Høgskolen i Gjøviks rapportserie, 2004 nr. 5

Nøkkeltall i miljøvernpolitikken:

Metodikk for bruk av satellittbilder til kvalitetssikring av kartlagte rekreasjonsarealer og nærturterreng: - Nøkkeltallene 2RM4-1 og 2RM4-2

Dagrun Vikhamar (Høgskolen i Gjøvik)
og Leif Kastdalen (Høgskolen i Hedmark)
Avdeling for teknologi

Gjøvik 2004
ISSN 0806-3176

Forord

SatNat-prosjektet ble formelt etablert som et samarbeidsprosjekt mellom Direktoratet for naturforvaltning og Norsk Romsenter den 12. februar 2001. Prosjektet omhandler bruken av satellittdata i naturforvaltning, og er delt opp i 9 delprosjekt. Delprosjekt 6 omhandler nasjonale nøkkeltall. Denne rapporten undersøker muligheten av å bruke satellittdata for å få et bedre datagrunnlag for kartlegging av nøkkeltall knyttet til nærturterreng og rekerasjonsområder.

Innholdsfortegnelse

Sammendrag **5**
1 Innledning **6**
1.1 Nøkkeltall 6
1.2 Definisjoner av nøkkeltallene 2RM4-1 og 2RM4-2 7
2 Problemstilling **8**
3 Bakgrunnsinformasjon om jordobservasjon **9**
3.1 Jordobservasjonssatellitter 9
3.2 Tolkning av satellittbilder 10
4 Metodikk **11**
4.1 Forsøksområde 11
4.2 Metode 1: Objektsegmentering og klassifikasjon av SPOT 5 bilde 14
4.2.1 Datasett 14
4.2.2 Eksperimenter 14
4.2.3 Resultater 16
4.2.3.1 Eksempler på feilkartlagte arealer 16
4.2.3.2 Korreksjon av feilkartlagte arealer 19
4.2.3.3 Statistikk for rekreasjonsarealer og nærturterreng 20
4.2.3.4 Kontroll av klassifikasjonsresultatet 21
4.2.3.5 Forbedring av klassifikasjonen 25
4.2.4 Oppsummering 27
4.3 Metode 2: Regresjonstreanalyse for klassifikasjon av Landsat ETM+ bilder 29
4.3.1 Datasett 29
4.3.2 Eksperimenter 30
4.3.3 Resultater 33
4.3.3.1 Evaluering av regresjonstrær 33
4.3.3.2 Kart over grått areal 34
4.3.3.3 Statistikk for rekreasjonsarealer og nærturterreng 41
4.3.4 Oppsummering 42
5 Diskusjon **43**
5.1 Kvalitetssikring av beregnede nøkkeltall 43
5.2 Sammenligning av metoder for kvalitetssikring av beregnede nøkkeltall 43
5.2.1 Metode 1 43
5.2.2 Metode 2 44
5.3 Oppsummering 44
6 Konklusjoner **45**
Linker **47**

Sammendrag

De grå arealene (bygninger, veier, parkeringsplasser) øker på bekostning av de grønne arealene. Dette endrer miljøet i byer og tettsteder. Statistisk Sentralbyrå (SSB) har utviklet metodikk for å kartlegge nøkkeltall for friluftsliv som omfatter rekreasjonsarealer og nærturterreng (Engelien & Schøning, 2001). SSB's metode er basert på eksisterende registre og kartdatabaser. Disse områdene er ikke verifisert i detalj, og man ønsker å vite om de kan inneholde grå arealer eller om de utelukkende består av grønne arealer. Denne rapporten fokuserer på disse problemene og er utarbeidet på oppdrag fra Direktorat for naturforvaltning innenfor et program kalt SatNat. SatNat er et femårig program (2001-2005) som omhandler metodeutvikling for bruk av satellittbilder i naturforvaltning. Programmet er et samarbeid mellom Norsk Romsenter og Direktorat for naturforvaltning.

Rapporten gir to forslag til hvordan man kan kvalitetssikre kartlagte rekreasjonsarealer og nærturterreng ved hjelp av satellittbilder. Datagrunnlaget besto av satellittbilder fra sensorene Landsat 7 ETM+ og SPOT 5 HRG. Flybilder og digitale FKB kartdata over veier og bygninger er brukt som referansedata. Ulike bildebehandlingsprogrampakker er benyttet til de to metodene.

Målsettingen med arbeidet var å kvalitetssikre rekreasjonsarealer og nærturterreng kartlagt med SSB's metodikk. Dette ble gjennomført ved to forskjellige analysemetoder:

1) Objektsegmentering og klassifikasjon av høyoppløselig SPOT 5 HRG satellittbilde.
2) Regresjonstreanalyse for klassifikasjon av Landsat 7 ETM+ satellittbilder.

Metode 1 er egnet for å finne arealtypene (grønt areal, grått areal, vann) innenfor kartlagte områder, mens metode 2 gir et grovt kart som viser konsentrasjonen av grått areal (tettstedskart).

1 Innledning

1.1 Nøkkeltall

I stortingsmelding nr. 8 (1999-2000) om regjeringens miljøvernpolitikk og rikets miljøtilstand, la regjeringen fram nasjonale mål for de miljøvernpolitiske resultatområdene. Målene skal følges opp gjennom å definere og avklare et sett med nasjonale nøkkeltall som viser utviklingen når det gjelder miljøtilstanden og viktige faktorer som påvirker denne (http://odin.dep.no/md/norsk/publ/stmeld/022005-040006/index-dok000-b-n-a.html).

De nasjonale nøkkeltallene vil være sentrale i de framtidige stortingsmeldingene om Regjeringens miljøvernpolitikk og rikets miljøtilstand. De vil også være viktige i andre sammenhenger, blant annet i «Miljøstatus i Norge» på Internett og obligatorisk internasjonal rapportering. Bakgrunnen for arbeidet er stortingsmelding nr. 24 (2000-2001) og nr 25 (2002-2003): ”Regjeringens miljøvernpolitikk og rikets miljøtilstand”, og ønske om nasjonale nøkkeltall for miljøvernpolitikken. (http://odin.dep.no/md/norsk/publ/stmeld/022001-040006/index-dok000-b-n-a.html , (http://odin.dep.no/md/norsk/publ/stmeld/022001-040020/index-dok000-b-n-a.html)

Stortingsmelding nr 25 definerer nøkkeltall slik: ”*Nøkkeltallene skal på en representativ måte vise utviklingen i miljøtilstanden og viktige faktorer og hvilke samfunnssektorer som påvirker miljøtilstanden innenfor hvert resultatområde, og dokumentere om de nasjonale målene i miljøvernpolitikken nås*.”

Den sier også: ”*Overvåking er en form for kunnskapsoppbygging som blant annet står sentralt for å vurdere utviklingen i miljøtilstanden (der nasjonale nøkkeltall brukes) slik at det kan reageres raskt ved akutte situasjoner, og for å ha løpende kunnskap om den langsiktige utvikling.*”

Nøkkeltall brukes på forhold som miljøvernforvaltningen ønsker kvantifiserte framstillinger av. Nøkkeltallene kan være målte eller beregnede parametere, indikatorer, statistikk, kartografiske framstillinger etc. Nøkkeltallene vil legge grunnlaget for hva slags miljøovervåkning miljøvernforvaltningen skal gjennomføre framover, hva slags miljøstatistikk som skal innhentes, og for hva sektormyndighetene skal rapportere til miljøvernforvaltningen (sektorvis miljørapportering). Gode nøkkeltall er derfor av stor betydning for oppfølgingen av miljøvernpolitikken (http://www.sft.no/arbeidsomr/resultatoppfolging/utskriftversjon_dbafile2478.html)

1.2 Definisjoner av nøkkeltallene 2RM4-1 og 2RM4-2

Denne rapporten omhandler to nøkkeltall for friluftsliv gitt i stortingsmelding nr 24 og 25. Bakgrunnen for utarbeidelsen av disse nøkkeltallene er et nasjonalt strategisk mål for friluftsliv (Stortingsmelding nr. 24 og 25):

"Alle skal ha mulighet til å drive friluftsliv som helsefremmende, trivselskapende og miljøvennlig aktivitet i nærmiljøet og i naturen for øvrig."

De to nøkkeltallene er definert slik:

- Rekreasjonsarealer (nøkkeltall 2RM4-1): *"Andel av boliger, skoler og barnehager som har trygg tilgang på leke- og rekreasjonsareal (minst 5 dekar) i en avstand på 200 meter"*.
- Nærturterreng (nøkkeltall 2RM4-2): *"Andel av boliger, skoler og barnehager som har tilgang på nærturterreng (større enn 200 dekar) i en avstand på 500 meter"*.

2 Problemstilling

Statistisk sentralbyrå har utarbeidet metodikk for en årlig beregning av disse to nøkkeltall på landsbasis (Engelien & Schøning, 2001). Åpne arealer med potensial for rekreasjon og nærturterreng avgrenses med bruk av eksisterende landsdekkende registre og kartdatabaser.

Problemstillingen som forsøkes besvart i denne rapporten er å undersøke hva disse arealene faktisk inneholder. Fokus er spesielt rettet mot tilstedeværelse av grå arealer, og det gis forslag til metodikk for å korrigere arealene til å bare inneholde grøntområder.

3 Bakgrunnsinformasjon om jordobservasjon

3.1 Jordobservasjonssatellitter

Bilder fra jordobservasjonssatellitter har i økende grad blitt benyttet i forskning og overvåkning de siste 20-30 år. Spesielt har antallet og utvalget av jordobservasjonssatellitter økt betraktelig de siste 10 år. De fleste sensorer er designet litt forskjellig avhengig av hva de er ment å brukes til. Romlig, spektral og temporal oppløsning gir en grov beskrivelse av en sensor. Før en starter en analyse bør en derfor studere disse egenskapene slik at man velger et satellittbilde fra en egnet sensor. Den enkelte sensors egenskaper er avgjørende for hvor godt egnet den er til å observere et spesifikt tema.

I denne rapporten er det valgt å fokusere på bilder fra to ulike optiske sensorer (Tabell 1): Landsat 7 ETM+ (Enhanced Thematic Mapper) og SPOT 5 (Système Pour l´Observation de la Terre) HRG (Haute Résolution Géometrique), heretter referert til som Landsat ETM+ og SPOT 5 bilder. Begge sensorer gir både pankromatiske og multispektrale bilder. Pankromatiske bilder måler reflektert synlig sollys. Multispektrale bilder måler reflektert sollys i et utvalg av synlige, nærinfrarøde og midlere infrarøde bølgelengder. Landsat ETM+ sensoren observerer også termisk utstråling fra jorden slik at overflatetemperatur kan beregnes.

Tabell 1. Egenskaper for de to satellittsensorene som er brukt i analysen i rapporten. Spektral oppløsning angir antall kanaler det er observert reflektert sollys eller termisk utstråling fra jorden. Romlig oppløsning angir areal som et piksel dekker på bakken. Temporal oppløsning angir hvor ofte satellitten avbilder eksakt samme område. Opptaksbredde beskriver størrelsen på hele satellittbildet i forhold til dekning på bakken.

Satellittsensor		Spektral oppløsning (kanaler)	Romlig oppløsning (m)	Temporal oppløsning (dager)	Opptaksbredde (km)
SPOT 5	multispektral	4	10	26	60
HRG	pankromatisk	1	5	26	60
Landsat 7	multispektral	7	30	16	185
ETM+	pankromatisk	1	15	16	185

3.2 Tolkning av satellittbilder

Klassifikasjon er en vanlig måte å tolke et satellittbilde. Hvert piksel tilordnes en klasse basert på en statistisk eller en fysisk modell. Grunnlaget for disse modellene er at hver enkelt overflate (vann, vegetasjon, snø etc.) karakteriseres ved en spektral signatur, dvs. dens evne til å reflektere sollys i de enkelte bølgelengder. I de bølgelengdene hvor to overflater har helt forskjellig spektral signatur kan en best skille overflatene fra hverandre. Klassifikasjon av et piksel gjøres i det tilfellet utelukkende ved hjelp av *spektral* informasjon (pikselvis klassifikasjon). I tillegg kan man utnytte informasjon om et piksels *naboskapsrelasjoner* til å forbedre klassifikasjonen (kontekstuell klassifikasjon). Med naboskap menes piksler som ligger ved siden av eller i nærheten av pikselet som skal klassifiseres. Videre kan man bruke naboskapsrelasjoner til å definere grupper av piksler med samme spektrale egenskaper. Disse gruppene omgjøres til objekter i form av polygoner. Hvert polygons spektrale egenskap kan så benyttes til å klassifisere enkeltobjektene (objektvis klassifikasjon).

