med bart fjell som følger av at vannet er regulert. Sammenlignes Figur 7.14 med Landsat 7 bildet er det synlig at områdene i øst som har fått predikert høy lavandel egentlig er områder med jord/stein og snø.



Figur 8.3 Feil prediksjon av lav rundt Sønstevatn, fra IR bilder

Ett problem som har blitt oppdaget i alle tre bildene er at skyer har blitt predikert som lav, men det er ikke like godt synlig i alle tre bildene. I lavbildet fra Kodak bildet har det blitt predikert store mengder lav i det sørvestre hjørnet hvor det ligger slør av dis over et stort område. For det tilsvarende området i IR bildet ser det ikke ut som om vi får den samme feilen. Hva dette kan skyldes er det vanskelig å si noe om, men en grunn kan være at regresjonsmodellene fra Kodak bildene legger større vekt på andre attributter en modellene fra IR bildene. Skyer eller dis registreres i Landsats bånd 5, slik at hvis modellen

legger stor vekt på dette båndet så er det stor mulighet for at det kommer til syne i de predikerte bilde.

Den totale mengden lav som blir predikert i alle tre regresjonsmodeller ser ut til bli større enn den reelle mengden, det kommer blant annet av at modellene har lett for å predikere lav der det ikke skal være lav, som for eksempel i vann og på snø.

8.3.3 Forbedringer i metoden

En ting som må forbedres i forhold til fleksibiliteten er muligheten for å legge til og fjerne datalag i skriptene, slik at det blir lettere å eksperimentere med ulike kombinasjoner. Slik det var nå ble det vanskelig på grunn av at skriptene ble skrevet for kun å passe de valgte datalagene. En bør også se på muligheten for å redusere tiden det tar å utføre en del av opprasjonene for databehandlingen, dette gjelder spesielt predikeringen av lav på et helt satellittbilde. I analysen av satellittbildet i CubistSatNat med utgangspunkt i de 28440 pikslene fra IR, brukte programmet ca. 48 timer (på en PIII 400mhz med 256MB ram). Ser her at hvis en skulle ha brukt flere piksler så ville tidsforbruken ha økt yterligere. Redusert tidsforbruk er en viktig siden det ofte er snakk om store datamengder som skal behandles, og skal en rekke å ta med alle dataene i analysen må de forskjellige opprasjonene gå raskt.

Dette med en fleksibel bruk av forskjellige datalag kommer inn i tilfeller som for eksempel regresjonsmodellene til Duncan bildene. For disse regresjonsmodellene ser det ser ut til at høydeatributten er med på å begrense muligheten for å predikere lavmengden i områder som ligger i andre høyder enn de som finnes for testdataene. Et tilsvarende problem er også gjeldene for vegetasjonskoeffisienten, selv om den er en sterk indikator på hva slags vegetasjon som kan finnes på forskjellige steder. Er det mulig å forsøke flere kombinasjoner av datalag kan det være mulig å oppnå bedre regresjonsmodeller ved å bruke andre kombinasjoner av datalag eller å fjerne datalag som kan se ut til å ha liten betydning.

For å unngå å få med uønskede områder i analysen kan et regresjonstre trenes til å predikere null lav for områder som med sikkerhet kan sies ikke er lav uten å ødelegge prediksjons nøyaktigheten over andre områder [7]. Dette kan være områder som vann og veier. For å få dette til må det lages masker som tilsvarer null prosent lav. Maskene kan lages ved en styrt eller ikke styrt klassifisering. En annen ting som bør ses på er muligheten for å fjerne eller redusere virkningen av skyer i satellittbildet.

9 Konklusjon

Som nevnt i hypotesen (kap.1.1) for denne oppgaven var et av målene i denne oppgaven å se på muligheten for å benytte Landsat 7 data til å kartlegge reinens tilbud av lavbeite. Resultatet av de analysene som er gjort i viser at muligheten er tilstede, men for at det skal være mulig å gi et konkret svar kreves det som nevnt i kapittel 8.3.3, utprøving av en del forbedringer av metoden. Før dette er gjort kan ikke hypotese 1 forkastes.

