



Utredning 2006-5

Satellittdata til kartlegging av arealdekke

Utpøving av tilgjengelige kartdata for klassifisering av Sør-Trøndelag



Miljøsamarbeid



Naturområder
og arealbruk



Dyr og planter



Friluftsliv



Satellittdata til kartlegging av arealdekke

Utprøving av tilgjengelige kartdata for klassifisering av Sør-Trøndelag

Utredning 2006-5

Utgiver:

Direktoratet for naturforvaltning

Dato:

August 2006

Antall sider:

59

Emneord:

Arealdekke, satellittdata, kartdata, datamining, Sør-Trøndelag

Keywords:

Land cover, satellite data, digital map data, data-mining techniques, Sør-Trøndelag county

Bestilling:

Direktoratet for naturforvaltning,
7485 Trondheim
Telefon: 73 58 05 00
Telefaks: 73 58 05 01
www.dirnat.no/publikasjoner

TE 1199

Refereres som:

Lieng, E., L. Kastdalen og
J.P. Bolstad. Satellittdata til kart-
legging av arealdekke. Utprøving av
tilgjengelige kartdata for
klassifisering av Sør-Trøndelag.
DN utredning 2006-5.
Copyrightangivelse for Landsat
satellittbilder vist i rapporten:
©NASA, USGS.

Forsidefoto:

Utsnitt av SatNat arealdekkekart
for Sør-Trøndelag og Nord-Østerdalen

Ekstrakt:

Dette prosjektet er en del av SatNat-programmet, et samarbeid mellom Norsk Romsenter og Direktoratet for naturforvaltning. I produksjon av arealdekkekart for Sør-Trøndelag fylke og Nord-Østerdalen er dataminingsteknikker som beslutningstrær valgt som klassifikasjonsmetodikk etter en metode utarbeidet ved Laboratory for Global Remote Sensing Studies (University of Maryland, USA). Klassifikasjonen baserer seg på en multitemporal analyse med opptil tre satellittbilder (vår, sommer og høst). Andre datalag slik som vegetasjonsindekser og ulike rastere avledet fra terrengmodeller er brukt direkte i klassifikasjonen.

Sør-Trøndelag fylke og Nord-Østerdalen var meget utfordrende å kartlegge fordi området har stor landskapsmessig variasjon og dermed liten homogenitet. Dessuten har trøndelagsområdene mye skyet vær slik at det foreligger få optimale satellittbildeopptak. Som referansedata for trening i klassifikasjonen ble en rekke eksisterende kartdatabaser hentet inn: N50 kartdata, DMK, 3Q, Landsskogtakseringen, bestandsskogtakseringer og ulike vegetasjonskart. I undersøkelsen ble det også samlet inn egne referansedata i felt og gjennom visuell tolkning av fly- og satellittbilder.

Det er avdekket flere svakheter ved bruk av eksisterende kartdata som referansekilde. Graden av romlig generalisering er ofte for stor, slik at det blir vanskelig å fange opp noen trend i dataene. Vegetasjonsklassifiseringen er bygd opp rundt en plantesosiologisk legende der karakterarter er styrende for mange av klasseinndelingene. Med satellittdata er det ikke mulig å fange opp enkeltarter. Bruken av slike referansedata blir dermed lite effektiv, og gir en tematisk inndeling sterkt begrenset av at klassifikasjonen ikke er tilpasset egenskapene til optiske sensorer. Likevel, selv om eksisterende referansedata inneholder både feil og romlige generaliseringer gjør dataminingsteknikken det mulig å utnytte slike data som kilde for treningsdata til klassifikasjon av satellittbilder. Men klassene som kartlegges blir et resultat av informasjonsinnholdet i referansedataene. I dette prosjektet laget vi seks temakart, som ble satt sammen til ett arealdekkekart bestående av 32 klasser. Alle kartdata fra prosjektet kan fritt benyttes og lastes ned fra hjemmesiden til Direktoratet for naturforvaltning (www.dirnat.no/satnat).

Abstract:

This project is a part of the SatNat program, a joint effort between the Norwegian Space Centre and the Norwegian Directorate for nature management. Data mining techniques such as decision and regression trees have been utilized in order to produce land cover maps of Sør-Trøndelag county and Nord-Østerdalen from a methodology established by Laboratory for Global Remote Sensing Studies (University of Maryland, USA). The classification is based on multitemporal analysis of up to three satellite images. Ancillary data such as vegetation indices, DEM derived raster layers and more is applied directly to the classification.

Sør-Trøndelag County is extremely challenging to map due to a very diverse landscape and low homogeneity. A short growing season and frequent cloud coverage leaves few optimal satellite images available.

Numerous map data bases were used as training for the classification. Topological maps, land inventory maps, forest stand maps, national forest inventory samples and vegetation maps. Reference data based on field observations and visual image interpretation of aerial and satellite images was collected in the project as well. Existing map data are not optimal as a source of reference data. It has proven difficult to extract information about the sought classes since the degree of spatial generalisation is mainly too big. Vegetation maps are based on a plant sociological legend where character species decides the class hierarchy. Satellite data can not identify single species. The use of this kind of reference data is therefore inefficient and the resulting map legend will be strongly restrained since the classification is not adapted to optical sensors. However, the use of data mining techniques shows that we can use reference data as a source for classification of satellite data although it contains both errors and spatial generalisation.

The mapped classes will be a result of the information in the reference data. In this project we made six thematic maps which we combined to a land cover map containing 32 classes. All map data from the project may freely be used and downloaded from the web-site of Norwegian Directorate for Nature Management (www.dirnat.no/satnat).

Forord

Norsk Romsenter (NRS) og Direktoratet for naturforvaltning (DN) etablerte i 2001 et 5-årig program (SatNat) for å utvikle produkter og produksjonsgang for bruk av satellitt-data innen naturforvaltningen. Programmet er inndelt i flere prosjekt, hvorav et prosjekt omhandler metodeutvikling for å lage fylkesvise arealdekkkart. Til dette formålet ble Sør-Trøndelag og Østfold fylker valgt som prøveområder. Denne rapporten beskriver analysene foretatt i Sør-Trøndelag og deler av Nord-Østerdalen. Arealdekkkartleggingen av Østfold er rapportert i DN-utredning 2004-3.

Arbeidet er gjennomført av Einar Lieng (Høgskolen i Hedmark), Leif Kastedalen (Norsk Romsenter/ Høgskolen i Hedmark) og Jan-Paul Bolstad (Direktoratet for naturforvaltning). Til analysene er det benyttet data fra ulike institusjoner. Satellittbilder (IRS-1C, SPOT og Landsat ETM+) og N50 kartdata er levert av Statens Kartverk og Geodatasenteret AS. Bestandsskogsdata for ulike kommuner i Sør-Trøndelag er velvillig stilt til disposisjon fra Skogeierforeningen Nord, mens NIJOS har bidratt med data fra Landsskogtakseringen og

3Q for Sør-Trøndelag. NTNU Vitenskapsmuseet har deltatt i innsamling av referansedata og delvis benyttet egne data som kilde til klasseinndelingen. I samarbeid med Høgskolen i Telemark har NTNU Vitenskapsmuseet utformet et system for kategorisering av vegetasjonselementer i felt eller ved tolking av høyoppløselige bildedata.

Matthew C. Hansen (nå Geographic Information Science Center of Excellence, South Dakota State University; tidligere tilknytning Laboratory for Global Remote Sensing Studies, University of Maryland) har gitt opplæring i hvordan datamineringsteknikker kan benyttes i analyse av satellittdata, og tilbrakt en måned i Norge for å bistå med selve analysene.

Vi takker alle som har bidratt i dette prosjektet.

Trondheim, juni 2006

Yngve Svarte
(sign.)

Direktør
Artsforvaltningsavdelingen



Innhold

1 Innledning	6
2 Målsetting.....	7
3 Studieområde.....	8
4 Metodikk.....	9
4.1 Tolkning og analyse av optiske satellittbilder.....	9
4.1.1 Hva optiske satellittsensorer måler	9
4.1.2 Spektrale klasser kontra informasjonsklasser	9
4.1.3 Hovedtyper av klassifikasjonsmetoder	10
4.1.4 Klassifikasjon med beslutningstre og regresjonstre	10
4.1.5 Bakgrunn for valg av analysemetodikk.....	11
4.1.6 Multi-Resolution Land Characteristics (MRLC)	11
4.1.7 MODIS Global Land Cover dataset.....	11
4.2 Inndeling i arealklasser	11
4.3 Validering av klassifikasjonen.....	12
5 Datagrunnlag	15
5.1 Referansedata	15
5.1.1 Data for skogareal.....	15
5.1.2 Data for skog, åpen mark og dyrket mark	17
5.1.3 Data for fjellområder.....	18
5.1.4 Andre tilgjengelige kartdata	18
5.1.5 Høyoppløselig bildemateriale	19
5.1.6 Punktbasert skjermdigitalisering.....	20
5.2 Satellittbilder og avledede vegetasjonsindekser.....	20
5.2.1 Valg av satellittbilder.....	20
5.2.2 Geometrisk korleksjon og referansesystem for bilder.....	21
5.2.3 Normalisert vegetasjonsindeks (NDVI)	22
5.2.4 Andre indekser	23
5.3 Basiskart.....	23
5.3.1 N50 arealtema.....	23
5.3.2 N50 terrengmodell og avledede rasterlag	23
5.3.3 Vegetasjonssoner	23
6 Produksjonsprosessen.....	24
6.1 Etablering av skyfrie multitemporale datasett	24
6.2 Inndeling i regioner	24
6.3 Uttrekk av referansedata.....	26
6.4 Beslutningstreanalyser	26
6.5 Sammensetting av klassifiserte raster	28

7	Resultat	29
7.1	SatNat kartene.....	29
7.2	Samsvar mellom eksisterende kartdata	29
7.3	Lokaliseringsavvik mellom piksel og referansepunkt	30
7.4	Skala i referansedata.....	31
7.5	Temakart for skogareal.....	32
7.5.1	Tretetthet	32
7.5.1	Treslag.....	32
7.5.1	Skogtilstand	33
7.5.2	Skogtypekart	34
7.6	Temakart for fjell og åpne områder i lavlandet	35
7.6.1	Inndeling i fjell og lavland	35
7.6.2	Klassifikasjon av åpen mark.....	35
7.7	Temakart for myr og jordbruksmark	37
7.7.1	Myr	37
7.7.2	Jordbruksmark.....	37
7.8	Arealdekkekart Sør-Trøndelag og Nord-Østerdalen	38
8	Diskusjon	41
8.1	Valg av klasser	41
8.2	Klassifikasjonsmetoden.....	41
8.3	Satellittdata	42
8.4	Treningsdata	42
8.4.1	Treningsdata og beslutningstreanalyser.....	42
8.4.2	Egnethetsvurdering av eksisterende kartdata	43
8.5	Tematisk nøyaktighet	46
8.6	Mulige forbedringer i analysemetoden	48
8.7	Vegen videre til et operasjonelt system	49
9	Konklusjoner og anbefalinger.....	50
10	Referanser.....	51
11	Vedlegg	54
11.1	Validering av klassifikasjon (feilmatriser)	54
11.1.1	SatNat Treslag	54
11.1.2	SatNat Skogtilstand	55
11.1.3	SatNat Åpen mark	57



1 Innledning

Norsk Romsenter og Direktoratet for naturforvaltning etablerte i fellesskap programmet (SatNat) i 2001. Hensikten med SatNat-programmet var å utvikle metoder der satellittdata var et hjelpemiddel i bruk og forvaltning av naturressurser. To prosjekt i dette programmet har vært rettet mot utprøving og utvikling av metoder for arealdekkekartlegging av store areal.

Det er tidligere utført arbeid i Norge med heldekkende kartlegging av store areal, både ved visuell tolking (Hjeltnes m.fl. 1996) og med ikke-styrt klassifikasjon (Johansen m.fl. 1995). I Sverige er det nylig gjennomført en metodeutvikling og produksjon av et arealdekkekart for hele landet (Lantmäteriet 2003), som et ledd i EUs etableringen av CORINE Land Cover for alle medlemsland. Metodikk for CORINE kartlegging er også prøvd ut i Norge (Arnoldussen m.fl. 1999).

Arbeidet som er gjennomført i SatNat må sees på som komplimenterende til disse arbeider, da det her er valgt å prøve en styrt klassifikasjon basert på dataminerings-

teknikker og på bruk av eksisterende datakilder for trening av klassifikasjonen. Dette er en relativt ny tilnærming i analyser av bildedata.

Ved oppstart av dette prosjektet var det ingen vedtak i Norge om etablering av CORINE Land Cover. I 2004 fikk NIJOS i oppgave å lage et slikt kart basert på den topografiske kartserien N50 og det digitale markslagskartet (DMK) som hovedkilde, supplert med tolking av satellittdata i fjellområdene. Arbeidet med metodeutprøvingen i SatNat har fokusert på å utvikle en robust og automatisert metode for heldekkende arealkartlegging med satellittdata, og hvor kartdataene også var egnet til produksjon av utbredelseskart for viltarter ved hjelp av habitatmodeller. Utviklingsarbeidet har foregått i Østfold og Sør-Trøndelag, og denne rapporten dokumenterer arbeidet i Sør-Trøndelag.

2 Målsetting

Prosjektets hovedmålsetting:

- Utvikle en kostnadseffektiv og objektiv metodikk til produksjon av arealdekkekart for store arealer (fylker) basert på analyser av satellittbilder.

Prosjektets delmål:

- Uprøve og validere egnethet av å bruke eksisterende kartdata som referanse (data til trening og validering) til klassifikasjon av data fra Landsat-satellittene.
- Utarbeide arealdekkekart med tematisk inndeling relevant for bruk til modellering av potensielle vilthabitat.

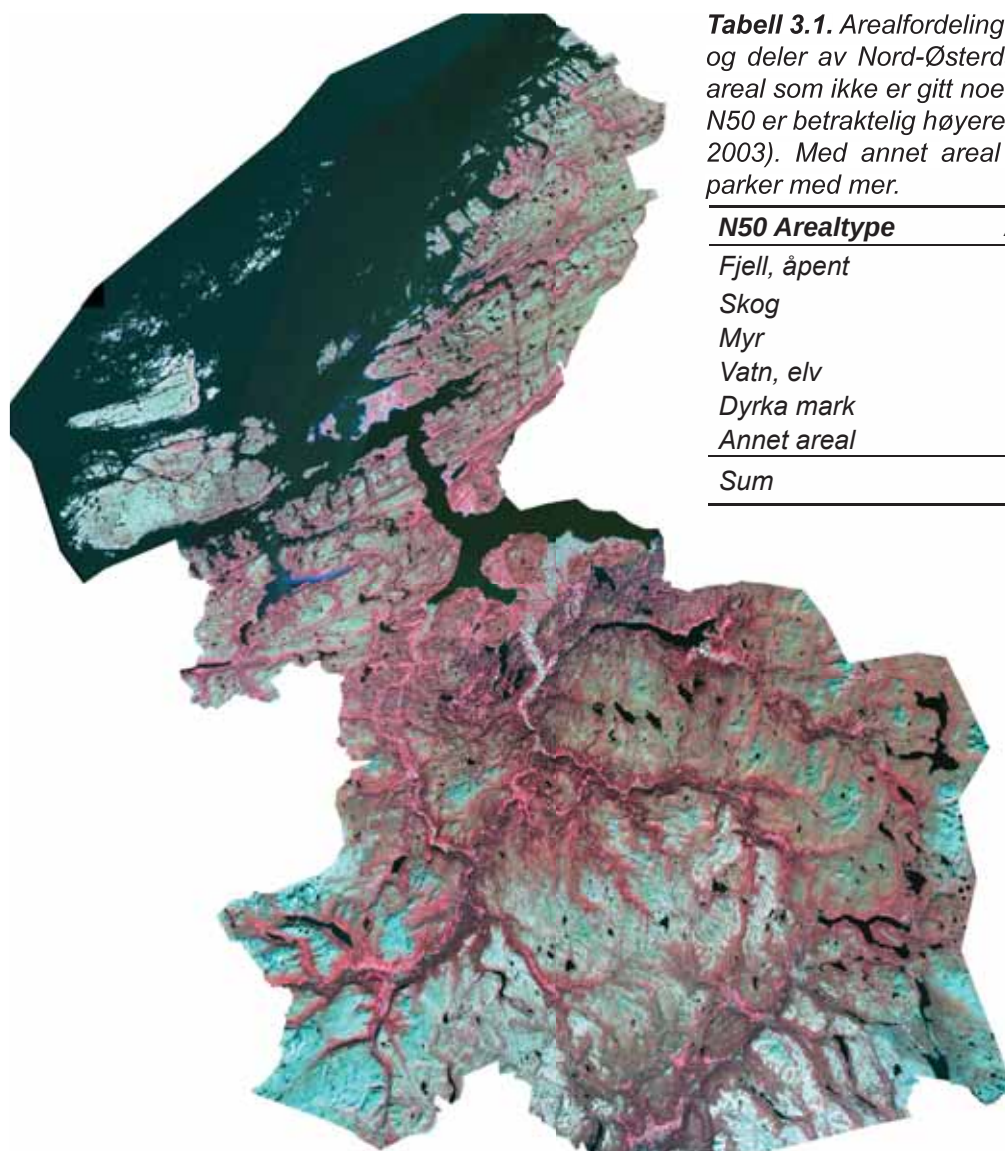
Et mål er at slike kart kan gi grunnlag for arealmessige vurderinger for områder på størrelse med kommuner og fylker. Med kostnadseffektiv metodikk menes det at metodikken skal kunne implementeres i automatiserte rutiner, slik at det minimaliserer manuelt arbeid og graden av subjektiv tolking. Videre vil gjenbruk av eksisterende landsdekkende databaser kunne representere en tidsbesparelse sammenlignet med at feltregistreringer blir innhentet i hvert kartleggingsprosjekt.

3 Studieområde

For utprøving av metodene beskrevet i denne rapporten ble Sør-Trøndelag fylke og den øvre delen av Nord-Østerdalen (deler av Tynset, Tolga og Os) valgt som testområde. Dette utgjør et landareal på 22 100 km², hovedsakelig bestående av fjell og skog. Arealet er langt større enn det areal som tidligere har blitt kartlagt (Østfold, 4 200 km²) etter tilsvarende metoder (Vikhamar m. fl. 2004). Arealet ble valgt såpass stort for å belyse problematikken som oppstår når det må brukes flere satellittbilder for

å dekke et kartleggingsområde og for å fange opp et stort spekter av landskapstyper.

For å dekke hele Sør-Trøndelag og Nord-Østerdalen var det nødvendig å bruke mange satellittscener. Kartleggingsområdet er også landskapsmessig meget variert (tabell 3.1). Det strekker seg fra kystområdene i vest til høyfjellområder på over 2200 meter og innehar vegetasjonssoner fra boreonemoral til høgalt (Moen 1998). Klimaet varierer fra sterkt oseanisk til svakt kontinentalt.



Tabell 3.1. Arealfordeling i studieområdet Sør-Trøndelag og deler av Nord-Østerdalen. Med "fjell, åpent" menes areal som ikke er gitt noe arealtema i N50. Skogarealet i N50 er betraktelig høyere enn produktivt skogareal (SSB 2003). Med annet areal menes bebyggelse, steintipp, parker med mer.

N50 Arealtype	Antall km ²	Fordeling
Fjell, åpent	9 076	41,1 %
Skog	8 505	38,5 %
Myr	2 498	11,3 %
Vatn, elv	1 091	4,9 %
Dyrka mark	865	3,9 %
Annet areal	73	0,3 %
Sum	22 108	100,0 %

Figur 3.1. Mosaikk av bilder fra Landsat og IRS satellittene over studieområdet (Sør-Trøndelag og deler av Nord-Østerdalen).

4 Metodikk

4.1 Tolkning og analyse av optiske satellittbilder

4.1.1 Hva optiske satellittsensorer måler

Et optisk satellittbilde er et digitalt bilde som består av piksler eller bildeelementer. Pikselen inneholder en verdi (vanligvis mellom 0 og 255) som representerer reflektert sollys målt av satellittsensoren. Pikselverdien kan omregnes til radians eller reflektans ved hjelp av kalibreringsverdier oppgitt for den enkelte sensor. Med radians menes energien i de elektromagnetiske signaler som satellitten måler og med reflektans menes forholdet mellom innstrålt og avgitt lys fra en overflate.

Hovedbidraget til den målte radiansen er reflektert sollys innenfor et areal på bakken som tilsvarer sensorens romlige oppløsning. Ulike overflatetyper har ulike reflektansegenskaper. Dette illustreres i figur 4.1 ved hjelp av spektrale signaturer til jord, vegetasjon og vann.

I tillegg til at ulike overflatetyper reflekterer sollyset forskjellig påvirker også atmosfæren og topografien lysrefleksjonen. Ulike gasser og partikler i atmosfæren reflekterer og absorberer sollys. Dette er effekter som varierer over tid, og som en ideelt sett bør korrigere for. Videre har geometrien mellom solens posisjon, sensorens posisjon og terrengets helning og helningsretningen sterk innflytelse på målt lysrefleksjon.

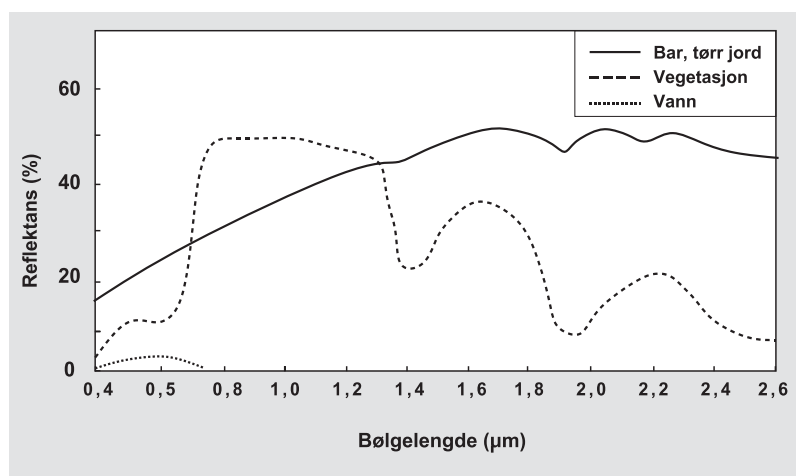
4.1.2 Spektrale klasser kontra informasjonsklasser

Når målsetningen med en analyse av satellittbilder er å lage et temakart er det viktig å først definere hvilke klasser som ønskes vist på kartet. Slike klasser kalles *informasjonsklasser*. Utfordringen med å lage et temakart som resulterer i ønskede informasjonsklasser er at de ofte ikke tilsvarer de *spektrale klassene* som finnes i et satellittbilde. Spektrale klasser er naturlige grupperinger av refleksjonsverdiene i satellittdataene.

En søker dermed en metodikk for å identifisere spektrale klasser som representerer ønskede informasjonsklasser. Flere spektrale klasser utgjør ofte en enkelt informasjonsklasse. Et eksempel er et tilfelle der en skal kartlegge ulike skogtyper og man ønsker informasjonsklassene barskog og løvskog. Løvskog kan finnes i satellittbildet som de spektrale klassene: skyggelagt tett løvskog, belyst tett løvskog, skyggelagt glissen løvskog, belyst glissen løvskog.

Klassifikasjon av satellittdata består i å kategorisere og tilordne en informasjonsklasse til hver piksel i et satellittbilde. Resultatet blir et temakart. Grunnlaget er den spektrale signaturen til hver enkelt overflate. To overflater skilles best fra hverandre ved å benytte de bølgelengder hvor overflatene har størst forskjell i spektral signatur.

Klassifikasjon av en piksel som gjøres utelukkende ved hjelp av denne ene pikselens spektrale informasjon kalles pikselvis klassifikasjon. En klassifikasjon kan bli bedre dersom det inkluderes informasjon fra pikselens nabo-skap. Med naboskap menes piksler som ligger ved siden av eller i nærheten av pikselen som skal klassifiseres. Dette refereres til som tekstur. Klassifikasjonen kalles da gjerne kontekstuell klassifikasjon. Det er også mulig å beregne andelen av en type innenfor en piksel. Dette refereres til som subpiksel klassifikasjon.



Figur 4.1. Spektrale signaturer til jord, vegetasjon og vann i et spekter fra blått (0,4 µm) til midlere infrarødt lys (2,6 µm).



4.1.3 Hovedtyper av klassifikasjonsmetoder

Vanligvis blir klassifikasjonsmetoder skilt i hovedgruppene styrt og ikke-styrt klassifikasjon (Campbell 1996, Hastie m. fl. 2001). Ved styrt klassifikasjon behøves a priori kunnskap om hvilke klasser som ønskes kartlagt. Treningsdata er da viktig for å trene opp klassifikatoren. Til dette behøves informasjon om områder i bildet hvor det er sikkert at disse klassene finnes. Denne informasjonen brukes så til å styre inndelingen i klasser gjennom bruk av ulike algoritmer. Algoritmene vil da forsøke å finne optimale sammenhenger mellom responsvariabelen (informasjonsklassene) og de uavhengige variabler som legges inn i analysene (Tan m. fl. 2006). Dette kan være båndene fra satellittdataene eller data fra andre kartkilder.

Ikke-styrt klassifikasjon er en motsatt tilnærming. Satellittbildet inndeles statistisk i naturlige spektrale klasser. Etter klassifikasjonen identifiserer man hvilken informasjonsklasse hver enkelt spektral klasse tilhører. Statistiske separabilitetsmål brukes som støtte i dette arbeidet. De vanligst brukte klassifikasjonsalgoritmene for begge disse hovedtilnærmin-

gene (styrt og ikke-styrt klassifikasjon) er basert på parametriske modeller for å beregne klassesignaturene, men optiske data følger ikke alltid slike statistiske fordelinger.

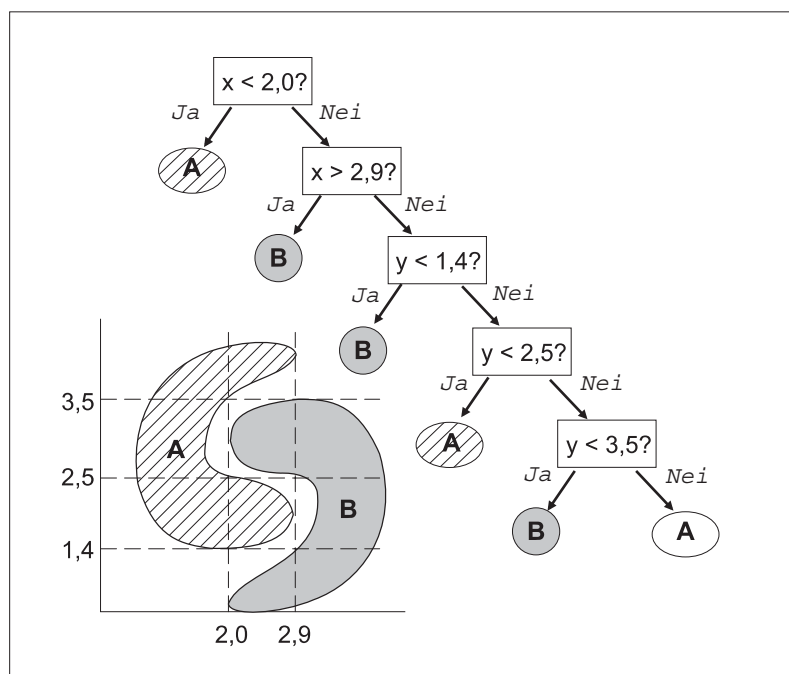
Bruken av styrt og ikke-styrt klassifikasjon har endret seg mye siden teknologien kom i 1970-årene. I starten ble styrt klassifikasjon brukt mye. På 90-tallet ble ikke-styrt klassifikasjon nesten enerådende. Med økt tilgang til høyoppløselige digitale bilder fra fly og satellitt, forbedret diskplass og nye analysemetoder har styrt klassifikasjon igjen fått økt fokus.

Et alternativ til en maskinmessig klassifikasjon er visuell tolkning ved digitalisering på skjerm eller papirkart. Denne metoden brukes ofte i grovere målestokker. Løsningen gir kart som er visuelt innbydende, men den er arbeidskrevende. For kartlegging av store områder blir den romlige detaljeringen begrenset, dersom arbeidsmengden skal holdes på et akseptabelt nivå. NIJOS lager nasjonale arealressurskart i målestokk 1: 250 000 (AR250, Moum 1999) på denne måten.

4.1.4 Klassifikasjon med beslutningstre og regresjonstre

For å lage arealdekkekart i Sør-Trøndelag har vi valgt styrt klassifikasjon med beslutningstrær som metodikk. Beslutningstrær til bruk i klassifisering av arealdekke ble valgt til analysene fordi det i lignende arbeider har vist seg å være en lovende metode (Homer m. fl. 2002, Hansen m.fl. 2000). Beslutningstrær og regresjonstrær tilhører en type analysemetoder som kalles datamining (Campbell 1996, Hand m. fl. 2001, Witten & Frank 2005), og er basert på en logisk oppsplitting i ønskede klasser. Oppsplittingen skjer ut fra et prinsipp at splittingen øker forklaringsgraden, og det brukes logiske regler for hver gruppesplitt (figur 4.2).