4 Metodikk

En vurdering av kostnader ved innkjøp av satellittbilder i forhold til den romlige oppløsningen i hvert bilde var grunnlag for valg av datasett og metodikk. En generell oppfatning er at man oppnår bedre resultater dess bedre den romlige oppløsningen er. Dette er ikke nødvendigvis tilfelle ettersom den spektrale informasjonen er avgjørende for identifikasjon av klasser. Satellittbilder er også dyrere jo høyere romlig oppløsning de har. To tilnærmingsmåter og datasett er valgt:

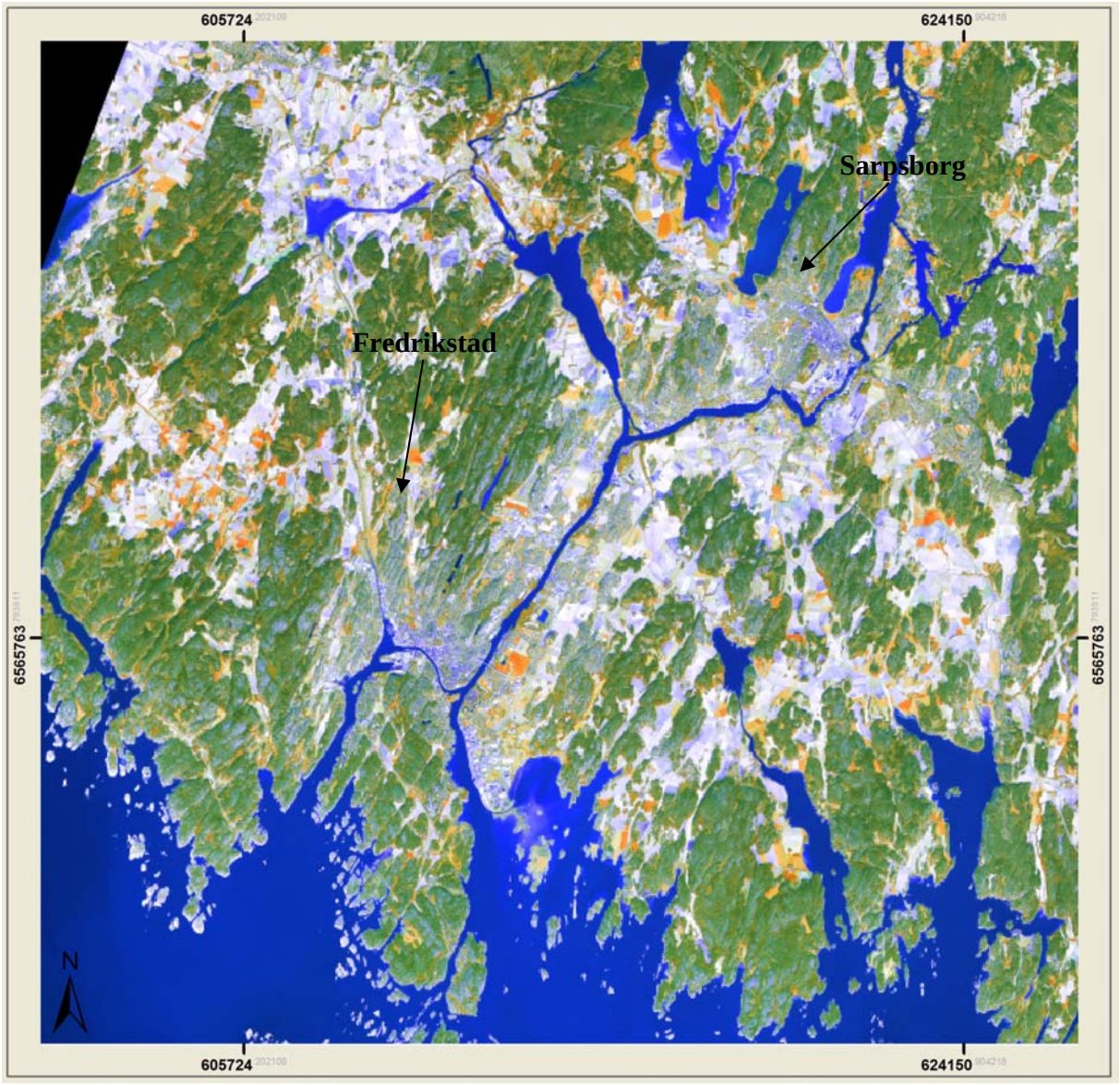
1) Objektsegmentering og klassifikasjon av SPOT 5 satellittbilde.
2) Klassifikasjon av Landsat ETM+ satellittbilder med regresjonstreanalyse.

I forhold til SPOT 5 som har inntil 2.5 m romlig oppløsning regnes Landsat ETM+ som et lavere oppløsnings satellittbilde. SPOT 5 er også svært mye dyrere i innkjøp enn Landsat ETM+. Valg av datasett og metodikk gir mulighet til å undersøke om man oppnår like gode resultater til to forskjellige kostnader.

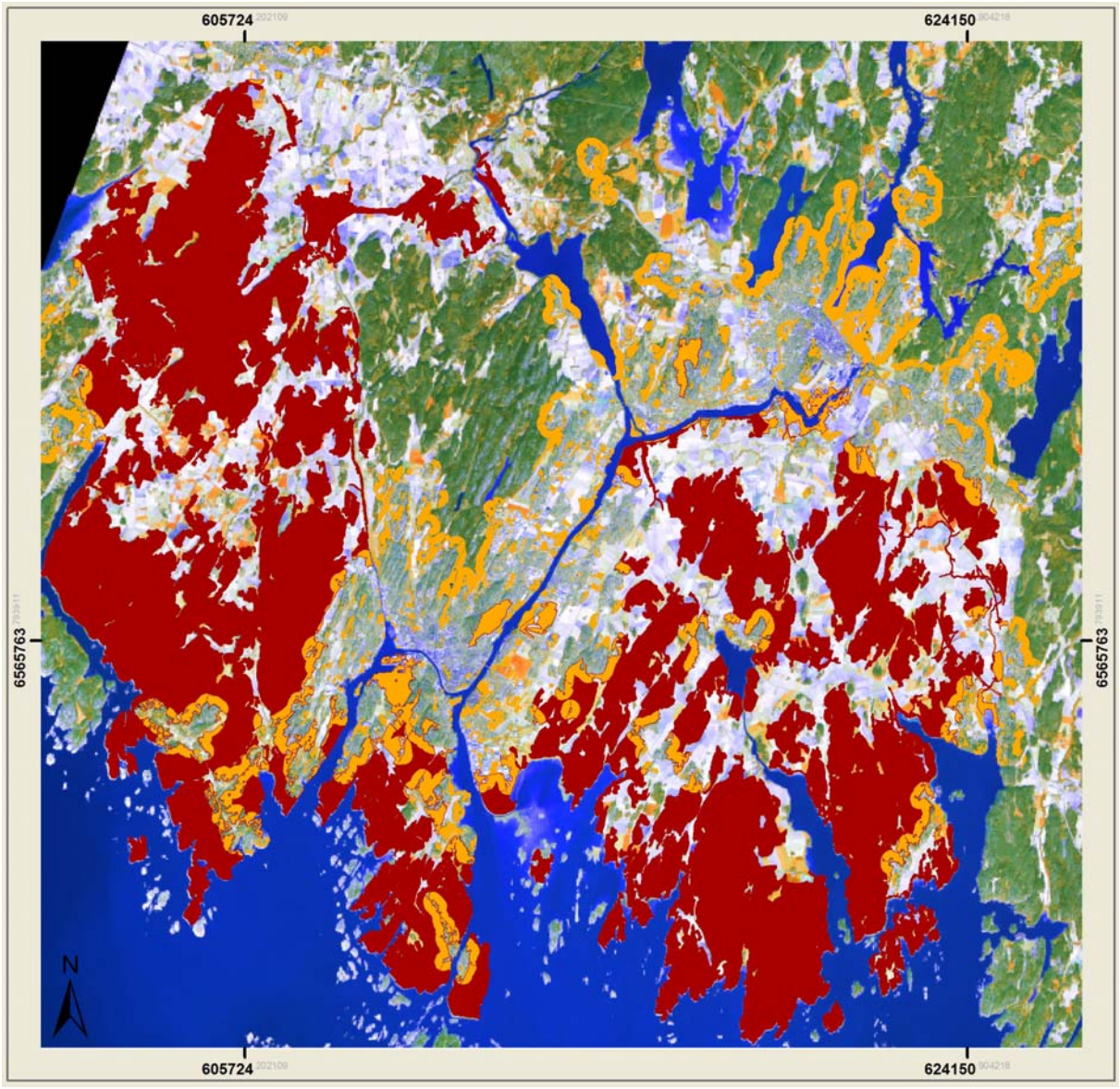
Metode 1 er en studie som viser arealtypene (grønt areal, grått areal, vann) innenfor kartlagte områder, mens metode 2 resulterer i et grovt kart som viser konsentrasjonen av grått areal (tettstedskart). Begge metoder vil gi et kvantitativt estimat på andel grått areal innenfor kartlagte nærturterreng og rekreasjonsarealer. Metode 2 gir et mye grovere anslag enn metode 1, men er enklere og raskere å gjennomføre for en uerfaren person.

4.1 Forsøksområde

Et område omkring Fredrikstad og Sarpsborg kommuner innenfor Østfold fylke ble valgt som forsøksområde i de to studiene (Figur 1). Rekreasjonsarealer og nærturterreng er vist i henholdsvis oransje og rødt i Figur 2.



Figur 1. Forsøksområdet for metode 1 og 2 er de kartlagte rekreasjonsarealer og nærturterreng innenfor bildeutsnittet omkring Fredrikstad og Sarpsborg. Bildet er et utsnitt fra et SPOT 5 satellittbilde fra 9. september 2002. © Statens kartverk 2003 SPOT(r) © CNES 2002 © Metria 2002.



Figur 2. Rekreasjonsarealer og nærturterreng for 2002 er vist i henholdsvis oransje og rødt. Områdene er kartlagt av Statistisk sentralbyrå (Engelien & Schøning, 2001). Bildeutsnittet er det samme som i Figur 1. © Statens kartverk 2003 SPOT(r) © CNES 2002 © Metria 2002.

4.2 Metode 1: Objektsegmentering og klassifikasjon av SPOT 5 bilde

4.2.1 Datasett

Datasettet som ble valgt til dette eksperimentet var:

- En satellittscene fra 9. september 2002 fra SPOT 5 satellitten:

    1. Multispektralt bilde:
        - tre bånd med 10 m romlig oppløsning (grønt, rødt og nærinfrarødt).
        - et bånd med 20 m romlig oppløsning (midlere infrarødt).
    2. Pankromatisk bilde:
        - 1 bånd med 5 m romlig oppløsning (synlig lys).
    3. Fusjonert bilde med 2,5 m romlig oppløsning, laget fra det multispektrale og det pankromatiske bildet.
    4. Fargeforbedret bilde i 2,5 m romlig oppløsning, levert av Miljøenheten i Arendal, Statens kartverk.

- Flybilder (ortofoto) over Fredrikstad kommune med scanningsoppløsning 0,2 m per piksel.
- Datasett fra Statistisk sentralbyrå over kartlagte rekreasjonsarealer og nærturterreng. I rapporten refereres dette datasettet til som SSB's polygoner for rekreasjon og turterreng.

4.2.2 Eksperimenter

Analysen ble gjort i flere trinn med bruk av eCognition programvare (Baatz et al., 2002). eCognition er et objektorientert klassifikasjonssverktøy. Det inneholder moduler for både bildesegmentering og klassifikasjon av de identifiserte segmentene. Et slikt verktøy er hensiktsmessig ved bruk av høyoppløselige bilder hvor en ikke ønsker å analysere enkeltpiksler, men i stedet inndeler bildet i objekter (større eller mindre sammenhengende arealer). Eksempelvis vil en vei, en parkeringsplass eller en stor bygning bli representert med et objekt og ikke mange enkelt piksler.

Her gis et forslag til metodikk for hvordan SSB's polygoner for rekreasjon og turterreng kan korrigeres til ikke å inneholde grått areal. Metodikken er også illustrert i avsnitt 4.2.3.2. Trinn 1 er

utført i Erdas Imagine bildebehandlingsprogramvare (Erdas Field Guide, 2002). Trinn 2-5 er utført i eCognition programvare. Trinn 6-7 er utført i ArcGIS (Esri, 2002) som er et geografisk informasjonssystem og Erdas Imagine. Analysen besto av 7 trinn:

1. Utklipning av SSB's polygoner for rekreasjon og turterreng fra SPOT 5 scenen. Bare arealer innenfor disse polygonene ble analysert videre.
2. Segmentering av SPOT 5 scenen innenfor SSB's polygoner for rekreasjon og turterreng. Et segment er et objekt (polygon) som innehar spektrale egenskaper, formegenskaper og naboskapsrelasjoner. Eksperimentering viste at en inndeling i relativt små segmenter var mest hensiktsmessig. Parametre som ble brukt for segmenteringen: Scale parameter 10,0; shape 0,2; color 0,8; smoothness 0,9 og compactness 0,1.
3. Enkeltsegmentene ble initielt klassifisert basert på en nærmeste nabo klassifikasjonsalgoritme. Dette er en styrt klassifikasjonsmetode hvor brukeren må spesifisere både antall og typer klasser, i tillegg til å identifisere treningsområder for hver spesifiserte klasse. I denne analysen ble treningsområder for de enkelte klassene identifisert ved å studere flybilder.
4. Objektegenskaper ble utnyttet til å forbedre den initielle klassifikasjonen. Objektegenskaper inkluderer både form og spektrale egenskaper. Naborelasjoner mellom objekter er også inkludert. Formegenskaper kan utnyttes til å skille ut objekter med relativt lik spektral signatur, men ulik form. Tilsvarende kan naborelasjoner benyttes til å skille objekter med lik spektral signatur, men som opptrer med spesifikke naboer. Eksempelvis var det i denne analysen problemer med å skille grått areal (veier, parkeringsplasser, bygninger) fra berg-i-dagen basert på spektrale egenskaper. Ved å inkludere en betingelse på at berg-i-dagen alltid har skog som naboobjekter kunne man relativt bra separere disse to klasser.
5. Sammenslåing av små tilgrensende polygoner til større polygoner av samme klasse. På denne måten ble små objekter til større sammenhengende enheter. Eksempelvis ble en gresslette til et polygon i stedet for å bestå av mange mindre polygoner.
6. Statistikk på arealtyper innenfor SSB's polygoner for rekreasjon og turterreng ble avledet.
7. Grått areal ble fjernet fra de opprinnelige SSB's polygoner for rekreasjon og turterreng. Nye kart ble generert.

