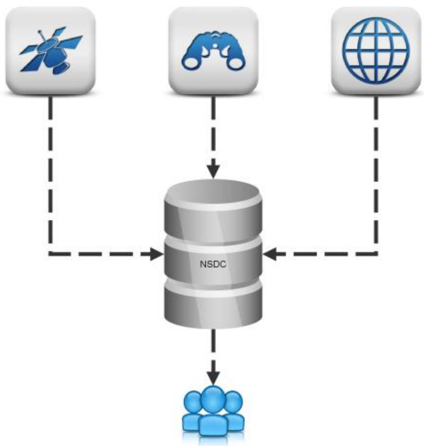
Report ANB-11-05

Preparations for acquisition and application of optical satellite data for Norway Digital – Part 2



March 2014

Arnt Kr. Gjertsen, Norwegian Forest and Landscape Institute

Jon Arne Trollvik, Norwegian Mapping Authority

Anna Birgitta Ledang, Norwegian Institute for Water Research

Johan Danielsen, Norwegian Environment Agency

Ragnvald Larsen, Norwegian Environment Agency

Report ANB-11-05

Preparations for acquisition and application of optical satellite data for Norway Digital – Part 2

March 2014

Arnt Kr. Gjertsen, Norwegian Forest and Landscape Institute

Jon Arne Trollvik, Norwegian Mapping Authority

Anna Birgitta Ledang, Norwegian Institute for Water Research

Johan Danielsen, Norwegian Environment Agency

Ragnvald Larsen, Norwegian Environment Agency

Gjertsen, A. K., Trollvik, J. A., Ledang, A. B., Danielsen, J., and Larsen, R., 2014. Preparations for acquisition and application of optical satellite data for Norway Digital - Part 2, Report ANB-11-05, Norwegian Mapping Authority, March 2014, 127 p.

VERSION
1.0

DATE
12 March 2014

CONTRACTOR COMPANY
Norwegian Space Centre

CONTRACTOR
Norwegian Mapping Authority

CONTACT PERSONS
Jon Arne Trollvik
Arnt Kristian Gjertsen

CONTACT ADDRESSES
Jon.Arne.Trollvik@kartverket.no
Arnt.Kristian.Gjertsen@skogoglandskap.no

COVER PICTURE
Landsat 8 image over Lake Mjøsa and Hamar, Norway

KEY WORDS
Sentinel-2, Landsat 8, collaborative ground segment, satellite data centre, Norway Digital, national users, user requirements

Acknowledgements

We have been in contact with several individuals and organizations and we are grateful for contributions to the development of this report. We would like to mention Carsten Brockmann from Brockmann Consult GmbH who has been of great help by clarifying details on BEAM and the add-on toolbox for Sentinel-2 data that is being developed on behalf of ESA. Thanks also to Andreas Prebensen Korsnes at the Norwegian Mapping Authority who has made the comparison between the national DTM and SRTM DEM. KSAT in Tromsø has explained and demonstrated how a satellite ground station operates and has also given us help to describe modern data-base technology suitable for large raster datasets. We thank Graeme Bell at the Geomatics section of the Norwegian Forest and Landscape Institute who helped us to outline technical details of possible data management solutions. And finally, we thank the Norwegian Space Centre for granting money to this survey and for valuable advice and support.

Contents

Foreword

This report is a continuation of the work presented in the report "Preparations for acquisition and application of optical satellite data for Norway Digital" (Trollvik et al., 2012). The main goal has been to specify the requirements for a national satellite data centre for optical satellite data from the Sentinel-2 and Landsat series Earth observation satellites. The main objective of a national satellite data centre is to facilitate easy access to and use of Sentinel-2 and Landsat 8 data for Norwegian users.

The report is prepared by a group of specialists from Norway Digital: Arnt Kr. Gjertsen, Norwegian Forest and Landscape Institute, Anna Birgitta Ledang, Norwegian Institute for Water Research, Johan Danielsen and Ragnvald Larsen, Norwegian Environment Agency, and Jon Arne Trollvik, Norwegian Mapping Authority. Kai Sørensen, Norwegian Institute for Water Research has also contributed to the report. Trollvik has been the project manager.

The group has met at a regular basis and the work to write the report was divided among the members. Danielsen and Larsen are the authors of Chapter 2 and Chapter 3.5; Jon Arne Trollvik is the author of Chapter 5; Anna Birgitta Ledang is the author of Chapter 6.3; and Arnt Kristian Gjertsen is the author of Chapter 6 (except 6.3). Graeme Bell at the Geomatics section of Norwegian Forest and Landscape Institute has written Chapter 3.7. In other parts of the document several authors have contributed through the process. Arnt Kristian Gjertsen has edited the document.

The document is divided in two parts. *Part I* is the main part describing the use cases of remotely sensed data in Norway today and a proposed national satellite data centre to serve the users when ESA's new Sentinel-2 satellites start to deliver images in 2015. The part ends with a chapter that describes a pilot project needed to clarify technical and economical questions. *Part II* is a technical description of data and processing tools. The appendices at the end give details from the use cases and in addtion some technical descriptions and examples.

March 2014

Abbreviations

Abbreviation	***Description***
A/D	*Analogue-to-digital converter*
AOP	*Atmospheric Optical Properties*
ASTER	*Advanced Space-borne Thermal Emission and Reflection Radiometer*
ATCOR	*Atmospheric and Topographical Correction*
BEAM	*Earth Observation Toolbox and Development Platform*
BOA	*Bottom of the atmosphere*
BRDF	*Bidirectional reflectance distribution function*
Byte	*Basic unit of data storage. 1 byte (SI symbol is B) consists of 2^3 = 8 bits, which was sufficient to describe the alphanumeric characters of the Latin letters and the Arabic digits. 1000 B is 1 kB.*
CNES	*Centre National d'Etudes Spatiales*
Copernicus	*Name for successor of GMES*
CP	*Control Point*
CZCS	*Coastal Zone Colour Scanner*
DBMS	*Database management system*
DEM	*Digital Elevation Model*
DHuS	*Data Hub System, the data access hub ESA offers for member states collaborative partners*
DLR	*Deutsches Zentrum für Luft- und Raumfahrt*
DN	*Digital Number. Information from satellite sensors are usually stored as integers in an interval determined by the radiometric resolution of the sensor. An 8-bit system stores data in the interval [0, 255] and a 12-bit system stores data in the interval [0, 4095].*
DSM	*Digital Surface Model*
DTM	*Digital Terrain Model*
EMW	*Electro-magnetic waves*
ENVISAT	*Environmental Satellite*
EO	*Earth Observation*
ESA	*European Space Agency*
EU	*European Union*
FOS	*Flight Operations Segment*
GB	*1 gigabyte is 1000^3 B (bytes)*
GCP	*Ground Control Point*
GDEM	*Global Digital Elevation Model*
geoNorge	*A national internet portal that promotes sharing of geospatial data among partners in Norway Digital*