Resultatet blir en trestruktur, der greinene (kalt noder) representerer en klasse. Når greiner ender uten nye forgreninger gir økt splitting ingen forbedret forklaring. Ved å etablere regler fra et utvalg med treningsdata, kan reglene etterpå benyttes effektivt på store datamengder. Metoden er derfor egnet for analyser av satellittdata som dekker store områder. Beslutningstreet tildeler en klasse til hvert eneste piksel i det kartet som lages. Forskjellen på beslutningstre og regresjonstre er at den første



Figur 4.2. En enkel illustrasjon som viser hvordan et beslutningstre deler opp datasett i mindre enheter ved hjelp av regler. Ved beslutningsregler kan kompliserte datasett skilles fra hverandre. Merk også at i et beslutningstre må reglene utføres i en bestemt rekkefølge.

gir distinkte klasser som resultat, mens den andre gir som resultat en kontinuerlig variabel. Regresjonstre brukes derfor mye til å beregne hvor stor andel av en piksel som utgjøres av en spesiell type, for eksempel hvor stor andel (0-100 %) av en piksel som dekkes av lav.

En sammenligning av ulike metoder for kartlegging av arealdekke med satellittbilder er beskrevet i f.eks. Pal & Mather (2003) og Hansen m.fl. (1996). Artiklene fokuserer spesielt på beslutningstreanalyser.

4.1.5 Bakgrunn for valg av analysemetodikk

Avsnittet presenterer to eksempler på arealdekkekartlegging som har vært sentrale i valg av metodikk ved utarbeidelse av kartet over Sør-Trøndelag. I forberedelsene til dette prosjektet ble det tatt kontakt med personer sentrale i Gap Analysis Program i USA (www.gap.uidaho.edu) og to arbeidsseminar ble holdt med personer fra GAP-miljøet. Senere ble det også tatt kontakt med Laboratory for Global Remote Sensing Studies ved University of Maryland (USA). Disse miljøene arbeidet spesielt med vegetasjonskartlegging av større areal basert på satellittdata.

4.1.6 Multi-Resolution Land Characteristics (MRLC)

I USA er to heldekkende kartlegginger av arealdekke satt i gang først og fremst basert på data fra Landsat-satellittene. Den første generasjons kartlegging i 1992 baserte seg på klustering (ikke-styrt klassifikasjon). Her ble det brukt data fra Landsat5-satellitten og hver delstat ble kartlagt hver for seg. Et klassifikasjonsskjema med 21 klasser var grunnlaget for kartleggingen. Samlet kom kartet ut med en nøyaktighet per piksel på 45 %. Denne tilnærmingen er beskrevet nærmere på webadressen: <http://landcover.usgs.gov/natl/landcover.asp>.

I andre generasjons kartlegging fra 2001 (noe arbeid gjenstår fortsatt i 2005) er beslutnings- og regresjonstre analyser benyttet (Homer m. fl. 2002). I stedet for ett arealdekkekart over USA produseres det i denne kartleggingen en mengde datalag som kan nyttes i videre GIS-analyser og presentasjoner. Referansedata fra

flybilder (Huang m. fl. 2001), IKONOS satellitten, data fra landbruk og skogbruk med mer er benyttet for å lage kartene.

I stedet for å kartlegge hver delstat hver for seg ble USA delt inn i 60- 70 økologiske regioner (Homer & Gallant 2001). Dette gav en mer effektiv kartlegging i form av at klassene innenfor hver region er jevnere fordelt, det ble mindre spektralt overlapp mellom klassene og det gav bedre arealmessig utforming. Nøyaktigheten i klassifikasjonen ble også forbedret sammenliknet med klusteringsteknikken. Denne tilnærmingen er beskrevet nærmere på webadressen:

http://www.mrlc.gov/mrlc2k_nlcd.asp.

Metoden ble brukt for arealdekkekartleggingen av Østfold (Vikhamar m.fl. 2004).

4.1.7 MODIS Global Land Cover dataset

For Sør-Trøndelag ble samme tilnærming benyttet, men analyseprosedyren fulgte mer den metodikk som var utarbeidet ved Laboratory for Global Remote Sensing Studies (University of Maryland, USA). Denne metoden ble brukt for å lage et globalt arealdekkekart. Metoden var mer fleksibel for tilpasninger og det var enklere å undersøke reglene som ble dannet, og dermed lettere å analysere noder der to eller flere klasser hadde omtrent samme sannsynlighet for å bli valgt. Ved slike tilfeller kunne endring av noden til en annen klasse gi en forbedret resultat i kartet.

Dr. Matthew C. Hansen fra Laboratory for Global Remote Sensing Studies var på besøk i Norge sommeren 2004 for å arbeide med kartet over Sør-Trøndelag. Hansens metodikk er i hovedsak lagt til grunn for denne rapporten. Metodikken er beskrevet nærmere på webadressen:

<http://glcf.umd.edu/data/modis/vcf/>.

4.2 Inndeling i arealklasser

Det er gjort mange forsøk på å få til enighet om et felles klassifikasjonssystem i Norge. Noen systemer er laget for kartlegging i felt, andre mer for kartlegging med flybilder. "Vegetasjonstyper i Norge" (Fremstad 1998) er



laget for å kunne inndelegge vegetasjonen i typer når en er i felt. En er da avhengig av å kunne identifisere enkeltarter, noe som ikke er mulig i fly- og satellittbilder. Kartleggingen av biologisk mangfold (DN-Håndbok 13 1999) baserer seg også i stor grad på identifikasjon på bakkenivå. Landsskogtakseringen til NI-JOS er en annen kartlegging som baserer seg på observasjoner i felt.

Ved kartlegging av vegetasjon over større arealer i Norge er det tradisjonelt brukt sort-hvite flybilder, med inventering i felt og kartlegging i stereoinstrument. ØK/DMK og 3Q fra NI-JOS og bestandstakster for skog-eierforeningene er laget på denne måten. Legenden i kartene vil være en kombinasjon av egenskaper observert i felt og innmålinger utført i stereoinstrument.

For kartlegging med satellittbilder er det forsøkt å lage en norsk versjon av CORINE Land Cover, en kartlegging av arealdekke basert på satellittbilder som alle EU-land må gjennomføre (Arnoldussen m.fl. 1999). I et forsøk på å enes om en norsk standard ble Nasjonalt system for arealdekkeklassifikasjon (NaSAK) utarbeidet (Gjertsen m.fl. 2003, Hjeltne & Øien 2003 og SOSI standard for skogbestandsregistreringer, Statens Kartverk). NaSAK er et resultat fra et prosjekt på utvikling av et norsk klassifikasjonssystem for arealkartlegging og bygger igjen på det internasjonale LCCS (Land Cover Classification System) til FAO (http://dwms.fao.org/lccs/index_en.asp).

I arbeidet med arealdekkekartlegging i SatNat-programmet er det blitt utarbeidet et eget system for registrering av arealtyper basert på feltkontroll eller flybildetolkning (Hjeltne & Øien 2003). Dette systemet er siden blitt noe revidert. Systemet legger større vekt på å gruppere vegetasjonsenheter slik de skilles ut ved tolking av flyfoto eller høyoppløselige satellittbilder enn grupperinger etter plantesosiologiske kriterier. Systemet bygger på NaSAK og er spesielt utformet for å være en rask metode til å kategorisere under registreringsarbeid.

Vi har lagt hovedinndelingen i dette systemet til grunn for arealtypeklassifiseringen i Sør-Trøndelag, men har måttet modifisere det etter inndelinger brukt i de referansedata vi har benyttet.

4.3 Validering av klassifikasjonen

Nøyaktighetsvurdering

Når det blir utført en arealkartlegging må produktet valideres. En kartlegging bør ikke betraktes som ferdig før dens nøyaktighet er utredet (Lillesand & Kiefer 2000). Den vanligste metoden for å validere et kartprodukt er ved bruk av feilmatriser, dvs. en tabellarisk sammenlikning mellom klasser i et datasett mot kjente referansedata.

I utgangspunktet skal valideringsdataene ikke inneholde feil og de skal være representative for det området som kartlegges. I virkeligheten vil situasjonen være annerledes. Å benytte valideringsdata uten feil er nærmest umulig å få til når det gjelder kartlegging til distinkte klasser i et landskap med kontinuerlig overganger mellom enheter. Et kartprodukt basert på styrt klassifikasjon med statistiske algoritmer kan sjelden gi større nøyaktighet i forhold til det datasett det valideres mot enn nøyaktigheten mellom valideringsdataene og treningsdataene.

Tematisk unøyaktighet - feilmatrise

Uansett hvor stor innsats vi legger i å få god kvalitet på valideringsdata så vil det ligge en del subjektiv tolkningsvurdering bak klasseinndelingen. Det vil for eksempel være nærmest umulig å gi en riktig anslag over dekning av ulike vegetasjonsenheter innen en 30 x 30 meters rute når en er på bakken. På samme måte vil aldri en piksel i en satellittscene treffe akkurat det arealet der refleksjonen, som har tildelt pikselen en verdi, har kommet fra. Det er alltid noe geometrisk unøyaktighet i plasseringen av bilder.

Et godt resultat i en feilmatrise viser at klasseinndelingen i et kartprodukt er i samsvar med den inndeling som er brukt i valideringsdataene og at klassene kan skilles med den metoden som er brukt. De sier derimot ikke så mye om hvordan klassifikasjonen virker andre steder i kartet enn der treningsdata er hentet fra. Valideringsdata må dermed både være tilstrekkelig mange til at resultatet ikke kommer fra tilfeldigheter og de må være romlig representative

slik at de fanger opp variasjonen innen de enkelte klasser.

Riktigheten av valideringsdata er ofte vanskelig å avgjøre. Dersom treningsdata er av god kvalitet er det vanlig å trekke ut en tilfeldig andel til kontroll av det ferdige produktet. Siden vi for det meste har brukt data innsamlet til andre formål som grunnlag for treningen er det vanskelig å vite hvor riktige disse er både når det gjelder hva de enkelte klasser representerer og hvor tidsriktig informasjonen er sett i forhold til de satellittbilder vi har brukt. Flere av datasettene vi har hatt tilgjengelige har beskrevet den samme klassen og vært lokalisert til samme område. For å undersøke hvor god overensstemmelse vi kunne forvente mellom kartprodukt fra satellittkartlegging og kilder som var tilgjengelige som referanse-data, beregnet vi feilmatriser mellom noen av disse kartkildene.

I en feilmatrise benyttes begrepene produsent og brukers nøyaktighet. Begrepet produsentens nøyaktighet er mest interessant for den som lager et kart, da det sier i hvor stor grad et areal av en bestemt type ble klassifisert til den typen. Det sier derimot ikke noe om hvor mange andre arealer av andre typer som er klassifisert til den samme typen. En bruker vil ofte være mer opptatt av hvor stor sannsynlighet det er for at et gitt område (i terrenget) er blitt klassifisert til den rette typen. Brukerens nøyaktighet beskriver dette. Det er også ønskelig å beregne en nøyaktighet for alle klassene samlet. Den enkleste måten er å beregne andelen av punkter som er korrekt klassifisert og uttrykke dette i en verdi for samlet nøyaktighet.

I beregninger av verdi for Kappa (Congalton&Green 1999) og Tau (Ma & Redmond 1995) er produsent og brukes nøyaktighet inkludert og disse beregninger gir samtidig mulighet for statistisk sammenlikning av kartprodukter. En Kappa analyse beregner en Khat-verdi, som vanligvis ligger mellom 0 og 1. En Kappa verdi på for eksempel 0,8 forteller at klassifikasjonen har unngått 80 % av feilene som ville ha oppstått ved en tilfeldig klassifikasjon (Congalton&Green 1999). Tau koeffisienten er noe lettere å tolke. For eksempel vil

en Tau-koeffisient på 0,8 bety at 80 % av punktene var klassifisert korrekt (sammenliknet med om alt var tilfeldig klassifisert). Begge koeffisienter er tilnærmet normalfordelte og Z-statistikk kan benyttes for å undersøke forskjeller mellom kartprodukt. Vi har valgt å ta med begge koeffisienter. Kappa og Tau, sammen med verdien for samlet nøyaktighet, gir et bra bilde på hvor godt en kartlegging samsvarer med de data den valideres mot (tabell 4.1).

Dersom samlet nøyaktighet i en klassifikasjon er betraktelig høyere enn Khat og Tau gir dette en indikasjon på at det er en skjev fordeling av referansepunktene mellom klassene. Klasser med få observasjoner og lav nøyaktighet blir da overdøvet av en eller flere klasser med mange observasjoner og høyere nøyaktighet. Et eksempel på dette er å finne i kapittel 7.7.2, tabell 7.15.

Geometrisk unøyaktighet

Satellittbildene som er benyttet i analysen har en oppløsning på 30 meter og en midlere feil på om lag 20 meter. Punktdatasettene som benyttes til validering vil også ha et lokaliseringsavvik i størrelse 5 – 15 meter. Dette gjør at referansepunktet kan bli plassert utenfor avgrensingen av den sammenfallende klassifiserte pikselen. For å kompensere for effekten av denne geometriske feilen har vi beregnet både en feilmatrise basert på nøyaktig lokalisering (senter av piksel) og en matrise der det er gitt treff dersom senterpunkt eller en av de åtte pikslene omkring inneholder samme klasse som valideringsdataene (Vikhamar m. fl. 2004). Senterpunkt er et mye strengere kriterium enn naboskap. Naboskap vil alltid øke treffprosenten.

Tabell 4.1. Retningslinjer for tolking av Kappa og Tau verdier.

Verdier	Kvalitetsmål
≤ 0	Dårlig samsvar
0 – 0,2	Noe samsvar
0,2 – 0,4	En del samsvar
0,4 – 0,6	Brukbart samsvar
0,6 – 0,8	Stort samsvar
0,8 – 1,0	Meget sterk samsvar (sjeldent)



For at vi skal kunne kontrollere at denne økningen ikke er kunstig høy har vi beregnet variasjonen i pikselverdier i datasettet. Har et kart to klasser og maksimal homogenitet (ingen blanding av piksler) vil feilverdiene ved senter og naboskap være tilnærmet like. Ved maksimal variasjon i pikselverdier (sjakkbrett), vil det være maksimal forskjell på senter og naboskapsverdi. Med to klasser vil da senter teoretisk gi 50 % treff og naboskap 100 % treff ved tilfeldige data.

Ved å benytte et diversitetsfilter til å telle opp antall klasser i et 3x3 område rundt referansepunktene kan dette tallfestes. En summering av diversiteten ved alle referansepunktene vil gi en indikasjon på om forbedringen i treff ved naboskapsanalyse skyldes variasjonen i pikselverdier i det klassifiserte bildet eller en antatt lokaliseringfeil. Maksimal homogenitet i et kart vil gi en gjennomsnittlig verdi for diversitet nær 1, mens maksimal variasjon i pikselverdier gir en verdi nær antall klasser i kartet. Tabell 4.2 viser diversitet i det klassifiserte datasettet for Skogtilstand.

For å undersøke nærmere effekten av naboskapskriteriet har vi beregnet hvor stor del av økningen i treffprosent som skyldes geometriske unøyaktigheter og hvor mye som skyldes variasjonen i temakartet, har vi for hvert kart gjort en senter og en naboskapsanalyse basert på tilfeldige data (se også Kap. 6.3). Økningen i treffprosent på

Tabell 4.2. Diversitet i det klassifisert datasettet skogtilstand. Et 3x3 vindu rundt hvert referansepunkt i det klassifiserte bildet er undersøkt for diversitet, altså antall ulike klasser. Med fem klasser og hovedtyngden av pikslene (75 %) har en til to klasser som naboer kan nøyaktighetstallet ved naboskaps betraktning forsvares.

Antall verdi 1	208	27,3 %
Antall verdi 2	359	47,2 %
Antall verdi 3	159	20,9 %
Antall verdi 4	33	4,3 %
Antall verdi 5	2	0,3 %
Antall punkt:	761	
Gjennomsnitt	2,03	

disse tilfeldige dataene vil være et uttrykk for hvor stor virkning kartets heterogenitet har på økningen i treffprosenten ved bruk av kriteriet naboskap. Ved å trekke forbedringen som oppnås på et reelt kart der klassene er omkodet til en tilfeldig klassifisering (med samme antall klasser) fra tilsvarende økning i de reelle dataene kom vi fram til hvor stor forbedring i treffprosent som utelukkende skyldes kombinasjonen av kartets romlige struktur og antall klasser. Denne forskjellen er benyttet til å korrigere naboskapsberegningene slik at feilmatrisen uttrykker treffprosenten *korrigert* for geometrisk unøyaktighet.

5 Datagrunnlag

Kapittelet inneholder en beskrivelse av de ulike datasett som er brukt i analysene. Først gis en oversikt over de ulike kildene for referansedata som ble analysert og hvilken bearbeiding som ble gjort før dataene kunne brukes som treningsdata for klassifikasjonen. Deretter følger en beskrivelse av satellittdataene og andre heldekkende støtdata som ble brukt i Sør-Trøndelag og Nord-Østerdalen.

5.1 Referansedata

5.1.1 Data for skogareal

5.1.1.1 Bestandsskogstakst

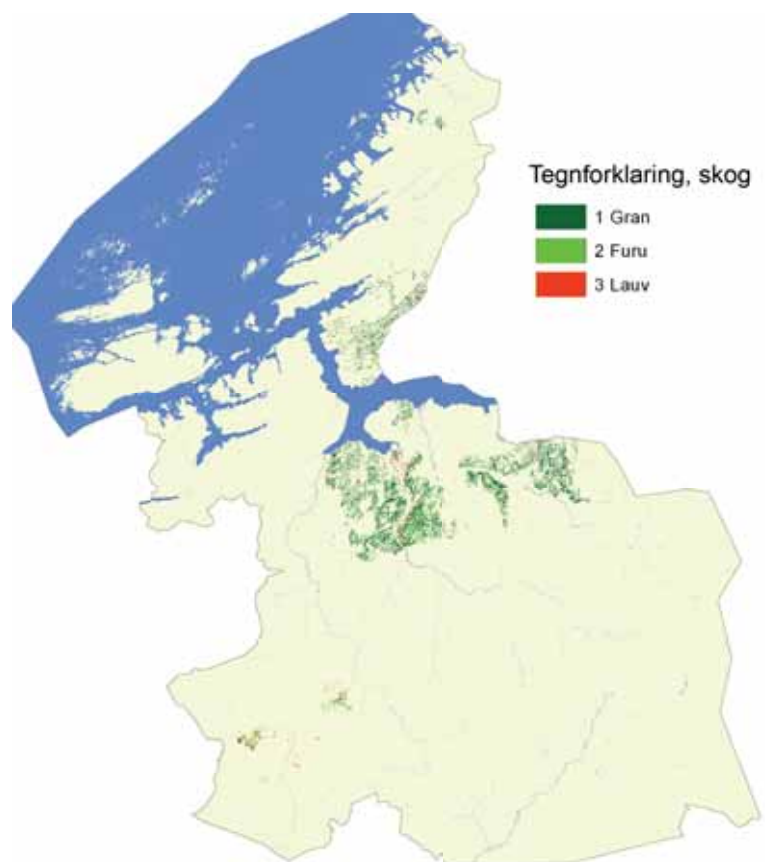
Bestandsskogstaksten er en takst for skog-eieren. I hovedsak er bestandsdataene fra skogbruksplanleggingen konsentrert om det produktive skogarealet. Bestand er behandlingsenheten som skogeieren forholder seg til, og den skal være relativ ensartet med hensyn på bonitet, tetthet, alders- og treslagssammensetning. Der spesielle arealtyper grenser inntil det produktive skogarealet kan disse være integrert som bestand. I enkelte tilfeller kan for eksempel myr, ikke produktiv skog og vann danne bestand når dette er viktig informasjon for skogbehandlingen i området. Bonitering og avgrensningen mot uproduktive arealer er ofte tatt fra Digitalt markslagskart (DMK).

Bestandsdataene inneholder informasjon om skogens tetthet, men denne er uttrykt som trevolum og ikke kronedekning, som er det satellitten måler. Vi utledet kronedekning fra Landbruksdepartementets godkjente funksjoner for volumberegning. For gran og furu la vi til grunn Tomters funksjoner basert på volum /daa og middelhøyde, og for lauv brukte vi Galaaens funksjon der også hogstklasse var inkludert (Landbruksdepartementet 2003).

Videre forsøkte vi å justere for at lauvskog, spesielt yngre lauvskog var klart underrepresentert i bestandsdataene. Kun 7 % av bestandsdataene var lauvdominert, mens gran

og furudominert skog hadde henholdsvis 72 % og 22 % av observasjonene. En del lauvskog ble digitalisert inn manuelt basert på flybilder. Før bestandsdata kunne brukes til trening av satellittdata måtte det brukes mye tid på omkodning og siling. Bestandsdata vi hadde tilgjengelig dekket et areal på hele 1778 km² i Sør-Trøndelag (figur 5.1). Av dette inneholdt 452 km² informasjon om volum. Dette utgjør 5 % av skogen fra N50 i studieområdet og 11 % av produktiv skog i Sør-Trøndelag. Samlet gir dette mye data, men arealet er sterkt konsentrert sentralt i fylket og dekker for det meste lavereliggende, produktiv skog.

Bestandsskogdataene ble benyttet som treningsdata for å etablere datalag som beskrev tettheten av skog, dominerende treslag (gran, furu, lauv) og hogstklasser. Siden en stor del av bestandene var blitt hogd mellom kartleg-



Figur 5.1. Dekning av kartdata fra bestandsskogtaksering i Sør-Trøndelag. Omkodet til forenklede klasser for treslag.



ging og tidspunktet da opptak av satellittbilde fant sted, måtte det mye siling av data til for å fjerne denne typen støy. Dette silingsarbeidet var meget tidkrevende.

5.1.1.2 Landsskogtakseringen

Landsskogtakseringen utføres av Norsk institutt for jord- og skogkartlegging (NIJOS) og er en utvalgsregistrering innenfor små flater med radius 8,92 meter (250 m²). Takseringen har som oppgave å skaffe ressurs- og miljødata for skogarealene i Norge (<http://skog.nijos.no>). Registreringene startet i 1919, og senere er syv mer eller mindre fullstendige landsomfattende registreringer blitt gjennomført. Landsskogtakseringens data er samlet inn fra prøveflater som er lagt ut i et systematisk nett. Over hele landet med unntak av Finnmark er det lagt ut permanente prøveflater i et forband på 3x3 km.

Selv om det ikke finnes trær innenfor en prøveflate vil flata etableres og få informasjon om vegetasjon og hogstklasse for den nærliggende skogen. Hvert fylke får mellom 1000

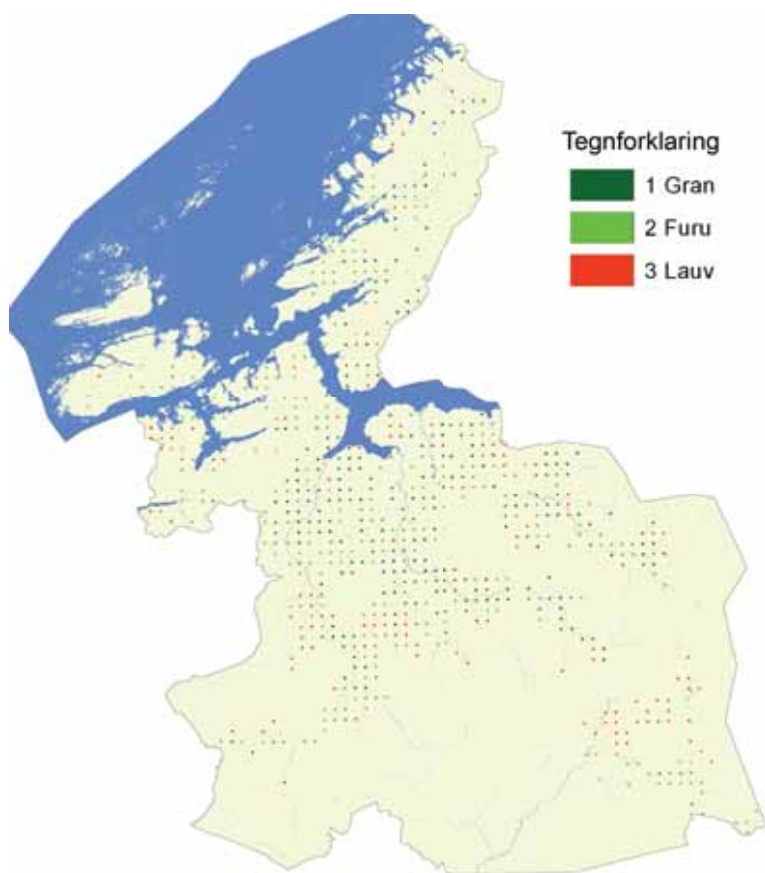
og 1500 permanente flater innefor arealet som er vurdert til å være skog. Flatene oppsøkes hvert 5. år, og danner grunnlaget for fylkesvis skogstatistikk. Alle markslag under barskogsgrensa omfattes, men det er bare på skogsmark at det blir gjort en detaljert beskrivelse.

Takseringene av Sør-Trøndelag fylke i perioden 1996 - 2000 inngår som ledd i analysene (figur 5.2). Registreringen er basert på 1032 permanente prøveflater med skoginformasjon, som er lagt ut over fylket i henhold til et bestemt system. 99 av disse er delte flater som grenser til andre areal typer (skogkant). Fra hver prøveflate benyttet vi informasjon om treslagssammensetning, bonitet, høyde og trevolum. I forhold til et arealmessig lite fylke som Østfold (Tomter & Eriksen 2001) som har 965 permanente flater (hvorav 79 er delte) med skoginformasjon, er det dermed relativt få punkt tilgjengelig for Sør-Trøndelag. Av flatene i produktiv skog har Østfold faktisk flere enn Sør-Trøndelag, henholdsvis 742 mot 501 flater. Med et areal på 4140 km² produktiv skog (SSB 2003), nesten det dobbelte av skogarealet i Østfold (2292 km²) blir det for lite data til å fange opp den variasjonen som er i Sør-Trøndelags skoger.

5.1.1.3 Tilleggsdata for uproduktiv skog

Siden dekningen av digitale bestandsdata var konsentrert til de sentrale deler av Sør-Trøndelag var denne datakilden ikke tilstrekkelig til å fange opp de ulike skogtyper, spesielt ikke i den uproduktive skogen mot fjellet. Landsskogtakseringens punktdata blir også for få til alene å være egnet som referansedata for tolking av satellittdata. To alternative kilder til referansedata ble derfor utprøvd. En gikk ut på å kombinere informasjon fra DMK og Landsskogtakseringen med satellittdata. Den andre var bruk av vegetasjonskart.

For i det hele tatt å kunne få treningsdata på treslagsfordeling i områder med lite referansedata beregnet vi prosent gran, furu og lauv basert på sammenhengen mellom disse variablene fra Landsskogtakseringen og data fra heldekkende datasett vi hadde tilgjengelig. Vi kjørte multiple regresjoner med prosentandelen fra Landsskogtakseringen som avhengig variabel og data fra DMK, satellittdata, høydemodellen og geografisk plassering (x og y koordinater) som forklaringsvariabler.



Figur 5.2. Dekning av punkter fra Landsskogtakseringen i Sør-Trøndelag. Omkodet til forenklete klasser for treslag.

Fra DMK (Bjørndal m.fl. 2004) brukte vi ATIL = 24 (barskog) eller 25 (blandingsskog) som utvalg for furu og gran, og ATIL = 25 (blandingsskog) eller 26 (lauvskog) som utvalg for lauvskog. De kategoriske variablene fra ASKOG (verdiene 11, 12, 13, 14) ble brukt som kontrast. Fra Landsat-dataene brukte vi rød kanal, vegetasjonsindeksen NDVI og forskjeller i NDVI mellom mai og juli, og fra høydemodellen de avledede produktene varmekoeffisient og vegetasjonsgeografisk sone. Alle kontinuerlige variabler. I utvelgelsen av modeller brukte vi Akaike Informasjonskriterium (AIC, Akaike 1973).

Vi gjorde slike analyser på behandlingsområde 2 som dekker blant annet hele Fosen-halvøya og kommunene Agdenes og Snillfjord (se kapittel 6.2.1). For skogen i fjellnære områder forsøkte vi å bruke vegetasjonskartet fra NIJOS (se kapittel 5.1.3.1). Dette ble kodet om til gran, furu og lauvskog etter hovedtypen. Dette gav viktige treningsdata spesielt for å kunne skille bjørk og furu i fjellnære skoger.

temet 3Q startet i 1998, etter 5 år med forprosjekter og evalueringer. 3Q har bruksområder knyttet til landbruksforhandlinger, lokal og nasjonal landbrukspolitikk, samt innrapportering til internasjonale organisasjoner som OECD. I overvåkningssystemet 3Q registreres informasjon om jordbruksarealer, skogsarealer, bebygde arealer, vann, is, snø m.m. De fleste dataelementer er registrert som polygoner, men noen er registrert som linjer eller punkt. Prøveflatene er 1 x 1 kilometer store kvadrater, som er lagt ut i et forband på 3 x 3 kilometer.

Det ferdige 3Q kartet er et resultat av fotogrammetrisk konstruksjon med visuell tolking av flybilder (vanlige fargebilder) i målestokken 1: 12 500. Generelt skal alle arealenheter som grenser mot jordbruksareal være minimum 100 m². Som regel må andre arealenheter være minimum 1000 m². 103 ulike arealtyper er kartlagt. Tilstandsregistreringen skal gjentas hvert femte år. I Sør-Trøndelag hadde vi 28 prøveflater tilgjengelig av de 95 etablerte prøveflatene.