4.2.3 Resultater

Resultater fra analysen presenteres først med å vise noen eksempler på gråe arealer som er kartlagt som rekreasjonsareal eller nærturterreng. Videre følger et avsnitt med illustrasjon av hvordan SSB's polygoner for kartlagte rekreasjonsareal og nærturterreng ble korrigert. Siste avsnitt inkluderer statistikk over arealtyper innenfor testområdet.

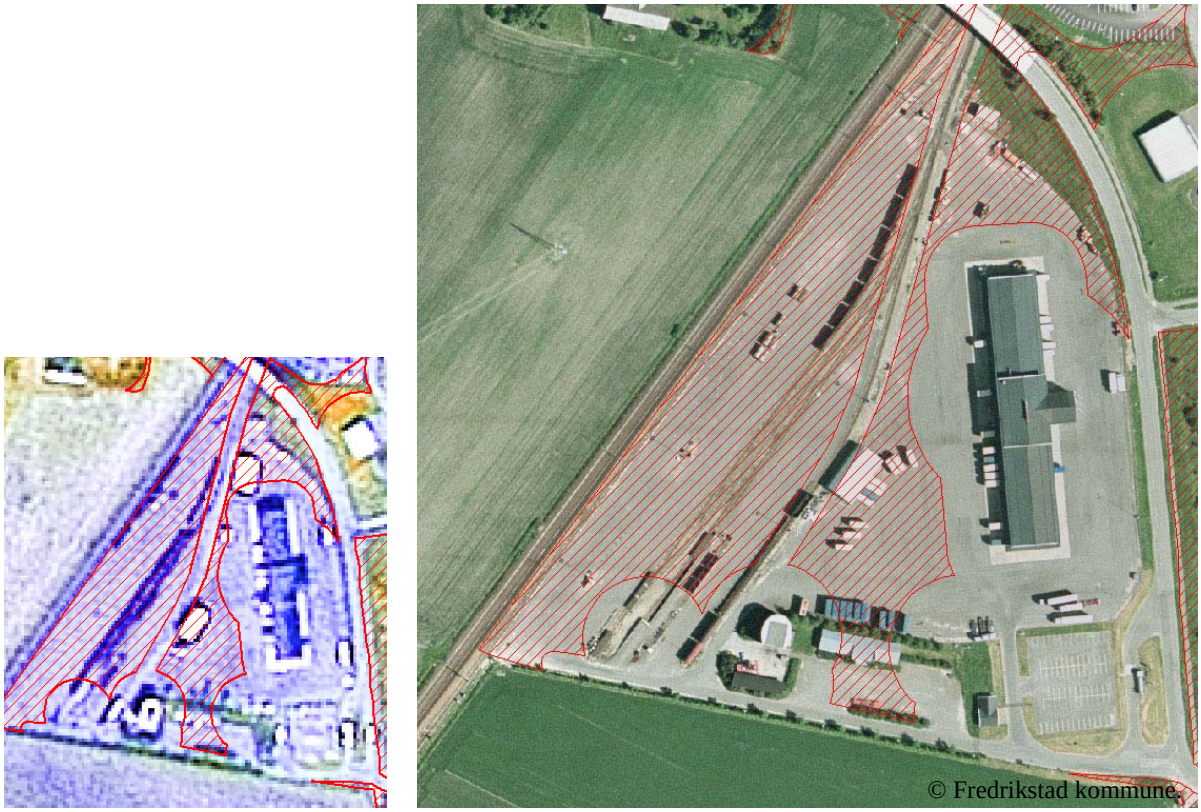
4.2.3.1 Eksempler på feilkartlagte arealer

Her presenteres eksempler på grått areal som er observert innenfor SSB's kartlagte rekreasjonsareal og nærturterreng, og som typisk ikke er friluftsområder (grøntområder eller vann) (Figur 3-6):

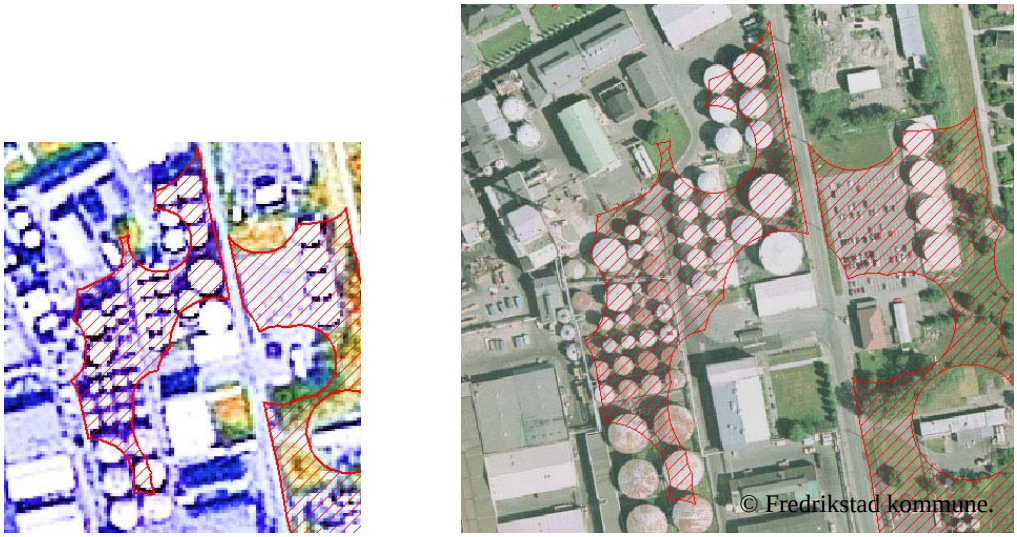
- Parkeringsplasser
- Industribygg
- Boligområder

Grunnen til at denne type arealer finnes innenfor de kartlagte rekreasjons- og nærturterreng er mest sannsynlig at de ikke er registrert i de landsdekkende registre og kartdatabaser som SSB har benyttet i kartleggingen. Det kan være en tidsforsinkelse mellom byggetidspunkt og registrering i databasene. De kartlagte arealene er beregnet for 2002. Satellittbildet som er benyttet i denne analysen er også fra 2002.

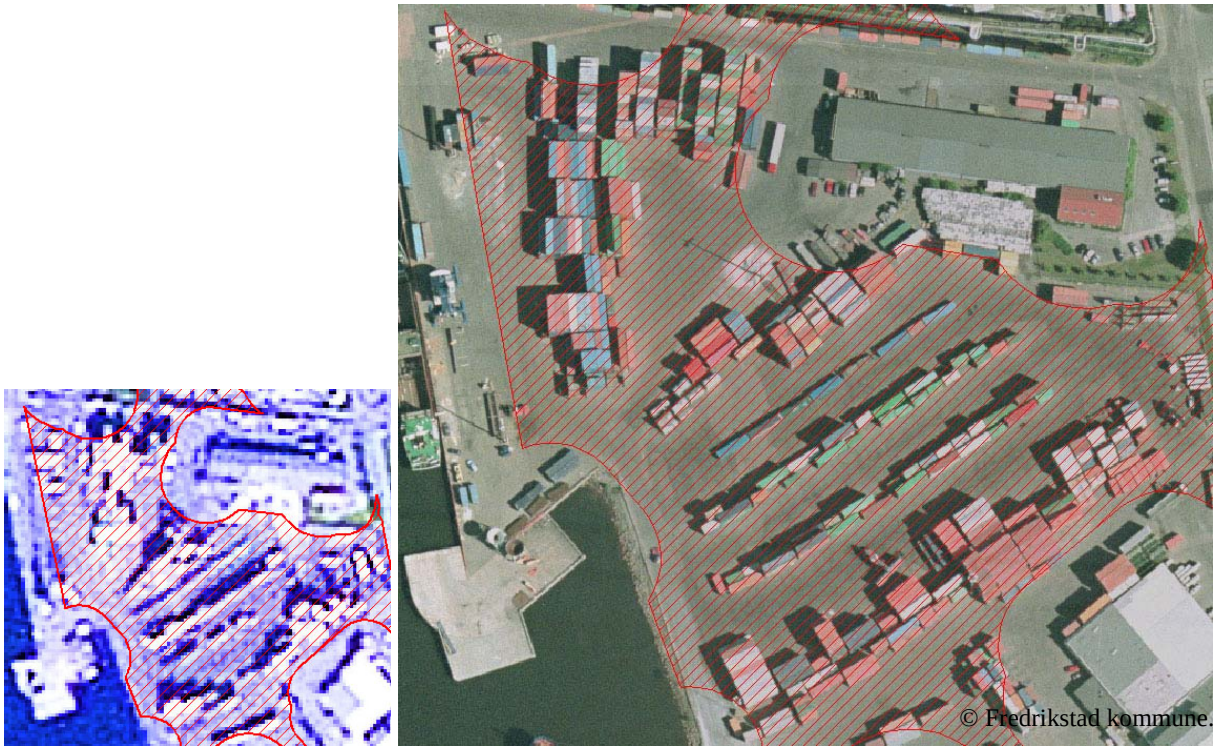
Det må presiseres at de kartlagte arealene generelt er korrekt kartlagt som friluftsområder. Feilkartlegging ser ut til å opptre hyppigst i bynære områder, dvs. oftere innenfor rekreasjonsarealene enn i kartlagt nærturterreng.



Figur 3. Parkeringsplasser kartlagt som rekreasjonsarealer. SPOT 5 bilde til venstre, og flybilde til høyre. Skravert areal er kartlagt rekreasjonsareal.



Figur 4. Industribygg kartlagt som rekreasjonsarealer. SPOT 5 bilde til venstre, og flybilde til høyre. Skravert areal er kartlagt rekreasjonsareal.



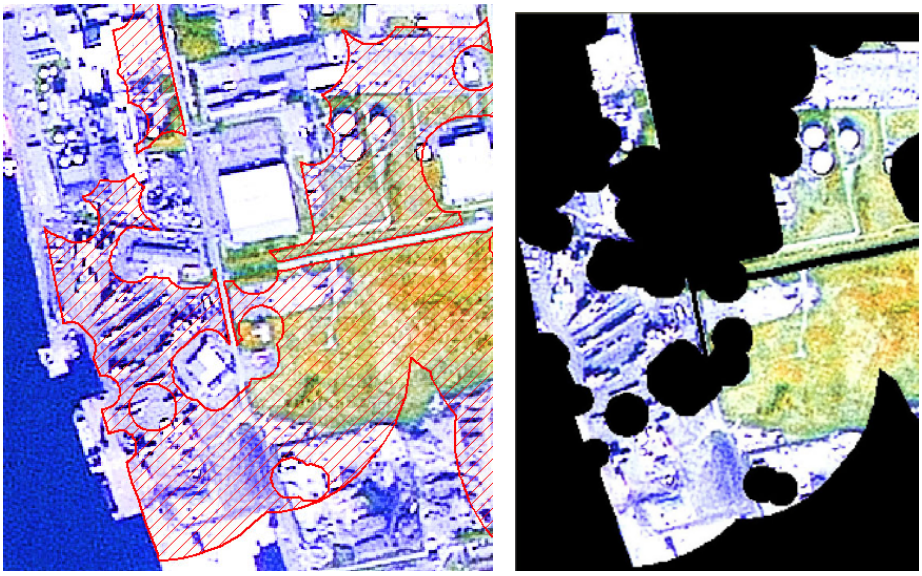
Figur 5. Industriområde kartlagt som rekreasjonsarealer. SPOT 5 bilde til venstre, og flybilde til høyre. Skravert areal er kartlagt rekreasjonsareal.