GMES	*Global Monitoring for Environment and Security*
GPS	*Global Positioning System*
kB	*Kilobyte, which is 1000 bytes and 8 x 1000 bits*
IFOV	*Instantaneous field of view*
IR	*Infrared*
L1B	*Level 1B. Sentinel-2 dataset stored as top of the atmosphere radiance in sensor geometry*
L1C	*Level 1C. Sentinel-2 dataset stored as top of the atmosphere reflectance in UTM geometry (orthorectified).*
L1CN	*National version of ESA's L1C*
L2A	*Level 2A. Sentinel-2 dataset stored as bottom of the atmosphere reflectance in UTM geometry*
L2AN	*National version of ESA's L2A*
LIDAR	*Light Detection and Ranging*
LUT	*Look-up-tables*
MACCS	*Multisensory Atmospheric Correction and Cloud Screening*
MB	*1 megabyte is 1000^2 B (bytes)*
MERIS	*Medium Resolution Imaging Spectrometer. One of ten instruments aboard ESA's Envisat earth observing satellite. The Envisat mission ended in 2012, following an unexpected loss of contact with the satellite*
MODIS	*Moderate-resolution Imaging Spectro-radiometer. One of the main instruments aboard the Terra and Aqua satellites*
MODTRAN	*MODerate resolution atmospheric TRANsmission*
MS	*Multi-spectral*
MTCD	*Multi-temporal cloud detection*
NCGS	*National Collaborative Ground Segment for Sentinel-2*
NDVI	*Normalized difference vegetation index*
NIR	*Near Infrared*
Norway Digital	*The Norwegian Geospatial Data Infrastructure*
NSDC	*National Satellite Data Centre*
NSDI	*Norwegian Spatial Data Infrastructure*
OGC	*Open Geospatial Consortium*
PDS	*Payload Data System*
Rasdaman	*Raster data manager*
RGB	*Red, green, and blue. Designates the three bands in the visible portion of the solar spectrum*
SOS	*Successive Order of Scattering*
SRTM	*The Shuttle Radar Topography Mission*
SWIR	*Short wave infrared*
TB	*1 terabyte is 1000^4 B (bytes)*

TIR	*Thermal infrared*
TOA	*Top of the atmosphere*
USGS	*United States Geological Survey*
UTM	*Universal Transverse Mercator coordinate system. The projection is conformal; it preserves angles and approximates shapes but distorts scale: thus distance and area will be distorted*
WCS	*Web Coverage Service*
WMS	*Web Map Service*
WRS-2	*Worldwide Reference System, a cataloging system used to organize Landsat data into scenes, which are defined by and named after path and row number. WRS-2 is used for Landsat 7 and 8.*

Oppsummering

Norge digitalt er et nasjonalt samarbeid etablert i 2005 med oppgave å etablere, ivareta og distribuere digitale geografiske datasett. I 2009 ble det opprettet en faggruppe på fjernmåling med hovedfokus på utfordringene med å bruke data fra de nye jordobservasjonssatellittene Sentinel-2 og Landsat 8. ESA planlegger å skyte opp to Sentinel-2 satellitter innen 2016, og til sammen vil de dekke Norge med opptak minst hver tredje dag. Dataene vil bli gratis tilgjengelig for brukerne. Den amerikanske satellitten Landsat 8 ble skutt opp i februar 2013. Den har en opptaksperiode på 16 dager, men over Norge vil naboscener ha relativt stort overlapp så den reelle dekningsfrekvensen er på ca. 8 dager. Sentinel-2 har en geometrisk oppløsning på 10 og 20 meter og Landsat 8 har oppløsning på 30 meter. Opptak fra satellittene er egnet til å kartlegge og overvåke status og forandringer over land og kystvann, og det ble på et tidlig tidspunkt antatt at eksisterende tjenester basert på jordobservasjonsdata vil ha stor nytte av de nye satellittene og at nye tjenester vil oppstå.

En sentral oppgave for faggruppen er å vurdere løsninger for best mulig utnyttelse av Sentinel-2 data for nasjonale brukere. I en rapport fra 2012 beskrev gruppen Sentinel-2 og planene ESA har for lagring og prosessering av data. Den nye amerikanske satellitten Landsat 8 ble også drøftet. I rapporten ble utfordringene for norske brukere diskutert, og den konkluderte med noen oppfølgingsspørsmål som trengte videre bearbeiding. Noen av disse har blitt gjennomgått i denne rapporten.

Med tanke på den høye frekvensen av bildeopptak og prinsippet om tilgang til gratis data vil bruken av satellittdata trolig øke kraftig. Dagens bildeorienterte bruk vil skifte mer og mer mot bruk av tidsserier av bildedata (bildestakk eller datakube): altså et skifte av fokus fra analyser av enkeltopptak til analyser basert på datakuber med tidsdimensjonen i tillegg til de romlige dimensjonene. Utviklingen vil stille store krav til kapasiteten i prosesseringen ettersom datavolum vil øke betraktelig. Bruk av tidsserier stiller dessuten høye krav til preprosesseringen av dataene:

- Den geometriske nøyaktigheten må være høy slik at opptakene blir nøyaktig samlokalisert.
- Pikselverdiene må korrigeres for atmosfærens effekt på målingene slik at en kan beregne reflektans fra bakken. Med reflektansverdier kan målinger fra ulike datoer og sensorer sammenlignes.
- Områder i opptakene som er dekket av skyer og skyskygger må maskes ut før atmosfærisk korreksjon kan utføres.

I rapporten har vi beskrevet syv brukstilfeller for å dokumentere kravene til en nasjonal satellittdatatjeneste for effektiv bruk av data fra Sentinel-2 og Landsat 8. Brukstilfellene betraktes som representative for nesten alle brukstilfellene i Norge digitalt i dag og kan deles opp i fire kategorier:

- bildefremvisning
- kartlegging

- endringsanalyse
- assosiativ analyse

Brukstilfellene viser retningen til satellittfjernmåling i dag med økende fokus på endringsanalyse og bruk av tidsserier for å overvåke utvikling (f.eks. fenologisk utvikling). Brukstilfeller uten fokus på endringsanalyse har også behov for data fra flere opptak for å danne skyfrie (eller nesten skyfrie) tidskompositter. Brukstilfellene demonstrerer at det er behov for å bruke flere opptak gjennom et lengre tidsrom – uker, måneder eller år – og derfor også behov for atmosfærekorrigerte data som representerer reflektans til objektene på bakken.