Når det gjelder bruken av flyfoto som treningsdata, ser det ut fra mine undersøkelser til at det er mulig, men for at resultatet av tolkningen skal bli bra er en avhengig av å ha god kunnskap om hvordan ting på bakken gjengis i ulike bildetyper.

Hvilken av de tre flybilde typene som er best egnet til bruk som treningsdata er ikke lett å si siden de ga opphav til tre så vidt forskjellige lavkart.

Den endelige konklusjonen på dette sier at det her må det undersøkes mer før enn kan ta dette inn i mer regulær bruk.

10 Litteratur:

[1] Lieng, E. og Kastdalen, L., 2002, ”Produksjon av arealdekkekart fra satellittdata ved hjelp av flybårne digitale kamera”. Kartdagene 13.-15.mars 2002

[2] Lindgaard, A., Klokk, T., Bolstad, J.P. og Kastdalen, L. 2002
”Mapping and Monitoring Biodiversity in Norway – use of satellite data (SatNat)”

[3] ERDAS Imagine 8.6 - Tour Guide, 2003, ”Frame Sampling Tools” s.521-557

[4] Hjemmesiden til USGS Landsat prosjekt:
http://landsat7.usgs.gov/programdesc.html

[5] Mehmed Kantardizic, (2003). ”Data Mining: Concepts, models, methods and Algorithms” s.139-153. ISBN 0-471-225852-4

[6] ”ERDAS Imagine 8.6 - Field Guide”, 2003, Unsupervised Training
s. 213-217

[7] Richards, J.A. og Jia, X.,1999, ”Remote Sensing Digital Image Analysis” 3rd ed Springer

[8] “IMAGINE Subpixel Classifier White Paper”.
http://www.discover-aai.com/imsp_wp.htm, Sist besøkt 28/5-03.

[9] Yang, L m.fl., 2002, ”An approach for mapping large-area impervious surfaces: Synergistic use of Landsat 7 ETM+ and high spatial resolution imagery”
http://landcover.usgs.gov/publications.html Sist besøkt 12/3-03

[10] Huang, C. m.fl., 2001, ”A strategy for estimating tree canopy density using Landsat 7 ETM+ and high resolution images for large areas”
http://landcover.usgs.gov/publications.html Sist besøkt 12/3-03

[11] Gaare, E. og Tømmervik, H., 2000
“Overvåking av lavbeite i Finnmark” – NINA Oppdragsmelding 638.

[12] Lieng, E., Slaymaker, A.M. and Kastdalen, L. …. “Use of airborne small-format digital cameras as tools for collecting ground truth”

[13] Dietz, H. og Steinlein, T., 1996, “Determination of plant species cover by means of image analysis”. Journal of Vegetation Science 7: 131-136.

[14] Nordberg, M. L. og Hemström, P.
”Studier av linjära strukturer i fjällen med hjälp av satellitbilder”
http://www.natgeo.su.se/naturgeo/rsgis/vegchange/mountain/pilot/
sist besøkt 10/4-03

[15] Jensen, J.R., 2000, ”Remote sensing of the environment. An earth resource perspective”. Prentice Hall.

[16] http://www.kodak.com/global/en/professional/products/cameras/dcs760/dcs760Index.jhtml sist besøkt 30/4-03

[17] Moen, A. og Odland, A., 2002, ”Retningslinjer for detaljert vegetasjonssonekart, Sør-Trøndelag som prøvefylke” Notat

[18] ”An Overview of Cubist”, http://rulequest.com/cubist-win.html sist besøkt 22/5-03

[19] Vanha-Majamaa, I. m.fl., 2000, “Digitized photographs in vegetation analysis – a comparison of cover estimates”. Applied Vegetation Science 3: 89-94,

[20] ”ERDAS Imagine 8.6 - Field Guide”, 2003, “Polynomial Transformation” s. 339-340