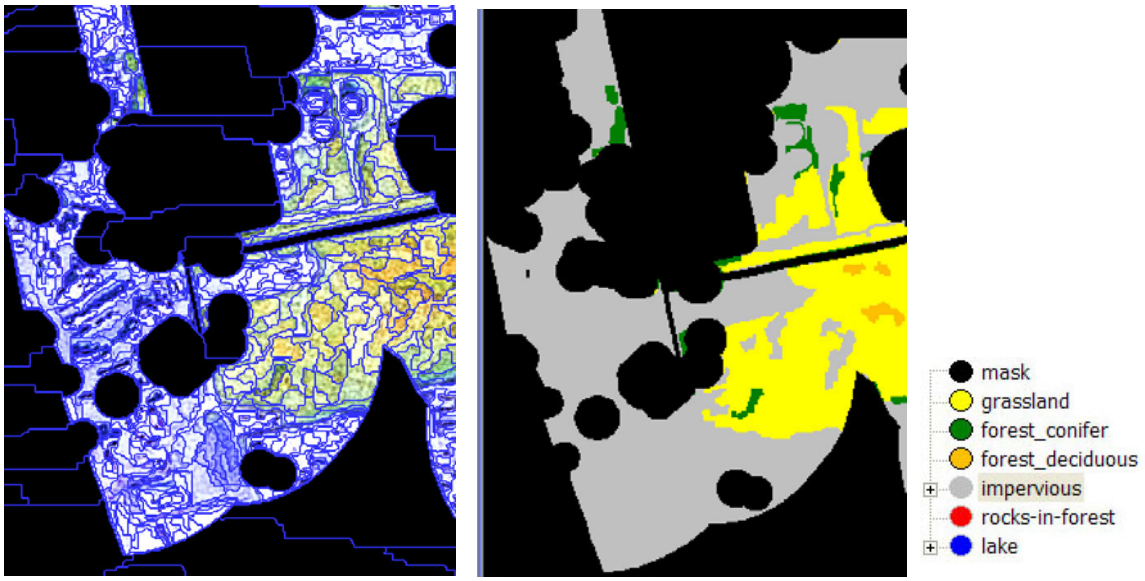
Figur 6. Boligområde kartlagt som nærturterreng. SPOT 5 bilde til venstre, og flybilde til høyre. Skravert areal er kartlagt nærturterreng.

4.2.3.2 Korreksjon av feilkartlagte arealer

Her illustreres metodikken som ble benyttet til å identifisere og videre fjerne grått areal som er feilkartlagt til rekreasjonsareal eller nærturterreng (figurene 7-9). Utsnittet til illustrasjonen dekker et industriområde sør for Fredrikstad by.



Figur 7. Trinn 1: Utklipning av SSB's kartlagte rekreasjonsareal og nærturterreng fra SPOT 5 bildet. © Statens kartverk 2003 SPOT(r) © CNES 2002 © Metria 2002.



Figur 8. Trinn 2: SPOT 5 bildet segmenteres til små enkeltobjekter (venstre bilde). Trinn 3 - 5: Enkeltobjektene klassifiseres, og senere slås enkeltobjektene sammen til større sammenhengende objekter (høyre bilde). © Statens kartverk 2003 SPOT(r) © CNES 2002 © Metria 2002.



Figur 9. Skravert areal viser områder kartlagt som grått areal. Gult viser områder som skal beholdes som rekreasjonsareal eller nærturterreng. Skravert areal utgjør polygoner som fjernes fra opprinnelig kartlagt areal basert på SSB's metode. Korrigert kart (gult areal) kan lagres som en ny ArcView Shape fil.

4.2.3.3 Statistikk for rekreasjonsarealer og nærturterreng

Resultater fra analysen av arealtyper innenfor SSB's kartlagte rekreasjonsareal og nærturterreng vises i tabellform (Tabell 2 og Tabell 3). Klassifikasjonen fokuserer på kartlegging av grått areal, mens kartlegging av andre arealtyper er blitt mindre prioritert. Relativt grove arealtypeklasser ble dermed valgt for klassifikasjon. Det er brukt 5 klasser:

1. grått areal
2. gress/eng (plen i parker, åpen engmark)
3. skog (løvskog og barskog)
4. berg-i-dagen/glissen skog
5. vann

Klassen berg-i-dagen/glissen skog inneholder både svaberg ved kysten og glissen skog med mye berg-i-dagen.

Statistikken viser at feilkartlagt grått areal er 13% innenfor rekreasjonsarealene, mens det er 2% innenfor nærturterrenget. Dette betyr at SSB's polygoner for rekreasjon og turterreng ble korrigert med henholdsvis 13% og 2%.

Tabell 2. Statistikk over arealtyper innenfor SSB's kartlagte *rekreasjonsarealer* i testområdet. Arealer er oppgitt i km^2 og prosentandel av rekreasjonsarealene innenfor testområdet.

Arealtype	Areal (km^2)	Areal (%)
Grått areal	5,3	**13**
Gress/eng	5,6	14
Skog	25,6	64
Berg-i-dagen/glissen skog	1,0	3
Vann	3,0	7
Totalt rekreasjonsareal	40,4	100

Tabell 3. Statistikk over arealtyper innenfor SSB's kartlagte *nærturterreng* i testområdet. Arealer er oppgitt i km^2 og prosentandel av nærturterrenget innenfor testområdet.

Arealtype	Areal (km^2)	Areal (%)
Grått areal	4,1	**2**
Gress/eng	4,4	3
Skog	107,7	66
Berg-i-dagen/glissen skog	46,8	28
Vann	1,2	1
Totalt nærturterreng	164,2	100

4.2.3.4 Kontroll av klassifikasjonsresultatet

Metodikk

Flybilder over Fredrikstad kommune ble brukt til å innhente bakkesannhet for å kontrollere klassifikasjonsresultatene. Feltobservasjoner ble ikke utført. Metoden som ble valgt for å lage kontrollpunkter for de enkelte klassene var å generere totalt 600 punkter med en tilfeldig fordeling innenfor SSB's rekreasjonsarealer og turterreng. For rekreasjonsarealene ble 200 punkter fordelt tilfeldig innenfor alle klasser, og 100 punkter ble fordelt tilfeldig bare innenfor klassen med gråe arealer. Hensikten var å fokusere spesielt på klassifikasjonsnøyaktighet av gråe arealer. Tilsvarende prosedyre ble fulgt for SSB's polygoner for turterreng. Hvert enkelt punkt ble lagt

over flybildene og manuelt tilskrevet dens sanne klasse basert på visuell tolkning. Resultatet av denne kontrollen er presentert i to feilmatriser, henholdvis en for rekreasjonsarealer og en for nærturterreng (Tabell 4 og Tabell 5). En god del punkter var lokalisert utenfor flybildedekningen, slik at antall punkter per matrise er færre enn 300.

En feilmatrise viser hvor godt et klassifikasjonsresultat er, både totalt og for enkelt kategorier. Her sammenlignes klassifikasjonen mot bakkesannhetsdata, oppsummert for alle kontrollpunktene. Tabellen er kvadratisk. De anvendte klassene vises både i radene og kolonnene. Kolonnene representerer klassifiserte kategorier, mens radene representerer bakkesannhet. Optimalt bør alle punkter befinne seg langs diagonalen i tabellen. Alle punkter som havner utenfor diagonalen er feilklassifiserte punkter. Tre parametre er vanlige å beregne fra en feilmatrise (Lillesand & Kiefer, 1994):

- Total klassifikasjonsnøykatighet.
- Producers's accuracy.
- User's accuracy.

Total klassifikasjonsnøyaktighet beregnes ved å summere antall riktig klassifiserte punkter (punkter langs diagonalen) og dividere på totalt antall kontrollpunkter. Producers's accuracy er et mål på hvor godt kontrollpunktene fra hver bakkesannhetskategori ble klassifisert til sin sanne klasse. Eks: Hvor sannsynlig er det at et punkt på en parkeringsplass faktisk ble klassifisert til grått areal? Dette regnes ut ved å dividere antall korrekt klassifiserte punkter i hver kategori med antall kontrollpunkter som fantes i den kategorien (summerte punkter i hver kolonne). User's accuracy angir sannsynligheten for at et punkt klassifisert til en gitt klasse faktisk er denne klassen. Eks: Hvor sannsynlig er det at et punkt klassifisert til grått areal faktisk også er grått areal (parkeringsplass, vei, bygning)? Dette regnes ut ved å dividere antall korrekt klassifiserte punkter i hver kategori med antall punkter klassifisert til denne kategorien (summerte punkter i hver rad). For en god statistikk bør en ha minst 50-100 punkter per klasse. Det kravet er det bare kategoriene grått areal og skog som oppfyller for rekreasjonsarealer, mens for nærturterreng er det de tre kategoriene grått areal, skog og berg-i-dagen/glissen skog som oppfyller kravene. Av den grunn er det disse klassene som omtales videre.

Total klassifikasjonsnøyaktighet

Total klassifikasjonsnøyaktighet ble 64% og 66% for henholdvis rekreasjonsarealer og nærturterreng.

Klassen grått areal

For rekreasjonsarealer får kategorien grått areal 95% producer's accuracy og 49% user's accuracy. For nærturterreng er tallene i samme størrelsesorden, henholdsvis 95% og 45%. Det er altså svært sannsynlig at sanne grå arealer faktisk er klassifisert til grått areal. Problemet er at kategorien er blitt for stor, og dermed inneholder andre klasser.

Berg-i-dagen/glissen skog er den dominerende kategorien som er feilklassifisert til grått areal. 42-45% av klassene grått areal inneholder altså berg-i-dagen/glissen skog. Det er særlig svaberg langs kysten og svært glissen skog med berg-i-dagen som utgjør et problem. Ved å ekskludere en sone langs kysten fra forsøksområdet blir resultatene betydelig bedre (se avsnitt 4.2.3.5). Det er altså ingen tilfeldige feil, men systematiske feil knyttet til spektral overlapp mellom svaberg, berg-i-dagen og gråe menneskeskapte arealer (asfaltveier, grusveier, parkeringsplasser). Feilklassifikasjon av glissen skog med mye berg-i-dagen ble forsøkt løst ved å bruke objektrelasjoner, i tillegg til de spektrale egenskapene, til å separere klassene. Eksempelvis var områder med berg-i-dagen oftere totalt omgitt av skog enn veier og parkeringsplasser. Denne informasjonen ble brukt til å sette begrensninger på naboskap mellom objekter i programmet eCognition.

Klassen berg-i-dagen/glissen skog

For nærturterreng er både user's accuracy og producer's accuracy blitt svært lav for klassen berg-i-dagen/glissen skog. Ettersom denne klassen inneholdt både svaberg og blanding av svaberg og skog (glissen skog) var overgangen til klassen skog ikke helt klar. 42% av klassen var i realiteten skog. Denne problemstillingen var underordnet hovedproblemstillingen som var å identifisere grått areal. Forbedringer er derfor ikke forsøkt gjort.

Klassen skog

Skog oppnår desidert best klassifikasjonsnøyaktighet av alle klassene. For rekreasjonsarealer er user's accuracy 83% og producer's accuracy 94%. For skog innenfor rekreasjonsarealene er det altså svært sannsynlig at all skog virkelig er klassifisert til skog. Klassen er noe for stor, og inneholder noe feilklassifisert gressområder og glissen skog/berg-i-dagen klasse.

For nærturterreng er user's accuracy og producer's accuracy henholdsvis 95% og 83%. Dette betyr at klassen bare minimalt inneholder andre klasser, men at noen faktiske skogområder er klassifisert til andre klasser, og da nesten uten unntak berg-i-dagen/glissen skog klassen. Årsaken er en diffus subjektiv vurdering ved flyfototolkningen av hva som er glissen skog og hva som er skog.

Tabell 4. Feilmatrise for rekreasjonsarealer. Tallene i den øverste tabellen angir antall punkter.

		Bakkesannhet fra flyfoto					
		Gress	Skog	Grått	Berg-i-dagen	Vann	Antall punkter
Klassifiserte data	Gress	**11**	0	2	1	0	14
	Skog	7	**63**	0	6	0	76
	Grått	6	0	**36**	31	0	73
	Berg-i-dagen	0	0	0	**1**	0	1
	Vann	0	4	0	0	**6**	10
	Antall punkter	24	67	38	39	6	174

Klasse	User's accuracy	Producer's accuracy
Gress	79%	46%
Skog	83%	94%
Grått	49%	95%
Berg-i-dagen	100%	3%
Vann	60%	100%

Total klassifikasjonsnøyaktighet: (111/174)*100% = 64%

Tabell 5. Feilmatrise for nærturterreng. Tallene i den øverste tabellen angir antall punkter.

Klassifiserte data	Bakkesannhet fra flyfoto					
	Gress	Skog	Grått	Berg-i-dagen	Vann	Antall punkter
Gress	**1**	1	0	0	0	2
Skog	3	**99**	0	2	0	104
Grått	4	0	**38**	39	4	85
Berg-i-dagen	5	20	2	**22**	0	49
Vann	0	0	0	0	**2**	2
Antall punkter	13	120	40	63	6	242

Klasse	User's accuracy	Producer's accuracy
Gress	50%	8%
Skog	95%	83%
Grått	45%	95%
Berg-i-dagen	45%	35%
Vann	100%	33%

Total klassifikasjonsnøyaktighet: (162/242)*100% = 66%

4.2.3.5 Forbedring av klassifikasjonen

Det ble gjort et eksperiment på å syneliggjøre at svaberg langs kysten ble systematisk feilklassifisert til grått areal. En sone innenfor en avstand på 100 meter fra kystlinjen ble fjernet fra polygonene for rekreasjonsareal og nærturterreng. Hensikten var å utlede feilmatriser for de samme arealene som før, men at kontrollpunkter innenfor denne sonen ble ekskludert. To feilmatriser ble beregnet pånytt for gjenværende punkter, henholdvis en for rekreasjonsarealer og en for nærturterreng (Tabell 6 og Tabell 7).