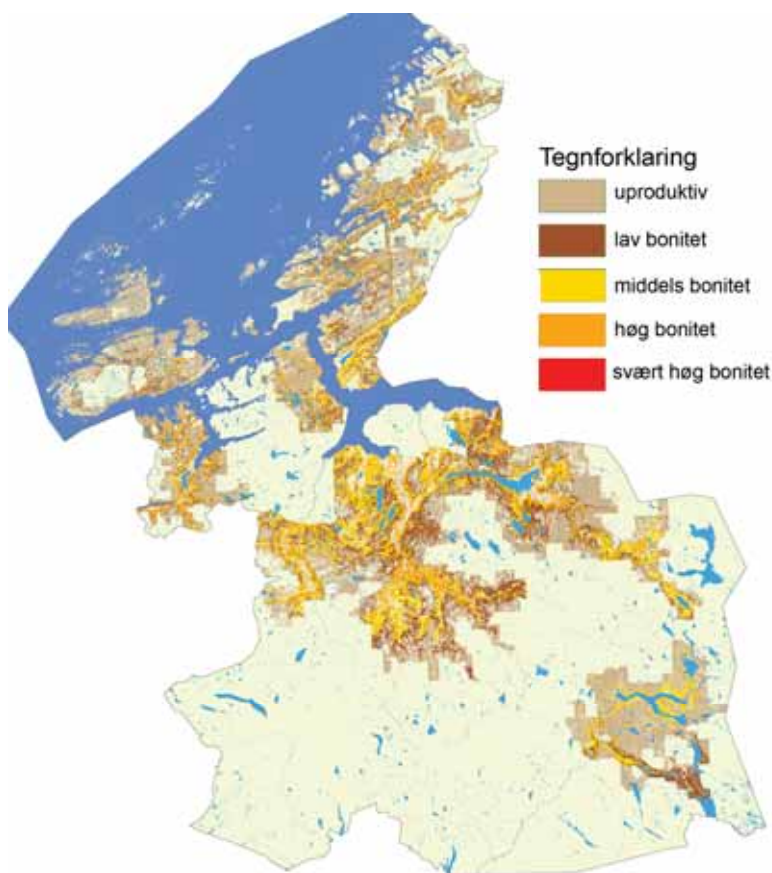
5.1.2 Data for skog, åpen mark og dyrket mark

5.1.2.1 Digitalt Markslags Kart (DMK)

Digitalt Markslags Kart (DMK) er et nasjonalt datagrunnlag om arealtilstand og arealkvalitet for jord- og skogbruksområder og annet areal kartlagt for Økonomisk kartverk (ØK, nå: N5) i målestokk 1:5 000. Markslaget gir informasjon om dyrkingstilstand og driftsforhold på jordbruksareal, produksjonspotensialet (bonitet) i skog og på anna areal, og arealtilstand for alle andre arealkategorier (anna areal). Kartserien omfatter det meste av arealet under skoggrensa og oppdateres periodisk. Mer informasjon på <http://www.nijos.no>. Da dette prosjektet pågikk manglet store deler av Sør-Trøndelag (figur 5.3), men ved utgangen av 2005 var alle kommuner unntatt Oppdal, Rennebu og Holtålen ferdig kartlagt med DMK.

5.1.2.2 3Q

3Q er et datasett som også produseres av NIJOS. 3Q er et verktøy for tilstandsovervåking og resultatkontroll for jordbruk og kulturlandskap (Fjellstad m.fl. 2004). Overvåkningssys-



Figur 5.3. Dekning av DMK fra NIJOS. Egenskap "ASKOG" er vist (bonitet i skog).



5.1.3 Data for fjellområder

5.1.3.1 Vegetasjonskart fra NIJOS

Vegetasjonskartlegging i målestokk 1: 20 000 - 50 000 er beskrevet av Rekdal & Larsson (2005). Slike vegetasjonskart ble ansett som en bra kilde for referansedata i fjellet. De er basert på feltarbeid og visuell tolking av svarthvite flyfoto i fotogrammetrisk arbeidsstasjon og de har et stort antall vegetasjonsklasser. Vi har sett på bruken av kartene Oppdal, Trollheimen og Gaula (Balle 2000), et kartlagt areal på totalt 4 622 km² (figur 5.4). En klassifisering av satellittdata basert på et direkte uttrekk av treningsdata fra vegetasjonskartet ble ikke vellykket. I en sammenlikning med satellittbildene var tydelig at de digitaliserte polygonene var sterkt romlig generalisert.

Det ble gjort en hel rekke forsøk på å trekke ut "essensen" i vegetasjonskartet. En omkodning til bare seks klasser ble valgt. Alle blandede klasser i kartet ble luket bort. Nyten av tilleggssymbolene i kartet ble vurdert, men

ikke funnet særlig nyttige da dette ikke virket systematisk gjennomført. Ved å studere vegetasjonskartet mot billedata fra SPOT5, IRS-1C og fly, kunne vi for hver vegetasjonsklasse sette opp hvilke minimums- og maksimumsverdier NDVI-indeksen hadde i sommeropptaket fra Landsat. Flere runder med siling der usikre data ble fjernet samtidig som variasjonen ble ivarettatt var nødvendig for i det hele tatt å kunne bruke disse dataene. 20 % av vegetasjonskartet var igjen etter siling. Noe tilleggssymbolisering måtte til.

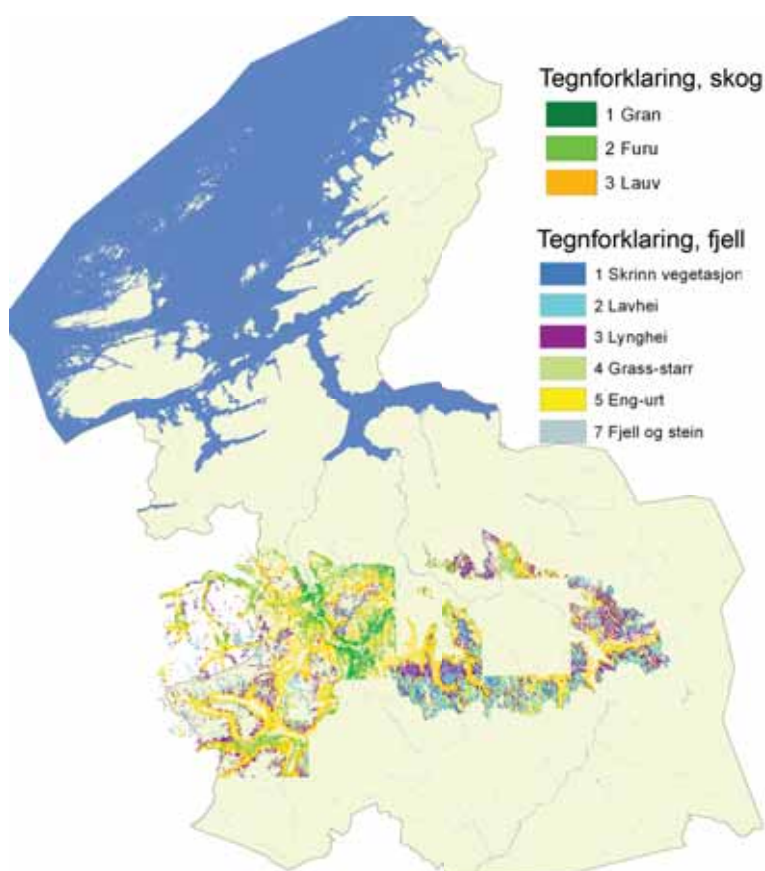
Vegetasjonskartet gav til slutt data som kunne brukes til trening av satellittdata. Det var arbeidskrevende å jobbe med data som gav så lite samsvar med referansepunkt og billedata. Antallet klasser ble skuffende få, da en viss grad av gjenkjenning av klassene i Landsat-bildet var nødvendig på grunn av usikkerheten i treningsdataene. Det ble heller ikke mulig å skille intakte lavmatter fra slitte lavmatter.

5.1.4 Andre tilgjengelige kartdata

5.1.4.1 Vegetasjonskart fra andre institusjoner

Vi innhentet analoge vegetasjonskart med tanke på å hente treningsdata fra disse. Vegetasjonskartene Flekkvika, Gauldalsvidda, Gåvålia, Nord-Fosen, Sandaneset og Garbergselva, er skannet til TIFF-format og georeferert til WGS 84 UTM sone 32. Kartene er basert på ØK eller 1: 50 000 kart. Det er deretter tegnet inn vegetasjonsavgrensinger som er fargelagt eller markert med symbol i henhold til legenden. Da dette var gamle papirkart viste det seg vanskelig å georeferere kartene slik at den geometriske nøyaktigheten ble god nok til bruk sammen med nyere kart og billedata.

Videre viste det seg å være tidkrevende å benytte ikke-vektoriserte tematiske kart selv når de var kartreferert. Dette skyldes at det var vanskelig å lese farge og tegnkoder i kartene og at det må digitaliseres ut flater eller punkt manuelt. I satellittbildene kunne en også se stor variasjon innenfor hver avgrensning i vegetasjonskartene. De plantesosiologiske legendene i disse kartene er heller ikke lett å overføre til fysiske egenskaper egnet for automatisert fly- og satellittbildekartlegging.



Figur 5.4. Dekning av vegetasjonskart fra NIJOS. Omkodet til forenklete klasser for treslag og fjellvegetasjon.

5.1.4.2 Kart over naturtyper

Det ble plukket ut en del flater fra Naturbase (<http://dnweb5.dirnat.no/nbinnsyn/>) til trening. Disse flatene var sterkt generalisert, men etter visuell kontroll mot fly- eller satellittdata ble noen av dem lagt til treningsdataene. En del bestander av furuskog ble hentet inn på denne måten.

5.1.5 Høyoppløselig bildemateriale

Høyoppløselige fly- og satellittbilder ble kun benyttet til visuell tolkning og kontroll i dette prosjektet. Ved å sammenlikne disse bildene med Landsat kunne en identifisere bestander av for eksempel myr, lauv- og ungsog. Bruk av automatisk klassifikasjon på slike bildedata er for tiden under utprøving andre steder i landet.

5.1.5.1 Egne opptak av flybilder

I august 2001 ble det foretatt stripeflyginger med småfly utstyrt med digitale kamera og en GPS/INS løsning over Sør-Trøndelag (Lieng m. fl. 2002). To kameratyper ble utprøvd, et Kodak 760 digital speilreflekskamera og et Duncan-Tech firebånds kamera. Stereobetraktning var mulig med overlappende bilder fra begge kameraene, men geometrisk hadde bildene en utilfredsstillende presisjon. På grunn av at bildene kunne ha geometriske unøyaktigheter på opp til 100 meter, ble de kun benyttet til innsamling av punkter i felt og til skjermdigitalisering der det forelå annet materiale til støtte for å få en kartriktig plassering. Det ble imidlertid laget to ortofoto av Kodak bildene over større områder, Gaula deltaet og Svorka nordøst for Meldal.

5.1.5.2 Nye digitale flyfotokamera

Det har siden 2003 vært gjort forsøk med bruk av nye digitale spesialkamera for vertikalfotoografering fra fly i Norge. Fra 2004 var slike kamera i operasjonell drift. Dataene har en spektral informasjon som gjør dem meget egnet til fangst av referansedata, og med nøyaktige GPS- og INS-målere kan disse foto georefereres til en meters nøyaktighet uten bruk av bakkeinnmålte støttepunkt. Denne datakilden

er derfor meget lovende for bruk til å etablere treningsdata for satellittkartlegging. Det ble ikke mulig å prøve slike bildedata i Sør-Trøndelag, men de er benyttet i kartleggingsprosjekt på Hardangervidda og i Setesdal/Ryfylkeheiene.

5.1.5.3 FKB ortofoto

Tilgangen til ortofoto har blitt betydelig forbedret i 2005 med etableringen av "Norge i Bilder" som Web Map Server (WMS). Dette har kommet mot slutten av kartleggingen i Sør-Trøndelag og har bare blitt delvis utnyttet. Farge og svarthvite ortofoto er ikke alltid like gode til tolkning av vegetasjon, men i dette prosjektet er de benyttet blant annet til vurdering av tett og åpen skog, tresatt og ikke tresatt myr (se <http://www.norgeibilder.no/>).

5.1.5.4 Høyoppløselige satellittbilder

Med høyoppløselige satellittbilder menes data hvor den romlige oppløsningen er bedre enn 6 meter. Dette omfatter satellittene SPOT5 og IRS med pankromatiske bånd på 5 og 5,8 meter, og de meget høyoppløselige satellittene hvor det pankromatiske båndet har 0,6-1 meters oppløsning. For satellitter i oppløsningsområdet 5-6 meter er det behov for mange feltdata som støtte for å kunne gjøre visuell tolking, for satellitter med oppløsning på 1 meter eller bedre kan mye tolkes bare ved å studere bildene. De meget høyoppløselige satellittene (IKONOS, Quickbird og Orbview) har i tillegg til den økte visuelle gjenkjenning også radiometriske egenskaper som er de samme som Landsat bånd 1-4.

En stor del av Sør-Trøndelag er dekket av satellittbilder med 5 meters oppløsning fra SPOT og IRS satellittene. Sammen med observasjoner gjort i felt er disse blitt benyttet til å digitalisere flere bakkesannheter. Fjellvegetasjon, menneskeskapte inngrep og vekststadier i skogbruket er tema som kan tolkes tilfredsstillende i disse bildene. Fjellbjørkeskog og åpen skog i nordboreal sone kan være heller vanskelig å tolke i en slik oppløsning, grunnet forvekslingen som oppstår med mark- og buskvegetasjon.

I to mindre områder innenfor studieområdet ble det i 2004 bestilt opptak med QuickBird



satellitten for andre prosjekter. Vi fikk tilgang til disse bildene mot slutten av 2004, men det tok noe tid før de var geometrisk korrigert. Dermed ble bildene ikke tatt i bruk før mot slutten av prosjektet. De høyoppløselige satellittbildene var til stor nytte for å innhente nye treningsdata og til verifisering, spesielt i Østerdalen der vi hadde lite referansedata.

5.1.6 Punktbasert skjermdigitalisering

Dag Inge Øien ved NTNU Vitenskapsmuseet i Trondheim har sammen med andre samlet inn punkter fra en hel rekke lokaliteter i Sør-Trøndelag basert på SatNat-kodesettet (Hjeltnes & Øien 2003). Hensikten med dette arbeidet var å skaffe detaljerte data til trening og validering av klassifiseringen. Siden dataene også skulle brukes til validering var det viktig at sikkerheten i klassifikasjonen var meget god. Samtidig var det behov for et stort datamateriale for å kunne fange opp all variasjon som var innen kartleggingsområdet. Det ble derfor valgt en kombinasjon av feltregistreringer og registreringer fra andre kilder som eksisterende vegetasjonskart, flyfoto (både de som ble tatt i prosjektet og andre), samt IRS og SPOT5 satellittbilder.

Problemet med punkt samlet inn i felt er at det er vanskelig å ta hensyn til den romlige skalaen informasjonen skal representere. En

observasjon beskriver ikke gjennomsnittet av en Landsat piksel (0,9 dekar), men ofte et mindre areal. For å nyttiggjøre seg av slike observasjoner er et detaljert bildegrunnlag til stor hjelp når informasjonen skal overføres til en lavere oppløsning. For å få en effektiv skjermdigitalisering av punkt fikk vi laget til et eget script i ArcView og utviklet et registreringssystem som i større grad enn eksisterende systemer var egnet for bruk sammen med fly- og satellittbasert bildedata (Hjeltnes & Øien 2003). Totalt samlet Vitenskapsmuseet inn 6174 punkt. I ettertid viser det seg at punktene har en skjev fordeling både geografisk og klassemessig (over- og underrepresentasjon av klasser). Den tematiske og geografiske nøyaktighet er også mangelfullt dokumentert. Det er dermed ikke mulig å rangere punktene etter kvalitet og fjerne usikre observasjoner. Datasetset er likevel det beste vi har for åpne områder og er derfor brukt for validering.

5.2 Satellittbilder og avledede vegetasjonsindekser

5.2.1 Valg av satellittbilder

I denne rapporten er det valgt å fokusere på multispektrale opptak fra den optiske sensoren ETM+ (Enhanced Thematic Mapper) om bord

Tabell 5.1. Satellittdata benyttet i kartleggingen av Sør-Trøndelag. Termiske og pankromatiske bånd ble ikke benyttet i analysene. For IRS-1C LISS ble i tillegg ikke bånd 4 benyttet, grunnet lavere romlig oppløsning og mye støy.

Sensor	Ant. bånd	Path	Row	Dato	Årstid
L7-ETM+	6	198	16	19.mai.02	Vår
L7-ETM+	6	199	15,16	07.mai.01	Vår
L7-ETM+	6	200	15,16	18.apr.03	Vår
L7-ETM+	6	198	16	23.aug.02	Sommer
L7-ETM+	6	199	15	24.jun.01	Sommer
L7-ETM+	6	199	16	06.aug.99	Sommer
L7-ETM+	6	200	15	05.aug.02	Sommer
L7-ETM+	6	201	15,16,17	21.jul.00	Sommer
IRS-1C LISS	3	18	20,21	21.jul.02	Sommer
L5-TM	6	199	15,16,17	09.aug.03	Sommer
L7-ETM+	6	199	16	25.sep.00	Høst
L7-ETM+	6	200	15,16	02.okt.00	Høst

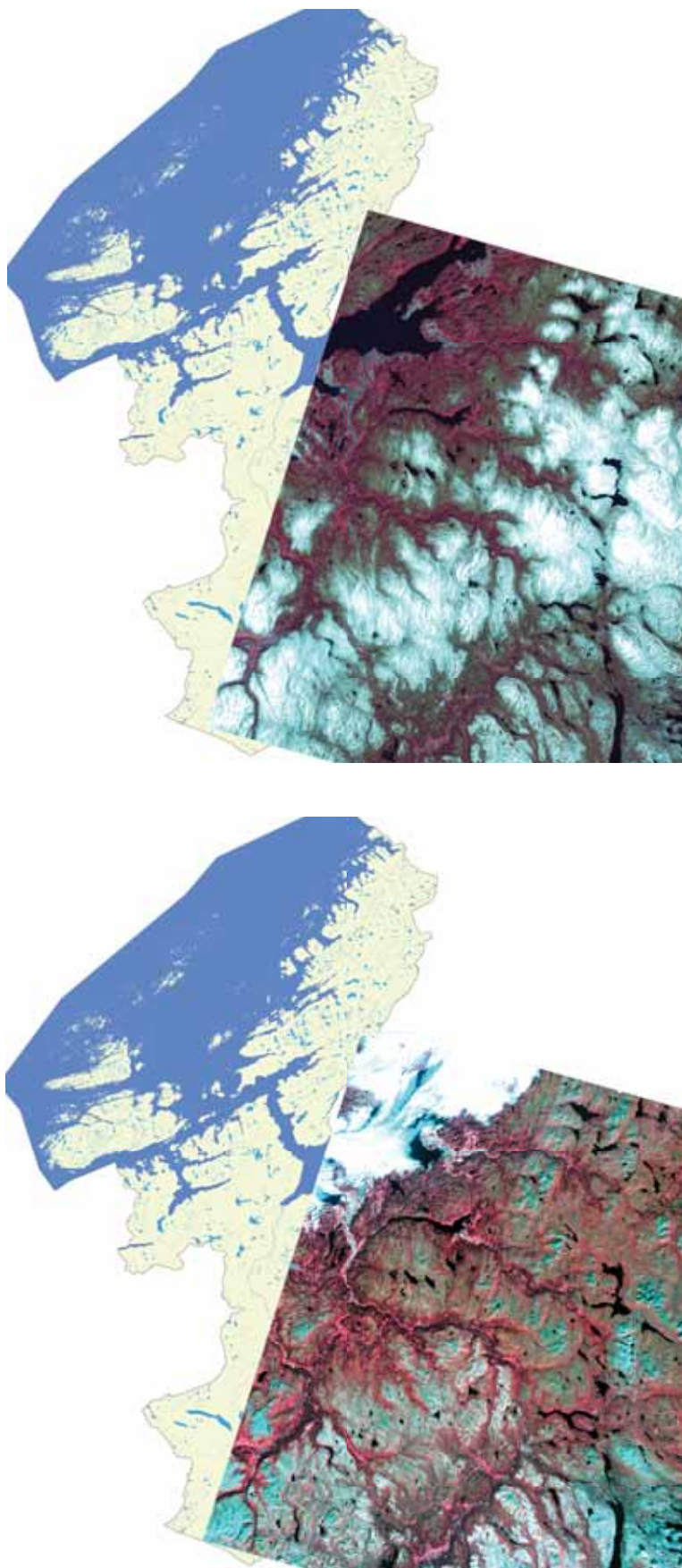
i Landsat7-satellitten. Det er også benyttet data fra ett opptak av Thematic Mapper sensoren i Landsat5 og data fra ett opptak med den indiske satellitten IRS-1C. I multispektrale bilder måles reflektert sollys i et utvalg av synlige, nærinfrarøde og midlere infrarøde bølgelengder. ETM+ sensoren måler refleksjon fra en større del av spekteret enn IRS. Pikseloppløsningen på bakken i de multispektrale Landsat dataene er 30 m x 30 m, mens IRS har 23 m x 23 m.

Dessverre har Landsat7 ikke levert bilder av tilfredsstillende kvalitet siden mai 2003 på grunn av en teknisk feil. Bilder levert etter dette er ikke korrigert for satellittens bevegelse forover og har et sikksakk stripemønster på tvers i bildene (se <http://landsat.usgs.gov/>). En etterfølger for Landsat 7 er ikke ventet i drift før i 2009-2010. IRS og SPOT satellittene er inntil videre et alternativ med 10 til 20 meters oppløsning i de multispektrale bånd og en innkjøpspris fra leverandør på under 10 kroner per km².

Ved å benytte opptak fra flere datoer kan det gjøres en *multitemporært* analyse. Hensikten med dette er å utnytte de fenologiske endringer i vegetasjonen til å bedre klassifikasjonen. Vår klassifikasjon baserer seg på multitemporale analyser med opptil tre Landsat ETM+ bilder fra vår, sommer og høst (figur 5.5). For å få til et oppdatert arealdekkkart, ble kun opptak fra 1999 eller nyere valgt ut. Det var vanskelig å plukke ut skyfrie opptak fra Sør-Trøndelag fra denne perioden. Til sammen ble 20 scener funnet brukbare (tabell 5.1).

5.2.2 Geometrisk korreksjon og referansesystem for bilder

Hver scene er geometrisk korrigert i forhold til N50-kartdata og andre Landsat-scener. For ETM+ brukes PAN-båndet som grunnlag for den geometriske korreksjonen. Normalt er den geometriske nøyaktigheten innenfor 15 meter, men større unøyaktigheter forekommer. Dette skyldes i hovedsak feil i N50 (gamle skannede kartblad) og 16. linjes unøyaktighet i Landsat-sensoren. Sensoren ETM+ danner en bildefil ved å skanne 16 linjer av gangen på tvers av bevegelsesretning i en sveip (Lillesand & Kiefer 2000). Uprosesserte Landsat data kjennetegnes ved at blokker av 16 linjer i bildet er forskjøvet i forhold til hverandre. Ved prosesse-



Figur 5.5. Landsat Satellittbilder fra 19. mai 2002 (øverst) og 23. august 2002 (nederst) gav vår og sommer opptak som muliggjorde kartlegging av de østlige deler av Sør-Trøndelag og Nord-Østerdalen.

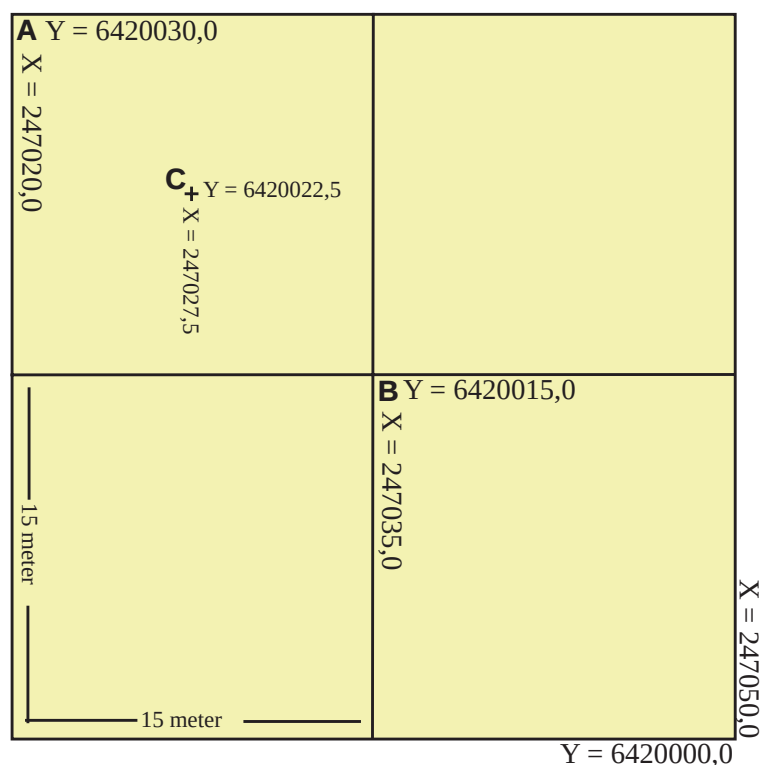


ring til formatet 1G, som er vanlig å bestille fra leverandør, blir sveipene plassert rett i forhold til hverandre slik at feilene blir minimert. Men 1-4 pikslers avvik kan tidvis forekomme. Dette kan lett avsløres ved visuell kontroll; legg to Landsat opptak oppå hverandre i et bildebe-handlingsprogram og bytt raskt mellom dem. 16. linjefeil gir en stripevis bevegelse av bildet på tvers av skjermen.

For å kunne sette sammen multitemporære opptak, er en avhengig av at bildene har høy geometrisk nøyaktighet, samme oppløsning og har samme nullpunkt som koordinatrefe-ranse. Derfor er det viktig å være klar over om rasterdata referer seg til **kant eller sentrum** på pikselen. Når en jobber med satellittbilder med forskjellig oppløsning er det mest hen-siktsmessig å referere hvert bilde i forhold til koordinatverdi til kanten av et piksel og ikke til sentrum. Da vil for eksempel fire piksler med 15 meters oppløsning falle innenfor et 30

meters piksel (figur 5.6). For terrengmodeller er imidlertid ofte pikselens midtpunkt referanse, da en gjerne tenker seg senterpunktet som et målepunkt med en høydeverdi.

I forkant av kartleggingen av Sør-Trønde-lag ble alle satellittdata geometrisk korrigert på nytt og referert til kant av piksel. Når kant brukes som referanse skal **en divisjon av ko-ordinatverdien til kanten av en piksel med pikselens oppløsning i meter gi et heltall**. Har noen datalag i analysen senterreferanse vil de bli flyttet 0,5 piksel når lag blir kombinert. For 30 meters bilder er dette å tilføre en feil på 21 meter i bildet, altså en dobling av bildets geometriske feil. Det vil ofte være nok til at en bommer på smale korridorer i landskapet og vannkant forflyttes inn i skog etc. Dette må derfor taes hensyn til når bilder bestilles fra distributør, når det foretas geometrisk opp-retting eller annen bildetransformasjon (eng. resampling) eller ved konvertering fra vektor til raster formatet.



Figur 5.6. Referering av bildedata til et standardisert rutenett. Illus-trasjon av en Landsat piksel (30 meters ETM) inndelt i fire piksler (15 meters PAN). A markerer kant av PAN og ETM pikslene med koordinater som er dividerbare med pikslenes oppløsning, altså 15 og 30 meter. B markerer senter av ETM pikselen og C senter for PAN pikselen. Verktøy som ArcGIS eller Erdas Imagine bruker koordinatene for sentrum av en piksel når bildets utstrekning oppgis.

5.2.3 Normalisert vegetasjonsin-deks (NDVI)

Vegetasjonsindekser er mye brukt i gro-vere vegetasjonskartlegging (Jensen 2000). Det finnes ulike typer vegetasjonsindekser, men felles for de fleste er at de bruker røde og nærinfrarøde kanaler. Dette skyldes at det fin-nes en direkte sammenheng mellom reflektans i det nærinfrarøde lyset og plantenes biomasse. Tilsvarende er det en invers sammenheng mellom reflektansen i den røde kanalen og plantenes biomasse. Det er spesielt klorofyll i blader som absorberer rødt sollys, mens det i nærinfrarødt foregår refleksjon i bladenes cellestruktur. Disse spesifikke reflektansegen-skapene til vegetasjon gjør at en kombinasjon av røde og nærinfrarøde kanaler er svært egnet til vegetasjonskartlegging. I denne rapporten er den normaliserte vegetasjonsindeksen (NDVI) valgt å benyttes:

$$NDVI = (NIR - R)/(NIR + R),$$

hvor NIR er nærinfrarødt og R er rødt (Tucker 1979). NDVI er normalisert slik at den varierer mellom -1 og 1. Jo nærmere verdien er 1, desto mer sannsynlig er det at en piksel inneholder

mye vegetasjon. I beslutningstreet blir NDVI omskalert til 8 bit, altså verdier mellom 0 og 255. 128 tilsvarer da NDVI lik 0.