I rapporten fra 2012 ble det konkludert med at data fra ESA ikke vil tilfredsstille alle disse kravene, og det er to alternativer for å løse denne utfordringen:

1. Alle brukerne laster ned data fra ESAs arkiver og tar selv hånd om all preprosessering
2. Brukerne samarbeider og etablerer et nasjonalt satellittdatasenter som laster ned data og gjennomfører all preprosessering

Forestående utvikling med sterk økning i datavolum og antall satellittopptak over Norge og dreining mot analyse av tidsserier taler for det andre alternativet. Nødvendig tilrettelegging og preprosessering av store datamengder vil gi store kostnader for den enkelte bruker og det vil derfor være rasjonelt å samordne tilretteleggingen av dataene for alle brukerne. ESA støtter slike tiltak ved å tilrettelegge for bedre datatilgang for nasjonale kollaborative bakkesegmenter.

Bruk av en detaljert og nøyaktig høydemodell er nødvendig for sikre høy geometrisk nøyaktighet til ortokorrigerte datasett (ortofoto) slik at de er best mulig forberedt for bruk i endrings- og tidsserieanalyser. Det nasjonale rutenettet SSBgrid, definert av Statistisk Sentral Byrå, foreslås brukt som et fast gitter for ortokorrigeringen. På denne måten vil piksler fra alle ortofotoene bli liggende i et felles gitter slik at pikslene som dekker det samme området på bakken blir liggende nøyaktig oppå hverandre. Pikselobservasjonene må konverteres til reflektans før data fra ulike opptaksdatoer og sensorer kan kombineres i analyser. Atmosfærisk korreksjon må derfor utføres på datasettene som en del av tilretteleggingen for brukerne.

In-situ data spiller en viktig rolle i tolking av satellittbilder. Nettstedet geoNorge er utviklet for Norge digitalt og er en portal for å søke etter geografisk stedfestet informasjon i Norge. Portalen vil derfor være et naturlig sted å legge inn informasjon om in-situ data som kan være relevant i tolking av satellittdata.

Det nasjonale satellittdatasenteret for Sentinel-2 og Landsat 8 data må tilfredsstille følgende krav:

1. Senteret må inkludere et arkiv med lagringskapasitet for å lagre alle Sentinel-2 datasett over Norge på nivå 1B, 1CN and 2AN, dvs. rå data i sensorgeometri (1B), ortokorrigerte data resamplet til SSBgrid (1CN), og ortokorrigerte data som i tillegg er atmosfærisk korrigert der pikseldata representerer reflektans på bakken (2AN). For data fra Landsat (dvs. Landsat 8 og etterfølgere) foreslås tilsvarende produksjon og lagring av produktnivåer. Metadata som beskriver datasettene må også lagres.

2. Senteret må laste ned og prosessere satellittbilder så fort disse er tilgjengelig fra internasjonale arkiver (hos ESA og USGS) og umiddelbart gjennomføre geometrisk og atmosfærisk korreksjon. Den geometriske rektifisering må være basert på den beste tilgjengelige nasjonale høydemodellen og prosesseringstrinnene bør være automatisert og kreve liten manuell innsats.
3. Senteret bør støtte produksjon av tidskompositter, dvs. datasett satt sammen av opptak over en tidsperiode der piksler som ikke er dekket av skyer og skyskygger blir benyttet.
4. Senteret må støtte en web-basert tjeneste for effektiv distribusjon av data og bør støtte uttak av tidsserier av data over brukerdefinerte tidsrom.

Tekniske løsninger for datatilgang, algoritmer for rask og robust preprosessering, datalagring, og distribusjon må undersøkes og testes. En automatisk metode for ortorektifisering må utvikles, muligens basert på matching mot et referansebilde for å minimalisere manuelt arbeid. Det er fremdeles uklart om ESA vil tilby kollaborative bakkesegmenter programvare for å prosessere fra level 1B til 1C.

To metoder for atmosfærisk korreksjon er beskrevet i rapporten:

1. MACCS er basert på en multi-temporal analyse av opptak fra to eller tre datoer og utnytter den relative stabiliteten i reflektansen til objekter over korte perioder på noen få dager til å bestemme atmosfærens egenskaper som påvirker strålingen gjennom atmosfæren.
2. ATCOR er basert på analyse av opptak fra én enkelt dato og bruker en fysisk modell av strålegangen som går gjennom atmosfæren og ned til bakken, reflekteres fra et objekt på bakken, og går opp gjennom atmosfæren og mot satellittens sensor. ATCOR kan også korrigere for belysningsvariasjoner i kupert terreng.

ESA har gitt Deutsches Zentrum für Luft- und Raumfahrt (DLR) i oppdrag å utvikle en spesialversjon av ATCOR som en modul i BEAM, som er et verktøy for å prosessere data fra ESAs satellitter.

Vi har estimert datavolumet til minst 8,5 TB/år for ett produktnivå og 25,5 TB/år for tre produktnivåer (1B, 1CN, 2AN). Korresponderende nødvendig datalagringskapasitet er beregnet til hhv. 43 TB/år og 130 TB/år. Over en periode på 10 år vil det totale volumet være minst 432 TB for ett produktnivå og 1 296 TB for tre produktnivåer. Dette er et stort volum som setter høye krav til en lagringsløsning som samtidig kan tilfredsstille forventningene fra brukerundersøkelsene. Kostnaden for å etablere et senter er estimert til 4 MNOK og den årlige driften til 1,2 MNOK.

På dette stadiet er det vanskelig å gi presise estimater for alle kostnader. Aktuelle tekniske løsninger for geometrisk rektifisering og atmosfærekorreksjon må testes. En automatisert rektifiseringsmetode basert på et skyfritt referansebilde og matching må utvikles for å unngå eller minimalisere manuelt arbeid. Det er fortsatt uklart om ESA vil gjøre programvaren for geometrisk prosessering fra nivå 1B til nivå 1C tilgjengelig for brukerne. I hvor stor grad atmosfærisk korreksjon kan automatiseres er foreløpig uklart og må utprøves. Hvilken datalagringsløsning som gir den ønskede funksjonaliteten må avklares. Nye løsninger basert på arkitektur for å håndtere multi-dimensjonale rasterdatasett kan muligens være aktuelle. Resultatene fra testene vil bidra til utarbeiding av et kostnadsbudsjett for en framtidig operasjonell drift av et nasjonalt satellittdatasenter.

Faggruppen har konkludert med at et pilotprosjekt må gjennomføres for nærmere å vurdere og teste tekniske løsninger. Data fra Sentinel-2 vil ikke være tilgjengelig før siste halvår av 2015, men vi antar at et pilotprosjekt kan baseres på datasett som ESA beskriver på deres nettsted (http://due.esrin.esa.int/s2t5.php). I tillegg antar vi at datasett fra Landsat 8 kan brukes for å teste ut de fleste utfordringene i preprosessering og lagring.