Resultatet for rekreasjonsarealene er en forbedring av total klassifikasjonsnøyaktighet fra 64% til 72%. For nærturterreng forbedres total klassifikasjonsnøyaktighet fra 66% til 77%.

Som forventet økte klassifikasjonsnøyaktigheten for temaet grått areal betraktelig, men mest for nærturterreng fordi mer kystlinjer befinner seg innenfor nærturterreng sammenlignet med

rekreasjonsareal. For nærturterreng økte user's accuracy fra 45% til 70%. Det betyr at det er en større sannsynlighet for at klassen grått areal faktisk inneholder grått areal. Fortsatt resterende feilklassifikasjon for klassen grått areal er som tidligere nevnt svært glissen skog med berg-i-dagen. Av 55 kontrollpunkter som befant seg innenfor buffersonen langs kysten, var 30 punkter på svaberg. Disse var feilklassifisert til grått areal, og gjorde den opprinnelige kontrollanalysen skjev. Tilsvarende for buffersonen innenfor rekreasjonsarealene var det 11 av 22 kontrollpunkter som befant seg på svaberg, og som var feilklassifisert til grått areal.

Denne analysen viser hvordan systematiske feil kan fjernes slik at det totale klassifikasjonsresultatet blir forbedret. Samtidig må man være klar over at selve buffersonen langs kysten bør analyseres på annen måte, ettersom den i dette tilfellet bare er ekskludert fra analysen.

Tabell 6. Feilmatrise for rekreasjonsarealer, ekskludert en 100 m sone langs kysten. Tallene i den øverste tabellen angir antall punkter.

		Bakkesannhet fra flyfoto					
		Gress	Skog	Grått	Berg-i-dagen	Vann	Antall punkter
Klassifiserte data	Gress	**10**	0	2	1	0	13
	Skog	5	**61**	0	6	0	72
	Grått	5	0	**31**	20	0	56
	Berg-i-dagen	0	0	0	**1**	0	1
	Vann	0	4	0	0	**6**	10
	Antall punkter	20	65	33	28	6	152

Klasse	User's accuracy	Producer's accuracy
Gress	77	50
Skog	85	94
Grått	55	94
Berg-i-dagen	100	4
Vann	60	100

Total klassifikasjonsnøyaktighet: (109/152)*100% = 72%

Tabell 7. Feilmatrise for nærturterreng, ekskludert en 100 m sone langs kysten. Tallene i den øverste tabellen angir antall punkter.

Klassifiserte data	Bakkesannhet fra flyfoto					
	Gress	Skog	Grått	Berg-i-dagen	Vann	Antall punkter
Gress	**1**	1	0	0	0	2
Skog	3	**93**	0	2	0	98
Grått	3	0	**32**	9	2	46
Berg-i-dagen	5	18	1	**16**	0	40
Vann	0	0	0	0	**1**	1
Antall punkter	12	112	33	27	3	187

Klasse	User's accuracy	Producer's accuracy
Gress	50	8
Skog	95	83
Grått	70	97
Berg-i-dagen	40	59
Vann	100	33

Total klassifikasjonsnøyaktighet: (143/187)*100% = 77%

4.2.4 Oppsummering

Objektbasert klassifikasjon av SPOT 5 satellittbilde ble utført innenfor et forsøksområde omkring Fredrikstad og Sarpsborg. Bare områder dekket av SSB's polygoner for rekreasjonsareal og nærturterreng ble undersøkt.

Klassifikasjonen viste at 13% av rekreasjonsarealene og 2% av nærturterrenget inneholdt grått areal. En kontroll på klassifikasjonen viste at disse tallene er altfor høye fordi ca. 42-45% av det kartlagte gråe arealet faktisk var: 1) svaberg langs kysten og 2) glissen skog med mye berg-i-dagen. Klassifikasjonen ble forsøkt forbedret ved å fjerne en 100 m sone langs kysten. Eksklusjon

av denne kystsonen førte til at den totale klassifikasjonsnøyaktigheten forbedret seg fra 66% til 77% for nærturterreng, og fra 64% til 72% for rekreasjonsarealer.

Selv om absoluttverdiene for innhold av grått areal er feil, er det likevel grunn til å tro at den relative forskjellen mht. innhold av grått areal innenfor rekreasjonsarealene og nærturterrenget er riktig. Korreksjonen var altså størst for rekreasjonsarealene. At det er større innslag av grått areal i rekreasjonsarealene kan skyldes at rekreasjonsarealene er lokalisert inni og omkring byer og tettsteder. Det er i disse områdene betydelig tettere med bygninger, veier, industriområder og parkeringsplasser enn det er i områder omkring nærturterrenget. Det foregår mer endringer knyttet til bygningsaktiviteter nær rekreasjonsarelene sammenlignet med nærturterreng. Av denne grunn kan en forvente mer feilkartlegging i rekreasjonsarealene enn i nærturterrenget. Det må også bemerkes at analysert areal for nærturterreng er omtrent fire ganger større enn for rekreasjonsareal.

4.3 Metode 2: Regresjonstreanalyse for klassifikasjon av Landsat ETM+ bilder

Metoden brukt i analysen går under betegnelsen CART ”Classification and regression tree analysis”. Dette er en statistisk klassifikasjonsmetode som finnes beskrevet i statistikk lærebøker (se f.eks. Breiman et al., 1984). Den kalles også beslutningstreanalyse og regresjonstreanalyse. Metoden er en subpixel analyse som resulterer i et klassifisert bilde hvor verdien til hvert piksel representerer en prosentvis andel av en overflate. For analysen i denne rapporten blir resultatet et kart der hvert piksel angir prosentandel grått areal av det bakkearealet den representerer. Kartet som helhet kan tolkes som en tettstedsmaske som i tillegg inneholder informasjon om tettstedskonsentrasjon.

I denne rapporten er det fulgt omtrent samme prosedyre for bruk av regresjonstreanalyse til kartlegging av grått areal som beskrevet av Yang et al. (2003). Analysen inkluderer en empirisk klassifikasjonsalgoritme som krever utstrakt bruk av treningsdata for å bygge opp beslutningstreet eller regresjonstreet. Her er det benyttet datasett for bygninger og veier fra Felles KartDatabase (FKB) for hele Fredrikstad kommune som treningsdata. Landsat ETM+ bilder er klassifisert for hele Østfold fylke basert på treningsdataene. Programvare brukt til analysen var Cubist (Rulequest research) og Erdas Imagine. Script for direkte integrasjon av Cubist i Erdas Imagine ble tilsendt fra USGS EROS Data Center (CART user guide, 2003).

4.3.1 Datasett

Datasettet som ble valgt til analysen var:

- Tre Landsat ETM+ bilder tatt opp over samme område, men på ulikt tidspunkt. Multispektralt bilde fra hvert tidspunkt inneholder 6 bånd fordelt på synlige, nærinfrarøde, midlere infrarøde og termisk infrarøde bølgelengder med 30 m romlig oppløsning per piksel. Dato for opptakene var:
    - Vårscene: 9. mai 2001.
    - Sommerscene: 5. juli 2001.
    - Høstscene: 11. oktober 1999.
- Flybilder (ortofoto) over Fredrikstad kommune med scanningsoppløsning 0,2 m per piksel.

- Utvalg av datasett fra Felles KartDatabase (FKB) for Fredrikstad kommune:
    - Vegsituasjon (grunnriss). Datasettet inkluderer bl.a. bilveier, sykkelveier og gangveier.
    - Bygninger (grunnriss).
- Digital terrengmodell i 25 m romlig oppløsning (Statens Kartverk) for Østfold fylke.
- Digitale temakart for markslag fra N50-kartserie (Statens Kartverk) for Østfold fylke.

Preprosessering av Landsat ETM+ bildene inkluderte geometrisk korreksjon, utklipning av Østfold fylke og radiometrisk kalibrering til reflektans (Irish, 2000; Markham & Barker, 1986). Videre ble det beregnet tasseled cap transformasjoner for hvert Landsat ETM+ bilde (Huang et al., 2002). Alle datasettene var i UTM (sone 32) WGS84 kartprojeksjon.

FKB-datasettet (veier og bygninger) for Fredrikstad kommune ble preprosessert til et referansekart som inneholdt en fasit på prosentandel grått areal innenfor arealer som geometrisk overlappet med Landsat ETM+ pikslene.

4.3.2 Eksperimenter

En eksperimentplan ble laget som inneholdt fem datasett (A-E) hvorav hvert etterfølgende datasett inkluderte et økende antall datalag (Tabell 8). Selve analysen med hvert enkelt datasett er skissert i Figur 10. Den består av fire trinn, hvorav det første trinnet er å avgrense datasettet til et treningsdatasett og et testdatasett. Treningsdatasettet brukes til å bygge et regresjonstre (en klassifikator), mens testdatasettet brukes til å kontrollere klassfikatoren i ettertid. Dette gjøres for de målepunktene evt. pikslene hvor en har referansekart. I dette tilfellet dekket referansekartet hele Fredrikstad kommune. Med pikseloppløsning på 30 m resulterte det i omtrent 441 500 piksler. Det ble valgt to samplingsstrategier for å danne treningsdatasett og testdatasett til klassifikasjonen:

1. Alle punkter fra Fredrikstad kommune benyttes (ca. 441 500 piksler).
2. 15 000 punkter velges med ”random generator”.

For begge strategier ble det selekterte datasettet delt inn slik at treningsdata utgjorde 80% mens testdata utgjorde 20%.

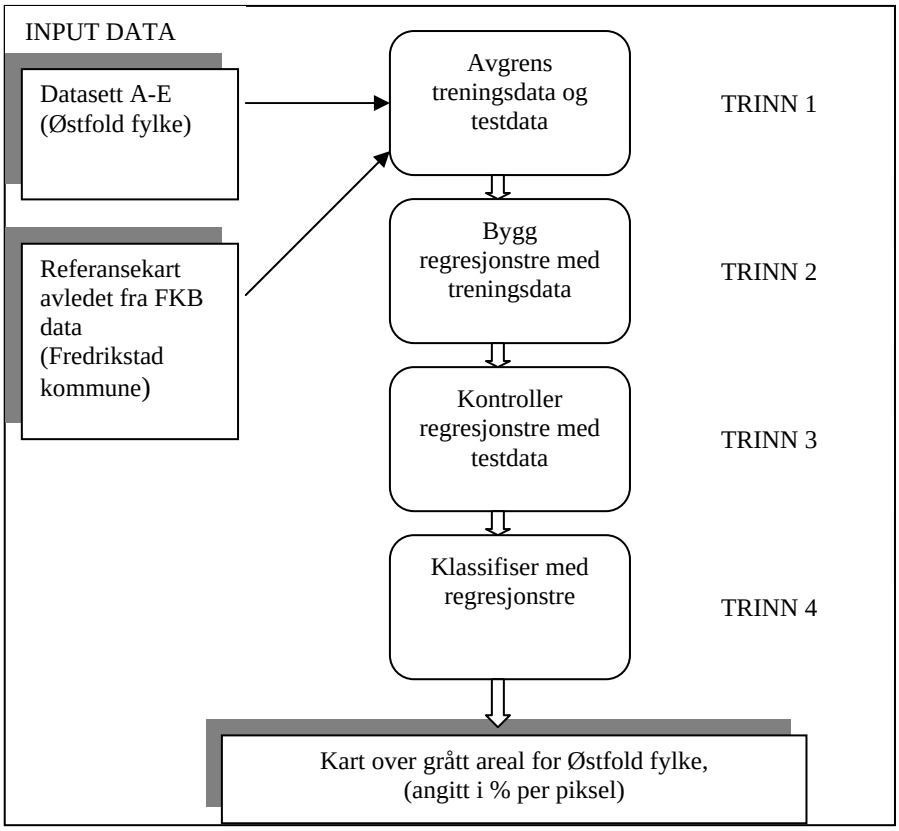
Trinn 2 i analysen bygger regresjonstreet. Et regresjonstre skal bestå av en samling regler. Programmet Cubist ble brukt til å analysere datasettet og danne regelbaserte lineære modeller. Reglene bygges opp ved hjelp av treningsdatasettet. For samplingsstrategi nr. 2 ble det satt krav om at den enkelte beslutningsregel var gyldig såfremt det fantes minst 10 tilhørende observasjoner. En slik betingelse begrenset antall beslutningsregler i regresjonstreet.

Trinn 3 er en nøyaktighetskontroll av det genererte regelsettet ved hjelp av uavhengige testdata. Regresjonstreet evalueres med statistiske parametre som gjennomsnittsfeil og korrelasjonskoeffisient.

Trinn 4 er selve klassifikasjonen hvor de genererte beslutningsreglene anvendes på områder hvor en ikke har referansedata. Dette er i realiteten en ekstrapolasjon. De nye områdene bør inneholde samme arealtyper som treningsdataene ettersom dette er en rent empirisk algoritme.

Tabell 8. Oversikt over datasettene A-D som dekket Østfold fylke. Datasett B inneholder datasett A, og datasett C inneholder datasettene A og B osv. Tallene angir antall tasseled cap bånd og temperaturbånd fra de enkelte Landsat ETM+ bildene.