H. Tømmervik, G. Bergersen, L. Kastdalen, T. Brobakk og E. Lieng. 2003 Utkast til ”Kartlegging av reinbeite i Forelhogna villreinområde”, NINA, I trykk

“Subpixel Impervious Surface Mapping” - http://resac.uconn.edu/publications/tech_papers/index.html

Vedlegg

Vedlegg A

Perl opptelling av klasser i data filer.

```
# NB! Sjekk desse for start ------------------------------

$dirPath = 'D:\HovedOppg\Cubist_Modeller\Duncan\Data';
$OutFileName = 'D:\HovedOppg\Cubist_Modeller\Duncan\Data\21150.log';
$prefix = "data_";

# NB! Sjekk desse for start ------------------------------

$lastElem = 75;

unless (open(OutFILE, ">$OutFileName"))
{
  die("Cant open output file $OutFileName");
}

# open selected dir
opendir(InDIR, $dirPath) || die("Cant open current directory");

$count = 1;
@inFiles = readdir(InDIR); # les alle filnamna inn i ei array naa,
                    # ellers vil vi faa dei nye filnamna inn seinare

for ($cou = 0; $cou<=$lastElem; $cou++) { $dataArr[$cou] = 0; }

# skriv kolonne titel ----------------------
print OutFILE ("Id\t");
for ($cou = 0; $cou <= $lastElem; $cou++){ print OutFILE ("P$cou\t");}
print OutFILE ("RUPE\n");

# tel opp  ----------------------------------
foreach $inFileName ( @inFiles )
{
  $bDropp  = 0;
  next if ( $inFileName !~ /\.txt/i  );
  $dataFile = $dirPath . "\\" . $inFileName;
  unless (open(InFILE, $dataFile)){die("Cant open input file $inFileName");}
  $recno = $inFileName;
  $recno =~ s/\.txt//i;
  $recno =~ s/$prefix//i;

  # leser forbi header
  while($line = <InFILE>) { last if ( $line =~ /NODATA_value/i  ); }

  while($line = <InFILE>) # les inn ei linje fr fila til fila er tom.
  {
    chop($line); # fjern siste karakter i linja dvs. linjeskift
    @lin = split(/[ ]/, $line); #splitt inn-tekst ved tabulator eller space
    foreach $elem ( @lin )
    {
      if( ( $elem == -9999) || ($elem == 0) )
      {
        $bDropp = 1;
        last;
```

```
        }

        if( ($elem >= 0) && ($elem <= $lastElem ) )
        {
          $dataArr_1[$elem]++;
        }
      }
    }
    close(InFILE);
    if ($bDropp == 1)
    {
      for ($cou = 0; $cou<=$lastElem; $cou++) { $dataArr_1[$cou] = 0; }
      next;
    }
    $sum = 0;
    for ($cou = 0; $cou<=$lastElem; $cou++)
     {
      $dataArr[$cou] = $dataArr[$cou] + $dataArr_1[$cou];
      $sum = $sum + $dataArr[$cou];
     }
    #print("$sum\n");
    for ($cou = 0; $cou<=$lastElem; $cou++) { $dataArr_1[$cou] = 0; }

    # skriv ut arrayen med recno fyrst
    print OutFILE ("$recno\t");
    for ($cou = 0; $cou <= $lastElem; $cou++)
    {
      print OutFILE ("$dataArr[$cou]\t");
      $dataArr[$cou] = 0;
    }
    print OutFILE ("\n");

    $count++;
}

close(OutFILE);
```