Summary

The Norwegian Geospatial Data Infrastructure (Norway Digital), established in 2005, is a national cooperation among institutions producing geospatial data. Norway Digital's mandate is to establish, maintain and distribute digital geographic datasets. In 2009 Norway Digital established a technical committee on remote sensing, and the main task was to study and review the challenges related to the use of data from the new Earth observation (EO) satellites Sentinel-2 and Landsat 8.

When the two planned Sentinel-2 satellites are both operational, Norway will be covered at least every third day by 10 meter resolution optical images. This represents access to an unprecedented large amount of data, which can be used to map and monitor status and change over land and coastal waters. It was anticipated at an early stage that existing services based on the use of EO data would strongly benefit from this development and that new services will emerge.

The main assignment for this work was to review what must be done in order to facilitate full national advantage of Sentinel-2 data. In the report delivered in 2012 (Trollvik et al. 2012), the committee described the Sentinel-2 mission and the data processing plans of the European Space Agency (ESA). The new American Landsat 8 satellite was also discussed. A discussion of the challenges for the Norwegian users was included along with some questions that needed further considerations. Some of these questions are examined in this report.

The high frequency of image acquisition and the free of charge data policy for Sentinel-2 will probably change how satellite data are used. Today's single scene oriented use will probably shift more and more to use of satellite image time series (i.e. image stacks or data cubes). Correspondingly, focus will shift from analysis of spatial structures based on single date datasets to analysis of spatio-temporal structures in dynamic scenes based on image stacks. This development will put high demands on data processing capacity as data volumes will increase greatly. Time series of images require that high standards are met in the preprocessing steps:

- The geometric accuracy must be high so that images from different dates can be precisely aligned.
- The radiance measurements must be corrected for the influence of a highly variable atmosphere to obtain surface reflectance values.
- Pixels covered by cloud and cloud shadow have to be detected and masked.

In this report seven use cases have been described in order to document the requirements on the services needed by the users for accurate and efficient use of Sentinel-2 and Landsat 8 data. The use cases are considered representative for nearly all use cases in Norway today and can be divided into four categories:

- Presenting images
- Mapping
- Change detection

- Associative analyses

The use cases demonstrate the current trend in image analysis of increased focus on change detection and use of time series data to monitor land cover dynamics, e.g. multi-year phenology. Use cases not focused on change detection also show a need for images from several dates in order to make cloud free or near cloud free time composites. Thus, all use cases demonstrate the need for data from multiple dates. However, use of multi-temporal datasets requires atmospherically corrected data converted to surface reflectance.

In the first report, it was concluded that ESA will not provide data fulfilling all these requirements. To resolve this dilemma two different solutions are possible:

1. The users will take care of all the necessary preprocessing.
2. The users will collaborate and establish a central processing unit that takes care of all necessary preprocessing and thus takes the burden of preprocessing from each user.

The National Satellite Data Centre suggested in this report, is a proposal for the second solution, and the main motivation is to facilitate easy use and access of data from Sentinel-2 and Landsat for Norway Digital partners. ESA supports such initiatives by providing improved data access for national collaborative ground segments.

A detailed and accurate elevation model is necessary in order to obtain high geometric accuracy of orthorectified datasets and make them suitable for change detection and time series analysis. SSB-grid, a national grid defined by Statistics Norway, is suggested as a geometric frame for the orthorectified datasets. In this way, pixels from all datasets will be defined in the same grid and pixels representing the same object on the ground will be accurately aligned. The users need surface reflectance to combine and compare observations from several dates and image sensors, and as a consequence the data have to be corrected for the effects of the atmosphere.

In situ data are important to support interpretation of remotely sensed data. The internet portal geoNorge is part of Norway Digital and is a web service for searching for geospatial data in Norway. The portal would therefore be a natural place to record information about in situ data relevant for interpretation of satellite data.

The national satellite data centre should meet the following requirements:

1. The centre must include an archive with sufficient storage capacity to contain all Sentinel-2 level 1B, 1CN and 2AN datasets, and Landsat 8 data on the same levels, including metadata for all datasets stored.
2. The centre must download and process satellite images as soon as they are available from international archives and immediately perform geometric and atmospheric corrections producing product levels 1CN (orthorectified) and 2AN (orthorectified and atmospherically corrected). The geometric rectification must be based on the best available national elevation model and the processing steps must be automated and require very little manual input.
3. The centre must support production of cloud and cloud shadow free time composites.

4. The centre must support distribution of the data to the users through a web service and should support server-side processing for different time composites.

Technical solutions for data access, fast and robust preprocessing algorithms, data archiving, and distribution have to be further examined and tested. An automated geometric rectification method has to be developed, possibly based on matching with a reference image, to avoid or minimize manual work. It is still unclear whether ESA will provide collaborative ground segments with software to process from level 1B to 1C.

Two methods for atmospheric corrections are described in this report:

1. MACCS is based on analysis of multi-date (two or three dates) datasets and exploits the relative stability of surface reflectance over a short period of a few days to estimate atmospheric properties.
2. ATCOR is based on analysis of single-date datasets and uses a physical model of the path of solar radiation traveling through the atmosphere, reflected from an object on the ground, and traveling through the atmosphere in direction to the sensor. ATCOR can also correct for the effect rugged terrain has on the illumination of the nearby pixels.

At the time of writing, we are informed that ESA plans to provide ATCOR as a module of the BEAM software, a toolbox provided by ESA to enable the users to process satellite data provided by the agency.

We estimate that the yearly increase in data volume will be at least 8.5 TB for one product level and 25.5 TB for three product levels and the corresponding required data storage 43 TB and 130 TB respectively. Over a 10 years period the storage requirements will be at least 432 TB for one product level and 1 296 TB for three product levels. This is a large volume and puts high demands on the data management solution to provide the level of service described in the use cases. A crude cost estimate for establishing the centre is 4 million NOK and for yearly running 1.2 million NOK.

At this stage, it is difficult to give a detailed description of all requirements and technical solutions for a national satellite data centre. Accordingly, a follow-up pilot project must be designed in order to test technical solutions and estimate costs to build a data management solution and to preprocess data from Sentinel-2 and Landsat 8. Specifically, solutions for orthorectification, atmospheric correction, data management platform, and discovery and access need to be tested.

By doing so, we believe that still more technical questions and challenges implementing the proposed services will arise. Thus, the pilot project will provide a more solid foundation to estimate the cost of operating a national data centre for Sentinel-2 and Landsat data. Sentinel-2 data will not be available in 2014, but we believe that the pilot project can be based on processing datasets provided by ESA as described on their website (http://due.esrin.esa.int/s2t5.php). Additionally, we believe Landsat 8 datasets can be used to explore most of the challenges presented by Sentinel-2 datasets.