Datasett	Landsat ETM+	Tasseled Cap	Temperatur	Digital terrengmodell	N50 temakart
A	Sommer	6	-	-	-
B	Vår	6	-	-	-
	Sommer	6	-		
	Høst	6	-		
C	Vår	6	1	-	-
	Sommer	6	-		
	Høst	6	1		
D	Vår	6	1	h.o.h terrenghelning	-
	Sommer	6	-		
	Høst	6	1		
E	Vår	6	1	h.o.h terrenghelning	Markslag
	Sommer	6	-		
	Høst	6	1		



Figur 10. Skisse for klassifikasjon med regresjonstreanalyse.

4.3.3 Resultater

4.3.3.1 Evaluering av regresjonstrær

Resultater fra eksperimentene oppsummeres i Tabell 9 og Tabell 10. For hvert datasett A-E er gjennomsnittfeil og korrelasjonskoeffisient oppgitt både for trenings- og testdataene. Det tilstrebes å få en lav gjennomsnittsfeil og en korrelasjonskoeffisient nærmest mulig 1.0. Av tabellene ser man at uavhengig av samplingsstrategi gir datasett E alltid gir best resultat, mens datasett A alltid gir dårligst resultat. Generelt ser man at både gjennomsnittsfeilen blir mindre og korrelasjonskoeffisienten blir større når antall datalag økes i analysen. Dette skyldes særlig at inkludering av temperaturkanal, digital terrengmodell og temakart gjør at hovedgrupper av arealtyper lettere identifiseres (for eksempel vann og skog). Det dannes også flere beslutningsregler innenfor hver hovedgruppe.

Flesteparten av reglene i regelsettene gjelder observasjoner som har 0% grått areal. Dette er naturlig ettersom store deler av Østfold fylke inneholder 0% grått areal (inkludert vann og sjø). For disse områdene er gjennomsnittsfeilen svært lav: 0,6% (Tabell 13). For områder med grått areal større enn 0%, så øker gjennomsnittsfeilen med økende innhold av grått areal (fra 7-22%). Spesielt områder med høy konsentrasjon av grått areal vil kunne være underestimert. Dette skyldes at treningsdataene hadde en skjev fordeling. Områder med lav prosentandel grått areal var godt representert mens høye prosentandeler grått areal var dårlig representert. For å forbedre klassifikasjonen av områder med høy konsentrasjon av grått areal, bør en i framtidige analyser spesifikt inkludere flere treningsdata av denne typen.

Analysen viser at det er svært viktig å få en god fordeling av andel grått areal blant treningsdataene. Det er også viktig at alle typer kombinasjoner av grått areal og andre arealtyper innen forsøksområdet blir repressentert.

Tabell 9. Evaluering av hvert enkelt regresjonstre bygget med datasett A-E. Treningsdata (345 000 piksler) og testdata (96 500 piksler) er dannet ved samplingsstrategi nr. 1, dvs. alle piksler innenfor Fredrikstad kommune er brukt. Antall beslutningsregler for hvert regresjonstre er angitt i parentes.

Datasett (regler)	EVALUERING PÅ TRENINGSDATA		EVALUERING PÅ TESTDATA	
	Gjennomsnittsfeil (%)	Korrelasjons-koeffisient	Gjennomsnittsfeil (%)	Korrelasjons-koeffisient
A (22)	2,1	0,59	2,1	0,59
B (28)	2,0	0,65	1,9	0,64
C (27)	1,8	0,70	1,8	0,70
D (30)	1,7	0,71	1,7	0,70
E (24)	1,6	0,74	1,6	0,75

Tabell 10. Evaluering av hvert enkelt regresjonstre bygget med datasett A-E. Treningsdata (15 000 piksler) og testdata (15 000 piksler) er dannet ved samplingsstrategi nr. 2, dvs. at pikslene er valgt ut med en "random generator". Antall beslutningsregler for hvert regresjonstre er angitt i parentes.

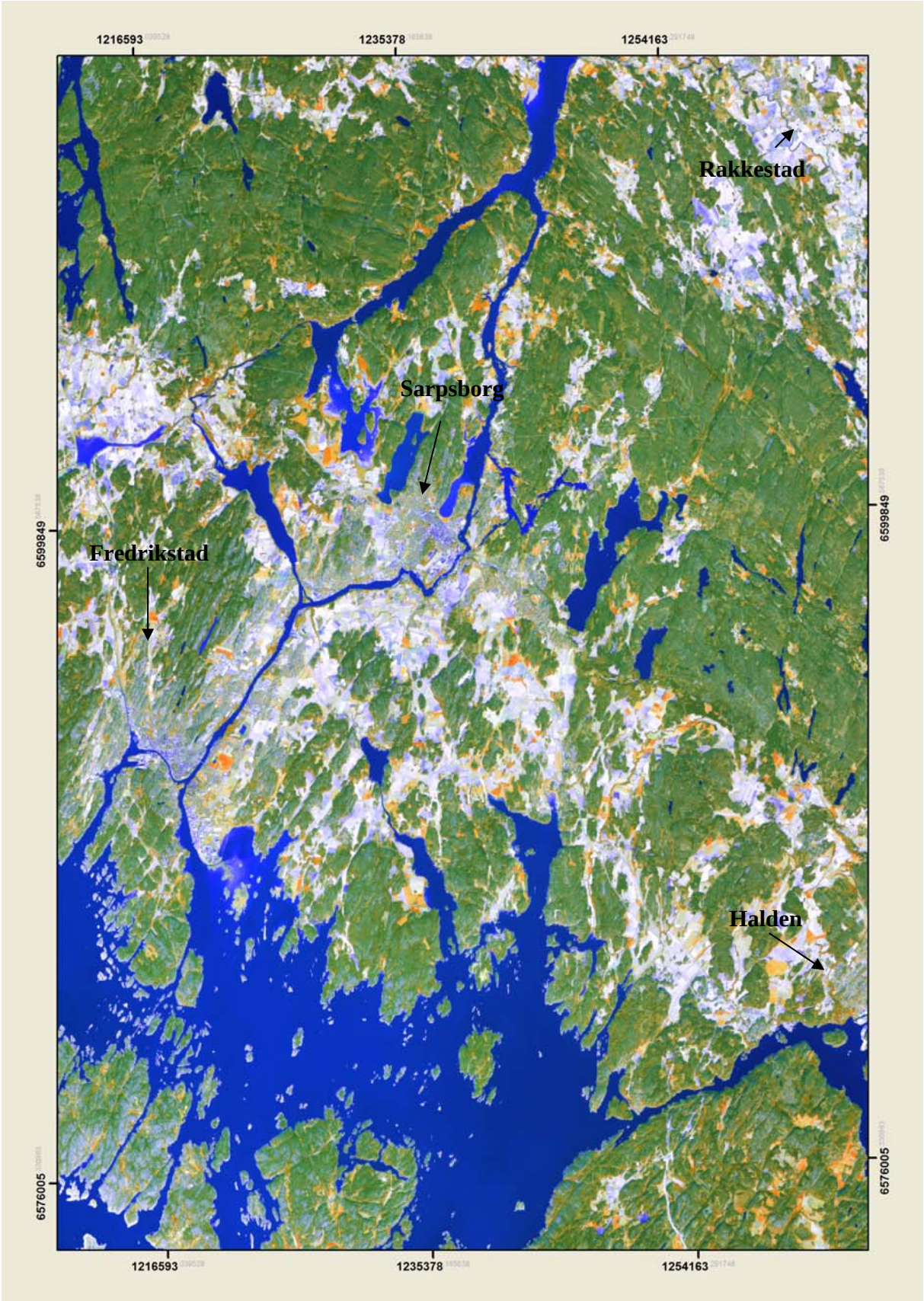
Datasett (regler)	EVALUERING PÅ TRENINGSDATA		EVALUERING PÅ TESTDATA	
	Gjennomsnittsfeil (%)	Korrelasjons-koeffisient	Gjennomsnittsfeil (%)	Korrelasjons-koeffisient
A (5)	4,6	0,69	3,0	0,45
B (5)	4,2	0,72	2,8	0,53
C (4)	3,7	0,82	2,6	0,65
D (21)	2,8	0,86	2,0	0,68
E (28)	2,6	0,87	1,8	0,71

Tabell 11. Gjennomsnittsfeil beregnet for kategorier av observert grått areal (datasett E i tabell 8). Antall datapunkter viser at de fleste observasjonene har 0% grått areal, mens det avtar sterkt med økende andel grått areal. Gjennomsnittsfeilen øker også med økende andel grått andel. Grått areal > 70% har for få datapunkter til å gi signifikante statistiske verdier.

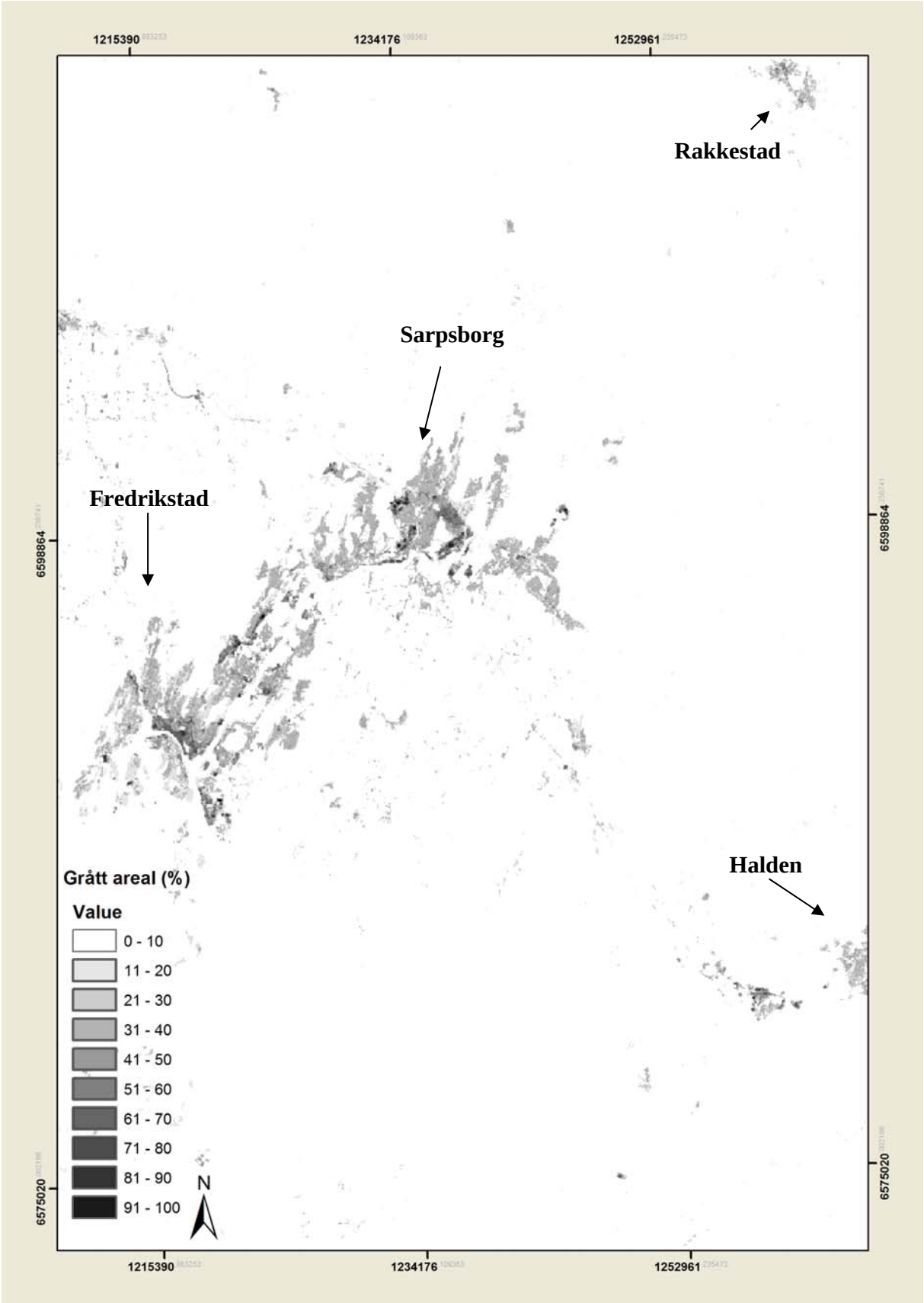
Observert grått areal (%)	Antall datapunkter	Gjennomsnittsfeil (%)
0	13079	0,6
1-10	648	7,1
11 - 20	421	12,9
21 - 30	319	14,6
31 - 40	181	14,9
41 - 50	54	18,0
51 - 60	24	22,8
61 - 70	19	13,8
71 - 80	6	27,4
81 - 90	0	-
91 – 100	4	15,5

4.3.3.2 Kart over grått areal

Kart over grått areal ble laget for hele Østfold fylke. Forskjellige utsnitt av kartene vises i Figur 11 til Figur 16. Et større område omkring Fredrikstad, Sarpsborg og Halden vises i figurene 11 og 12. For å illustrere mer detaljer av kartene vises et mindre område over Sarpsborg i figurene 13-16. Kartene er laget ved å bruke de to beste regresjonstrærne til å klassifisere hele Østfold: 1) Samplingsstrategi nr. 1 på datasett E og 2) Samplingsstrategi nr. 2 på datasett E.