5.2.4 Andre indekser

Ved å dividere $NDVI_{sommer}$ på $NDVI_{vår}$ dannes et rasterlag som blant annet har betraktelig høyere verdier i områder dominert av gras og lauv enn barskog. Denne indeksen har gitt god forklaringsgrad i beslutningstreanalysen. En indeks av $NDVI_{høst}$ dividert på $NDVI_{sommer}$ gav ikke samme uttelling. Indeks for fuktighet (eng. wetness) er en av flere andre indekser som kan vurderes senere for slike analyser (Collins & Woodcock 1996).

5.3 Basiskart

5.3.1 N50 arealtema

Statens kartverks topografiske kartserie N50 er benyttet for å dele kartleggingsarealet inn i hovedklassene skog, myr, jordbruksmark og åpen mark. N50 vann og markslag fra 2002 ble konvertert til raster med samme oppløsning og referanse som satellittbildene. Deler av Sør-Trøndelag består av eldre kartblad som ikke er à jour og har lav geometrisk nøyaktighet. Markslag blir heller ikke hyppig oppdatert. Mange steder vil en derfor oppleve at nyere jordbruksareal, industriområder og andre anlegg fortsatt er kartlagt som for eksempel skog eller myr i N50. Gjengroing mot høgfjellet og jordbruksareal bidrar også til at N50 markslag kan være misvisende.

5.3.2 N50 terrengmodell og avledede rasterlag

Opprinnelig ble det i SatNat prosjektet laget en 30 meters terrengmodell over Sør-Trøndelag ut fra N50 høydekurver, høydepunkt og vann-tema. Senere ble denne erstattet av en forbedret digital høydemodell fra Statens Kartverk med 25 m romlig oppløsning. Terrengmodellene er brukt til å avlede terrenghelning (eng. slope) og terrengeksposisjon (eng. aspect). Beregning av disse parametrene er gjort basert på målt høydeverdi i nabopikslene innenfor et 3 x 3

vindusfilter. Beregningsmetoder for terrengparametere er beskrevet i Wilson (2000).

Varmeinnstråling beregnes som en indeks basert på terrengets helning og helningsretning etter metodikk beskrevet i Parker (1988):

$$\text{Varmeindeks} = (\cos \alpha - 225) \times \tan \beta,$$

hvor α er terrengeksposisjon og β er terrenghelning, begge i grader. Varmeindeksen har kontinuerlige verdier som varierer mellom -1 og 1. Et tilsvarende raster ble laget for å modellere solinnstrålingen i satellittbildene. Solretningen (eng. sun azimuth) er å finne i metadataene som følger med bildene. Siden flere scener ble brukt samtidig valgte vi å bruke en gjennomsnittlig verdi for solretning, 164 grader.

5.3.3 Vegetasjonssoner

Et kart over vegetasjonssoner i Sør-Trøndelag er utarbeidet på bakgrunn av metodikk beskrevet i Moen & Odland (2002) og Lieng (2002). Kartet er en videreutvikling av vegetasjonssonekartene presentert i Moen (1998). Vegetasjonssoner er inndelt basert på høyde over havet, varmeinnstråling og grad av oseanisk eller kontinentalt klima. Både terrengmodellen avledet fra N50 kartdata, varmeindeksen og graden av oseanitet er alle kartdata med kontinuerlige variabler. Dermed får vi også et vegetasjonssonekart med *kontinuerlige* verdier. Dette kan så i etterkant inndeles i diskrete vegetasjonssoner hvis ønskelig. I analyser hvor disse kartdataene brukes sammen med satellittdata er det derimot ønskelig å bruke datasettet med kontinuerlige variabler da dette gir mulighet til en langt mer variert inndeling enn ett kart kun bestående av de seks sonene (boreonemoral til mellomalpin) som finnes i Sør-Trøndelag. I analysene er vegetasjonssoner benyttet spesielt for å skille typer av løvskog. Vegetasjonssoner kunne ikke benyttes som lag i beslutningstre analysen. Dette skyldes at treet vil finne for god forklaringsgrad i dette datalaget. Er det ingen observasjoner av furu blant treningsdataene i nordboreal kan treet f. eks. sette en regel om at furu ikke finnes dersom vegetasjonssoneverdien er over 3,8 (i mellomboreal sone). All furu i nordboreal sone vil da bli kartlagt som lauv eller gran.



6 Produksjonsprosessen

I kartleggingen av Sør-Trøndelag og Nord-Østerdalen inngikk mange trinn. Nedenfor følger en gjennomgang av arbeidet;

1. Satellittbilder som gir optimal dekning av området ble bestilt.
2. Bildene ble satt sammen til skyfrie multitemporale datasett.
3. Det ble gjort en vurdering om datasettene skulle splittes opp i økologiske regioner.
4. Referansedataene ble evaluert og klargjort for hvert datasett.
5. Delområdene (se Kap. 5.1.1) ble klassifisert med beslutningstre analyse.
6. Deretter ble det gjort en visuell editering, sammenslåing av noder og validering. Hvis vi ikke godkjente resultatet, gikk vi tilbake til punkt 3.
7. De klassifiserte datasett fra delområdene ble slått sammen til kart for hvert temalag og siden omkodet til ett heldekkende arealdekkekart for studieområdet.
8. Temalagene ble validert mot referansedata.

6.1 Etablering av skyfrie multitemporale datasett

De geokorrigerte satellittbildene ble importert til PCI Geomatica ver. 9.1 (heretter omtalt som PCI, <http://www.pci.on.ca/>) der all bildeanalyse ble utført. PCI har det fortrinn at en kan arbeide med et stort antall datalag i en bildefil og utføre kalkulasjoner mellom lagene på en effektiv måte. Hver bildefil som ble laget inneholdt flere satellittbildeopptak (hvert på tre til seks bånd) NDVI for hvert opptak, eventuelle andre indekser, varme- og solindeks og i tillegg om lag ti rasterlag for bruk til treningsdata og klassifikasjon. Bildeprosessering skjer på denne måten raskere i forhold til andre bildebehandlingsverktøy da en slipper å etablere nye filer hele tiden. Arbeidet blir dermed mer oversiktlig. Hvert datalag inneholder i tillegg historikk for hvilke prosesser det har gjennomgått. PCI muliggjør også bruk av forenklede 1-bits rasterlag (eng. bitmap) som rasjonaliserer bl.a. uttrekk av data.

For hvert satellittbilde (tabell 5.1) ble det laget en maske (bitmap) over sky- og skyggefritt areal ved manuelt å identifisere skygger, skyer og dis på bildene. Det finnes automatiske metoder for dette, men en manuell metode ble foretrukket siden en del dis og skygge var vanskelig å skille fra fjell og lav. For å dekke 99,5 % av Sør-Trøndelag skyfritt var det nødvendig med sju delområder (tabell 6.1). Det viste seg spesielt vanskelig å finne skyfrie opptak over Røros kommune. De resterende 0,5 % av fylket ble kartlagt visuelt med andre Landsat-opptak enn de som er referert til i avsnitt 5.2.1.

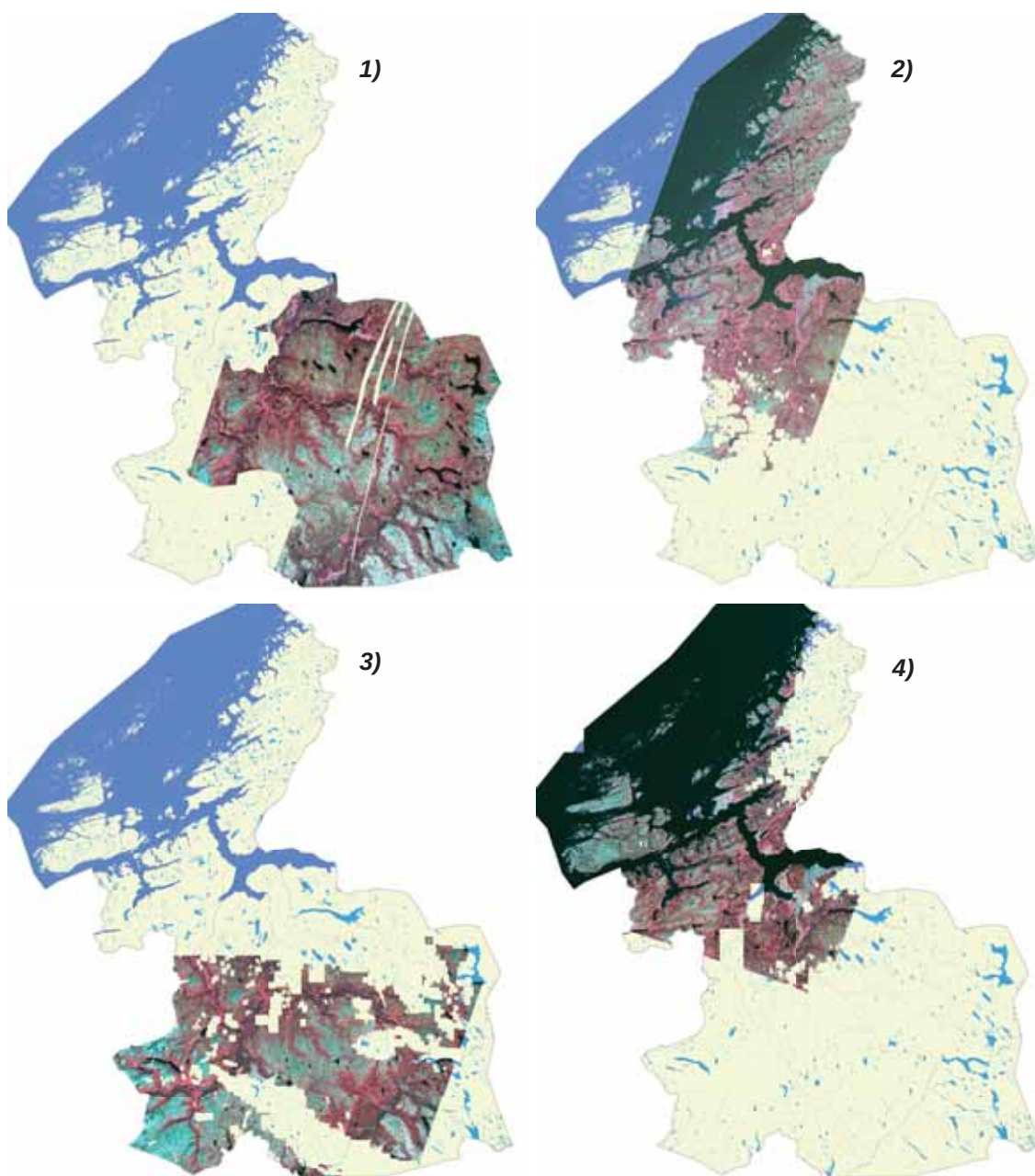
De sju områdene overlappet hverandre i stor grad og gjorde det mulig å bruke en del av de samme referansedataene til trening i flere enn ett område (figur 6.1). Områdene viste seg å ha ulik egnethet til klassifikasjon. Område 1 og 2 gav bedre samsvar med referansedataene enn område 3. Sommeropptakene ble statistisk valgt ut tidligere i oppbygningen av beslutnings-trærne enn vår-opptakene, og forklarer dermed de viktigste valgene i beslutningstreanalysen. Landsat7 opptaket fra 19.mai 2002 var imidlertid meget godt egnet til å kartlegge løvskog. Høstopptakene hadde minst innflytelse. Område 6 ble kun benyttet til å kartlegge dyrka mark og område 7 kun til fjell og åpen mark.

6.2 Inndeling i regioner

Lavskogen i øst viste seg å ha meget ulike spektrale (og fysiske) egenskaper enn resten av Sør-Trøndelag som er sterkt oseanisk preget. I forhold til de økologiske regioner fra USA beskrevet i avsnitt 4.1.6 er det klart at Sør-Trøndelag inneholder minst to slike regioner. For kartleggingen av Sør-Trøndelag betydde dette at en fikk dobbelt arbeid med å etablere referansedata og kjøre beslutningstre for vest og øst hver for seg. Ved en kartlegging av Norge burde landet deles inn i minst fem regioner. NIJOS har inndelt landet i 45 landskapsregioner (Elgersma & Asheim 1998), en tilpasning av dette datasettet er sannsynligvis egnet til formålet.

Tabell 6.1. Det ble etablert sju delområder for å få klassifisert hele studieområdet. For hvert delområde ble det laget en bildefil som inneholdt satellittdata fra alle opptakene som inngikk og støttedataene.

Område	Sted	Vår	Sommer	Høst
1	Østre del	L7 19.mai.2002	L7 23.aug.2002	
2	Fosen	L7 07.mai.2001	IRS-1C 21.jul.2002	L7 02.okt.2000
3	Oppdal	L7 07.mai.2001	L7 06.aug.1999	L7 25.sep.2000
4	Hitra	L7 18.apr.2003	L7 05.aug.2002	L7 02.okt.2000
5	Hele fylket		L5 09.aug.2003	
6	Frøya		L7 21.jul.2000	
7	Sørlig del		L5 09.aug.2003	L7 25.sep.2000



Figur 6.1. Oversikt over behandlings områdene 1 (øverst til venstre, se også figur 5.5), 2 (øverst til høyre), 3 (nederst til venstre) og 4 (nederst til høyre).



6.3 Uttrekk av referanse-data

Etablering av referansedata for trening og validering ble et meget tidkrevende arbeid i dette prosjektet, og utgjorde hovedtyngden av arbeidet med å lage kartene. Siden det ikke var noen godt egnede datakilder for bruk til treningsdata for kartleggingsområdet Sør-Trøndelag og Nord-Østerdalen var det også i utvelgelsen av treningsdata nødvendig med mange trinn.

I denne prosessen fulgte vi strategien beskrevet under og vist i figur 6.2 med temakartet treslag som eksempel.

1. Vi undersøkte først om det fantes punktdata som kunne nyttes som referansedata. Slike data ble så kvalitetskontrollert ved visuell sammenlikning av fly- eller satellittbilder. For studieområdet Sør-Trøndelag og Nord-Østerdalen ble alle vurderte punktdatasett satt til side for validering da antallet punkt var for lavt til bruk som treningsdata.
2. Deretter undersøkte vi alle arealdata fra kart. Aktuelle klasser ble valgt ut. Ny visuell tolking ble gjort mot billedata. Siling, det vil si en fjerning av antatt blandede klasser, ble foretatt (f. eks. bjørkekratt i lynghei). Som kilde for siling brukte vi terskelverdier for NDVI og rød kanal i billedata fra den sensor som dannet hovedkilden til satellittolkingen (IRS eller Landsat).
3. Vi vurderte så om dekningen av kartdataene var representative for området som skulle analyseres. Om dette ikke var tilfellet forsøkte vi å hente inn referansedata fra andre kilder. Der det var dekning med flyfoto eller høyoppløselige satellittdata (Quickbird, SPOT5) og aktuelle klasser kunne identifiseres visuelt i disse kildene, foretok vi en skjermdigitalisering av nye treningsdata. På de behandlingsområdene der vi hadde for liten dekning med treningsdata supplerte vi med data fra klassifiserte produkt der andre behandlingsområder overlappet. Dette ble gjort for område 4 (Frøya og Hitra) som hadde få treningsdata.
4. Dersom det var referansedata fra flere kilder

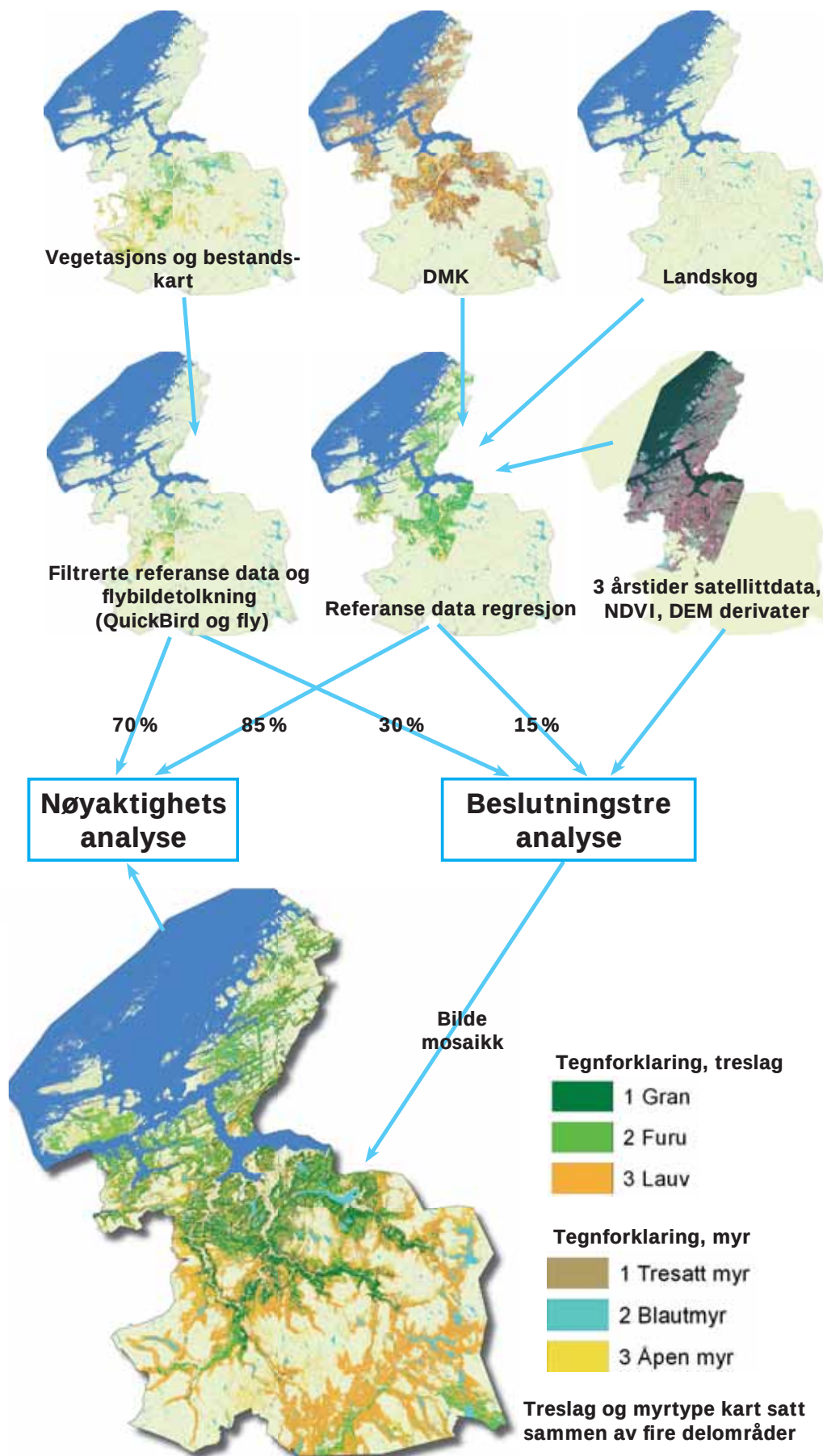
over ett område var det nødvendig å rangere dem ut fra kvaliteten. De ulike kilder ble slått sammen ved at de beste data fikk høyere prioritet og overskrev andre data der det var overlapp.

5. Antallet observasjoner per klasse måtte jevnnes ut i treningsdatasettet. Klasser med høy spektral variasjon ble gitt noe flere observasjoner, men ikke flere enn at en stor klasse fikk mer enn fire ganger antallet til klassen med færrest observasjoner.
6. Satellittdata, NDVI og parametere avledet av terrengmodell ble sammen med treningsdata skrevet ut til tekstfil for import i statistikkprogram (modellberegning). Antallet observasjoner som var tilgjengelig for trening ble begrenset til maksimalt 400 000.

6.4 Beslutningstre-analyser

Alle datalag ble eksportert på tekstformat fra PCI. For område 2 med opptak fra tre årstider ble dette 15 bildebånd, tre NDVI bånd, et bånd med sommer NDVI dividert på vår NDVI, varmeindeks og treningsdata, til sammen 21 kolonner. Typisk antall observasjoner som ble benyttet til oppbygning av hvert beslutningstre var 400 000. Da ble 30 % av tilgjengelige referanse data benyttet og 70 % gjensto til validering. Har en færre referansedata vil det være mer fordelaktig å benytte 70 % av observasjonene til trening og 30 % til validering (Vikhamar m.fl. 2004). Antallet treningspunkter ble valgt slik at datamengden skulle være lett håndterbar i en statistikkpakke. Det var i alle analysene tilstrekkelig antall observasjoner, fordelingen av observasjonene var imidlertid ikke ideell.

Statistikkprogrammene R med modul `rpart` (<http://cran.r-project.org/>) og S-PLUS med modul `tree` (Chambers & Hastie 1992, Insightful Corporation 2005) ble benyttet til å lage beslutningstreet i analysen. Programmene gav samme resultat, men S-PLUS kunne i tillegg levere et plott av treet på PDF-format. Dette var nyttig for å forstå klassifikasjonsprosessen. Hvert



Figur 6.2. Oversikt over prosessen for å fremstille kart over treslag.



beslutningstre inneholdt 20- 80 noder avhengig av antall klasser og kompleksiteten i dataene. Statistikkpakkene viser også et tall på grad av samsvar mellom treningsdata og beslutningstre. Verdier mellom 70 og 90 % var vanlig. Treet som S-PLUS leverte måtte konverteres til programmeringsspråket C som PCI kan lese. Dette ble foretatt med et Perl skript. S-PLUS treet ble også benyttet til å lage en fargetabell for nodene ved hjelp av Microsoft Excel. I PCI kjørte vi treet på rasterfila det ble tatt ut treningsdata fra. Dette gav et raster med nodene fra treet med farger importert fra fargetabellen.

Vi kontrollerte så noderasteret visuelt i PCI eller ArcMap. Dette gav mulighet for å endre noen noder eller vurdere om klassifikasjonen kunne godtas eller forkastes. Om forkasting ble valget, måtte vi gå tilbake til preprosesseringen av referansedataene og se om fordelingen og silingen av disse kunne forbedres. Var resultatet godkjent aggregerte vi nodene til klasser slik at de kunne valideres mot referansedataene.

Vi foretok analyser av 21 beslutningstrær for å lage kartet over Sør-Trøndelag. Hver av

disse analysene er imidlertid kjørt flere ganger for å få til et best mulig resultat.

6.5 Sammensetting av klassifiserte raster

De klassifiserte delområdene ble mosaikkert for å lage et endelig heldekkende temalag. Siden områdene hadde stort overlapp, kunne vi for disse arealene sette opp en prioriteringsrekkefølge som grunnlag i mosaikkingen. Vi valgte å basere prioriteringen på kvaliteten (forklaringsgraden) på hver beslutningstreanalyse, på den visuelle kontrollen og på resultatet av feilmatrixene.

Vi eksporterte til slutt klassifiserte data og data fra de kilder vi valgte som endelig validering til en tekstfil fra PCI. Med et Perl-skript beregnet vi feilmatrixer slik at vi fikk informasjon om den endelige kvaliteten på hvert temakart.

7 Resultat

7.1 SatNat kartene

Kart over Sør-Trøndelag er laget av satellittbilder med opptak fra opptil tre årstider. I fjellet er det vanligvis benyttet opptak fra to årstider. Effekt av terrenghelning og eksposisjon er ivarettatt ved at et datalag som beskriver solinnstråling (varmekoeffisient) er tatt med i analysen. Kartet over vegetasjonssoner er benyttet for inndeling i høydelag (eksempelvis for å skille varmekjær og boreal lauvskog). Der det finnes DMK (digitalt markslag kart) er dette i noen tilfelle benyttet til omkodning av klasser ut fra bonitet (eksempelvis for å skille hogstflate fra åpen skrinns furuskog).

Alle kartene er til fri benyttelse og kan lastes ned fra Direktoratet for naturforvaltnings nettside (www.dirnat.no/satnat). Filnavnen står under kapitlene for hver enkelt temalag. For generell referanse av kartene brukes betegnelsen: "SatNat arealdekkkart Sør-Trøndelag", og for de enkelte temalag benyttes: "SatNat Sør-Trøndelag" temalagets navn". Alle temakart er lagt ut på ERDAS Imagine format i UTM sone 32 kartprojeksjon og WGS84 datum og med 30 meters pikseloppløsning.

Vi har forkortet SatNat til SN og brukt fylkeskoden for Sør-Trøndelag (16) før temalagets type og navn.

7.2 Samsvar mellom eksisterende kartdata

Av tilgjengelige data for skogarealene i Sør-Trøndelag er det bare Landsskogtakseringens punktnett som er heldekkende og representativt fordelt innen studieområdet. En ulempe med disse dataene for validering av SatNat-kartene er at få punkter fra Landsskogtakseringen overlapper med de viktigste kildene for treningsdata – bestandskart og vegetasjonskart fra NIJOS. En annen er at de dekker kun et areal mindre enn 1/3 av et Landsat-piksel. De er likevel de beste data som er tilgjengelig for validering i skog.

For åpne områder er det ingen presise datasett tilgjengelig som dekker Sør-Trøndelag og Nord-Østerdalen. De mest aktuelle for validering er de punktdata som ble innsamlet for SatNat-prosjektet av NTNU Vitenskapsmuseet. Disse 6174 klassifiserte punktene er derimot klyngevis fordelt slik at de ikke gir et representativt bilde av kartleggingsområdet. Men noe bedre datasett for validering av klassifikasjon i åpne områder finnes ikke, så disse data er brukt i en sammenlikning med den satellittbaserte klassifikasjonen. De gir tross alt et visst bilde av kartleggingen i de områder datasettene overlapper.

For jordbruksmark er 3Q-datasettet det meste egnede for validering. Vi hadde tilgjengelig kartdata for 28 prøveflater i Sør-Trøndelag. 3Q-kartene overlappet 5387 klassifiserte Landsat-piksler. Disse ble nytt til validering av temakartet dyrka mark.

Siden disse tre kartkildene utgjør vårt mulige valideringsgrunnlag har vi sett nærmere på hvordan disse samsvarer med noen av de treningsdata vi har brukt. På grunn av at vi bare hadde 64 punkter der det var overlapp mellom bestandskartene og punkter fra Landsskogtakseringen valgte vi å redusere temakartet over tilstandsklasser til bare to nivå i sammenlikningen med landskogsdata (tabell 7.1).

Tabell 7.1. Feilmatrix som viser tilstandsklasser fra bestandsdata (Skogeierforeningen Nord) kontrollert mot tilsvarende informasjon i Landsskogtakseringen. Med ung skog menes hogstklassene I t.o.m. III og for eldre skog hogstklassene IV og V. I matrisen er inkludert produsent (inkluderende) og bruker (ekskluderende) nøyaktighet for hver klasse.

Landsskogtakseringen				
Bestandstakst	Ung skog	Eldre skog	Totalt	Bruker
Ung skog	18	8	26	69 %
Eldre skog	4	34	38	89 %
Totalt	22	42	64	
Produsent	82 %	81 %	Samlet:	81 %
Khat = 66%, Tau = 66%				



Tabell 7.2. Feilmatrix som viser dominerende treslag fra NIJOS vegetasjonskart og skogbrukets bestandskart sammenliknet med data fra Landsskogtakseringen.

Landsskogtakseringen					
Bestandskart	Gran	Furu	Lauv	Totalt	Bruker
Gran	38	9	9	56	68 %
Furu	8	22	7	37	59 %
Lauv	8	3	19	30	63 %
Totalt	54	34	35	123	
Prod.	70 %	65 %	54 %	Samlet:	64 %

Khat = 45 %, Tau = 53 %

Resultatet av denne sammenlikningen viste at i 81 % av tilfellene var det samsvar mellom klassene yngre og eldre skog i disse to datasettene. Dette er to grove relativt entydige klasser, hvor forskjellen i stor grad må tilskrives ulike registreringstidspunkt.

For utarbeidelse av temakartet over tre-type benyttet vi flere kilder. Bestandsdata og NIJOS vegetasjonskart var sammen med regresjonen med DMK kilder til treningsdata på dette kartet. I Sør-Trøndelag hadde vi for liten dekning av bestandsdata og for lite overlap mellom vegetasjonskart og punkter fra Landsskogtakseringen til at vi kunne teste disse kildene til treningsdata hver for seg i forhold til Landsskogtakseringen. Siden begge inngår i treningsdataene har vi slått dem sammen og gjort en samlet validering (tabell 7.2).

For klassifisering av åpne arealer i fjellet er NIJOS vegetasjonskart benyttet til trening. En feilmatrix mellom NIJOS vegetasjonskart og Vitenskapsmuseets punktdata for de tre

klassene lav, lyng og gras-urter gav kun en samlet forklaringsgrad på 38 %, og bare 28 % bedre treff enn en tilfeldig klassifisering (tabell 7.3).

Selv om datamaterialet er lite og det legges ulike arealer til grunn for fastsettelse av klasse-tilhørighet, gir det nok likevel et bilde på hvor stor variasjon det er mellom ulike tolkninger, og vanskene som da oppstår ved bruk av slike kilder som trening for styrt klassifikasjon av satellittdata.