Vedlegg B

Perl script som teller summerer antallet lav piksler og beregner prosent andel i et Landsat 7 piksel:

```
$dirPath = "D:\\HovedOppg\\Cubist_Modeller\\493\\";
$OutFileName = 'D:\HovedOppg\Cubist_Modeller\493\prosentAndel.log';

%param = ("0_7.log"=>[63,66,67,69,70,71]);

unless (open(OutFILE, ">$OutFileName"))
{
  die("Cant open output file $OutFileName");
}

printf OutFILE ("Id\tFilnamn\tlav_pros\n");

$count = 0;
foreach $fil ( keys %param )
{
   @ara = @{$param{$fil}}; # finn parameter liste som er lav
   $FullFil = $dirPath.$fil;
   open(InFILE, "$FullFil") || die "kan ikke opne $fil";

   $line = <InFILE>; # fjernar forste linje som er header
   while($line = <InFILE>)
   {
      @data = split(/[\t]/, $line);

      # finn summen av anatll pixel totalt
      for($j=1;$j<@data;$j++)
      {
         $tot = $tot + $data[$j];
      }

      # finn summen av anatll pixel med lav
      for($i = 0; $i<@ara; $i++)
      {
         $lav = $lav + $data[$ara[$i]];
      }
      $prosLav = ($lav/$tot)*100;

      printf OutFILE ("%s\t%s\t%4.2f\n",$data[0],$fil,$prosLav);

      $tot = 0; $lav = 0; $prosLav = 0;

   }
   close(InFILE);

  $count++;
}
close(OutFILE);

print "end prog\n";
```