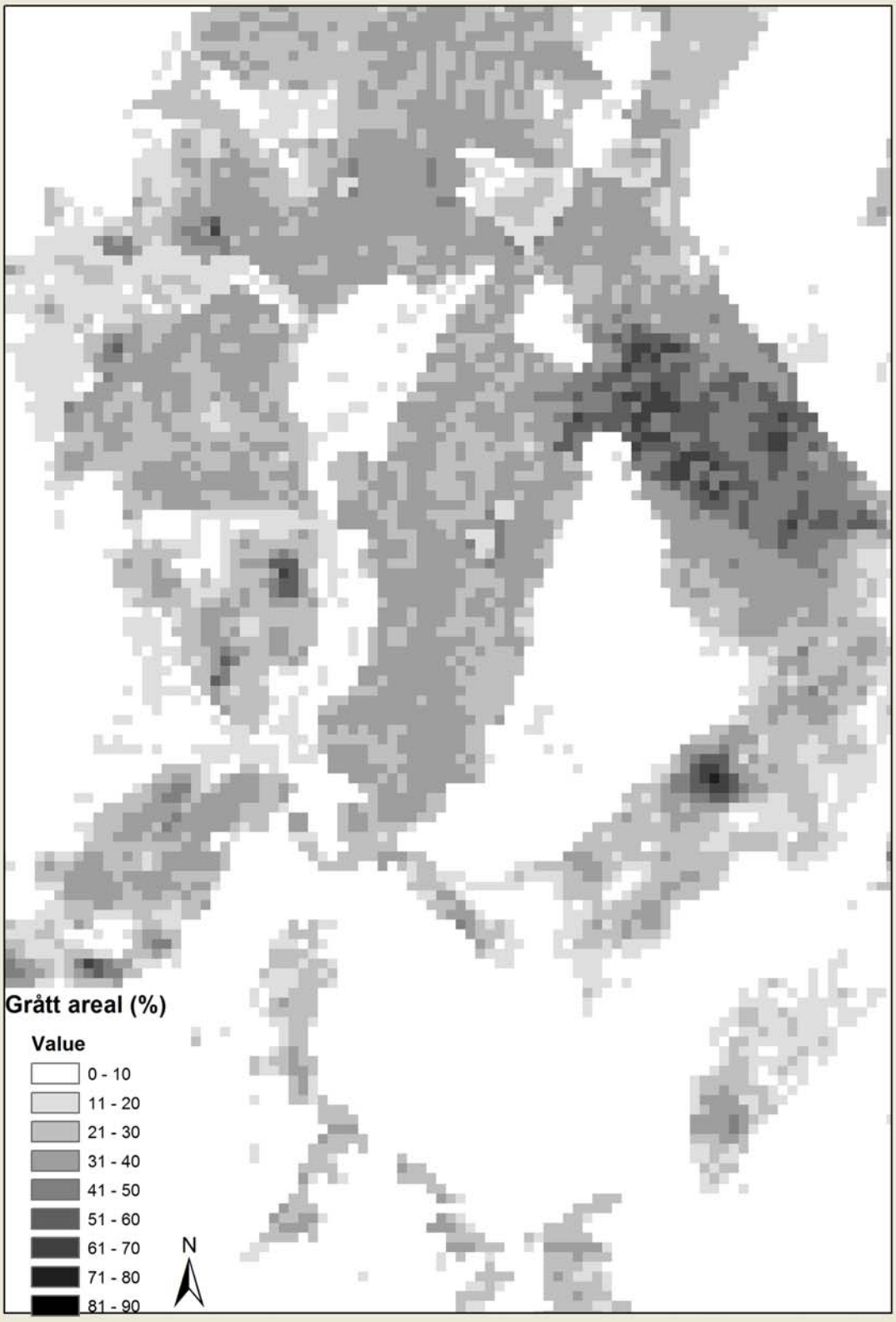
Figur 11. Et høyoppløselig SPOT 5 bilde viser deler av det klassifiserte området (SPOT 5 er ikke benyttet i analysen). © Statens kartverk 2003 SPOT(r) © CNES 2002 © Metria 2002.



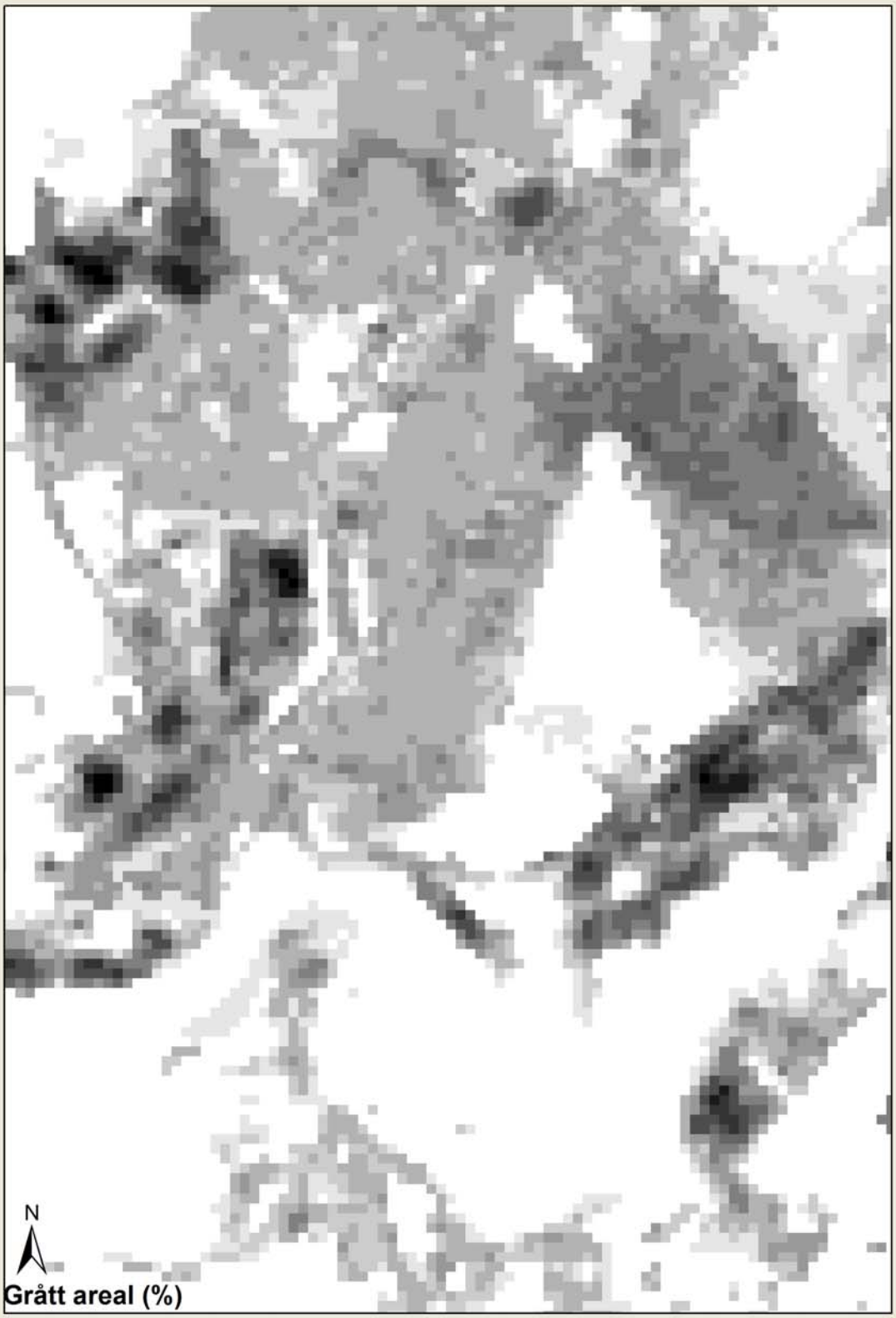
Figur 12. Kart som viser prosentandel grått areal per piksel. Klassifikasjonen er laget med regresjonstre på datasett E med 15 000 punkter som treningsdata (samplingsstrategi nr. 2). Kartutsnittet er det samme som illustrert i Figur 11.



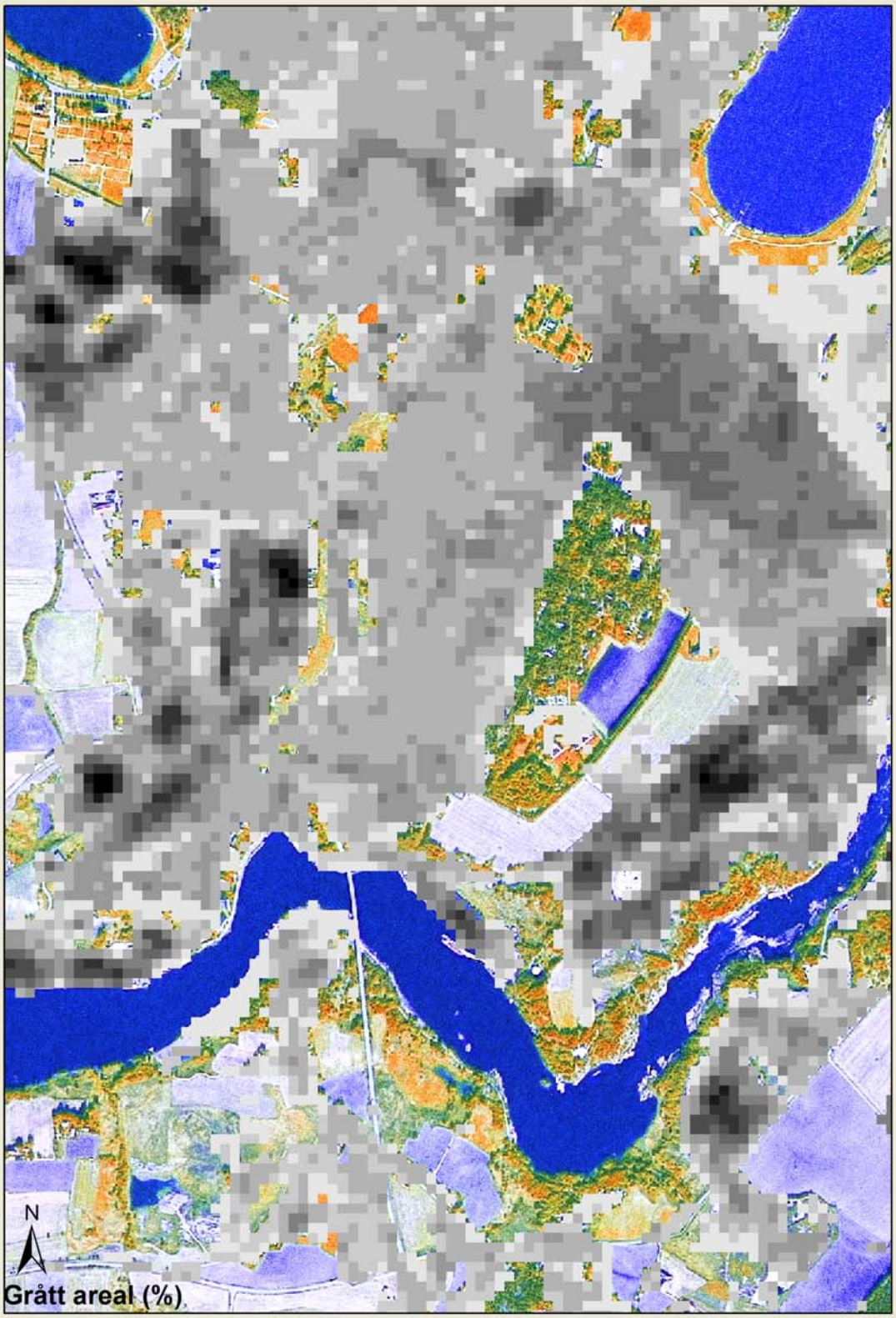
Figur 13. Et høyoppløselig SPOT 5 bilde viser Sarpsborg by (SPOT 5 er ikke benyttet i analysen). © Statens kartverk 2003 SPOT(r) © CNES 2002 © Metria 2002.



Figur 14. Kart som viser prosentandel grått areal per piksel for Sarpsborg by. Klassifikasjonen er laget med regresjonstre på datasett E med ca. 441 500 punkter fra Fredrikstad kommune som treningsdata (samplingsstrategi nr. 1). Kartutsnittet er det samme som illustrert i Figur 13 og Figur 15.



Figur 15. Kart som viser prosentandel grått areal per piksel for Sarpsborg. Klassifikasjonen er laget med regresjonstre på datasett E med 15 000 punkter som treningsdata (samplingsstrategi nr. 2). Kartutsnittet er det samme som illustrert i Figur 13 og Figur 14.



Figur 16. Bildet viser klassifikasjonen i Figur 15 lagt over SPOT 5 bildet i Figur 13. Der SPOT 5 bildet er synlig er det ikke estimert grått areal.

4.3.3.3 Statistikk for rekreasjonsarealer og nærturterreng

Dette avsnittet presenterer statistikk avledet fra kartene som ble laget med regresjonstrærne. Det er beregnet hvor mye grått areal som finnes innenfor SSB's kartlagte rekreasjonsareal og nærturterreng for forsøksområdet som ble beskrevet i avsnitt 4.1. Regresjonstreanalysen med datasett E, samplingsstrategi nr. 2 ga 1,8% grått areal innenfor rekreasjonsarealene (Tabell 12). Tilsvarende tall for nærturterreng ble 0,6% grått areal (Tabell 13).