7.3 Lokaliseringsavvik mellom piksel og referansepunkt

For å få et korrekt bilde av klassifiseringsnøyaktigheten har vi gjort beregninger på hvor mye det geometriske avviket mellom pikslene i de ulike temakartene og punktene som er brukt til validering påvirker samlet treffprosent og de statistiske verdiene for Kappa og Tau.

Vi har omkodet segmentene i kartet (sammenhengende piksler av én klasse) til tilfeldige klasser. Deretter har vi beregnet feilmatrixer for senterpunkt og for senterpunkt med nabopiksler. Endringen i disse matrixene, som er basert på kartenes reelle struktur og antall klasser, forteller hvor mye av forbedringen i treffprosent fra senter- til nabopikselanalyse som skyldes kartets heterogenitet. De viser dermed hvor mye det geometriske avviket mellom kart og valideringsdata påvirker nøyaktigheten i hvert enkelt tilfelle. Feilmatrixene for disse analysene finnes i vedlegg A, her er kun gjengitt prosentverdiene.

Differanseverdiene som kom fram ved denne analysen (tabell 7.4) er siden trukket fra prosenttallene beregnet i naboanalysene. Dermed kunne vi beregne korrigerte verdier for samlet treffprosent, Kappa og Tau i disse tre temalagene.

Kappa er en statistisk beregning som skal gi negative verdier eller verdier nær null når det er liten sammenheng mellom to datasett. Dette kommer tydelig fram fra senterpikselanalysene basert på tilfeldig klasseinndeling. Differansen er størst for skogkartene. I skogen finner en større sammenhengende arealer, mens arealtypene er mer mosaikkert i åpen mark.

Tabell 7.3. En sammenlikning av tre sammenslåtte hovedklasser fra NIJOS vegetasjonskart kontrollert mot tilsvarende klasser fra vegetasjonsdata samlet inn av NTNU Vitenskapsmuseet i Trondheim.

NTNU Vitenskapsmuseets data					
NIJOS Veg.kart	Lavmark	Lyngmark	Gras-urte mark	Totalt	Bruker
Lavmark	2	110	6	118	2 %
Lyngmark	7	104	19	130	80 %
Gras-urte mark	0	62	18	80	23 %
Totalt	9	276	43	328	
Prod.	22 %	38 %	42 %	Samlet:	38 %

Khat = 0 %, Tau = 28 %

Tabell 7.4. Effekten av den romlige fordeling på verdiene for samlet treffprosent, Kappa og Tau i temalagene Treslag, Skogtilstand og Åpen mark. Verdiene er beregnet basert på en tilfeldig tildeling av klasser til de sammenhengende segmentene i kartene og feilmatriser utregnet for både senterpiksel og senterpiksel med nabopikslar. Differansen uttrykker hvor stor del av forbedringen i de statistiske verdiene som skyldes avviket i geometrisk lokalisering mellom kart og valideringsdata.

Temalag	Statistisk begrep	Senterpunkt (%)	Naboskap (%)	Differanse
Treslag	Samlet	27	44	17
	Khat	-9	15	24
	Tau	20	34	14
Tilstandsklasser	Samlet	21	37	16
	Khat	1	19	18
	Tau	17	31	14
Åpen mark og fjell	Samlet	13	22	9
	Khat	3	12	9
	Tau	11	20	9

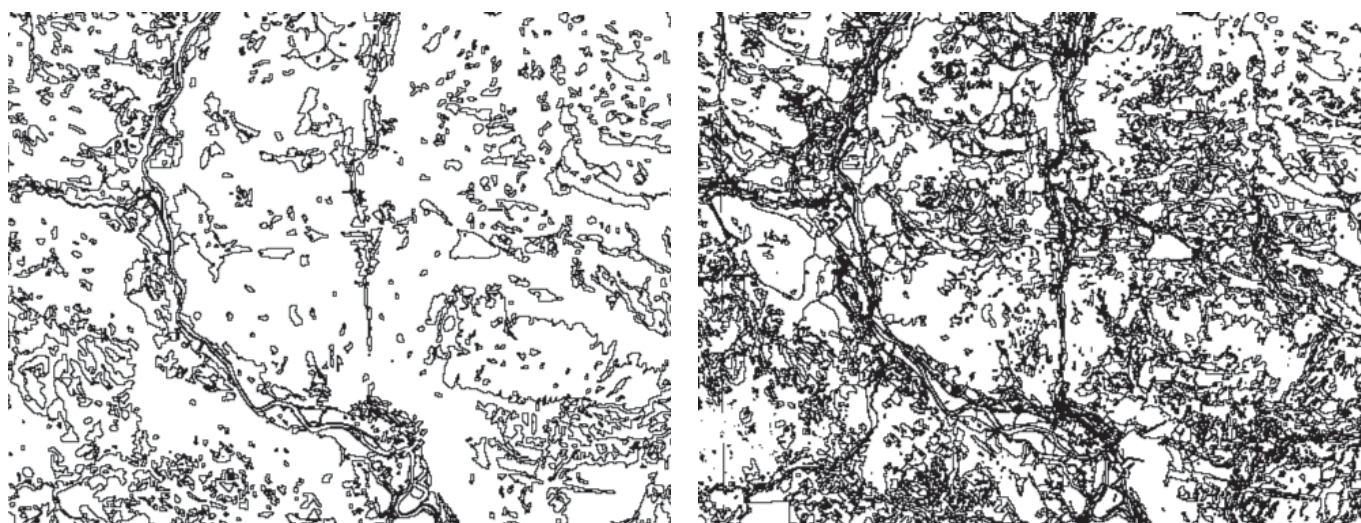
7.4 Skala i referansedata

Når en arbeider med digitale data kan en lett blande data med ulik detaljering og kvalitet. Vi ønsker her å belyse forholdet mellom detaljgrad i et digitalt kart med raster av ulik oppløsning. Målestokk er et naturlig begrep å forholde seg til. Analoge kart er laget for å trykkes i en bestemt målestokk. N50 og DMK er eksempler på digitale kart som har bakgrunn i analoge kart i målestokkene 1: 50 000 (M711) og 1: 5 000 (ØK).

I arbeidet med raster analyser må det velges en oppløsning på rasteret, ofte referert til som pikselstørrelse når en snakker om satellittbilder eller Ground Sampling Distance (GSD) for

flybilder. Pikselstørrelsen på datasettene som er analysert i kartleggingen av Sør-Trøndelag og Nord-Østerdalen er satt til 30 meter da dette er oppløsningen til Landsat 7 ETM+. Hvilken kartmålestokk tilsvarende så dette? Sammenliknes N50 med Landsat viser det at detaljgraden på arealtema er noenlunde lik. DMK blir altfor detaljert i forhold til oppløsningen til Landsat. DMK tilsvarende mer en bildeoppløsning til SPOT5 (10 meter eller bedre).

Om ATIL i DMK blir brukt som datalag i en 30 meters analyse, vil 40 % av pikslene inneholde mer enn én ATIL verdi, i følge tabell 7.5. Uttrykt i prosent av kantpikslar ser vi at DMK i 10 meters pikselstørrelse tilsvarende N50



Figur 7.1. Kantpikslar i N50 og DMK. Utsnitt fra Meldal av N50 markslag (venstre) og DMK ATIL (høyre) der grensene mellom arealtema er gjort om til raster med 30 meter oppløsning.



Tabell 7.5. Et uttrekk av N50 og DMK data ved Meldal i Sør-Trøndelag viser sammenhengen mellom detaljeringsgrad og rasteroppløsning. N50 markslag er et uttrekk av linjer med linjetema 4201 (markslag). I datasettet DMK ATIL (arealtilstand) er nabopolygon med lik ATIL verdi blitt slått sammen (GIS-funksjonen dissolve).

Datasett	Antall polygon	% kantpiksler 30 m	% kant 10 m	% kant 2,40 m
N50 markslag	1512	19.2 %	6.9 %	1.5 %
DMK ATIL	12152	39.8 %	16.4 %	4.5 %

i 30 meter pikselstørrelse (figur 7.1). DMK inneholder åtte ganger flere polygon enn N50 og har om lag fem ganger så stor punkttetthet i linjeelementene.

For rasteranalyser er det vanlig å velge en betraktelig grovere oppløsning enn det som er naturlig ved visuell tolkning. For å identifisere detaljer i flybilder er det ønskelig med 20 - 40 centimeter GSD. En tolker da hovedsakelig ut fra objektenes form og tekstur. Ved automatiske pikselbaserte metoder, slik som i analysene av Sør-Trøndelag og Nord-Østerdalen, benyttes farge til å klassifisere objektene. Det krever at pikselverdien representerer arealtypens gjennomsnitt og dermed er grovere oppløsning (5-30 meter) mer hensiktsmessig.

7.5 Temakart for skogareal

I SatNat prosjektet er det laget fire temakart over skogarealene i Sør-Trøndelag og Nord-Østerdalen. For utvelgelse av skogareal har vi benyttet skogtemaet i N50-kartet. Kart med inndeling i ulike mark- og busksjikt i skogen ble ikke utarbeidet grunnet manglende referansedata. Følgende kart er laget:

- Tretetthet
- Treslag
- Skogtilstand
- Skogtype

Kartene over tilstandsklasser og dominerende tretype er validert mot punktregistreringene i Landsskogtakseringen. For tretetthet hadde vi ikke tilstrekkelig med landskogspunkt til å gjøre en validering. Temakartet skogtype er satt sammen av temalagene tretetthet, tilstandsklasser og tretype.

7.5.1 Tretetthet

Klassifikasjon

Dette kartet er laget med utgangspunkt i verdien for volum hentet fra skogbrukets bestandsdata for barskogarealene. For lauvskogen er flybilder og visuell tolking av satellittbilder benyttet som grunnlag for å etablere treningsdata, da det er meget få registreringer av lauvskog i bestandsdataene. Dette gjelder også lavproduktiv skog.

På grunn av liten dekning med referansedata og stor grad av visuell tolking har vi begrenset dette temalaget til klassene:

- 1 Tett skog, dvs. skog med minst 75 % kronedekke eller trevolum over 25 m³
- 2 Åpen skog, dvs. skog med under 75 % kronedekke eller trevolum under 25 m³

Ungskog med lauvoppslag på god bonitet vil være i klassen tett. Lauvkratt og fjellbjørkeskog regnes derimot som åpen skog i dette kartet.

Kvalitetskontroll

Da kartet i stor grad ble laget ut fra tolkning av kronedekning i fly- og satellittbilder gir det en noen annen inndeling enn Landsskogtakseringen. En feilmatrix ble derfor ikke laget for dette datasettet.

Kartets filnavn er: SN16_Skog_Tretetthet.

7.5.1 Treslag

Klassifikasjon

Kartet er laget med referansedata fra flere kilder. I områdene, hvor vi hadde bestandsdata, er volum fra gran, furu og lauv kodet om til treslag. For også å få treningsdata i områder der vi ikke hadde bestandsdata beregnet vi denne

treslagsfordeling ut fra en multippel regresjon mellom volumprosent hentet fra Landsskogtakseringens dataene og data fra DMK, terrengmodell og satellittbilder (se kapittel 5.1.1.3). Regresjonene gav muligheten til å avlede nye treningsdata fra DMK-kartet (ATIL-informasjonen). I tillegg hentet vi for fjellnær skog inn noe data fra NIJOS sitt vegetasjonskart og noe fra digitalisering på QuickBird opptak fra Tynset i Hedmark.

Vi har klassifisert kartet etter inndelingen:

- 1 Granskog, dvs. skog med minst 50 % volum av gran eller større volum av gran enn lauv eller furu.
- 2 Furuskog, dvs. skog med minst 50 % volum av furu eller større volum av furu enn lauv eller gran.
- 3 Lauvskog, dvs. skog med minst 50 % volum av lauv eller større volum av lauv enn gran eller furu.

Kvalitetskontroll

I en sammenlikning med data fra Landsskogtakseringen fant vi 580 punkt lokalisert slik at de kunne brukes til validering (tabell 7.6).

Korrigerer vi de statistiske verdiene for det geometriske avviket som var mellom satellittdata og landskogspunkt (se Kap. 6.3), har temakartet Treslag en samlet overensstemmelse på 67 % og en Tau verdi på 64 %.

Verdiene for volum i bestandsdataene er basert på trevolum, mens satellitten først og fremst registrerer kronedekning. Dette gir et noe ulikt utgangspunkt for beregning av dominerende treslag.

Filnavn: SN16_Skog_Treslag.

7.5.1 Skogtilstand

Klassifikasjon

Dette kartet har utspring i bestandsdataenes hogstklasseinndeling, men en forenkling har vært nødvendig. I tillegg er klassen impediment blitt lagt til. Impediment er definert ut fra at variabelen ATIL i DMK er lik 28 (grunnlendt mark) som og at NDVI-verdiene er lave. Det er vanskelig å skille naturlig åpen skog fra områder med hogst, spesielt frøtrestilling og ungskog med lite lauvoppslag. Her er det kodet om med hjelp av DMK der dette finnes, åpen

Tabell 7.6. Feilmatrix mellom SatNat temakart Treslag og Landsskogtakseringen.

SN-kart	Landsskogtakseringen				Bruker
	Gran	Furu	Lauv	Totalt	
Gran	115	42	61	218	53 %
Furu	49	129	40	218	59 %
Lauv	29	15	100	144	69 %
Totalt	193	186	201	580	
Prod.	60 %	69 %	50 %	Samlet:	59 %
Khat	39 %		Tau	49 %	

skog (klasse 4) med middels bonitet og bedre er kodet om til ungskog (klasse 2).

Temalaget skogtilstand er klassifisert etter inndelingen i tabell 7.7.

Kvalitetskontroll

Totalt var det 761 punkter fra Landsskogtakseringen som overlappet med temakartet for skogtilstand (tabell 7.8). For hogstflater var det derimot for få (6 stk.) sikre referansedata til å gjennomføre en validering (Congalton & Green 1999). En viktig årsak til dette er at hogstflater gror igjen i løpet av noen år.

Sammenlikningen med data fra Landsskogtakseringen hadde temakartet Skogtilstand en samlet overensstemmelse på 53 % og en Tau verdi på 48 % etter korrigering for romlig avvik mellom satellittdata og referansepunkter. Sammenliknet med resultatet vi fikk ved valideringen av bestandsdata mot data fra Landsskogtakseringen gav temalaget Skogtilstand

Tabell 7.7. Klasser og kriterier for SatNat temakart Skogtilstand

Verdi	Klasse	Bestand HKL	Forklaring
1	Hogstflate	1.1, 1.2	Ferske hogstflater og åpen jord (veg, fylling, ras)
2	Ungskog	2.1, 2.2, 3.1, 3.2	Yngre barskog, men inkluderer det meste av lauvskog. Noe frodig lauv opp mot fjellet.
3	Voksen skog	4.1, 5.1	Voksen, tett barskog med lite lauvinnhold
4	Lavproduktiv skog	4.2, 5.2	Åpen, eldre skog. Oftest på lav bonitet. Skog i nordboreal sone.
5	Impediment		Fjell i dagen (spesielt på kysten), grunnlendt mark og lavhei innen N50-skogmaske

**Tabell 7.8.** Feilmatrise mellom SatNat temalag Skogtilstand og punkter fra Landsskogtakseringen.

<i>Landsskogtakseringen</i>							
SN-kart	Hogstflate	Ungskog	Voks. skog	Lavpr. skog	Impediment	SUM	Bruker
Hogstflate	0	4	8	1	24	37	0 %
Ungskog	3	92	69	22	79	265	35 %
Voksen skog	2	31	100	16	60	209	48 %
Lavproduktiv skog	1	17	49	15	27	109	14 %
Impediment	0	1	8	1	131	141	93 %
SUM	6	145	234	55	321	761	
Produsent	0 %	63 %	43 %	27 %	41 %	Samlet:	44 %
Khat	27 %		Tau	37 %			

lavere overensstemmelse. Dette skyldes delvis økningen i antall klasser (fra to til fem) som vil gi lavere treffprosent. Men det viser også at klassene som ble valgt er noe vanskelige å tolke, selv i flybilder. Hogstklassen er styrt av mange parametere, som volum, tetthet, trehøyde og bonitet.

Det kan også skyldes at det var større tidsforskjeller mellom Landsskogtakseringens feltregistreringer og satellittopptakene enn mellom Landsskogtakseringen og bestandsdataene. Dataene fra Landsskogtakseringen og bestandstakseringene ble feltinnsamlet mellom 1996 og 2000, mens satellittbildene er fra 1999 til

2003. De fleste bildene er imidlertid fra 2002 og 2003.

Filnavn: SN16_Skog_Tilstand

7.5.2 Skogtypekart

De tre nevnte kartene over kan kodes om ved hjelp av vegetasjonssoner og kartdata. Ulike tematisk kart med et uttall klasser kan avledes. Vi har gjort en slik omkoding og laget en inndeling som vist i tabell 7.9.

Filnavn: SN16_Skogtype

Tabell 7.9. Klasseinndeling og kriterier for SatNat temakartet Skogtyper. Kartet er laget ved å kombinere temakartene Tretetthet, Treslag og Vegetasjonssoner.

Verdi	Klasse	Forklaring
1	Hogstflater	Nylig avvirket skog og bar jord/stein
2	Ung barskog	Bardominert ungsog (HKL 2.1 – 3.2)
3	Tett gran	Tett skog og gran
4	Åpen gran	Åpen skog og gran
5	Tett furu	Tett skog og furu
6	Åpen furu	Åpen skog og furu
7	Impediment furu	Grunnlendt mark og særs åpen lavskog mot fjellet.
8	Varmekjær lauv	Tett lauv og lauv som grenser til dyrka mark i boreonemoral og sørboreal sone
9	Tett boreal lauv	Tett lauv i mellomboreal sone
10	Åpen boreal lauv	Kratt, bjørk, lauvoppslag på hogstflater i alle soner
11	Fjellbjørkeskog	Lauv mot fjellet i nordboreal sone. NB! Samme som klasse 10 kun avgrenset av en vegetasjonssone verdi

7.6 Temakart for fjell og åpne områder i lavlandet

7.6.1 Inndeling i fjell og lavland

Fjell og åpne områder i lavlandet er i N50 ikke skilt, men utgjør de arealene som ikke er definert til noe annet. I analysen av satellittdata har vi derfor ikke skilt fjell fra åpne områder i lavlandet.

Dette har vi isteden gjort ved å lage en maske for arealer over og under tregrensa. Kartmasken tregrensa er laget ved å legge et systematisk 1000 meter punktgrid over studieområdet, kategorisere punktene etter om de var over (fjell) eller under (lavland) tregrensa basert på informasjon i N50 og visuell kategorisering på skjerm. Dette punktdatasettet er brukt som trening for en klassifikasjon. Som forklaringsvariabler benyttet vi vegetasjonssonekartet og X- og Y-koordinater. Vi undersøkte også om en variabel som avstand fra kystlinja kunne være bedre egnet enn XY-koordinatene. Det var den ikke.

Som klassifikator undersøkte vi både beslutningstre og logistisk regresjon. De gav omtrent det samme resultatet. Siden en parametrisert modell vil gi et kartprodukt som romlig er noe mer glattet enn et beslutningstre valgte vi å bruke resultatet fra den logistiske regresjonen. Best resultat fikk vi med modellen der klasseinndelingen over eller under tregrensa var en funksjon av vegetasjonssone, X og Y-koordinater og en interaksjon mellom koordinatene. Dette gav en forklaringsgrad på hele 91,1 %.

Filnavn: SN16_Fjellmaske

7.6.2. Klassifikasjon av åpen mark

For behandlingsområde 1 og 3 (se kapittel 6.1.1) er beslutningstreanalysen gjort for de udefinerte arealene i N50-kartet og med bruk av NIJOS vegetasjonskart som kilde for treningsdata. Utvelgelse av treningsdata fulgte tilnærmingen beskrevet i kapittel 6.1.3. For behandlingsområde 2 (Fosen) fantes det ikke vegetasjonskart, der ble isteden DMK (ATIL-informasjonen) brukt som kilde. For behandlingsområde 4 kunne vi benytte noen av treningsdataene fra behandlingsområde 1 og 2. Men for noen klasser var vi nødt til å benytte klassifisert data fra område 2 der dette overlappet med område 4.

På grunn av vanskelighetene med å finne gode treningsdata for hele kartleggingsområdet måtte vi velge en klasseinndeling som både oppnådde akseptable treff i beslutningstreet og som var sammenliknbar for hele kartleggingsområdet (Sør-Trøndelag og Nord-Østerdalen). Vi kodet derfor om NTNU Vitenskapsmuseets punktdata, NIJOS vegetasjonskart og DMK slik tabell 7.10 viser.

Klassen skrinne vegetasjon er en vid klasse som i fjellet kan inneholde både mose, gras og ris (tabell 7.11). Mengden av vegetasjon er liten eller arealtypen har et stort innslag av jord, bart fjell og stein. Frostmark og skrinne snøleie inngår også her. I lavlandet domineres den av arealer med lite vegetasjon bestående av skrinne deknning med mose, gras og ris eller partier med flekkvis frodigere vegetasjon, men med stort innslag av jord, bart fjell og stein.

For klassen lavhei hadde vi ingen referanse-data i vestre delen av fylket (behandlingsområde

Tabell 7.10. Omkodning av referansedata for åpne arealer.

Klasse	Vitenskapsmuseets punktdata	NIJOS vegetasjonskart	Kode i DMK
Skrinn vegetasjon	1A1, 1B1, 4A1	1a	"Fjell i dagen", ATIL 29 med lav NDVI.
Lavhei	1A2, 1B2, 4A2	2c, 2d	Ikke brukt
Lynghei	1A3, 1B3, 4A3	2e, 2f	"Grunnlendt mark", ATIL 28 i et midlere NDVI intervall.
Gras-starr	1A4, 1B4	1b, 2b, 2g	"Myr", ATIL 11 med noe lav NDVI.
Eng-urt	1A4-5, 4A5-6	3a, 3b	"Dyrka mark" og "gjødsla beite", ATIL 22 og 23 med høy NDVI.
Vegetasjonsløst	7A-C	12b, 12c	"Fjell i dagen", ATIL 29 med meget lav NDVI.

**Tabell 7.11.** *Klasser og kriterier for SatNat temakart Åpen mark. Kartet kan inndeles i klassene fjell og lavland ved bruk av kartmasken SN16_Fjell.*

Verdi	Klasse	Forklaring fjell	Forklaring lavland
1	Skrinn vegetasjon	Lite vegetasjon, frostmark og snøleier. TF	Lite og skrin vegetasjon. Lyng, bart fjell, stein og gras. A
2	Lavhei	Lavdominert, kun i øst. TF	Lavdominert, kun i øst.
3	Lynghei	Lyngtyper, dvergbjørk TF	Lyngtyper, noe kratt.
4	Gras-Starr	Fattigere grastyper TF	Fattigere grastyper, myr.
5	Eng-urt	Eng (lågurt og høgstaude) og eng/blåbær bjørkekratt. TF	Beitemark, frodige grastyper. Også frodig lauvkratt.
6	Kystlynghei		Lyng langs kysten i lavlandet. Opp til 2 km fra kystlinja
7	Vegetasjonsløst	Ikke vegetasjon.	
8	Utbygd		Åpen jord, asfalt, veg, kaianlegg. Opp til 100 m fra veg.
9	Strand		Svaberg, strand i lavlandet. Opp til 100 m fra kystlinja.
10	Strandenger		Eng i strandsonen. Opp til 150 m fra kystlinja (N50 kystlinje i 2004).

3 og 4), men trolig er det meget liten utbredelse på denne klassen i vest.

Klassen lynghei inneholder både rishei og røsslyngheier. Kratt, gras og annen frodig vegetasjon er forsøkt fjernet fra denne klassen.

Klassen gras-starr er fattigere grasvegetasjon. Typisk myr vegetasjon og lav beiteverdi. I satellittbildene er den karakteristisk ved at den tørker ut på sensommeren eller tidlig på høsten.

Klassen eng-urt er rik vegetasjon. I denne klassen finner vi også viermark. Fra klassen er strandenger skilt ut ved å definere at dette er enger i lavlandet som ligger nærmere sjøen enn

150 meter. For Strandengene kan skrinne utgaver og små partier (lite areal) ha havnet i klassene "Skrinn vegetasjon" eller "Kystlynghei" på grunn av rasteroppløsningen på 30 meter.

Klassen naturlig vegetasjonsløs mark, dvs. fjell i dagen, stein, ur, grus, sand og bar jord er en samleklasse for areal uten vegetasjon. Fra denne har vi skilt ut en klasse som vi har kalt Anlagt flate. Dette er arealer i nærheten av veg (vbase 2004) som i N50 ikke er definert som utbygd, men som fra satellittbildene er tolket som vegetasjonsløse. Her skulle en tro at mye bart fjell ble feilklassifisert, men naturlig bart fjell på 30 m piksel nivå finner en lite av der veier er

Tabell 7.12. *Feilmatrix som viser klassifiserte piksler for åpen mark sammenliknet med punkter samlet inn av NTNU Vitenskapsmuseet.*

NTNU Vitenskapsmuseets data										
SN-kart	Skrint	Lavhei	Lyng	Gras	Urt	Kyst	Stein/ur	Strand	Totalt	Bruker
Skrint/Frost/snøleie	182	63	13	92	18	115	164	17	664	27 %
Lavhei	24	1634	13	158	2	0	54	0	1885	87 %
Lyng og ris mark	17	141	253	201	192	0	1	0	805	31 %
Gras/starr	49	72	84	360	51	3	4	2	625	58 %
Urt/eng	1	2	183	43	169	15	0	0	413	41 %
Kystlynghei	0	0	18	0	23	364	0	23	428	85 %
Stein/ur	7	1	0	4	1	0	1032	0	1045	99 %
Strandenger	0	0	0	0	0	20	0	20	40	50 %
Totalt	280	1913	564	858	456	517	1255	62	5905	
Prod.	65 %	85 %	45 %	42 %	37 %	70 %	82 %	32 %	Samlet:	68 %
Khat	61 %		Tau	63 %						

lagt. Vi har sammenliknet slike områder visuelt mot flybilder og ser at det stemmer bra at dette er anlagte flater. Videre har vi skilt ut strand og svaberg fra vegetasjonløs mark som egen klasse ved at dette er vegetasjonløs mark under 150 meter fra kystlinja (2004). Disse to klassene har vi ikke hatt data til å validere. De er derfor ikke med i feilmatriksen.

Kvalitetskontroll

For dette temalaget har vi valgt å bruke de punktdata som NTNU Vitenskapsmuseet samlet inn for SatNat-prosjektet som grunnlag for en nøyaktighetsmatrise (tabell 7.12). Resultatet av denne sammenlikningen gav en samlet nøyaktighet på 74 % etter korrigering for romlig avvik mellom satellittdata og referansepunkter. (se Vedlegg, tabell 11.13-19).

Filnavn: SN16_Openmark

7.7 Temakart for myr og jordbruksmark

For myr og jordbruksmark var det så lite referansedata tilgjengelig at vi valgte å ikke benytte datamineringsmetoder, men isteden basere klassifiseringen på alternative metoder. Siden det finnes en hel rekke opptak fra flere årstider over Sør-Trøndelag valgte vi å bruke en metode der vi direkte kunne utnytte de fenologiske forandringer som skjer gjennom vekstsesongen og de endringer som skjer på dyrkede areal ved gjødsling, høsting og kraftig beiting. Metoden går ut på å manuelt bestemme grenseverdier i satellittbildenes filverdier for klassene vi er ute etter. Dermed var det mulig å benytte seg av et stort antall satellittbildeopptak på en effektiv måte.

7.7.1 Myr

For klassen myr har vi tatt utgangspunkt i de areal som er definert som myr i N50 (tabell 7.13). Datasettet er fremskaffet ved å finne terskelverdier i sommeropptak fra Landsat og IRS, basert på visuell tolkning av bilder fra QuickBird satellitten og fra fly. Vegetasjonssoner er benyttet i ulike høydelag. Frodig gras (eksempelvis myr omgjort til gjødsla beite) kan ha havnet i tresatt myr da lauvkratt og gras har svært like spektrale egen-

Tabell 7.13. Klasseinndeling og kriterier for SatNat temakart Myr.

Verdi	Klasse	Forklaring
1	Tresatt myr	Mer enn om lag 5 trær pr. dekar.
2	Blautmyr	Vassdekt eller delvis vassdekt overflate
3	Åpen myr	Gras, lyng, lav eller åpen jord. Mindre enn 5 trær pr dekar.

skaper. I områder der det er smale korridorer av åpne myrer (under 50 meter bredde) kan disse blitt tolket til tresatt myr grunnet pikseloppløsningen på 30 meter (store deler av pikslene har innslag av trær). I forhold til kartlagt areal av myr med skog i DMK inkluderer klassen tresatt myr betraktelig større arealer i SatNat-kartleggingen. Utstrekning av blautmyrer vil dessuten være noe avhengig av vannivået i myrene da opptakene ble foretatt (21. juli – 23. august).

Temalaget er ikke validert siden det ikke fantes noen kilder som kunne nyttes til referanse.