1 Introduction and background

Norway Digital established a technical committee on remote sensing in 2009. In 2011 the committee started a study on the national needs to prepare for the use of data from the Sentinels, ESA's new Earth observation satellite missions. We have chosen to focus on the Sentinel-2 mission, which will consist of a constellation of two operational satellites to obtain high revisit capacity with a goal to provide cloud-free products over Europe every 15 to 30 days. Sentinel-2 will provide enhanced continuity of data so far provided by the Landsat and SPOT missions. The Landsat series of Earth observation satellites represents the longest Earth imaging program in history, and started with the launch of Landsat 1 in 1972. After the launch of Landsat 7 in 1999, the program suffered serious setbacks with technical difficulties with the new satellite and long delays in development of a successor. Finally, in 2007 it was decided to launch a new satellite, Landsat 8, which would be operated by the USGS. Together, Sentinel-2 and Landsat 8 will continue and further develop the services provided by the Landsat missions since 1972.

The Sentinel-2 mission will be a central part of EU's Copernicus program, providing multispectral data in 13 bands with geometric resolution of 10, 20 and 60 meters. Landsat 8 has eight multispectral (MS) bands and one panchromatic band and geometric resolution of 30 meters for the MS bands and 15 meters for the panchromatic band. In addition, Landsat 8 has two thermal infrared bands with 100 meters resolution. Together, the two Sentinel-2 satellites will have a revisit time every 5 days (10 days with one satellite) at the equator and 2–3 days (4–6 days with one satellite) over Norway, while Landsat 8 has a revisit time every 16 days at the equator and every 6 to 8 days over Norway. The data policies for Sentinel-2 and Landsat 8 secure free and open access, which means that the data will be free of charge for the users. Landsat 8 was launched in February 2013 and the first Sentinel-2 satellite is planned to be launched in 2015 (the launch window is currently set between 30 March and 28 June). Thus in the near future, the availability of medium resolution satellite data will be dramatically improved and will possibly change the way we will use such data: we will see a development from an image oriented use to a pixel oriented use, where cloud free pixels from all acquisitions with cloud cover less than ca. 50 percent are used for change detection and time series analysis.

The technical committee published its study in a report with the title "Preparations for acquisition and application of optical satellite data for Norway Digital" (Trollvik et al., 2012) and in a paper in a Norwegian journal on geomatics (Gjertsen and Trollvik, 2013). The study concluded as follows:

- the user needs must be examined in more detail and linked to product levels that a future national satellite data centre would provide,
- the technical details for the preprocessing from low level to high level products (geometrically and atmospherically corrected) should be described in more detail, and
- the committee should contact international bodies that are involved with similar tasks.

In this second report, we discuss in more detail some selected use cases and link them to data requirements on a proposed national satellite data centre. Three relevant tools are described that can process data to high levels and provide bottom of the atmosphere reflectance values and cloud/cloud shadow masks.Reflectance values are necessary for the use cases that involve change detection, time series analysis, and temporal composites. We also discuss the use of alternative elevation models for geometric rectification of the images, as errors in an elevation model will lead to positional errors of pixels in the final orthoimages. Positional errors will have negative effects for change detection and time series analysis that are central in many of the use cases. A national satellite data centre is proposed and its purpose, organisation, links to standards, integration with Norway Digital, and costs and storage requirements are described. It has proven difficult to describe all requirements and costs in detail, and the committee has concluded that it will be necessary to follow up this work with a pilot project where technical solutions are tested. Sentinel-2 data will not be available for 2014, but we assume the tests can be based on Landsat 8 datasets (as most of the challenges will be similar with either of the satellite missions) or datasets provided by ESA as described on their website (http://due.esrin.esa.int/s2t5.php).

Part I: Use cases and user requirements on a national satellite data centre

2 Current and potential use of satellite data

The future use of the Sentinel and Landsat satellite data relies on sound knowledge about several issues. This chapter highlights some of them starting with a general overview of different satellite system sensors, active and passive, can be part of the process towards relevant products. In view of the prospective pixel level use we also found it necessary to introduce the reader to the national grids standard.Image analysis, and in particular pixel related products, relies on digital elevation models for proper rectification. The sub-chapter concludes with an advice on how to establish a proper national elevation model. The chapter concludes with a presentation of potential use for satellite data.

2.1 Current use of satellite data

Currently, there are several different uses of data achieved from satellite based sensors in Norway and internationally. In this report we have chosen to list failed attempts as well as ongoing use in order to establish baseline knowledge of the use of satellite data. We also choose to mention briefly the use of satellite data from other countries. Our main reasons for including international experience with these data and non-operationalized uses are to:

1. Heighten awareness of new potential for cost efficient research, monitoring and surveying based on satellite data
2. Advocate the new possibilities presented by Sentinel-2 and Landsat 8 based on the potentially high frequency and availability of free of cost, spatial data
3. Heighten awareness of unsuitable approaches to utilization of data from satellite based sensors

A number of attempts to use data from satellite based sensors in Norway have failed or failed to be established as operative methods for mapping and monitoring. It is however important to point out that there are a number of explanations to why establishing an operative use of different satellite data and methodologies did not succeed:

1. Poor availability of data due to few cloud free scenes and high cost of downloading data from commercial satellites
2. High cost of utilization of data due to the need for time consuming corrections to identify bottom of atmosphere reflectance (BOA)
3. Low decision maker awareness of benefits from new methodology
4. Attempts on using satellite data failed to give satisfactory results due to poorly developed algorithms for one or more needed corrections of the raw satellite dataset
5. The attempts to develop relevant and useful products in a particular field of work have failed

Currently, the use of satellite data in Norway is well established in meteorology and in monitoring of oceanic and arctic conditions. A number of exploratory projects using satellite data in nature management have been reported in Norway as well as internationally. Some have shown promising results for mapping and monitoring of certain natural recourses.

Structuring of satellite data in this chapter is achieved by distinguishing between active and passive sensors. By passive sensors we mean sensors recording emission of electromagnetic waves (EMW) from the top of the Earth’s atmosphere whereas active sensors send EMW in the form of laser light and from radar beams towards the Earth’s surface and record the top of atmosphere return signals.

Satellite sensors record top of the atmosphere (TOA) radiance values. Making use of the data, require correcting for atmospheric conditions and landscape effects. This is crucial for obtaining bottom of the atmosphere (BOA) reflectance values. The data need to be orthorectified using ground control points (obtained from a reference image) and digital terrain models to correct for parallax displacements and finally to georeference the data to a map projection. Alas, few if any products are solely based on passive sensors. Data from passive sensors require data obtained fully or partly from active sensors (radar and laser) like digital elevation models (DEMs) and meteorological in order to produce BOA reflectance values. DEMs are however not as important over water as over land, unless future developments will use DEMs to correct for the challenges with high mountains and narrow fjords.