Tabell 12. Statistikk over grått areal innenfor SSB's kartlagte *rekreasjonsarealer* i forsøksområdet. Arealer er oppgitt i km^2 og prosentandel av totalt rekreasjonsareal innenfor forsøksområdet.

Arealtype	Datasett	Samplings-strategi	Areal (km^2)	Areal (%)
Grått areal innenfor rekreasjonsareal	E	1	0,8	**1,8**
		2	1,6	3,6
Totalt rekreasjonsareal			43,8	100

Tabell 13. Statistikk over grått areal innenfor SSB's kartlagte *nærturterreng* i forsøksområdet. Arealer er oppgitt i km^2 og prosentandel av totalt nærturterreng innenfor forsøksområdet.

Arealtype	Datasett	Samplings-strategi	Areal (km^2)	Areal (%)
Grått areal innenfor nærturterreng	E	1	1,0	**0,6**
		2	1,9	1,2
Totalt nærturterreng			164,2	100

Resultatene i Tabell 12-11 kan direkte sammenlignes med tall for grått areal i Tabell 2 og Tabell 3. Fasittall for grått areal innenfor forsøksområdet finnes ikke tilgjengelig. De to metodene som er testet ut resulterer i tall av litt forskjellig størrelsesorden. Metode 1 (Objektsegmenentering og klassifikasjon av SPOT 5 bilde) resulterte i en relativt høy andel grått areal i rekreasjonsområder (13%), mens andelen var 2% for nærturterreng. Felles for metode 1 og metode 2 er at de begge kartlegger mer grått areal i rekreasjonsareaelet enn i nærturterrenget. I avsnitt 4.2.3.4 ble det vist at metode 1 overestimerte grått areal betydelig (opptil 45% for mye). For metode 2 skjer en underestimering av andelen grått areal for piksler med høy konsentrasjon gråe områder. Dette kan indikere at de reelle tallene for grått areal ligger et sted mellom resultatet fra de to analysene. Det er ikke mulig å angi eksakte verdier for innhold av grått areal i SSB's kartlagte rekreasjonsareal

og nærturterreng. Det kan fastslås at arealene inneholder for mye grått areal. Resultatene tyder på at rekreasjonsarealene inneholder en større feil enn nærturterrenget.

4.3.4 Oppsummering

Regresjonstreanalyse med bruk av tre overlappende Landsat ETM+ satellittbilder og tilleggsdata ble utført for å kartlegge grått areal innenfor hele Østfold fylke. Total andel grått areal ble videre beregnet for SSB's rekreasjonsarealer og nærturterreng i et forsøksområde omkring Fredrikstad og Sarpsborg. Området tilsvarte det samme forsøksområdet som ble benyttet i metode 1 (objektklassifikasjon av SPOT 5 bilde beskrevet i avsnitt 4.2).

Regresjonstreanalyse er en subpixel analyse som i dette tilfellet resulterer i et klassifisert bilde hvor verdien til hvert piksel representerer en prosentandel grått areal. Resultatet et kart som kan tolkes som en tettstedsmaske med tilleggsinformasjon om tettstedskonsentrasjon.

Klassifikasjonen viste at 1,8% av rekreasjonsarealene og 0,6% av nærturterrenget inneholdt grått areal. Analysen indikerer at metoden underestimerer andelen grått areal i områder med høy konsentrasjon grått areal. Absoluttverdiene for andel grått areal er ganske ulike for metode 1 og 2. Metode 1 gir altfor høye verdier, mens metode 2 kanskje er litt lave. Det er ikke mulig å angi korrekte absoluttverdier. Felles for begge metodene er at de resulterer i mer grått areal i rekreasjonsarealet enn i nærturterrenget.

Ved en vurdering av metodikk for å forbedre kartleggingen av disse nøkkeltallene bør en prioritere rekreasjonsarealene. Analysen indikerer at det er her feilene er størst.

5 Diskusjon

5.1 Kvalitetssikring av beregnede nøkkeltall

Rapporten skisserer to ulike metoder for å undersøke korrektheten på kartlagte rekreasjonsarealer og nærturterreng. Metode 1 kan benyttes på høyoppløselige satellittbilder for kartlegging av grått areal, og videre korrigere de kartlagte arealer til ikke å inneholde gråe arealer. Metode 2 kan benyttes på lavere oppløsningssatellittbilder, og resulterer i et kart som viser konsentrasjon av tettsteder. Et slikt kart kan brukes til å finne innslag av grått areal i kartlagte nærturterreng og rekreasjonsarealer, men gir et grovere anslag enn Metode 1. Metode 2 kan også brukes for en annen anvendelse: oppdatering av tettstedsmasker og utvikling av urbane områder (endringsanalyser).

5.2 Sammenligning av metoder for kvalitetssikring av beregnede nøkkeltall

5.2.1 Metode 1

Kompetansekrav og tidsbruk

Metoden stiller store store krav til kompetanse i:

- Programvare.
- Fjernanalyseforståelse.
- Klassifikasjon.

Denne metoden krever en større forståelse av fjernanalyse enn metode 2. Metoden forutsetter forståelse av klassifikasjon med objekter og spektrale egenskaper til ulike overflatetyper. I denne analysen ble eCognition programvare benyttet fordi den inneholder en totalpakke for segmentering og objektklassifisering. Andre programpakker som ArcMap og Erdas Imagine ble brukt i tillegg for å hente ut statistikk om rekreasjonsarealer og nærturterreng. Metoden er relativt tidkrevende for en uerfaren fordi den er brukerstyrt og krever erfaring. Ved opparbeidet erfaring vil tidsbruken avta drastisk. Det er også mulig å lage automatiske rutiner når en fast klassifikasjonslinje er etablert for et område. Segmentering av en høyoppløselig SPOT 5 scene er tidkrevende (bør kjøres over natten eller på PC som bare brukes til dette formål). En rask prosessor med stort minne er nødvendig.

5.2.2 Metode 2

Kompetansekrav og tidsbruk

Metode 2 stiller krav til kompetanse i:

- Programvare.
- Fjernanalyse.
- Regresjonstreanalyse

Dette er helt klart den metoden som stiller minst krav til kompetanse og tidsbruk. Gitt at kompetansekravene er oppfylt, vil anskaffelse av data utgjøre hovedtyngden av arbeidsmengden. Analysen er mye raskere enn metode 1 fordi metoden fungerer som en ”black box”. Store datasett mates inn. I boksen splittes dataene inn i grupper og et regresjonstre beregnes automatisk. Det er noen få strategiske valg som må gjøres for å oppnå godt resultat, men den tekniske gjennomføringen tar svært liten tid. Preprosessering av referansedata er mer tidkrevende enn selve klassifikasjonsprosessen. For en uerfaren bruker er denne metoden å anbefale. Programmene som ble brukt i analysen (Cubist, Erdas Imagine og CART) var enkle å ta i bruk).

5.3 Oppsummering

De to analysemetodene sammenlignes i tabellform (Tabell 14). Tabellen viser at datasettet brukt i metode 1 er dyrere enn datasettet i metode 2. Kompetansekravet er større til metode 1 enn til metode 2. Metode 1 er betydelig mer tidkrevende enn metode 2. Oppsummert er metode 2 best egnet til kvalitetssikring av nøkkeltalls-kartene, når betingelsene er at analysen skal være rask, enkel og prismessig rimelig.

Tabell 14. Oppsummering av de to metodene.

	Metode 1	Metode 2
Kostnader	Høy	Middels
Kompetanse	Høy	Middels
Tidsbruk	Høy	Lav
Egnet som operasjonell metode	Nei	Ja

6 Konklusjoner

Rapporten belyser kvalitetssikring av metodikk for beregning av nøkkeltallene 2RM4-1 (rekreasjonsarealer) og 2RM4-2 (nærturterreng). Basert på landsdekkende registre og kartdatabaser lager Statistisk sentralbyrå kart over rekreasjonsarealer og nærturterreng (Engelien & Schøning, 2001). I denne rapporten er to analysemetoder benyttet til å undersøke hva disse arealene faktisk inneholder i et forsøksområde omkring Fredrikstad og Sarpsborg. Fokus har vært spesielt rettet mot tilstedeværelse av grått areal.

Metode 1 brukte objektsegmentering og klassifisering av et høyoppløselig SPOT 5 bilde til å kartlegge grått areal. Klassifikasjonen viste at 13% av rekreasjonsarealene og 2% av nærturterrenget inneholdt grått areal. En kontroll på klassifikasjonen viste at disse tallene er altfor høye fordi ca. 42-45% av det kartlagte gråe arealet faktisk var: 1) svaberg langs kysten og 2) glissen skog med mye berg-i-dagen. Videre ble det vist at ved å fjerne en 100 m sone langs kysten ble den totale klassifikasjonsnøyaktigheten forbedret fra 66% til 77% for nærturterreng, og fra 64% til 72% for rekreasjonsarealer.

Metode 2 brukte regresjonstreanalyse med tre Landsat ETM+ bilder og tilleggsdata til å kartlegge grått areal. Klassifikasjonen viste at 1,8% av rekreasjonsarealene og 0,6% av nærturterrenget inneholdt grått areal. Analysen indikerte at metoden underestimerer andelen grått areal i områder med høy konsentrasjon grått areal.

Absoluttverdiene for andel grått areal er ganske ulike for metode 1 og 2. Metode 1 gir for høye verdier, mens metode 2 kanskje gir noe lave verdier. Det er ikke mulig å angi korrekte absoluttverdier. Felles for begge metodene er at de resulterer i mer grått areal i rekreasjonsarealet enn i nærturterrenget.

Ved vurdering av metodikk for å forbedre SSB's beregning av disse nøkkeltallene bør rekreasjonsareal prioriteres foran nærturterreng. Analysen indikerer at det er i rekreasjonsarealet feilen knyttet til grått areal er størst.

Oppsummert er metode 2 best egnet til kvalitetssikring av nøkkeltalls-kartene, når betingelsene er at analysen skal være rask, enkel og prismessig rimelig. Metode 2 har også størst potensiale til å utvikles til en operativ metode. Den gir kart på størrelse med en Landsat ETM+ scene, og behøver bare FKB-data som referansedatasett.

Metode 2 anbefales også for andre anvendelser, gjerne knyttet til tettstedsanalyser, der oppdaterte tettstedsmasker er ønsket. Kartproduktet inneholder informasjon om tettstedskonsentrasjon og bør være godt egnet til endringsanalyser (urbanitetsutvikling).

Referanser

ERDAS Field Guide (2002). Leica Geosystems, Atlanta, USA

CART Software User's Guide (2003). Earth Satellite Corporation.

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984). *Classification and regression trees*. The Wadsworth Statistics/Probability Series. California, USA, Wadsworth International group (358 pp).

Baatz, M., Benz, U., Dehghani, S., Heynen, M., Höltje, A., Hofmann, P., Lingenfelder, I., Mimler, M., Sohlbach, M., Weber, M. & Willhauck, G. (2002). *eCognition professional user guide*, Definiens Imaging GmbH, Munchen, Tyskland.

Engelien, E. & Schøning, P. (2001). *Friluftsliv og tilgjengelighet - metode for beregning av nøkkeltall*. 141, Statistisk sentralbyrå, Oslo-Kongsvinger (23 pp).

Esri (2002). *What is ArcGIS?*, ESRI, California, USA.

Huang, C., Wylie, B., Yang, L., Homer, C. G. & Zylstra, C. (2002). Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance. *International Journal of Remote Sensing*, *23*(8), 1741-1748.

Irish, R. R. (2000). *Landsat 7 science data user's handbook*. http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html.

Lillesand, T. M. & Kiefer, B. W. (1994). *Remote sensing and image interpretation*. USA, John Wiley & sons (750 pp).

Markham, B. L. & Barker, J. L. (1986). Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. *EOSAT Landsat Technical Notes*, *1*, 3-4.

Yang, L., Huang, C., Homer, C. G., Wylie, B. K. & Coan, M. J. (2003). An approach for mapping large-area impervious surfaces: Synergistic use of Landsat 7 ETM+ and high spatial resolution imagery. *Canadian Journal of Remote Sensing*, *29*(2), 230-240.

Linker

Miljøverndepartementet 1999. Regjeringens miljøvernpolitikk og rikets miljøtilstand St. meld. nr.8 (1999-2000). (http://odin.dep.no/md/norsk/publ/stmeld/022005-040006/index-dok000-b-n-a.html)

Miljøverdepartementet 2001. Regjeringens miljøvernpolitikk og rikets miljøtilstand - St.meld. nr. 24 (2000-2001). (http://odin.dep.no/ md/norsk/publ/stmeld/022001-040006/index-dok000-b-n-a.html)

Miljøverdepartementet 2003 Regjeringens miljøvernpolitikk og rikets miljøtilstand - St.meld. nr. 25 (2002-2003). (http://odin.dep.no/md/norsk/publ/stmeld/022001-040020/index-dok000-b-n-a.html)

The USGS National Land Cover Characterization 2001 project http://landcover.usgs.gov/natlandcover_2000.html