Filnavn: SN16_Myr

7.7.2 Jordbruksmark

Med jordbruksmark menes areal i N50 som er definert som dyrka mark eller åpen mark i lavlandet (beite). Det er forsøkt å dele jordbruksmarka inn i full- og overflatedyrka, gjødsla beite og beite (tabell 7.14). Disse klassene er antatt å kunne skilles fra hverandre ved at full- og overflatedyrka jord vil framtre som vegetasjonløs minst en gang i løpet av sesongen, gjødsla beite vil framtre som meget frodig og anna beite og eng vil ikke oppføre seg slik. Framgangsmåten har vært å benytte NDVI fra flere år og årstider. Lav NDVI tyder på areal som blir slått eller pløyd, men det kan også skyldes meget hardt husdyrbeite. Høy NDVI tyder på

Tabell 7.14. Klasseinndeling og kriterier for SatNat temakart Jordbruksmark

Verdi	Klasse	Forklaring
1	Full- og overflatedyrka	Åker- og engmark som er vegetasjonløs minst en gang i løpet av året. Korn, grønnsaker, eng som pløyes eller slåes med mer.
2	Gjødsla beite eng	Eng som ikke pløyes/slåes og har frodighet som tyder på gjødsling
3	Beite eng	Eng som ikke pløyes/slåes og har frodighet som ikke tyder på gjødsling

**Tabell 7.15.** Feilmatrise mellom SatNat temalag Jordbruksmark og 3Q kartdata.

	3Q			
SN-kart	Fulldyrka	Beite	Totalt	Bruker
Fulldyrka	3950	1034	4984	79 %
Beite	42	361	403	90 %
Totalt	3992	1395	5387	
Produsent	99 %	26 %	Samlet:	80 %
Khat	37 %		Tau	37 %

gjødslende areal.

Arealbruk kan endre seg fra år til år og kartet over jordbruksmark er basert på elleve satellittbildeopptak i perioden 1999 - 2003. Hvert delområde med opptak fra tre årstider har opptak fra opptil tre ulike år. Der det er overlapp mellom delområdene akkumuleres informasjonen. I de sentrale jordbruksområdene i fylket (Melhus, Trondheim, Klæbu, Skaun med flere) er det akkumulert sju opptak. Om noe fulldyrka mark ligger brakk ett år og blir først klassifisert som beite, kan den bli klassifisert som fulldyrka et annet år og får endelig status fulldyrka.

En del jordbruksareal er blitt bygd ut til veg og industri/boligformål. Slikt areal vil havne i klassen fulldyrka. Dette arealet kan identifiseres med en analyse for identifisering av grå areal som gjort av Vikhamar & Kastdalen (2005). Boliger med hage, gardstun og gardsveger havnet i klassen beite. Disse kan kodes om ved bruk av andre kartdata.

Små jordbruksparseller (smalere enn 60 meter) som er fulldyrka vil lett bli klassifisert som gjødsla beite om det er lauvkratt rundt. Det er spesielt jordbruksarealene på Frøya og Hitra som påvirkes av dette.

Det er ikke benyttet satellittbildeopptak fra juni i analysen (skyldes normalt mye skyer i denne perioden). Om en eng blir slått tidlig på sommeren og senere får vokse opp vil ikke dette bli fanget opp i kartet.

I en vurdering av kartet over jordbruksmark har vi valgt å sammenlikne det med kartdata fra 3Q-programmet (tabell 7.15). Det gav 79 % samsvar mellom kornareal i 3Q og klassen full- og overflatedyrka i SatNat kartet. Det er 90 % samsvar mellom beite i 3Q og gjødsla/ikke gjødsla i SatNat. Når det gjelder DMK er det slik at 96 % av jordbruksarealet i N50 er betegnet som full- og overflatedyrka i DMK. Beite i DMK er stort sett ikke kartlagt som jordbruk i N50, og klassene kan

heller ikke sammenliknes.

Her ser man at Khat og Tau slår ut med lave verdier grunnet skjev fordeling i referansedata (se kapittel 4.3). Samlet nøyaktighet er bra på hele 80 %, selv om det er lite samsvar mellom beitearealene i de to kartene.

Filnavn: SN16_Jordbruk

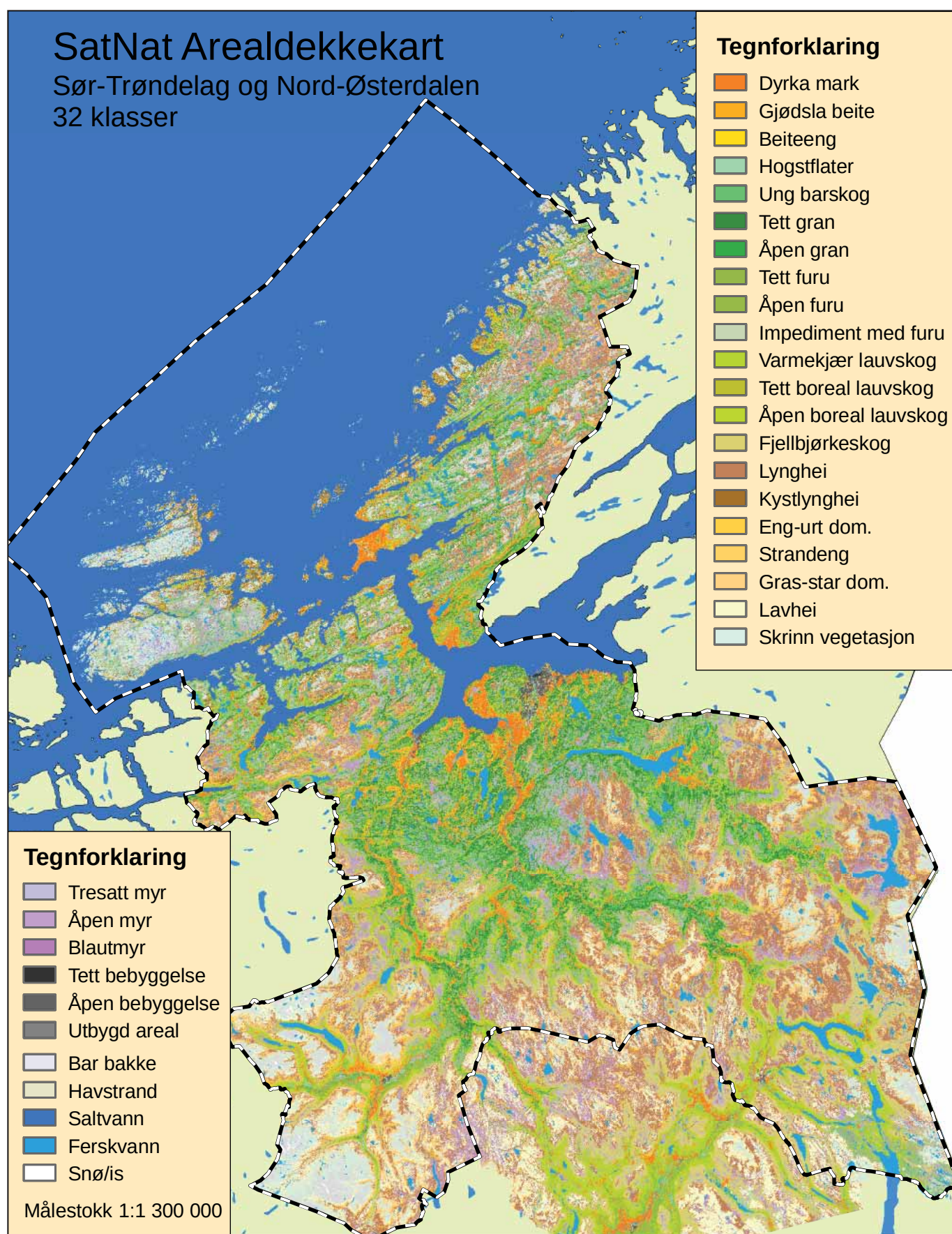
7.8 Arealdekkekart Sør-Trøndelag og Nord-Østerdalen

Vi har laget et heldekkende kart for Sør-Trøndelag og Nord-Østerdalen med 32 temaklasser ved å sette sammen temakartene skogtyper, åpen mark, myr og dyrka mark og trekke inn arealene for vann og utbygde områder fra N50 (figur 7.2). De ulike temakartene som arealdekkekartet for Sør-Trøndelag og Nord-Østerdalen er avledet fra er omkodet slik som spesifisert i tabell 7.16. Det fargevalg vi har benyttet er beskrevet i tabellens RGB-kode.

Vi har ikke bygget på noen eksisterende klassifikasjon siden klassene vi kunne inndeke kartet etter ble til en stor grad fastsatt gjennom de data vi hadde tilgjengelig som treningsdata. Bortsett fra at vi ikke har skilt ut fjell som egen hovedklasse har vi fulgt inndelingen i SatNat-klassifikasjonen til Hjeltnes og Øien (2003). Fjellmasken vi laget kan benyttes til å skille ut fjellvegetasjonen hvis ønskelig. Klasseinndelingen ligger nært opp til NaSAK (Gjertsen m. fl. 2003) og dermed det internasjonale systemet Land Cover Classification System (LCCS; Gregorio 2005). I tabell 7.16 er beskrevet hvilke NaSAK-koder som beskriver de ulike klassene.

Klassifikasjonen fanger også opp 10 av de 12 klassene som det europeiske klassifikasjonssystem CORINE Land Cover (CLC) er inndelt i på nivå 3 under gruppen Skog og halvnaturlige marker. Dette er den gruppen i CLC hvor det er lite informasjon å hente fra andre kartdata. De manglende klasser er Brent areal og Blandingsskog. Blandingsskog kan avledes fra dataene ut fra treslagsfordelingen, selv om vi valgte å inndeke skogen etter dominans av gran, furu eller lauv.

Flere av klassen går lengre i sin inndeling enn CLC og tilsvarer mer nivå 4 i svenske marktaledata (SMD), en klassifikasjon som bygger på CLC, men med flere tematiske nivå enn CLC. For de øvrige hovedgrupper i CLC vil mye av



Figur 7.2. SatNat arealdekkkart for Sør-Trøndelag og Nord-Østerdalen.



informasjonen til nivå 3 finnes i N50 og i DMK. N50-kartserien har ytterligere 6 kategorier i hovedklassen Anlagte flater og skiller ferskvann i stillestående og rennende. Vbase kan også brukes for å identifisere vegareal dersom det blir gjort en differensiert buffering etter vegstandard. Den

hovedklasse det mangler informasjon om til å etablere et fullt CLC nivå 3 er Våtmarker. I tillegg har vi ingen kartdata som skiller ut Estuarier og Kystlaguner under hovedklassen Vannområder. Det må en egen kartlegging og/eller GIS-analyser til for å skille ut disse arealene.

Filnavn: SN16_Arealdekke

Tabell 7.16. Klasseinndeling og kodevalg for SatNat arealdekkekart Sør-Trøndelag og Nord-Østerdalen.

Klassifiseringsnivå				Kartets	Fargeverdier			Andre system	
Nivå1	Nivå2	Nivå3	Nivå4	tallkode	Rød	Grønn	Blå	CORINE	NaSAK
Jordbruksmark	Dyrka mark			11	255	130	30	2.1.1	V11-A1B3E1
	Beite	Gjødsla beite		12	255	175	30	2.3.1	V11-A1B3E2
		Beiteeng		13	255	220	30	2.3.1	V11-A1B3E3
Skogsmark	Hogstflate/ kratt/ungskog	Hogstflater		21	140	255	190	3.2.4.2	V12-A1C1
		Ung skog		22	90	210	115	3.2.4.3	V12-A1C3
	Eldre barskog	Grandominert skog	Tett gran	23	55	140	0	3.1.2.1	V12-A1B1C2-3D1
			Åpen gran	24	50	170	0	3.1.2.1	V12-A1B2B3C2-3D1
		Furudominert skog	Tett furu	25	150	185	75	3.1.2.1	V12-A1B1C2-3D1
			Åpen furu	26	150	195	115	3.1.2.1	V12-A1B2-3C3D1
			Impediment med furu	27	200	215	180	3.1.2.3	V12-A1B3-4C3-4D1
	Eldre lauvskog	Varmekjær lauvskog Boreal lauvskog		28	185	220	0	3.1.1	V12-A1B1C3-4D2
			Tett boreal lauvskog	29	190	190	0	3.1.1	V12-A1B1C3-4D1-2
			Åpen boreal lauvskog	30	190	220	0	3.1.1	V12-A1B2C2-4D1-2
			Fjellbjørkeskog	31	220	210	115	3.1.1	V12-A1B2C4D2
Åpen mark	Lyngdominert	Lynghei	Ikke-kystlynghei	41	195	130	90	3.2.2	V12-A2B1-2C9
			Kystlynghei	42	165	110	0	3.2.2	V12-A2B1-2C9
	Grasdominert	Eng-urt dominert	Eng-urt dominert	43	255	210	70	3.2.1.2	V12-A3B1C8C13
			Strandenger	44	255	210	100	3.2.1.1	V12-A3B1C13
		Gras-star		45	255	210	130	3.2.1.1	V12-A3B1C13
	Lavdominert	Lavhei		46	250	250	200	3.2.2	V12-A4B1C14
	Skrinn vegetasjon			47	215	240	230	3.3.3	V12-A2-4B4C11H2
Myr	Tresatt myr			51	195	190	220	3.1.3.2	V22-A1B2-4D3E3
	Åpen myr			52	195	160	215	4.1.2.2	V22-A3B1C13-14E3
	Blautmyr			53	195	125	215	4.1.2.1	V22-A4B3-4E1
Anlagte flater	Bebyggelse	Tett bebyggelse		61	50	50	50	1.1.1	B11-A1B2E1G1
		Åpen bebyggelse		62	100	100	100	1.1.2	B11-A1B2E1G2-3
	Annen anlagt flate			63	130	130	130	1.2.2	B11-A2D1
Naturlig vegetasjonsløst	Bart fjell, stein, grus, sand			71	230	230	240	3.3.2	B12-A1C1-6
	Havstrand			72	230	230	200	3.3.1	B12-A1C5
Vann, is og snø	Saltvann			81	40	120	220	5.2.3	B22-A1B1E1
	Ferskvann			82	40	160	220	5.1.2	B22-A1B1E1-2
	Is/snø			83	255	255	255	3.3.5	B22-A2-3B3

8 Diskusjon

8.1 Valg av klasser

I dette prosjektet har vi fokusert på hvilke muligheter som ligger i å benytte eksisterende kartdata til en styrt klassifikasjon av satellittdata (Landsat). En stor del av arbeidet har vært undersøkelser av hvordan parametrene i kartdatabasene kan benyttes til å lage kart med en legende (tegnforklaring) som er kompatibel med etablerte norske vegetasjonssinndelinger og med internasjonale klassifikasjonssystem som CORINE Land Cover nivå 3, og som i tillegg inneholder relevante parametere for modellering av potensielle vilthabitat.

For at en arealklasse skal være mulig å kartlegge med satellittdata må visse betingelser være til stede. En hovedbetingelse er at det er spektrale forskjeller mellom klasser, enten i ett satellittbilde eller ved å bruke bilder fra ulike årstider. Med manuelle kartleggingsmetoder er det mulig å inndelegge i tema (klasser) basert på plantesosiologiske kriterier, som er basert på at visse karakterarter er observert i felt. Når automatiske metoder brukes på bildedata vil dette ikke være mulig i samme grad. Da vil en faktor som dominerende plantetype (trær/ikke-trær, lauv/bar, gras, lyng, skrinne vegetasjon, etc.) sammen med parametere som tetthet/frodighet, fuktighet og graden av stein/sand/grus være egnede basiskriterier for en klasseinndeling.

Et eksempel på en tilnærming mer basert på parametere er arbeidet med Ny Norsk Naturtypeinndeling (Økland 2005). Dette vil i langt større grad enn tidligere system bygge på gradienter og dermed passe bedre for kartlegging med optiske sensorer enn dagens plantesosiologiske system. Ved endringer forårsaket av for eksempel gjengroing og beiteslitasje vil slike parametrene endre seg og dermed bli mulig å fange opp ved maskinelle metoder.

Ved styrt klassifikasjon setter dataene klassifikatoren trenes med rammen for hvilke klasser et kart kan deles inn i. Inndelingen av klasser i arealdekket for Sør-Trøndelag og Nord-Østerdalen er derfor til stor grad styrt av informasjonen i treningsdataene.

8.2 Klassifikasjonsmetoden

Et viktig mål for denne satellittbaserte arealdekkklassifikasjonen var å utvikle metoder som var objektive, konsistente og effektivt reproduserbare i videre kartlegging av arealdekke. Ut fra erfaringer i nasjonale kartlegginger i Sverige, Finland og USA valgte vi å benytte en styrt klassifikasjon. En styrt klassifikasjon vil dessuten bli et supplement til tidligere utprøvinger av satellittbasert klassifikasjon basert på visuell tolking eller klustering (Arnoldussen m. fl. 1999).

Vi har i dette prosjektet ikke gjort sammenlikninger med andre metoder, men kun sett på hvordan datamineringssteknikken beslutningstre (Tan m. fl. 2006) kunne anvendes med utgangspunkt i bruken av eksisterende kartdata som referansedata. For sammenlikning mellom metoder viser vi til Pal & Mather (2003). Valget av metode ble i stor grad gjort ut fra erfaringer med at en styrt klassifikasjon med beslutningstre eller regresjonstre gir bedre resultat enn tradisjonelle parametriske klassifikasjonsmetoder på multispektrale satellittdata (Hansen m. fl. 1996, Homer m. fl. 2004), og at datamineringssteknikk er spesielt egnet til å utnytte informasjon fra eksisterende kartdata.

Fordelene med en klassifikasjon basert på datamining er at:

- store datamengder kan analyseres raskt og automatiseres
- det gir tolkbare klassifikasjonsregler
- metoden er ikke-parameterisk og derfor uavhengig av fordelingen på klassesignaturen
- datakilder med ulike format kan kombineres i analysen, f.eks. kategoriske og kontinuerlige (nominelle) data
- det kan lages kartprodukt klassifisert etter tematiske grupper eller uttrykt som kontinuerlige variabler (sub-piksel analyse).

Videre kan datamineringssteknikkene gi et konfidenskart, dvs. et kart som for hver piksel viser hvor godt metoden har klassifisert en piksel.



sel i forhold til de benyttede treningsdata, samt et nodekart. Nodekartet viser hvilke beslutningsregler som er benyttet for klassifisering av hver enkelt piksel.

Ulempen med metoden er at den krever et stort utvalg av godt balanserte treningsdata (høy kvalitet og representativ dekning). For å få det ønskede resultatet over hele kartleggingsområdet må dermed treningsdataene være godt fordelt både romlig og innenfor de tematiske klassene. Videre stiller metoden krav til at treningsdataene entydig beskriver de klasser en ønsker å få ut. Men godt balanserte data er også særs viktig i for eksempel kNN-metoden (Gjertsen & Eriksen 2004), som er benyttet som grunnlag for klassifisering av skogarealer i både den svenske og finske CORINE Land Cover.

Med egne treningsdata er beslutningstremetoden meget effektiv. Tildelingen av klasser til nodene skjer under klassifikasjonen. Nodene kan kontrolleres og valideringer gjøres på flere nivå under analysen. Det viste seg mulig å gjennomføre en klassifikasjon med et komplekst klassehierarki i løpet av noen få timer. Beslutningstreet er enkelt å tolke og feilsøking er mulig. Ved å undersøke de klassifiserte nodene visuelt og sammenlikne med treet er det mulig å finne ut for eksempel hvorfor to klasser ble blandet sammen. Virker nodene tilfeldig blandet tyder det på flertydigheter i referansedataene. En annen enkel kontroll er å se hvorvidt parameteren treet setter øverst (viktigste skillekriterier) virker fornuftig.

8.3 Satellittdata

I store deler av Norge er skyer et ekstra problem for bruk av optiske bildedata, og Sør-Trøndelag har vist seg å være et av områdene i landet med mest skyer. Landsat-satellitene var dermed den eneste relevante kilde til satellittdata for denne regionen, spesielt med tanke på å kunne gjøre multi-temporale analyser.

Satellitopptakene måtte innhentes fra en fem års periode for å få heldekning. Siden opptakene fordelte seg på flere år og i tillegg tatt fra ulike tidspunkt i hver årstid vurderte vi det som lite aktuelt å normalisere dem slik det er beskrevet i NLCD2001 (Homer m. fl. 2005), for på den måten kunne sette sammen ett vår-,

ett sommer- og ett høstbilde i stedet for de fem delområdene vi analyserte. For andre deler av landet er dette et noe mindre problem grunnet mindre problemer med skyer.

Vi erfarte at opptak fra siste halvdel av mai sammen med sommeropptak gir gode resultat i en analyse med tanke på skog. Opptak fra høsten har mindre verdi og kan eventuelt sløyfes. For fjellet gav vårbildene lite informasjon da det var mye snødekt areal i høyden.

8.4 Treningsdata

8.4.1 Treningsdata og beslutningstreanalyser

Å bruke eksisterende data, som i dette prosjektet, gir i utgangspunktet en kostnadsfordel og en tidsbesparelse. For eksempel er det med hell benyttet punktbasert data tilsvarende Landskogtakseringen som viktige treningsdata både i produksjon av SMD og NLCD 2001. En avgjørende faktor for å kunne bruke slike data er hvor mange treningsflater det vil bli innen hvert klassifikasjonsområde sett i forhold til den spektrale variasjonen i satellittbildene. Topografien i kartleggingsområdet vil spille en avgjørende rolle for dette, i det den både øker variasjonen i bildedataene og reduserer tilgangen på data fra flere aktuelle referansedatasett.

I litteraturen fant vi lite informasjon om hvordan antallet treningsdata påvirker kartleggingen med Landsatdata når det benyttes beslutningstremetoder. Pal & Mather (2003) gjorde en undersøkelse hvor de suksessivt økte antallet treningsdata og fant at nøyaktigheten i klassifikasjonen økte lineært med antallet treningsdata inntil ca 300 datapunkt per klasse. Treningsdata må gjenspeile variasjonen i hver klasse (Lillesand & Kiefer 2000), med tanke på ulike vekststadier, geografisk beliggenhet, variasjon i ulike høydelag og belysningsmessig påvirkning (sol og skyggeside). Vi må dermed anta at det optimale antallet treningsdata for en arealklasse vil variere ut fra klassens spektrale separabilitet i forhold til nærliggende klasser, terrengmessige faktorer og hvor sjeldent klassen opptrer i landskapet så vel som kvaliteten på de data som brukes for treningen.

I den svenske klassifiseringen for CORINE land Cover ble det for skogarealene benyttet en styrt klassifikasjon med treningsdata tilsvarende Landsskogtakseringens punktdata. For hver Landsat-scene gav det mellom 1000 og 4000 punkt til trening av klassifikatoren. På Østlandsområdet har NIJOS brukt Landsskogtakseringens punkter som trening for klassifikasjon av Landsatdata. Her er det store skogområder slik at innen en Landsat-scene var det tilgjengelig nesten 2300 flater fordelt på punkt fra flere fylker (1000- 1500 punkt etableres per fylke, se kapittel 5.1.1.2; Gjertsen & Eriksen 2004). I Sør-Trøndelag blir det færre landsskogsflater tilgjengelig for hver satellittscene da skogen er mer spredt fordelt mellom fjell og åpne områder, samtidig som vi har mer skyer i hver scene. Dermed gir ikke Landsskogtakseringen tilstrekkelig med informasjon til å fange opp variasjonen i vegetasjonen.

De ikke-parametriske metodene som beslutningstre og kNN er følsomme for store relative endringer i antall treningspikslar (kalibreringspunkter) mellom de ulike klassene (Gjertsen & Eriksen 2004). Vi valgte derfor å jevne ut antallet treningsdata slik at en arealmessig stor klasse ikke fikk mer enn fire ganger antallet treningsdata til klassen med færrest observasjoner. Hvis ikke dette ble gjort overrepresenterte beslutningstreet klassene med flest treningsdata. En alternativ strategi hvis sikre treningsdata er tilgjengelig er å duplisere observasjoner i klasser med lite data. Om treningsdata mangler for noen delområder kan data også fremskaffes fra klassifiserte produkter i tilgrensende delområder.

Vi erfarte for skogkartene at det var nødvendig å etablere egne treningsdatasett for lavskogen i øst. Siden vi manglet bestandstakst for de østlige områdene måtte dette digitaliseres. Andre mindre opplagte økologiske soneinndelinger ble ikke utprøvd, selv om det er en åpenbar forskjell mellom de kystnære områdene og innlandet. Vegetasjonssonekartene fungerte godt for å inndele i høydelag, men for Sør-Trøndelag gav vegetasjonsseksjonene en vel sterk påvirkning. Vegetasjonssoner og -seksjoner kan benyttes til å avgrense klassifikasjonsområder, men andre faktorer spiller også inn. Vi tror derfor at en slik inndeling må avgrenses manuelt.

8.4.2 Egnethetsvurdering av eksisterende kartdata

Vår erfaring er at Landsskogtakseringens punktdata, DMK, 3Q og skogbrukets bestandsdata bare delvis er egnede datasett for klassifikasjon av satellittdata. Ingen av datasettene kan benyttes direkte uten omkodning og siling av data. Klassene vi får ut er begrenset til treslag, bonitet og hogstklasser i produktiv skog og arealbruk på jordbruksmark. Parametere for tetthet og alder i bestandsdataene, som kunne gitt oss kontinuerlige verdier, hadde ikke tilstrekkelig romlig oppløsning som treningsdata på pikselnivå. I den lavproduktive delen av skogen var det ikke egnede referansedata. Landsskogtakseringen hadde for få punkter til å kunne benyttes til trening av en skogklassifikasjon, men viste seg godt egnet for validering.

Selv om vi hadde mye data for skog i Sør-Trøndelag var det ikke mulig å kartlegge marksjikt i skogen. De få registreringer som fantes var enten lite tilgjengelige (for eksempel skannede vegetasjonskart), informasjon om marksjikt mangler (bestandsdata har tidvis slike registreringer) eller er rett og slett ikke samlet inn på en måte som gjør dem sammenliknbare med andre kilder (ulike klassifikasjoner er benyttet). Ønskes informasjon om myrtyper, fjellvegetasjon og marksjikt i skogen er dette ikke å finne i eksisterende data med tilstrekkelig kvalitet. Noen egenskaper knyttet til åpen mark og myr i lavlandet kan hentes fra DMK.

Vi oppnådde gode nøyaktighetstall for noen areal typer. Blant annet for klassene lavhei, kystlynghei og stein/ur i temalag for fjell og åpen mark. Her ble treningsdata samlet inn manuelt kun med støtte i referansedata. Vegetasjonskartene fra NIJOS påførte så mye arbeid i forkant av klassifikasjonen at de ikke kan regnes til særlig nytte.

Med tanke på videre bruk av eksisterende kartdata som referansedata i satellittbasert arealdekkkartlegging har vi gjort en vurdering av de enkelte datakilder (tabell 8.1). Vi har gjort vurderinger knytt til areal typene skog, fjell, åpen mark og jordbruksmark for fem egenskaper. I kolonnen "mengde" har vi vurdert hvor mange observasjoner et datasett gir. Et punktdatasett gir maksimalt en observasjon per punkt og dermed liten mengde (karakter



Tabell 8.1 Rangering av referansedataenes egnethet til trening av beslutningstre for produksjon av kart i målestokk mellom 1: 20 000 og 1: 50 000. Type S = skog, Å = åpen mark, J = jordbruksmark og F = fjell. Karakter 1 = lite egnet/tilfører mye støy, 2 = noe egnet, 3 = godt egnet/bra overensstemmelse med fjernmålte bilder. Tabellen inkluderer også erfaringer fra Østfold (Vikhamar m.fl. 2004). Sum karakter er minimums karakter, om datasettet har et svakt punkt gir det samlet lav karakter.

Kilde	Type	Mengde	Skala	Temporal	Legende	Fordeling	Sum
DMK	S, Å	3	3	2	2	3	2
3Q	D	3	3	3	2	3	2
Bestandskart	S	3	3	2	3	2	2
Landskogtakseringen	S	2	3	3	3	3	2
NIJOS vegetasjonskart ¹	S, F	3	1	3	2	2	1
NTNU punktdata ²	F, Å	1	2	3	3	2	1

¹ NIJOS vegetasjonskart gjelder data fra de vegetasjonskart som produseres i målestokk 1: 20 000 - 50 000 beskrevet av Rekdal & Larsson (2005).

² Gjelder data levert av NTNU Vitenskapsmuseet etter metodikk beskrevet i Hjeltne og Øien 2003.

1), mens et flatedekkende kart gir et stort antall observasjoner (karakter 3).

Med skala mener vi hvordan målestokken i referansedataene passer med pikseloppløsningen i satellittdataene. Her har vi gjort en vurdering ut fra bruk av satellittdata med pikselstørrelse omtrent som Landsat. Høy grad av romlig generalisering eller for stor detaljering er begge negative faktorer. Et eksempel på sterk generalisering er kartdata som N50 markslag. All skog, uavhengig av tetthet og åpne partier, har samme tematiske verdi. En punktobservasjon av en karplante i en grasmyr er et eksempel på for stor detaljering. Sterkt romlig generaliserte og meget detaljerte kartdata har vi gitt verdi 1. En polygon som dekker 20 dekar skog med homogen utforming er bra egnet og får verdi 3. Et punkt som har fått en egenskap i forhold til arealet rundt (eksempelvis 0,9 dekar for Landsat skala) er også godt egnet.