Vedlegg C

Perl script for kobling av IDer og danning av data-fil til bruk i Cubist:

```
$dirIdFiler = 'D:\HovedOppg\Cubist_Modeller\kodak\ID';
$ErdasData  = 'D:\HovedOppg\Cubist_Modeller\kodak\ErdasData';
$inFileLav  = 'D:\HovedOppg\Cubist_Modeller\kodak\prosentAndel.log';

$OutFileName   = 'D:\HovedOppg\Cubist_Modeller\kodak\Alle.log';
$OutCubistName = 'D:\HovedOppg\Cubist_Modeller\kodak\Alle.data';

open(OutFILE, ">$OutFileName") ||  die("Cant open output file $OutFileName");

# open selected dir
opendir(InDIR, $dirIdFiler) || die("Cant open $dirIdFiler directory");

@inFiles = readdir(InDIR); # les alle filnamna inn i ei array naa,
                    # ellers vil vi faa dei nye filnamna inn seinare
$count = 0;
foreach $inFileName ( @inFiles )
{
  next if ( $inFileName !~ /\.txt/i  );
  $dataFile = $dirIdFiler . "\\" . $inFileName;
  open(InFILE, $dataFile) || die("Cant open input file $dataFile");

  $line = <InFILE>; # hopp over forste linje
  while($line = <InFILE>) # les inn ei linje fr fila til fila er tom.
  {
    chop($line); # fjern siste karakter i linja dvs. linjeskift
    @data = split(/[\t]/, $line); #splitt inn-tekst ved tabulator eller space
    $databaseId[$count] = [$data[0],$data[1],$data[2],$data[3]];
    $count++;
  }
  close(InFILE);
}
closedir(InDIR);

# Les inn erdas datafiler
# open selected dir
opendir(InDIR, $ErdasData) || die("Cant open $dirIdFiler directory");

@inFiles = readdir(InDIR); # les alle filnamna inn i ei array naa,
                    # ellers vil vi faa dei nye filnamna inn seinare
$count = 0;
foreach $inFileName ( @inFiles )
{
  next if ( $inFileName !~ /\.asc/i  );
  $dataFile = $ErdasData . "\\" . $inFileName;
  open(InFILE, $dataFile) || die("Cant open input file $dataFile");

  for(0..6)
          {
                  $line = <InFILE>;
          } # hopp over forste linje
  while($line = <InFILE>) # les inn ei linje fr fila til fila er tom.
  {
    chop($line); # fjern siste karakter i linja dvs. linjeskift
                  $line =~ s/^[ \t]+//; #fjernear spase fyst paa linja
    $line =~ s/[ \t]+\n$/\n/; #fjernar space og tab sist paa linja
```

```
    $line =~ s/[ \t]+/ /g; #byttar ut alle tab og space med ein space
    @data = split(/[ ]/, $line); #splitt inn-tekst ved tabulator eller space
    $databaseEr[$count] = [$data[0],$data[1],$data[2],$data[3],$data[4]

,$data[5],$data[6],$data[7],$data[8],$data[9],$data[10],$data[11]];
    $count++;
  }
  close(InFILE);
}
closedir(InDIR);


# Les inn lav mengder og ID med lav
open(InFILE, $inFileLav) || die("Cant open input file $inFileLav");

$line = <InFILE>; # hopp over forste linje
$count = 0;
while($line = <InFILE>) # les inn ei linje fr fila til fila er tom.
{
        chop($line); # fjern siste karakter i linja dvs. linjeskift
        @data = split(/[\t]/, $line); #splitt inn-tekst ved tabulator eller space
        $databaseLav[$count] = [$data[0],$data[1],$data[2]];
        $count++;
}
close(InFILE);


#-------------------------------------------------------------

for($i=0;$i<@databaseId;$i++)
{
  for($j=0;$j<@databaseEr;$j++)
  {
                $x1  = $databaseId[$i][1];
                $x2  = $databaseEr[$j][0];
                $y1  = $databaseId[$i][2];
                $y2  = $databaseEr[$j][1];
                $Id1 = $databaseId[$i][0];
                $F1  = $databaseId[$i][3];
                $F1 =~ s/\.shp//i;
                $F1 =~ s/Pol_//i;

                if( ($x1 == $x2) && ($y1 == $y2) )
                {
                        for($o = 0;$o<@databaseLav;$o++)
                        {
                                $Id2 = $databaseLav[$o][0];
                                $F2  = $databaseLav[$o][1];
                                $F2 =~ s/\.log//i;
                                if( ($Id1 == $Id2) && ($F1 eq $F2) )
                                {
                                        $maxAr = $databaseId[$i];
                                        $max = @$maxAr;
                                        for($p = 0;$p<$max;$p++)
                                        {
                                                print OutFILE ("$databaseId[$i][$p]\t");
                                        }
                                        $maxAr = $databaseEr[$j];
                                        $max = @$maxAr;
                                        for($p = 0;$p<$max;$p++)
```

```
                                        {
                                                print OutFILE ("$databaseEr[$j][$p]\t");
                                        }
                                        print OutFILE ("$databaseLav[$o][2]\n");
                                }
                        }
                }
  }
}

close(OutFILE);
sleep(5);

# Les inn resultatfila for aa lage Cubist data filer
open(OutFILE, $OutFileName) || die("Cant open input file $OutCubistName");
open(SluttFILE,">$OutCubistName") || die("Cant open input file $OutCubistName");

#$line = <OutFILE>; # hopp over forste linje
$count = 0;
while( defined($line = <OutFILE>)) # les inn ei linje fr fila til fila er tom.
{
        chop($line); # fjern siste karakter i linja dvs. linjeskift
        @data = split(/[\t]/, $line); #splitt inn-tekst ved tabulator eller space
        print SluttFILE ("$data[0],$data[6],$data[7],$data[8],$data[9],$data[10],");
        print SluttFILE ("$data[11],$data[12],$data[13],");
        printf SluttFILE ("%.4f,%.4f,%.2f\n",$data[14],$data[15],$data[16]);
        $count++;
        if($count == 860)
        {
                print("stop");
        }
}
close(OutFILE);
close(SluttFILE);
```

Vedlegg D

ArcView for uttrekk av flyfoto piksler innen for et område tilsvarende et Landsat 7 piksel:

```
fn_name = "Pol_21150.shp"
gridName = "Grid_21150"
outPrefix = "D:\HovedOppg\Cubist_Modeller\Duncan\Data\Data_"


' -------------------------------------------------------
' hentar ut pixlverdiane i bufferen til ei fil
' -------------------------------------------------------

' set the projection to null since graphics are projected
theView = av.GetActiveDoc
theprojection=Prj.MakeNull

thmBuf       = theView.FindTheme(fn_name)
ftbBuf       = thmBuf.GetFTab
fldShapeBuf  = ftbBuf.FindField("shape")
fldIdBuf     = ftbBuf.FindField("Id")


thmThemeIn = theView.FindTheme(gridName)
resGrid = thmThemeIn.GetGrid

count = 0
minVal = 1
maxVal = 3000
For each rec in ftbBuf
  count = count + 1

  if(count < minVal) then
    continue
  end

  tpoly = ftbBuf.ReturnValue(fldShapeBuf,rec)
  Id    = ftbBuf.ReturnValue(fldIdBuf,rec)
  if ( tpoly.Is(Polygon) ) then
    Grid.SetAnalysisExtent(#GRID_ENVTYPE_VALUE, tpoly.ReturnExtent)
    r = resGrid.ExtractbyPolygon(tpoly, theProjection, FALSE)
    outFile = outPrefix
    outFile = outFile + Id.AsString +".txt"
    r.SaveAsASCII(outFile.AsFilename)
  end
  if(count >= maxVal) then
    break
  end
end
```