2.1.1 Passive sensors

Passive sensors record TOA radiance of reflected visible (VIS) light, near infrared (NIR) light, short-wave infrared (SWIR) light, thermal infrared (TIR) light and passive microwave radiation. VIS and NIR measurements are utilized in analyses targeting chlorophyll content in vegetation and chlorophyll and particle content in marine and freshwater resources. These analyses are addressing a number of topics like vegetation cover, land use changes, forest resources and water quality.

Whereas VIS and NIR emission from the Earth is almost solely based on reflection of sunlight, TIR and passive microwave measures Brightness Temperature (BT) at the Earth surface. TIR is important for mapping heat emitted caused by underground fires in some parts of the world. Climate change is also an important issue for surface temperatures, both when it comes to land and ocean surface temperatures.

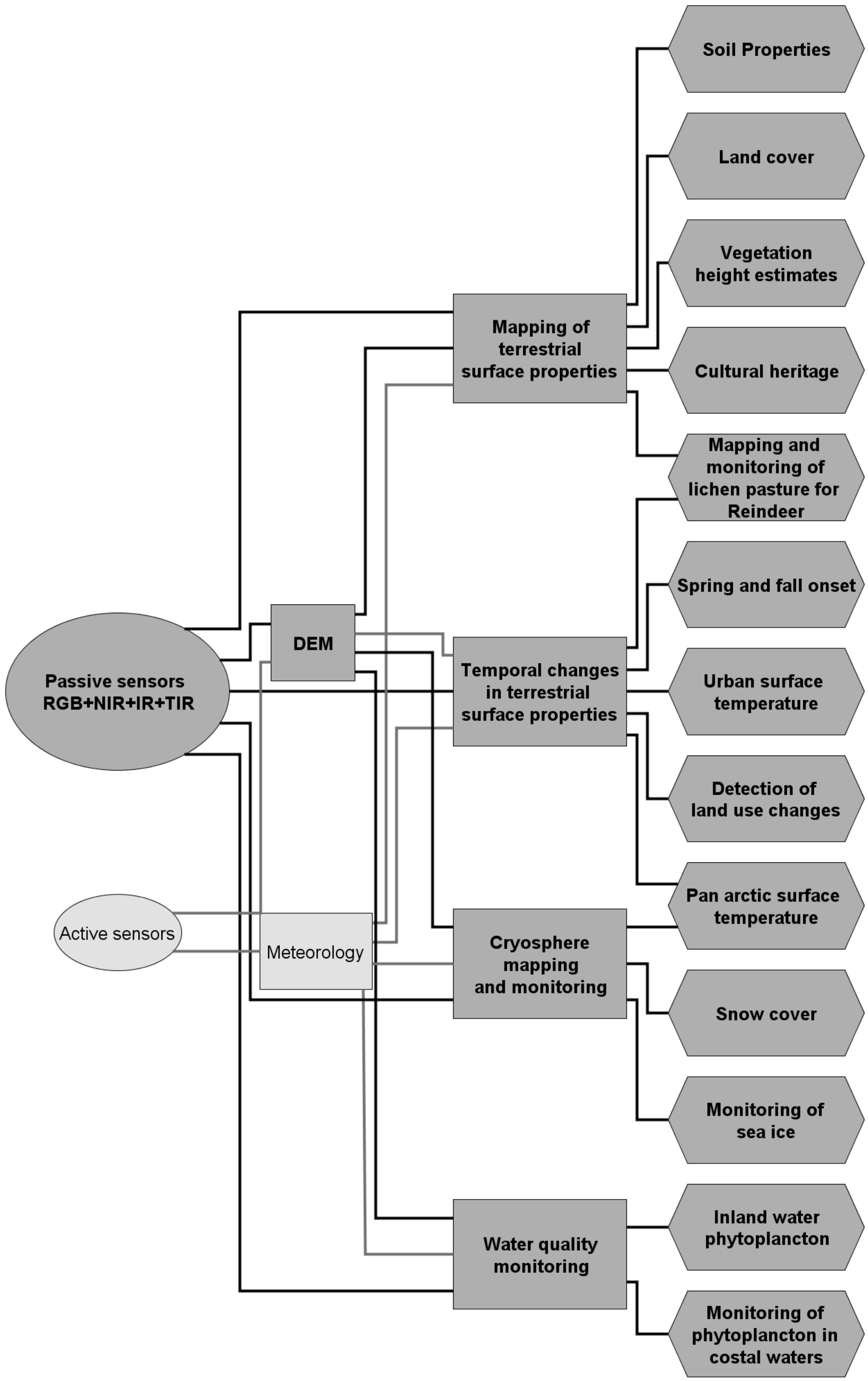


Figure 1 Current use of data from passive sensors operating from a satellite platform.

2.1.1.1 Mapping of area resources

Data from passive sensors using VIS, NIR, SWIR, and TIR light are particularly important in mapping of biological resources (Figure 1), mainly in interpreting occurrence and changes in the amount of chlorophyll in vegetation and surface waters. Biological resources will, unlike many abiotic resources, shift and change on geo-temporal scales. These data will as a result usually be valid within a limited space of time. The time they are valid may span from minutes, up to several years. Mapping of phytoplankton growth and particles in coastal waters represent an example of short lived validity. Analyses of volume of lichen heath as winter pasture for reindeer on the other hand represents an example of more long lived maps based on satellite images. Mapping of forest canopy cover and forest canopy status may represent a timescale in between. Forestry activities, windfall events, and changes in greenness of foliage due to different reasons may change forest canopy extent and distribution on a time scale of weeks up to a year. Repetition of, and comparison between a set of previous and future maps with comparative analyses, will convert mapping activities into monitoring activities.

Several attempts have been made at using satellite based sensor data for mapping of vegetation in Norway, the success has been to say the least, mixed. The lack of concurrence between mapping units in “classic” vegetation mapping and relative homogenous areas in satellite images represent a major challenge. Lack of information about ground vegetation in areas with closed canopy forest in sensor data is an obvious major challenge.

The development of methodology for mapping of the extent and volume of lichen heath as winter pasture for reindeer based on Landsat 5 and 7 images (30 m pixels) was however successful. The methodology has been augmented for use in areas with sparse coverage of lichen by supplementing the data from Landsat 5 and 7 with high resolution images (0.5 m pixels) from aerial photos. Basic methodology and augmented methodology both use a grid of ground reference points (Falldorf et al., 2014).

2.1.1.2 Monitoring

As outlined above, monitoring of natural resources on large to medium scale, based on satellite sensors, is already established. It is reasonable to assume that most monitoring activities are going to be based on mapping activities. Monitoring of algae bloom in coastal waters, forest canopy condition and lichen forage for wild and semi-domesticated reindeer are typical examples of such monitoring which is close to or, already established.