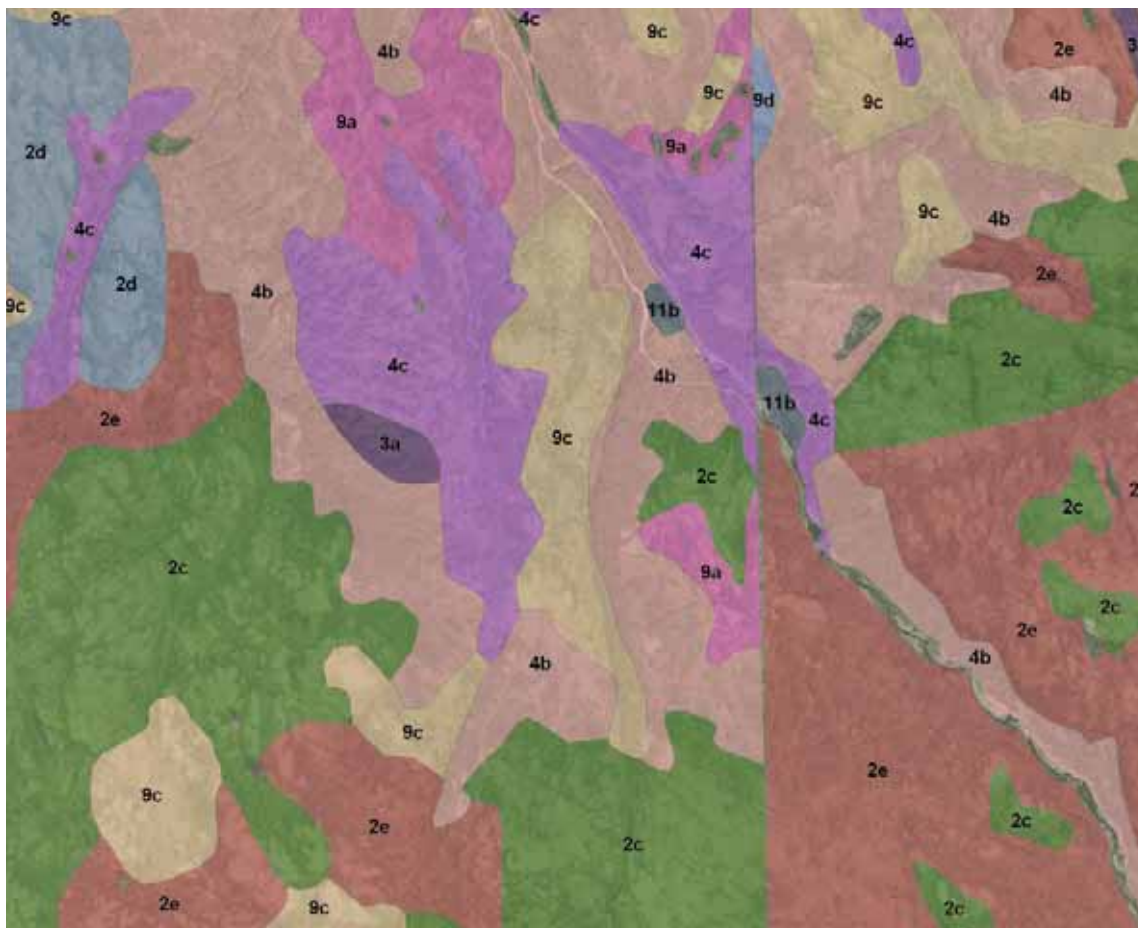
Den temporale oppdateringen av kartet har betydning for dets egnethet som referansedata. Vi har betegnet en oppdatering på 5 år eller oftere som meget bra (karakter 3) og data med oppdateringsintervall på 15–20 år som lite egnet. Dette er imidlertid avhengig av hvor raskt det skjer endringer i landskapstypen. Bestandstakster som er fem år gamle vil være sterkt endret, mens et kart over fjellvegetasjon endrer seg nesten ikke på fem år.

Videre har kartdataenes legende mye å si for hvor godt de er egnet til avgrensninger

i billedata. Et vegetasjonskart der indikatorarter bestemmer klassene har vi gitt karakter 1, mens et kart med tretype og kronedekning målt i flybilder får karakter 3 for legende. God romlig representativitet er viktig for å få med variasjonen innen et kartleggingsområde. Vi har derfor vurdert dataenes fordeling med tanke på kartlegging av arealer på størrelse med et fylke eller større. Om skog i alle kommuner i et fylke er kartlagt har vi gitt datakilden karakter 3 for fordeling.

Bruken av 3Q er omtalt i Vikhamar m.fl. 2004. Med treningsdata fra dette datasettet ble det laget et kart med fire klasser. Tre av klassene viste imidlertid dårlig samsvar med referansedata. Østfold har en arealmessig god dekning av 3Q flater, mens i Sør-Trøndelag er det heller få grunnet lite jordbruksareal. 3Q er kartlagt med flybilder ved hjelp av feltobservasjoner. Dette burde gjøre 3Q egnet til referansedata. Problemet med datasettet er at det ikke gir treningsdata til så mange av klassene en vil være ute etter. De kartlagte klassene i 3Q er også i stor grad avhengig av visuell gjenkjenning, slik som hagebruk. For norsk småskala landbruk er dessuten Landsat-data med 30 meters oppløsning vel grovt. Pløyd jord, bygninger og veier overstråler vegetasjonen. Åker har stor utstrekning mens de andre klassene utgjør små areal.

DMK med informasjonen i variablene ATIL og ASKOG kan imidlertid brukes som



Figur 8.1. Ortofoto med vegetasjonskart lagt oppå. Kartbladkanten, hvor to kartlegginger møtes, er godt synlig. Avgrensningene viser at vegetasjonskartlegging i denne målestokk gir stor grad av romlig generalisering og klasser som er vanskelige å bestemme entydig for kartleggeren.

datalag i analysene om de er heldekkende (ikke tilfellet når arbeidet med Sør-Trøndelag pågikk). Informasjonen i ATIL beskriver godt klasser som grunnlendt mark og fjell i dagen, men beskriver i liten grad tilstand i skogen. ASKOG gir god bonitetsinformasjon og kan nyttes til forbedring av et klassifisert produkt. DMK har mindre verdi brukt direkte som treningsdatasett. I kombinasjon med satellittdata og andre kartprodukt kan DMK derimot benyttes til å avlede nye treningsdata, for eksempel gjennom regresjoner slik som gjort i dette prosjektet.

Bestandskart får ikke full uttelling som referansedata grunnet hogsten som gjerne skjer etter en kartlegging. Dekningen vil ofte være mangelfull med tanke på fylkesvise kartlegginger. Bestandstakst utføres på produktiv skog, og dermed kartlegges normalt ikke fjellbjørkeskog og glissen furuskog. Taksten er imidlertid en

meget sikker datakilde om den er oppdatert og tilgjengelig. En ulempe med slike data er at dataene ikke kan skaffes fra en kilde, men ved henvendelser til hvert enkelt takstselskap. Selskapene benytter også ulike programvare, som fører til en del merarbeid med konvertering av data.

Landsskogtakseringens punkter (0,25 dekar) dekker et mindre areal enn en Landsat piksel (0,9 dekar), men passer bra for SPOT 5 (0,1 dekar). Problemet er at det blir få observasjoner med unntak av større skogområder på Østlandet. Kvaliteten på dem og plasseringen i et rutenett gjør imidlertid at Landsskogtaksering er et ypperlig datasett til validering.

Vegetasjonskartene som NIJOS utarbeider for beitekartlegging i fjell og fjellnære områder er lite egnet til referansedata på grunn av sin generalisering og legende som ikke tar hensyn til mengde og tetthet. Vegetasjonstypene



i kartet har også ofte ganske tilfeldige verdier i områder med blandet vegetasjon, noe som lett sees i kanter mellom kartbladene (figur 8.1). Blandede klasser lar seg vanskelig bruke som treningsdata (Congalton 1999; se punkt 5, kapittel 8.4 avsnitt 2).

Punktene som ble samlet inn av NTNU Vitenskapsmuseet for prosjektet er tatt med i tabellen selv om de ikke er basert på noen løpende kartlegging. Disse punktdataene har mange egenskaper som gjør dem egnet som treningsdata, men de er for få. Samtidig er de ikke innhentet tilfeldig eller etter et systematisk mønster. Dette gjør dem også bare delvis egnet til validering. Dersom slike punktdata hadde vært hentet ut fra høyoppløselige bilder tatt i et systematisk rutenett, slik som det er gjort i SatNat-prosjektet på Hardangervidda, ville de vært godt egnet.

8.5 Tematisk nøyaktighet

En vanlig metode å vurdere nøyaktigheten av tematiske kart er å sammenlikne dem med andre data som skal være en riktig klassifisering av hva som er "virkeligheten". Denne virkeligheten er i mange tilfeller nærmest umulig å fastsette (Foody 2002), spesielt i heterogene landskap hvor vi i stor grad forsøker å kategorisere kontinuerlige overganger.

Likevel, informasjon om nøyaktighet er viktig i bruken av kartprodukter. For å tolke nøyaktighetsberegninger er det nyttig å kjenne faktorene som påvirker resultatet. Congalton og Green (1993) har systematisert faktorene og satt opp følgende liste: (1) endringer i tiden mellom etablering av referansedata og satellitt-dataopptaket, (2) lokaliseringsfeil i referansedataene, (3) feil i innskrivingen av referansedata, (4) feil klasse tolket i referansedataene, (5) inkonsistent tilordning av klasse på grunn av heterogeniteten rundt referansedatapunktet, (6) avvik i geometrisk registrering mellom kart og referansedata, (7) feil i kartets inntegninger og (8) feil i kartets klassifisering. Av disse åtte faktorene, er det bare de to siste som har direkte med klassifikasjonsfeilene å gjøre.

I dette kartleggingsprosjektet har vi i nøyaktighetsvurderingene valgt data fra helt andre kilder enn de som ble benyttet for trening av klassifikatoren. Til informasjon om

"virkeligheten" benyttet vi data fra Landsskogtakseringen, fra NTNU Vitenskapsmuseet og fra 3Q-programmet, og sammenliknet dem med kartdataene ved bruk av feilmatriser. Dette har ikke vært data ideelle for formålet, men var de eneste relevante data vi kunne fremskaffe. Valideringsdataene har romlig sett ikke hatt den ønskede dekningen av kartleggingsarealet, så vi har valgt å benytte alle data som overlappet de satellittbaserte kartene. Ideelt burde vi trukket ut tilfeldige data for validering. For flere klasser fikk vi ikke det minimumsantallet på 50 punkt per klasse som Congalton og Green (1999) anbefaler for validering dersom kartproduktet har færre enn 12 klasser.

På kartene som er laget for Sør-Trøndelag og Nord-Østerdalen er det ikke gjort romlige generaliseringer. Minsteenheter er lik piksel-oppløsning fra Landsatdataene. Likevel, er den romlige skala mellom kartpikslene og valideringsdataene forskjellig. Mens Landsat-pikslene dekker et 900 m² areal er arealet for Landsskogtakseringen 250 m². 3Q-kartene er vektorbaserte kart med minste kartleggingsenhet under 100 m². For temaene beite og fulldyrka er derimot arealene i snitt ganske store i 3Q, så her påvirkes sammenlikningen lite av størrelsen på minste kartleggingsenhet.

Dataene fra Landsskogtakseringen, som ble benyttet til valideringen i skogområdene, var innhentet i perioden 1996- 2000, men satellittopptakene var fra 1999- 2003. Denne tidsforskjellen har nok ført til at vegetasjonen i enkelte arealtypen var endret når bildet ble tatt. Sterkest virkning har dette hatt for temakartet Tilstandstyper, og kan være en av årsakene til at overensstemmelsen med valideringsdataene er lavest for dette kartet. Her har også tidsforskjellen mellom satellittdataene og treningsdataene (bestandsdata) påvirket kartproduktets riktighet. Dataene som er benyttet for validering av åpen mark er derimot innsamlet mer i samme tidsrom som satellittdataene, samtidig som det skjer mindre endringer i disse arealene over tid enn i skogarealene.

Det vil alltid være lokaliseringsavvik mellom referanse- og kartdata registrert med basis i satellittdata. Vi har ikke evaluert den romlige eller tematiske nøyaktigheten på valideringsdataene. Landsskogtakseringens data er lokalisert med GPS, men først som fast rutine fra 1997. Gjertsen og Eriksen (2004) viser til

en undersøkelse NIJOS gjorde i 2003 hvor to uavhengige GPS-målinger på Landsskogtakseringens flater ble sammenlignet. Det viste at for ca. 7 % av flatene var differansen mellom de to målingene over 10 meter, og i ca. 3 % av tilfellene var differansen over 15 meter.

Benytter vi en validering av klassifiseringen basert kun på treff innen en piksel blir resultatet for konservativt når vi vet det er lokaliseringsavvik. For å unngå dette har mange valgt å presentere nøyaktighet basert på naboskapskriterier. Da beregnes nøyaktighet for 3x3 piksler rundt et referansepunkt. I en større undersøkelse av nøyaktigheten på 1992-utgaven av den satellittbaserte arealdekkeklassifiseringen for hele USA (NLCD 1992) ble det funnet at i snitt så økte nøyaktigheten med 8 -15 % ved å benytte et slikt naboskap eller homogenitetskriterium (Stehman m. fl. 2003, Wickham m.fl. 2004).

Vi har valgt å gjøre en egen beregning for hvert kart for å få frem hvor stor del av økningen i nøyaktighet som skyldes kartets struktur ved bruk av naboskapskriterium. Kriteriet vi har brukt er at hvis klassen finnes blant en av nabopikslene så tolkes klassifiseringen som riktig. Ved å beregne feilmatriser for bare senterpiksel og for naboskapskriteriet med utgangspunkt i de tematiske kartene, men la klassetilhørigheten for hver polygon (segment) bli tildelt tilfeldig, får vi fram endringen forårsaket av heterogeniteten i kartet (både romlige og antall temaklasser).

I våre beregninger stod det romlige avviket for mellom 9 % og 18 % av samlet nøyaktighet. Størst var avviket for skogkartene der kartene var validert mot Landsskogtakseringen. Vi må kunne anta at lokaliseringen av disse er relativ god, slik at forskjellen til kartet over åpen mark først og fremst kan tilskrives større romlig homogenitet i de validerte skogkartene.

Hvis data innhentes med tanke på at de skal brukes til validering kan det under registreringen merkes hva som er den primære klassen og hva som eventuelt er en sekundær klasse, enten ved at det er rom for tolking i selve punktet eller at det er en alternativ klasse i en av nabopikslene. I valideringen kan da benyttes et kriterium som gir treff for senterpiksel dersom klassetypen finnes blant den primære eller sekundære kategoriseringen. Wickham m.fl. (2004) har for eksempel valgt å presen-

tere nøyaktigheter basert på de tre kriteriene sentrum, sentrum med primær eller sekundær klasse og homogenitet som nabokriterium. Selv om feilen i kartets klassifisering kun er en av mange faktorer som påvirker resultatet i feilmatriser, gir matrisene likevel en god beskrivelse på overensstemmelsen mellom to datasett.

Sammenlikningen vi gjorde mellom kartdataene som ble benyttet til trening og dataene for validering gav overraskende store forskjeller, spesielt mellom NTNU Vitenskapsmuseets data og NIJOS sine vegetasjonskart i målestokk 1: 20 000 - 50 000. En årsak til de store forskjellene kan være at valideringsdataene fra Landsskogtakseringen og NTNU er punktmessig innhentet, mens treningsdataene ble hentet fra arealkart. Minste kartleggingsenhet er forskjellig og kan føre til ulik tolking av klasser i heterogene landskap.

Med overensstemmelser mellom eksisterende kartdata på bare 38 % – 81 % skulle en i utgangspunktet tro at overensstemmelsen mellom valideringsdataene og de satellittbaserte kartene ble dårligere enn dette. Men her har klart den vaskingen vi gjorde på treningsdataene ved visuell tolking i fly- og satellittbilder og terskling med NDVI-verdiene, samt beslutningstreets evne til å finne hovedtendensen i et materiale bidratt positivt for de satellittbaserte temakartene. Med unntak av tilstandstyper i skog gir satellittkartene den samme eller bedre overensstemmelse.

For temakartene Treslag og Åpen mark fant vi en samlet overensstemmelse med valideringsdataene på henholdsvis 67 % og 74 %. For tilstandsklasser i skog var overensstemmelsen kun på 53 %. Å sammenlikne disse tallene med andre klassifiseringer der systematiske valderinger er gjort bør kun gjøres med kart der den tematiske oppløsningen er noenlunde lik. For eksempel fant Wickham m. fl. (2003) at samlet nøyaktighet ble redusert fra ca 84 % på et overordnet arealklassenivå til 38 % til 70 % når de så på et nivå med flere vegetasjonsklasser.

Tilsvarende reduksjon rapporteres også i en evaluering av det svenske CORINE Land Cover produktet Svenska Marktäckedata (SMD, www.lantmateriet.se). Röst og Ahl-crona (2004) fant en samlet nøyaktighet på 71 %, men med stor variasjon mellom de ulike klasser. Marklund m. fl. (2004) har i fire län gjort en vurdering av Fjellklassifiseringen, en



mer detaljert pikselbasert klassifisering som forenkles (både romlig og tematisk) til SMD. De fant at den samlede nøyaktigheten var i gjennomsnitt 55 % for fjellklassifiseringen, og 66 % for SMD på samme areal. På pikselnivå rapporteres det også at samlet nøyaktighet for skogklassifiseringen ligger mellom ca 50 % i nordlige Sverige til ca 70 % i sørlige (Olsson i Röst og Ahlcrona 2004). For treslag og skogtilstand er tilsvarende analyser gjort av NIJOS for flere kommuner på Østlandsområdet (www.nijos.no). Gjertsen og Eriksen (2004) fant en samlet nøyaktighet for treslag på 64 % og 66 % for en inndeling av skogtilstand i tre hogstklasser.

I den andre nasjonale arealdekkeklassifikasjonen som nå gjennomføres i USA (NLCD 2001) basert på Landsatdata og med beslutnings-/regresjonstre som analysemetode foreligger nøyaktighetstall for tre testområder. Med en tematisk inndeling fra 12–18 klasser fikk de samlet nøyaktighet fra 70 % til 77 % (Homer m. fl. 2004). Men for testområdene ble det benyttet en annen metode for nøyaktighetsberegning (N-fold cross validation), så tallene er ikke direkte sammenliknbare med feilmatriser. I en nylig avsluttet klassifisering, som bygget på samme datagrunnlag og analysemetodikk som NLCD2001, men som tematisk var mer detaljert (85 klasser) ble det oppnådde en samlet nøyaktighet på 61 % (Lowry m. fl. 2005). Dette er et rimelig bra resultat.

Fra SMD kartleggingen i Sverige (Rost & Ahlcrona 2004) vises det til lav overensstemmelse i klassifiseringen av løvskog og blandingsskog, samt barskog på lavmark. Under kartleggingen i Sør-Trøndelag fant vi at dette var areal typer som manglet referansedata av tilstrekkelig dekning og kvalitet. For disse areal typene måtte det derfor hentes inn betydelige mengder med digitaliserte treningsdata fra fly- og satellittbilder. Dette resulterte i at vi fikk bedre nøyaktighet på for eksempel lauvskog enn de andre treslagene. Barskog på lavmark, en egen klasse i blant annet CLC, er mulig å kartlegge når det foreligger egne referansedatasett for lavmark. Vi har ikke gjort dette til en egen klasse i arealkartene.

Ut fra hva som rapporteres i litteraturen på nøyaktighet for tilsvarende type kartleg-

ging som i Sør-Trøndelag/Nord-Østerdalen og med tanke på at de treningsdata vi hadde tilgjengelig var langt fra optimale for oppgaven konkluderer vi med at metoden med bruk av beslutningstre har fungert godt.

8.6 Mulige forbedringer i analysemetoden

Ut fra erfaringene i dette prosjektet og andre prosjekt i SatNat-programmet ser vi en del faktorer som skal til for å gjøre kartproduktene bedre og hva som kan effektiviseres for å operasjonalisere produksjonsmetoden. Der som treningsdataene hadde gjort det mulig å gjøre analyser også med regresjonstre (kart med kontinuerlige verdier og ikke klasser) ville det gitt mer fleksibilitet i anvendelsene av kartproduktene.

Programvaren som ble benyttet var standard programvare som ble tilrettelagte med script for konvertering av regler mellom bildebehandlingsprogrammet og statistikkprogrammet. Statistikkfunksjonene for å lage regelsettene kunne vært integrert i bildebehandlingsverktøyet, men den manuelle håndteringen av data gav også fordeler ved gode muligheter for kontroll underveis i prosessen.

Effekten av ulik belysning og skygge kunne med fordel vært korrigert i bildene. Dette medfører imidlertid en betraktelig arbeidsmengde i forkant av klassifikasjonen. I hvert bilde må det gjøres en manuell identifisering av ulike arealer og korreksjonen medfører ofte en del uønsket støy. En effektiv metode for atmosfærisk og topografisk normalisering savnes. Fremtidige satellittplattformer vil gi mer informasjon om atmosfæriske forhold, slik at disse faktorene kan korrigeres bedre. Dermed vil også korreksjoner for bakkehelningen bli sikrere og opptak fra ulike tidspunkt kan analyseres samlet.

Satellitdata med høyere multispektral oppløsning, slik som for eksempel ASTER eller SPOT 5 XI med 10-15 meters pikselstørrelse, ville gitt et mer anvendelig produkt.

Kartproduktene ville da i større grad kunne utnyttes av aktører som forvalter mindre administrative enheter (større eiendommer, kommuner) til aktører med nasjonalt ansvar. Med SPOT 5 kan opptak bestilles og dermed sikre flere opptak over en kortere periode.

Det å jobbe med myr, skog og fjell hver for seg har nok vært lite hensiktsmessig. En alternativ tilnærming er å splitte landskapet opp i vann, vegetasjon og vegetasjonsløs mark først og så dele inn i undergrupper koblet med N50 markslag etterpå. En annen fremgangsmåte er å slå sammen fjell, myr og åpen mark til en klasse og heller tilføre avgrensningen for myrene i ettertid.

Det største forbedringspotensialet ligger likevel i hvordan treningsdata fremskaffes. Å benytte analyser av høyoppløselige bilde-data som kilde til trening av satellittdata som Landsat, Aster, Spot og liknende bilde-data som dekker store areal, vil medføre at treningsdata og analysedata begge er innhentet basert på spektrale verdier og dermed harmoniserer mer med hverandre.

8.7 Veggen videre til et operasjonelt system

I Norge er det nå gode erfaringer med styrte klassifikasjonsmetoder. Dataminingsteknikkene og kNN (Gjertsen & Eriksen 2004) er effektive og kan raskt få frem kartprodukt dersom referansedata er av god kvalitet. Å bedre kvaliteten på referansedata vil derfor være den viktigste komponent å jobbe med for å få bedre fylkesvise arealdekkkart og for å få raskere produksjonsgang. Referansedataene må dessuten være enhetlig innsamlet for hele landet og i alle landskapstyper.

Kartleggingen av Sør-Trøndelag og Nord-Østerdalen har vist at det er mulig å benytte eksisterende kartdata som trening til klassifikasjon av satellittdata, men også at det var store begrensninger i hvordan slike data kunne utnyttes. Blant annet var det nødvendig

å hente inn supplerende informasjon fra fly- og satellittbilder. For å lage relativt detaljerte fylkesvise kart med støtte fra satellittdata er en avhengig av treningsdata som gir entydig informasjon om parametere som kan skilles i satellittdata eller i satellittdata kombinert med andre heldekkende støtdata.

Eksisterende kartdata kan ha stor verdi som hjelpedatasett for klassifikasjon og til validering, særlig når de er heldekkende slik som DMK etter hvert begynner å bli, eller innsamlet i et systematisk rutenett slik som Landsskogtakseringens punktdata. Likevel, for både å kunne lage mer detaljerte kartprodukter og for å få en rask produksjonsprosess trengs høyoppløselige fly- eller satellittbilder med infrarød informasjon. Det oppstartede programmet for omløpsfotografering av hele landet kan gi slik informasjon dersom flyfotograferingen blir basert på de nye digitale flyfotograferingskameraene og opptakene kun gjort i den fotosynteseaktive delen av vekstsesongen. I fjellet vil det være et kort tidsrom. Flyfotoopptak i et regelmessig rutenett og med bruk av automatiserte klassifiseringsmetoder (Hjeltnes 2006) vil være enda bedre egnet. Det vil gi et tidsmessig bilde og data bedre egnet for overvåking.

Bestandskartleggingen har også et potensiale til å bli en bedre kilde til treningsdata i skogområder ettersom digital sensorer og mer automatiserte analysemetodene blir benyttet i økende grad. Det gir mulighet for utfigurering og klassifisering av enkelttrær (Leckie m. fl. 2003, Leckie m. fl. 2005). Slike kartdata vil kunne konverteres til presise treningsdata egnet for beslutnings- og regresjonstreanalyse på satellittdata med 5 – 30 meter pikseloppløsning, dvs. den type satellittdata som vil være aktuelle for heldekkende arealkartlegging av hele fylker. Med treningsdata basert på høyoppløselige bilde-data fra digitale multispektrale sensorer og analysert med segmenteringsteknikker vil vi i fremtiden kunne få langt mer presise arealdekkkart enn i dag.



9 Konklusjoner og anbefalinger

Ved hjelp av eksisterende kartdatabaser er det blitt laget kart over Sør-Trøndelag og Nord-Østerdalen omtrent uten bruk av feltarbeid. Dette kan medføre besparelser. Likevel, det har ikke forenklet arbeidet i særlig grad sammenliknet med om egne treningsdata ble innhentet i prosjektet, og det har gitt sterke føringer på klasseinndeling og nøyaktighet. Potensialet med bruk av eksisterende kartdata for produksjon av heldekkende arealdekkekart er klart til stede, men da må treningsdataene samsvare i større grad med den spektrale informasjonen i satellittbildene.

Ut fra erfaringene med arbeidet i dette prosjektet kan følgende hovedkonklusjoner kan settes opp:

- Styrkt klassifikasjon med bruk av dataminingsteknikker er rask og effektiv i bruk, men krever et representativt og stort antall referansedata av høy kvalitet. Siden vi for flere av analysene oppnådde bedre samsvar med valideringsdata enn det var mellom valideringsdata og treningsdata, og siden treningsdataene vi hadde tilgjengelig var langt fra optimale konkluderer vi med at analysemetoden som sådan fungerte godt.
- I valg av klassifikasjonssystem bør det arbeides for å etablere kriterier som legger større vekt på fysiognomiske faktorer enn dagens plantesosiologiske tilnærming. Videre bør det legges større vekt på muligheten for å få til en tidseffektiv kartlegging, hvor produkter fra automatiske klassifiseringsmetoder på detaljerte bilde data benyttes til trening for klassifisering på satellittdata med stort dekningsområde.
- Data fra Landsat-satellittene med 30 meter oppløsning er i groveste laget for en kartlegging av norsk natur med småskala arealbruk og heterogent landskap. Vi tror multispektrale data i oppløsningsskala 5-15 meter gir produkter mer egnet for norske forhold (dvs. at de fanger opp bruksområder fra store eiendommer via kommune og fylkesnivå til nasjonalt nivå).
- I en operasjonalisering av produksjonsprosessen bør det også utprøves om en atmosfærisk korreksjon av satellittdataene kan gi en gevinst, enten ved forbedret kvalitet på kartproduktet eller raskere produksjon ved at mange scener kan mosaikkes og analyseres samlet.
- Vi vil anbefale at det jobbes videre med å etablere et system for innhenting av referansedata, et system som bygger på detaljert tolking av bilde data fra høyoppløselige digitale sensorer, siden slike sensorer både gir god visuell gjenkjenning og spektral informasjon til å skille ut enheter og måle parametere som enten er til nytte i utarbeidelse av arealplaner, viktige i økonomisk beregninger eller for økologiske spørsmål.
- En større grad av koordinert datainnsamling mellom nasjonale offentlige aktører og bedre integrering av data fra private skogtakseringsselskap vil gi muligheter for en mer effektiv kartlegging og overvåking av Norges arealdekke enn dagens situasjon gir.