Vedlegg E

ArcView script for uttrekk av koordinater i et rutenett med ID:

```
fn_InFile = "Pol_38654.shp"
strOutFile ="D:\HovedOppg\Cubist_Modeller\Duncan\testdata\Data_38654.dbf"

theView    = av.GetActiveDoc
thmPoly    = theView.FindTheme(fn_InFile)
ftbPoly    = thmPoly.GetFTab

' make new theme
fileNameOut   = FileName.Make(strOutFile)
ftbOutput = VTab.MakeNew( fileNameOut, dBASE )
ftbOutput.AddFields({Field.Make("ID",#FIELD_LONG, 16, 0)})
ftbOutput.AddFields({Field.Make("X_coord",    #FIELD_LONG, 25, 0)})
ftbOutput.AddFields({Field.Make("Y_coord",    #FIELD_LONG, 25, 0)})
ftbOutput.AddFields({Field.Make("FileName",   #FIELD_CHAR, 25, 0)})

' find output field
fldIdOut     = ftbOutput.FindField("ID")
fld_X_Out    = ftbOutput.FindField("X_coord")
fld_Y_Out    = ftbOutput.FindField("Y_coord")
fld_Name_Out = ftbOutput.FindField("FileName")


' find polygon field
fldShapePoly   = ftbPoly.FindField("shape")
fldID_Poly     = ftbPoly.FindField("Id")


For each rec in ftbPoly
  id_poly  = ftbPoly.ReturnValue(fldID_Poly,rec)
  aPolygon = ftbPoly.ReturnValue(fldShapePoly,rec)
  aPoint = aPolygon.ReturnCenter
  X = aPoint.GetX
  Y = aPoint.GetY
  newPt=ftbOutput.AddRecord
  ftbOutput.setvalue(fldIdOut,  newPt, id_poly)
  ftbOutput.setvalue(fld_X_Out,  newPt, X)
  ftbOutput.setvalue(fld_Y_Out,  newPt, Y)
  ftbOutput.setvalue(fld_Name_Out,newPt, fn_InFile)

End
```

Vedlegg F

Lista under viser hvilke klassifiserte bilder som er brukt med hvilke klasser som er lav.
I kolonnen ”Brukte bilder” betyr M at bildet er brukt til å bygge enn regresjonsmodeller, mens T betyr at bildet er brukt til å teste modeller.