2.1.1.3 Associative analysis

Sensor data from satellites as well as aerial photos can be divided into subareas consisting of clusters of relative homogenous adjacent pixels. The likelihood of association between subunits with homogenous pixels and high probability for occurrence of certain resources has been targeted as a promising path of analyses. This was targeted in studies made during the SatNat program that analyzed the associations between homogenous clusters of pixels and specific habitat for certain wildlife species (pers. comm. Johan Danielsen, Norwegian Environmental Agency).

2.1.2 Active sensors

Radar and laser sensors provide different information than passive sensors. Radar satellites record data from the Earth's surface independently of sunlight. Data from radar satellites can be used to produce digital elevation models (DEM) and for use in meteorology, forestry, cryo-science, marine vessel detection, and oil spills (Figure 2).

Satellite based lasers (LIDAR) provide data for DEMs, meteorology and atmospheric condition (including detection of particular gasses in the atmosphere), archaeology, biology and conservation, geology and soil science.

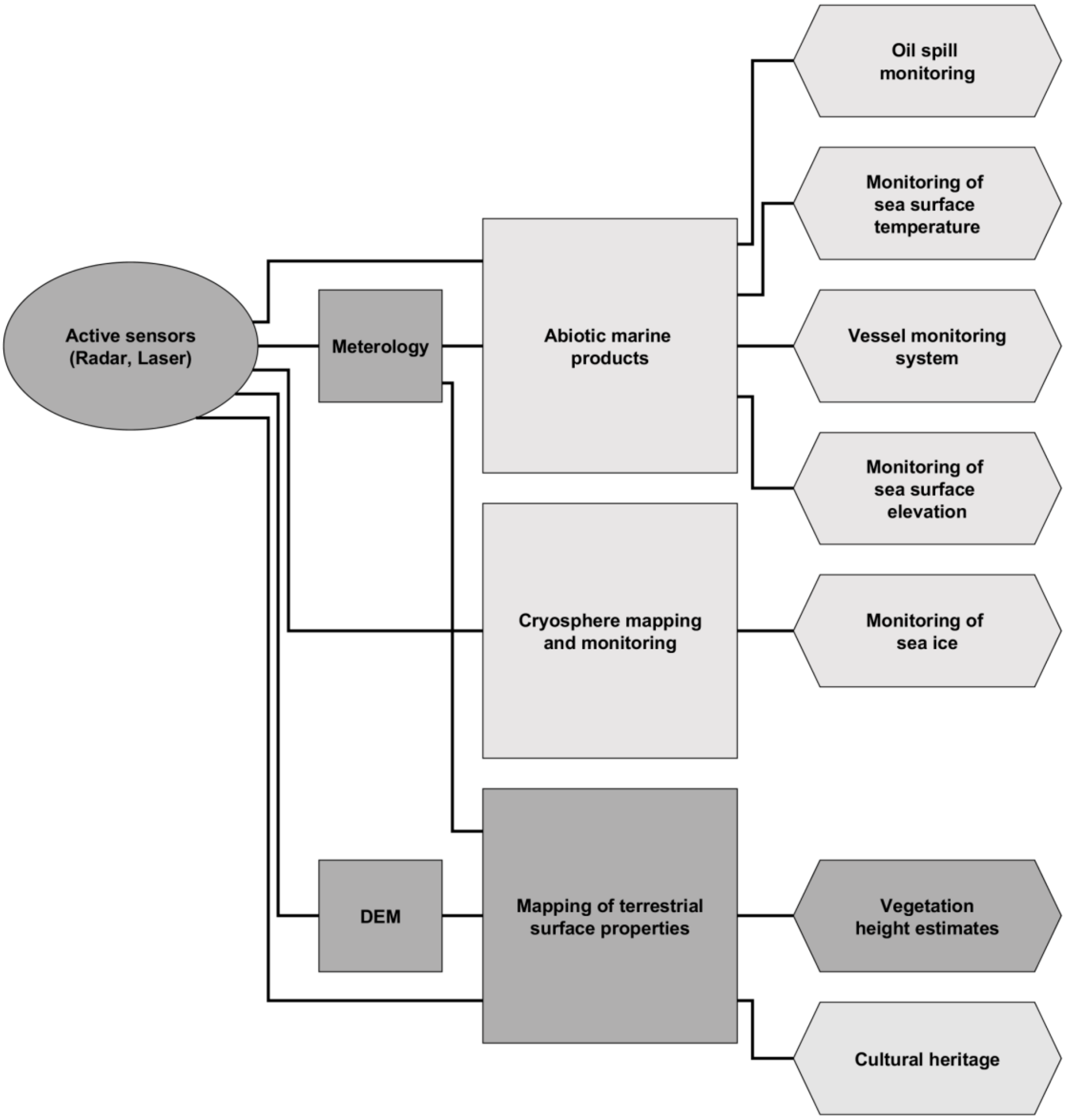


Figure 2 Current uses of data from active sensors operating from a satellite platform.

2.2 Discussion of the different national use cases

The Sentinel-2 mission promises a large amount of data with a wide range of uses. Preparing for the storage of such data and a system facilitating the use of them is a complex task. Use cases are widely accepted and recommended as a forceful and effective tool when developing new and complex systems serving a multitude of needs. Our use cases reveal that an archive with extensions must meet a set of common needs. Refining and correcting the data to the highest level possible, still useful for common use, is the most cost efficient national approach. Reprocessing the same data to the same level in different projects and/or institutions will not only be a waste of time and money: common solutions will also decrease the risk for preprocessing errors that can lead to different interpretations. This might occur due to different quality of toolboxes and data sets used for geometric and radiometric corrections.

The use cases are developed by designated users of data from the planned archive. The Sentinel-2 and Landsat 8 satellites will provide users with large amounts of free data. We find it likely that this is an unprecedented opportunity to design a national satellite data centre (NSDC). A wisely designed NSDC will provide useable data on the highest applicable levels for use and reuse by a number of stakeholders.

2.2.1 Categories of use cases

Based on our discussions we have decided to focus on a total of seven use cases developed by representatives for the various participating agencies in this project. The first six use cases can be divided into four categories:

1. *Presenting images*
   Image of areas with a resolution with same pixel size as the satellite based recordings or resampled to a larger pixel size. Use case 1, “Norway in images”, exemplifies this category.
2. *Mapping*
   Documentation of occurrence and distribution of resources having a known occurrence in areas which can be distinguished from other areas based on the sensor data from Sentinel-2 and Landsat 8. Use case 2, “Lichen pastures in wild reindeer habitats in Norway” and use case 6, “Quality of coastal water”, exemplifies this category.
3. *Change detection*
   Automatized, semi-automatized or manual analyses which may be bi-temporal or multi-temporal. This is basically monitoring activities and may and may not be based on mapping. The analyses may target single pixels or predefined pre-mapped homogenous groups of pixels. Use case 2, “Lichen pastures in wild reindeer habitats in Norway”, use case 4, “Monitoring of forest damage” and use case 6, “Quality of coastal water”, exemplifies this category.
4. *Associative analyses*
   Mappings of which resources are associated with pixels or continuous groups of pixels recorded to have certain reflectance values. Use case 5, “Habitat modeling” exemplifies this category.