10 Referanser

- Akaiki, H. 1973.** Information theory and an extension of the maximum likelihood principle. In: B.N. Petran and F. Csàaki (Editors), International symposium of Information Theory. Akadèmiai Kiad, Budapest, Hungary, pp. 267-281.
- Arnoldussen, A. H., Jansen, I. J. & Johansen, B. 1999.** Utprøving av CORINE Land Cover i tre utvalgte delområder i Norge. Rapport 01/99 NIJOS, Ås, Norge. 47s.
- Balle, O. 2000.** Vegetasjonskartlegginger i Norge. Kartlegginger fordelt på fylke/kommune, 5. utgave. Rapport 15/00 NIJOS, Ås, Norge.
- Bjørndal, I., Bjørkelo, K., Nilsen, B., Nystuen, I., Strand, G.H. & Thorvaldsen, K. 2004.** Kodeverk og symbolbruk i DMK og avleidd produkt. Rapport 36/04 NIJOS, Ås, Norge. 21s.
- Campbell, J. C. 1996.** Introduction to remote sensing. Taylor & Francis, London, Great Britain, 2nd edition.
- Chambers, J. M. & Hastie, T.J. 1992.** Statistical Models in S, Wadsworth and Brooks, Ch. 9.
- Collins, J.B. & Woodcock, C.E. 1996.** An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. Remote Sensing of Environment, 56:66-77.
- Congalton, R. G. & Green, K. 1999.** Assessing the accuracy of remotely sensed data: Principles and practices. Lewis Publishers, USA. 137s.
- DN-Håndbok 13. 1999.** Kartlegging av naturtyper. Verdisetting av biologisk mangfold. Håndbok 13. Direktoratet for naturforvaltning, Trondheim, Norge.
- Elgersma, A. & Asheim V. 1998.** Landskapsregioner i Norge – landskapsbeskrivelser. Rapport 02/98 NIJOS, Ås, Norge.
- Fjellstad, W. J., Puschmann, O., Dramstad, W.E., Engan, G., Reid, S.J. & Stokland, J.N. 2004.** 3Q - Tilstandsovervåking og resultatkontroll i jordbrukets kulturlandskap – Rapport for 2001-2002 – Hordaland, Sogn og Fjordane, Møre og Romsdal, og Sør-Trøndelag. Norsk institutt for jord- og skogkartlegging, Ås, Norge.
- Fremstad, E. 1998.** Vegetasjonstyper i Norge. NINA Temahefte 12, NINA, Trondheim, Norge.
- Foody, G.M. 2002.** Status of land cover classification accuracy assessment. Remote Sensing of Environment 80: 185-201.
- Gjertsen, A. K., Stokland, J., Jansen, I. J., Schøning, P. & Strøm, G. D. 2003.** Nasjonalt system for arealdekke-klassifikasjon (NaSAK), Rapport 03/04 NIJOS, Ås, Norge.
- Gjertsen, A.K. & Eriksen, R. 2004.** Test av MSFI-metoden. Nøyaktighetsanalyse på datasett fra Østfold og Hobøl. NIJOS rapport 2/2004.
- Gregorio, A.D. 2005.** Land Cover Classification System. Classification concepts and user manual. FAO, Rome
- Hand, D., Mannila, H. & Smyth, P. 2001.** Principles of data mining. Massachusetts Inst. of Technology, Cambridge, Massachusetts, USA.
- Hansen, M., Dubayah, R., & Defries, R. 1996.** Classification trees: an alternative to traditional land cover classifiers. International Journal of Remote Sensing, 17:1075–1081
- Hansen, M. C., Defries, R., Townshend, J. R. G., & Sohlberg, R. 2000.** Global land-cover classification at 1 km spatial resolution using a classification tree approach. International Journal of Remote Sensing, 21:1331–1364.



- Hastie, T., Tibshirani, R. & Friedman, J. 2001.** The Elements of Statistical Learning. Datamining, Inference, and Prediction. Springer, NY, USA.
- Hjeltnes, A., Brandtzæg, B.A., Lie, K. & Kleven, G. 1996.** Arealprosjektet i Vestfold - Sluttrapport. Arealklassifisering og dataformidling. Fylkesmannen i Vestfold.
- Hjeltnes, A. & Øien, D. I. 2003.** Klassifikasjon og flybildetolkning i SatNat. Upublisert notat, Høgskolen i Telemark og NTNU Vitenskapsmuseet i Trondheim.
- Hjeltnes, A. 2006.** Høyoppløselige bilder som grunnlag for overvåking av endringer i fjellvegetasjon. HiT skrift 1/2006. Høgskolen i Telemark, 55s.
- Homer, C.G., & Gallant, A. 2001.** Partitioning the conterminous United States into mapping Zones for Landsat TM land cover mapping, USGS Draft White Paper, available at <http://landcover.usgs.gov>; sist tilgang 1 april 2006.
- Homer, C. G., Huang, C., Yang, L., & Wylie, B. 2002.** Development of a CIRCA 2000 landcover database for the United States. In: Proceedings of the ASPRS Annual Conference, Washington D.C., USA. American Society of Photogrammetry and Remote Sensing.
- Homer, C., Huang, C., Yang, L., Wylie, B. & Coan, M. 2004.** Development of a 2001 National Land-Cover Database for the United States. PE&RS 70: 829-840.
- Huang, C., Yang, L., Wylie, B., & Homer, C. 2001.** A strategy for estimating tree canopy density using Landsat ETM+ and high resolution images over large areas. In: Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry, volume 1, Denver, Colorado, USA.
- Insightful Corporation. 2005.** S-PLUS Functional Guide. Seattle, WA, USA.
- Jensen, J. R. 2000.** Remote sensing of the Environment. An earth resource perspective. Prentice Hall, NJ, USA.
- Johansen, B.E., Johansen, M-E. & Karlsen, S.R. 1995.** Vegetasjons- og beitekartlegging i Finnmark og Nord-Troms. NORUT IT AS, Tromsø.
- Landbruksdepartementet. 2003.** Endringer i forskrift om tilskudd til skogbruksplanlegging. <http://odin.dep.no/lmd/norsk/dok/hoeringer/020081-080004/dok-bn.html>; sist tilgang 1.april 2006
- Lantmäteriet. 2003.** GSD – Marktäckedata. <http://www.lantmateriet.se/>
- Leckie, D.G., Gougeon, Walsworth, N., Paradine, D. 2003.** Stand delineation and composition estimation using semi-automated individual tree crown analysis. Remote Sensing of Environment 85: 355-369.
- Leckie, D.G., Gougeon, F.A., Tinis, S., Nelson, T., Burnett, C., Paradine, D. 2005.** Automated tree recognition in old-growth conifer stands with high-resolution digital imagery. Remote Sensing of Environment 94: 311-326.
- Lieng, E. 2002.** Vegetasjonssoner - teknisk notat. Statens kartverk, Aust-Agder, Norway, upublisert notat.
- Lieng, E., Slaymaker, D.M., & Kastdalen, L. 2002.** Use of airborne small-format digital cameras as tools for collecting ground truth. 29th International Symposium on Remote Sensing of Environment (ISRSE), Buenos Aires, Argentina.
- Lillesand, T.M. & Kiefer, R.W. 2000.** Remote sensing and image interpretation. 4th edition. John Wiley & Sons, Inc. New York.
- Lowry, J. H, Jr., R. D. Ramsey, K. Boykin, D. Bradford, P. Comer, S. Falzarano, W. Kepner, J. Kirby, L. Langs, J. Prior-Magee, G. Manis, L. O'Brien, T. Sajwaj, K. A. Thomas, W. Rieth, S. Schrader, D. Schrupp, K. Schulz, B. Thompson, C. Velasquez, C. Wallace, E. Waller and B. Wolk. 2005.** Southwest Regional Gap Analysis Project: Final Report on Land Cover Mapping Methods, RS/GIS Laboratory, Utah State University, Logan, Utah.

- Ma, Z. & Redmond, R.L. 1995.** Tau coefficients for accuracy assessment of classification of remote sensing data, Photogram. Eng. Remote Sens., 61, 435–439.
- Moen, A. 1998.** Nasjonalatlas for Norge: Vegetasjon. Statens kartverk, Hønefoss, Norway.
- Moen, A. & Odland, A. 2002.** Retningslinjer for detaljert vegetasjonssonekart, Sør-Trøndelag som prøvefylke. Norges Teknisk-Natur-vitenskapelige Universitet, Vitenskaps-museet, Norway, upublisert notat.
- Moum, S.O. 1999.** Arealressurskart i liten målestokk. Artikkel i Kart og plan, Vol. 59, 1999.
- Pal, M. & Mather, P. M. 2003.** An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment, 86:554–565.
- Parker, K. C. 1998.** Environmental relationships and vegetation associates of columnar cacti in the northern Sonoran desert (Arizona, USA). Vegetatio, 78(3):125–140.
- Rekdal, Y. & Larsson, J. Y. 2005.** Veiledning i vegetasjonsskartlegging. M 1:20 000 – 50 000. Rapport 05/05 NIJOS, Ås, Norge.
- Rost, T. & Ahlcrona, E. 2004.** Tematisk noggrannhet i Svanska Marktäckedata. Rapport Lantmäteriet SCMD-0020, versjon 2
- Statistisk Sentralbyrå (SSB). 2003.** Landa-real fordelt på jordbruksareal, produktivt skogareal og annet areal, etter fylke. km² og prosent. http://www.ssb.no/emner/10/04/20/nos_skogstat/nos_d320/tab/tab-2.1.html; sist tilgang 1. april 2006
- Stehman, S.V., Wickham, J.D., Smith, J.H. & Yang, L. 2003.** Thematic accuracy of the 1992 national Land-Cover Data for the eastern United States. Remote Sensing of Environment 86: 500-516.
- Tan, P-N., Steinbach, M. & Kuumar, V. 2006.** Introduction to datamining. Pearson Education, Boston, USA.
- Tomter, S.M. & Eriksen, R. 2001.** Landskogtakseringen 1995-99. Fylkestakst Østfold. NIJOS-ressursoversikt 01/01. Norsk institutt for jord- og skogkartlegging, Ås, Norway.
- Tucker, C. J. 1979.** Red and infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8:127–150.
- USGS. 2002.** MRLC 2001 Image preprocessing procedure. EROS Data Center, U. S. Geological Survey, Sioux Falls, U.S.A.
- Vikhamar, D., Fjone, G., Kastdalen, L. & Bolstad, J.P. 2004.** Satellittdata til kartlegging av arealdekke. Utprøving av beslutningstremetodikk i Østfold fylke, DN-utredning 2004-3. 64s.
- Vikhamar, D. & Kastdalen, L. 2005.** Impervious surface mapping in Southern Norway, 31st International Symposium on Remote Sensing of Environment, St. Petersburg, Russia.
- Wickman, J.D., Stehman, S.V., Smith, J.H. & Yang, L. 2004.** Thematic accuracy of the 1992 national Land-Cover Data for the western United States. Remote Sensing of Environment 91: 452-468.
- Wilson, J. P. 2000.** Terrain analysis. Principles and applications. John Wiley & sons, New York, USA.
- Witten, I.H. & Frank, E. 2005.** Datamining. Practical Machine Learning Tools and Techniques. 2ed. San Francisco, CA, USA
- Økland, R. 2005.** Ny norsk naturtypeinndeling. Notat 5/12-2005.Artsdatabanken.



11 Vedlegg

11.1 Validering av klassifikasjon (feilmatriser)

11.1.1 SatNat Treslag

Tabell 11.1. Feilmatrise som viser klassifiserte piksler i temakartet for treslag (i rader) validert mot punkter fra Landsskogtakseringen (i kolonner).

	Gran	Furu	Lauv	Sum	Bruker
Gran	115	42	61	218	53 %
Furu	49	129	40	218	59 %
Lauv	29	15	100	144	69 %
Sum	193	186	201	580	
Prod.	60 %	69 %	50 %	Samlet:	59 %
Khat	39 %		Tau	49 %	

Tabell 11.2. Feilmatrise som viser klassifiserte piksler i temakartet for treslag (i rader) validert mot punkter fra Landsskogtakseringen (i kolonner) basert på kriteriet naboskap.

	Gran	Furu	Lauv	Sum	Bruker
Gran	174	14	30	218	80 %
Furu	16	183	19	218	84 %
Lauv	8	4	132	144	92 %
Sum	198	201	181	580	
Prod.	88 %	91 %	73 %	Samlet:	84 %
Khat	76 %		Tau	78 %	

Tabell 11.3. Diversitet i det klassifiserte temakartet for treslag. Et 3x3 vindu rundt hvert referansepunkt i det klassifiserte bildet er undersøkt for diversitet, altså antall ulike klasser. Med bare tre klasser vil nøyaktighe-
ten ved å se på 3x3 piksler lett bli kunstig høy, da 62 % av punktene inneholder flere klasser som naboer.

Antall verdi 1	218	37.6 %
Antall verdi 2	274	47.2 %
Antall verdi 3	88	15.2 %
Antall punkt:	580	
Gjennomsnitt	1.78	

Tabell 11.4. Feilmatrise som viser tilfeldig omkodede segmenter i temakartet for treslag (i rader) validert mot punkter fra Landsskogtakseringen (i kolonner).

	Gran	Furu	Lauv	Sum	Bruker
Gran	66	76	71	213	31 %
Furu	71	63	44	178	35 %
Lauv	81	79	29	189	15 %
Sum	218	218	144	580	
Prod.	30 %	29 %	20 %	Samlet:	27 %
Khat	-9 %		Tau	20 %	

Tabell 11.5. Feilmatrise som viser tilfeldig omkodede segmenter i temakartet for treslag (i rader) validert mot punkter fra Landsskogtakseringen (i kolonner) basert på kriteriet naboskap.

	Gran	Furu	Lauv	Sum	Bruker
Gran	100	58	57	215	47 %
Furu	52	100	33	185	54 %
Lauv	66	60	54	180	30 %
Sum	218	218	144	580	
Prod.	46 %	46 %	38 %	Samlet:	44 %
Khat	15 %		Tau	34 %	

Tabell 11.6 Korrigering av nøyaktighetstall for lokaliseringsavvik. Korreksjonen er endringen i samlet nøyaktighet, Khat og Tau mellom tabell 11.5 og tabell 11.4.

	Senterpunkt (%)	Naboskap (%)	Korreksjon (%)	Korrigert verdi (%)
Samlet	59	84	17	67
Khat	39	76	24	52
Tau	49	78	14	64

11.1.2 SatNat Skogtilstand

Tabell 11.7. Feilmatrise som viser klassifiserte piksler (i rader) i temakartet for skogtilstand validert mot punkt fra Landsskogtakseringen (i kolonner). Hogstflater har vokst opp til ungskog i perioden mellom land-skog data ble målt og satellittbilder ble tatt.

	Hogstflate	Ungskog	Voksen skog	Lavproduktiv skog	Impediment	SUM	Bruker
Hogstflate	0	4	8	1	24	37	0 %
Ungskog	3	92	69	22	79	265	35 %
Voksen skog	2	31	100	16	60	209	48 %
Lavproduktiv skog	1	17	49	15	27	109	14 %
Impediment	0	1	8	1	131	141	93 %
SUM	6	145	234	55	321	761	
Produsent	0 %	63 %	43 %	27 %	41 %	Samlet:	44 %
Khat	27 %		Tau	37 %			

Tabell 11.8. Feilmatrise som viser klassifiserte piksler (i rader) i temakartet for skogtilstand validert mot punkt fra Landsskogtakseringen (i kolonner). Beregnet treff etter nærmeste nabo kriterium.

	Hogstflate	Ungskog	Voksen skog	Lavproduktiv skog	Impediment	SUM	Bruker
Hogstflate	0	0	2	1	4	7	0 %
Ungskog	3	125	35	17	63	243	51 %
Voksen skog	2	9	169	9	26	215	79 %
Lavproduktiv skog	1	10	24	28	21	84	33 %
Impediment	0	1	4	0	207	212	98 %
SUM	6	145	234	55	321	761	
Produsent	0 %	86 %	72 %	51 %	64 %	Samlet:	70 %
Khat	58 %		Tau	62 %			



Tabell 11.9. Diversitet i det klassifiserte datasettet. Et 3x3 vindu rundt hvert referansepunkt i det klassifiserte bildet er undersøkt for diversitet, altså antall ulike klasser. Med fem klasser og hovedtyngden av pikslene (75 %) har en til to klasser som naboer kan nøyaktighetstallet ved naboskaps betraktning forsvares.

Antall verdi 1	208	27.3 %
Antall verdi 2	359	47.2 %
Antall verdi 3	159	20.9 %
Antall verdi 4	33	4.3 %
Antall verdi 5	2	0.3 %
<hr/>		
Antall punkt:	761	
Gjennomsnitt	2.03	

Tabell 11.10. Feilmatrix som viser tilfeldig omkodete segmenter (i rader) i temakartet for skogtilstand validert mot punkt fra Landsskogtakseringen (i kolonner).

	Hogstflate	Ungskog	Voksen skog	Lavproduktiv skog	Impediment	SUM	Bruker
Hogstflate	1	31	42	8	70	152	1 %
Ungskog	0	32	46	14	54	146	22 %
Voksen skog	1	26	46	5	78	156	29 %
Lavproduktiv skog	2	30	55	13	54	154	8 %
Impediment	2	26	45	15	65	153	42 %
SUM	6	145	234	55	321	761	
Produsent	17 %	22 %	20 %	24 %	20 %	Samlet:	21 %

Tabell 11.11. Feilmatrix som viser tilfeldig omkodete segmenter (i rader) i temakartet for skogtilstand validert mot punkt fra Landsskogtakseringen (i kolonner) etter nærmeste nabo kriterium

	Hogstflate	Ungskog	Voksen skog	Lavproduktiv skog	Impediment	SUM	Bruker
Hogstflate	2	26	34	7	55	124	2 %
Ungskog	0	51	36	12	46	145	35 %
Voksen skog	1	23	93	4	66	187	50 %
Lavproduktiv skog	1	27	41	22	42	133	17 %
Impediment	2	18	30	10	112	172	65 %
SUM	6	145	234	55	321	761	
Produsent	33 %	35 %	40 %	40 %	35 %	Samlet:	37 %

Tabell 11.12 Diversitet i datasettet med tilfeldig tildeling av verdier til segmentene i temakartet Skogtilstand. Et 3x3 vindu rundt hvert referansepunkt i det klassifiserte bildet er undersøkt for diversitet, altså antall ulike klasser. Med fem klasser og hovedtyngden av pikslene (87 %) har fire til fem klasser som naboer kan nøyaktighetstallet ved naboskaps betraktning ikke forsvares. Sammenliknet med tabell 11.9 ser man at fordelingen er snudd opp ned.

Antall verdi 1	0	0.0 %
Antall verdi 2	6	0.8 %
Antall verdi 3	87	11.4 %
Antall verdi 4	376	49.4 %
Antall verdi 5	292	38.4 %
<hr/>		
Antall punkt:	761	
Gjennomsnitt	4.25	

11.1.3 SatNat Åpen mark

Tabell 11.13. Feilmatrixe som viser klassifiserte piksler (i rader) i temakartet for fjell og åpen mark validert mot satnat innhentede punkt (i kolonner). Senter av piksel.

	1	2	3	4	5	6	7	8	9	10	Sum
1 Frost/snøleie	182	63	13	92	18	115	164	0	0	17	664
2 Lavhei	24	1634	13	158	2	0	54	0	0	0	1885
3 Hei	17	141	253	201	192	0	1	0	0	0	805
4 Gras/starr	49	72	84	360	51	3	4	0	0	2	625
5 Urt/eng	1	2	183	43	169	15	0	0	0	0	413
6 Kystlyng	0	0	18	0	23	364	0	0	0	23	428
7 Stein/ur	7	1	0	4	1	0	1032	0	0	0	1045
8 Utbygd	0	0	0	0	1	3	0	0	0	2	6
9 Strand	0	0	3	0	0	15	0	0	0	3	21
10 Strandenger	0	0	0	0	0	20	0	0	0	20	40
Sum	280	1913	567	858	457	535	1255	0	0	67	5932

Tabell 11.14. Feilmatrixe som viser klassifiserte piksler (i rader) i temakartet for fjell og åpen mark validert mot satnat innhentede punkt (i kolonner). 3x3 naboskap analyse.

	1	2	3	4	5	6	7	8	9	10	Sum
1 Frost/snøleie	246	11	13	25	17	8	85	0	0	9	414
2 Lavhei	4	1842	6	57	2	0	32	0	0	0	1943
3 Hei	12	36	387	94	153	0	1	0	0	0	683
4 Gras/starr	17	22	29	658	49	0	4	0	0	2	781
5 Urt/eng	1	1	111	24	218	0	0	0	0	0	355
6 Kystlyng	0	0	18	0	16	516	0	0	0	16	566
7 Stein/ur	0	1	0	0	1	0	1133	0	0	0	1135
8 Utbygd	0	0	0	0	1	0	0	0	0	1	2
9 Strand	0	0	3	0	0	11	0	0	0	3	17
10 Strandenger	0	0	0	0	0	0	0	0	0	36	36
Sum	280	1913	567	858	457	535	1255	0	0	67	5932

Tabell 11.15. Nøyaktighet i forhold til referansedata i %. Fra feilmatrixene over i temakartet for fjell og åpen mark. Det var ikke etablert referansedata for klasse 8 og 9.

Producer	Senter	Naboskap	User	Senter	Naboskap
1 Frost/snøleie	65.0 %	87.9 %	1 frost/snøleie	27.4 %	59.4 %
2 Lavhei	85.4 %	96.3 %	2 Lavhei	86.7 %	94.8 %
3 Hei	44.6 %	68.3 %	3 hei	31.4 %	56.7 %
4 Gras/starr	42.0 %	76.7 %	4 gras/starr	57.6 %	84.3 %
5 Urt/eng	37.0 %	47.7 %	5 urt/eng	40.9 %	61.4 %
6 Kystlyng	68.0 %	96.4 %	6 kystlyng	85.0 %	91.2 %
7 Stein/ur	82.2 %	90.3 %	7 stein/ur	98.8 %	99.8 %
8 Utbygd	N/A	N/A	8 Utbygd	N/A	N/A
9 Strand	N/A	N/A	9 Strand	N/A	N/A
10 Strandenger	29.9 %	53.7 %	10 Strandenger	50.0 %	100.0 %
Samlet	67.7 %	84.9 %			
Khat	60 %	81 %			
Tau	62 %	82 %			



Tabell 11.16. Diversitet i temakartet for fjell og åpen mark. Et 3x3 vindu rundt hvert referansepunkt i det klassifiserte bildet er undersøkt for diversitet, altså antall ulike klasser. Se kapittel 4.3. 80 % av punktene er omgitt av en eller to klasser. Med ti klasser kan ikke forbedringen av samlet nøyaktighet fra 67,7 % til 84,9 % forklares utelukkende fra heterogeniteten i dataene. Dette viser at en feilmatrix basert på 3x3 naboskap gir i dette tilfellet et fornuftig resultat.

Antall verdi 1	2050	34.6 %
Antall verdi 2	2660	44.8 %
Antall verdi 3	1032	17.4 %
Antall verdi 4	181	3.1 %
Antall verdi 5	9	0.2 %
Antall verdi 6	0	0.0 %
Antall verdi 7	0	0.0 %

Antall punkt: 5932
Gjennomsnitt 1.89

Tabell 11.17. Feilmatrix som viser tilfeldig omkodede segmenter (i rader) i temakartet for fjell og åpen mark validert mot satnat innhentede punkt (i kolonner). Senter av piksel.

	1	2	3	4	5	6	7	8	9	10	Sum
1 Frost/snøleie	11	73	54	78	32	88	38	0	0	8	382
2 Lavhei	34	184	68	64	38	27	82	0	0	2	499
3 Hei	30	211	80	114	97	31	44	0	0	7	614
4 Gras/starr	29	197	70	84	40	10	49	0	0	1	480
5 Urt/eng	38	456	69	121	7	76	77	0	0	3	847
6 Kystlyng	25	116	24	70	60	41	34	0	0	11	381
7 Stein/ur	14	127	60	72	64	79	337	0	0	18	771
8 Utbygd	20	210	43	88	34	30	137	0	0	2	564
9 Strand	14	114	31	75	15	36	305	0	0	8	598
10 Strandenger	65	225	68	92	70	117	152	0	0	7	796
Sum	280	1913	567	858	457	535	1255	0	0	67	5932

Tabell 11.18. Feilmatrix som viser tilfeldig omkodede segmenter (i rader) i temakartet for fjell og åpen mark validert mot satnat innhentede punkt (i kolonner). 3x3 naboskap analyse.

	1	2	3	4	5	6	7	8	9	10	Sum
1 Frost/snøleie	37	60	50	65	31	79	35	0	0	7	364
2 Lavhei	29	426	61	56	37	19	75	0	0	2	705
3 Hei	28	150	129	101	93	31	38	0	0	7	577
4 Gras/starr	27	164	57	179	40	9	43	0	0	1	520
5 Urt/eng	34	410	67	116	25	72	71	0	0	3	798
6 Kystlyng	22	101	19	65	59	84	31	0	0	9	390
7 Stein/ur	12	105	58	53	63	74	398	0	0	18	781
8 Utbygd	18	181	36	77	32	30	129	0	0	2	505
9 Strand	12	104	29	66	10	34	295	0	0	8	558
10 Strandenger	61	212	61	80	67	103	140	0	0	10	734
Sum	280	1913	567	858	457	535	1255	0	0	67	5932

Tabell 11.19. Nøyaktighet i forhold til referansedata i %. Fra feilmatrixene over med tilfeldig omkodede segmenter. Det var ikke etablert referansedata for klasse 8 og 9.

<i>Producer</i>	<i>Senter</i>	<i>Naboskap</i>	<i>User</i>	<i>Senter</i>	<i>Naboskap</i>
1 Frost/snøleie	3.9 %	13.2 %	1 Frost/snøleie	2.9 %	10.2 %
2 Lavhei	9.6 %	22.3 %	2 Lavhei	36.9 %	60.4 %
3 Hei	14.1 %	22.8 %	3 Hei	13.0 %	22.4 %
4 Gras/starr	9.8 %	20.9 %	4 Gras/starr	17.5 %	34.4 %
5 Urt/eng	1.5 %	5.5 %	5 Urt/eng	0.8 %	3.1 %
6 Kystlyng	7.7 %	15.7 %	6 Kystlyng	10.8 %	21.5 %
7 Stein/ur	26.9 %	31.7 %	7 Stein/ur	43.7 %	51.0 %
8 Utbygd	N/A	N/A	8 Utbygd	N/A	N/A
9 Strand	N/A	N/A	9 Strand	N/A	N/A
10 Strandenger	10.4 %	14.9 %	10 Strandenger	0.9 %	1.4 %
Samlet	12.7 %	21.7 %			
Khat	3 %	12 %			
Tau	11 %	20 %			

Utredninger oversikt

2004

2004-1:	Store og Lille Hovvatn i Aust-Agder - en samlerapport etter 25 år med forsøringsundersøkelser og kalking	50,-
2004-2:	Sportsfiske etter sjøørret i Trondheimsfjorden. Deltakelse i fisket og estimering av fangst	50,-
2004-3:	Satellittdata til kartlegging av arealdekke. Utprøving av beslutningstremetodikk i Østfold fylke	50,-
2004-4:	Vannkjemisk og biologisk utvikling i innsjøen Vegår i Aust-Agder etter 17 år med kalking	50,-
2004-5:	Moser i kulturlandskapet og registreringer i åkerkanter og beitemark i Trondheimsfjordområdet.	50,-
2004-6:	Bestandsstatus for laks i Norge 2003. Rapport fra arbeidsgruppe	50,-
2004-7:	Vossolaksen - bestandsutvikling, trusselfaktorer og tiltak	50,-

2005

2005-1:	Report of the workshop on introduced marine organisms: Risks and management measures	50,-
2005-2:	Nasjonal overvåking av marint biologisk mangfold i kystsonen	50,-
2005-3:	Bleka i Byglandsfjorden - bestandsstatus og tiltak for økt naturlig rekruttering	50,-
2005-4:	Bestandsstatus for laks i Norge 2004. Rapport fra arbeidsgruppe	50,-
2005-5:	Effekter av kalking på biologisk mangfold. Undersøkelser i Tovdalsvassdraget 1999-2001	nettversjon
2005-6:	Endringer i norsk flora	50,-
2005-7:	Utvärdering av Norges nationella övervakningsprogram för stora rovdjur	nettversjon
2005-8:	Satellittbilder til kartlegging av arealdekke endringer - en utprøving av metodikk for utvalgte inngrepstyper	50,-
2005-9:	Reetablering av laks på Sørlandet. Årsrapport fra reetableringsprosjektet 2001-2003	50,-
2005-10:	Reetablering av laks på Sørlandet. Årsrapport fra reetableringsprosjektet 2004	50,-

2006

2006-1:	Effects on the marine environment of ocean acidification resulting from elevated levels of CO ₂ in the atmosphere	50,-
2006-2:	Effekter av klimaendringer på økosystemer og biologisk mangfold	50,-
2006-3:	Bestandsstatus for laks	50,-
2006-4:	Reetablering av laks på Sørlandet. Årsrapport fra reetableringsprosjektet 2005	50,-
2006-5:	Satellittdata til kartlegging av arealdekke Utprøving av tilgjengelige kartdata for klassifisering av Sør-Trøndelag	50,-

Utredning er utarbeidet av andre på oppdrag av DN eller i et samarbeid med DN. Innholdet har karakter av råd til DN.

Rapport er utarbeidet av DN, og gir uttrykk for direktoratets forslag eller standpunkter.

Notat er enklere oversikter, sammenstillinger, referater og lignende.

Håndbok gir veiledning og konkrete råd om forvaltning av naturen, som regel til bruk for lokale forvaltningsorganer

Temahefte gir en popularisert framstilling av et tema.

Mer info:
www.dirnat.no/publikasjoner

Direktoratet for naturforvaltning (DN) er det sentrale, utøvende og rådgivende forvaltningsorganet innenfor bevaring av biologisk mangfold, friluftsliv og bruk av naturressurser. DN's visjon, **For liv i naturen og natur i livet**, er et uttrykk for dette. DN er administrativt underlagt Miljøverndepartementet.

Myndigheten til å forvalte naturressurser er gitt gjennom ulike lover og forskrifter. Ut over lovbestemte oppgaver har direktoratet også ansvar for å identifisere, forebygge og løse miljøproblemer ved samarbeid, rådgivning og informasjon overfor andre myndigheter og grupper i befolkningen.



Direktoratet for
naturforvaltning