Filnavn	**Shp-fil**	**Kla**	**sser**	**med**	**lav**				**Brukte bilder**	I bilder det ikke står noe bak, er det ikke funnet lav
geo3953713_70.img	pol_39713.shp	69	70							
geo2357956_70.img	pol_23956.shp								**M**	
geo2225668_70.img	pol_22668.shp								**M**	
geo2157152_70.img	pol_21152.shp	65	66	67	69					
geo2084552_70.img	pol_20552.shp	63	66	67	69	70			**M**	
geo1982963_70.img	pol_19963.shp	56	61	67	68	69			**M**	
geo1967020_70.img	pol_19020.shp	62	63	67					**M**	
geo1866750_70.img	pol_18750.shp	61	64	67	68	69	70			
aug22-5770451_70.img	pol_57451.shp	3	4	5	9	15				
aug22-5754640_70.img	pol_57640.shp	60								
aug22-3631312_70.img	pol_3631312.shp	63	67							
aug22-3302434_70.img	pol_33434.shp	58	63	67	68	69	70			
aug22-5786123_70.img	pol_57123.shp	5	12	18	25				**M**	
aug22-5831852_70.img	pol_58852.shp	3	5	9	14	26			**T**	
aug22-5962954_70.img	pol_59954.shp	58	62	63	65	68			**M**	
aug22-5980207_70.img	pol_59207.shp	57	60	63	68					
aug22-6055319_70.img	pol_60319.shp	7	10	15	20				**M**	
aug23-1964385_70.img	pol_19385.shp	58	61	65	68	70				
aug23-2144897_70.img	pol_21897.shp	60								
aug23-2165454_70.img	pol_21454.shp	65							**M**	
aug23-2191150_70.img	pol_21150.shp	3	4	12					**T**	
aug23-2214599_70.img	pol_22599.shp	63	64	67	70					
aug23-2227116_70.img	pol_22116.shp	66	68	69	70					
geo4151883_70.img	pol_41883.shp	63	64	66	67	69	70		**M**	
geo4120128_70.img	pol_41128.shp	2	8							
geo4060441_70.img	pol_40441.shp	59	63	64	67	68	69		**M**	
geo4035800_70.img	pol_40800.shp	61	68							
geo3995086_70.img	pol_39086.shp	1	3	15						
geo3780843_70.img	pol_37843.shp									
geo3748430_70.img	pol_37430.shp								**M**	
geo2524107_70.img	pol_25107.shp								**M**	
geo2493802_70.img	pol_24802.shp									
geo2480501_70.img	pol_24501.shp	1	2	3	4	5	6			Alle klassene er vanskelig å skille fra stein
geo3891654_70.img	pol_38654.shp	61	64	65	66	68	69	70	**T**	
500_70.img	pol_500.shp	66	67	68	69					
510_70.img	pol_510.shp	64	66	67	68	69	70			
515_70.img	pol_515.shp	63	64	66	68	70				
520_70.img	pol_520.shp	66	68	69	70				**M**	
530_70.img	pol_530.shp	64	66	67	69					
610_70.img	pol_610.shp	70								
621_70.img	pol_621.shp	65	67	69	70				**M**	
630_70.img	pol_630.shp	64	66	68	69					
640_70.img	pol_640.shp	67	68	69						70 har fellestrekk

										med vei
649_70.img	pol_649.shp	69	69	70						70 har fellestrekk med vei
660_70.img	pol_660.shp	64	65	69						70 har fellestrekk med vei
671_70.img	pol_671.shp	65	66	67	68	69				70 har fellestrekk med vei
680_70.img	pol_680.shp	69								70 har fellestrekk med vei
690_70.img	pol_690.shp									
699_70.img	pol_699.shp									
493_70.img	pol_493.shp	66	67	68	69	70			**T**	70 har fellestrekk med vei
uv1_1_70.img	pol_1_1.shp	65	67	68	69	70				
uv1_10_70.img	pol_1_10.shp	63	68	69	70					
uv1_15_70.img	pol_1_15.shp									
uv1_4_70.img	pol_1_4.shp	62	63	67	69	70				
uv0_6_70.img	pol_0_6.shp	60	64	67	68	69	70			
uv1_8_70.img	pol_1_8.shp	64	67	69	70					
uv0_0_70.img	pol_0_0.shp	61	63	65	68	69	70			
uv1_13_70.img	pol_1_13.shp	64	66	68	69	70				
uv0_13_70.img	pol_0_13.shp	65	68	69	70					
uv0_16_70.img	pol_0_16.shp	65	66	68	69	70				
uv0_2_70.img	pol_0_2.shp	62	63	64	68	69	70			
uv0_4_70.img	pol_0_4.shp	61	62	63	64	67	68	69	70	
uv0_7_70.img	pol_0_7.shp	62	65	66	68	69	70			**M**
uv0_9_70.img	pol_0_9.shp	62	65	66	67	68	69	70		