In addition to the four categories, a use case describing the processing of satellite data for storage and use in a national archive has been made.

In this report we offer a brief presentation of the different use cases (Figure 3). As a few use cases are within more than one category we find it impractical to divide the presentation into the above mentioned categories. As a consequence we present the use cases one by one below and denote which categories they belong to.

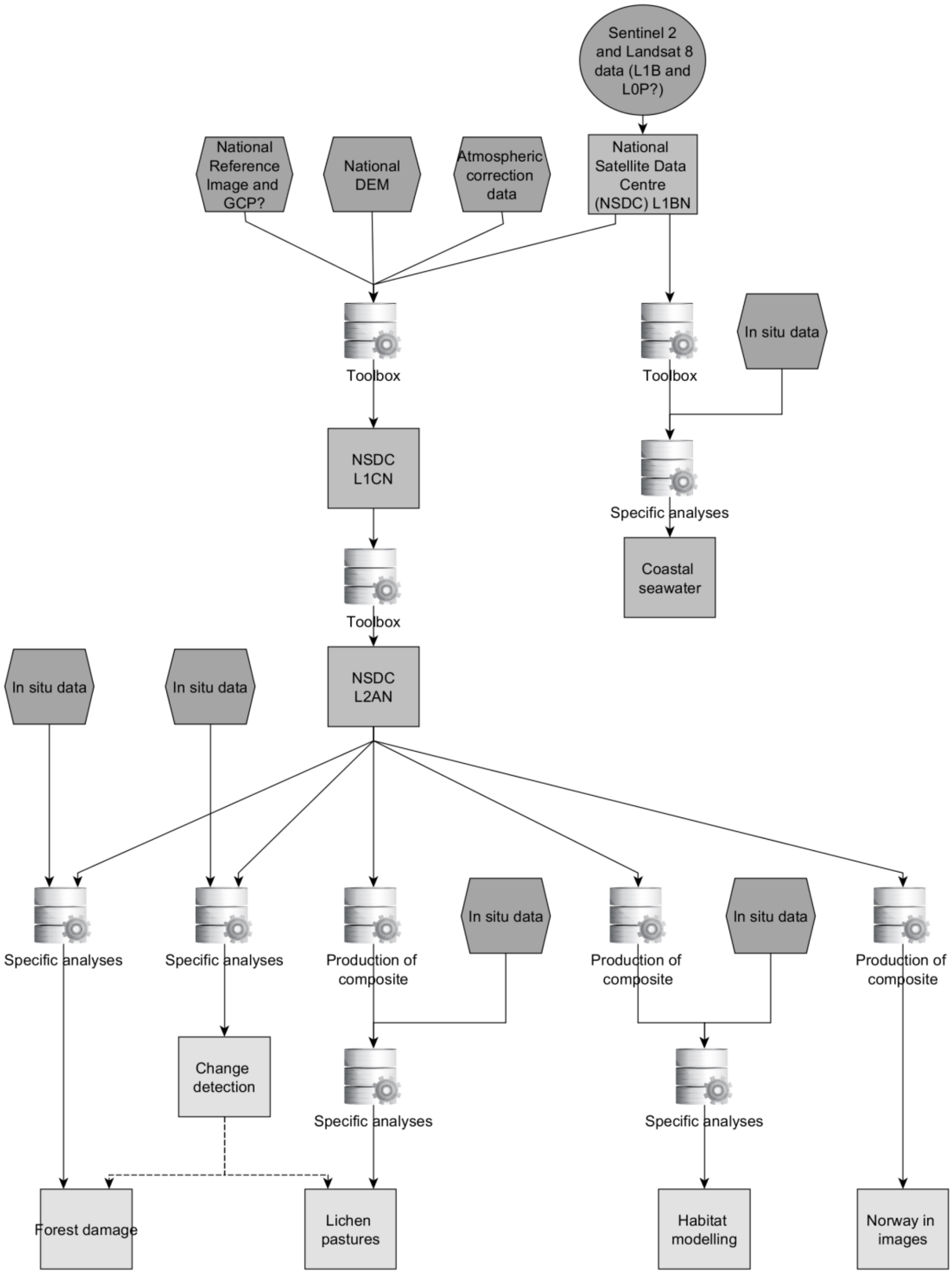


Figure 3 The use cases related to production flow. The use cases are shown in light grey boxes.

2.2.1.1 Norway in images, satellite image mosaic, use case 1

This product will be using images from Sentinel-2 to create a cloud- and cloud shadow free composite recorded during a predefined timeframe. The use case is a purely *presentation of images* (Category 1) based on recordings in the visible band. The result is a presentation of composites made from cloud- and cloud shadow free pixels.

2.2.1.2 Lichen pastures in wild reindeer habitats in Norway, use case 2

The methodology is already developed and implemented based on complete cloud free scenes from Landsat 5 and 7 (see "Current use of satellite data"). Images from Sentinel-2 and Landsat 8 can be used in the production of cloud- and cloud shadow free composite recorded during a predefined timeframe for the areas in question. This use case can be characterized as *mapping* (Category 2) as well as *change detection* (Category 3). The monitoring activity can be pixel orientated or include more elaborate analyses targeting lichen volume and distribution. The methodology utilizes recording in VIS, NIR and SWIR bands. Annual automatized mapping of the volume and distribution of lichen is possible when based on composites from cloud- and cloud shadow free pixels.

2.2.1.3 Detection of changes in reflectance, use case 3

Automatized detection of areas where pixels in cloud- and cloud shadow free parts of scenes have changed reflectance values between two consecutive images of an area. This use case describes a baseline service aiming to provide a wide range of stakeholders with firsthand cost efficient and common ground information on potential land cover changes. The use case represents an automatized *change detection* (Category 3) analyses.

2.2.1.4 Monitoring forest damage, use case 4

This monitoring activity will use data from Sentinel-2 and Landsat 8 to reveal change in reflectance values indicating deteriorating tree health in forested areas. The method is based on continuous pixel-level analyses of cloud- and cloud shadow free pixels in available scenes during the growing season. The use case represents automated *change detection* (Category 3) based on measuring greenness of foliage, using vegetation indices (e.g. NDVI).

2.2.1.5 Habitat modeling, use case 5

This use case describes the utilization of cloud- and cloud shadow free pixels in scenes recorded during the growing season for making composites. These are then analyzed in order to identify areas with relatively homogenous reflectance signature in adjacent pixels. Known occurrence of habitats for different species are associated to those areas providing prediction models. This use case describes an *associative analysis* (Category 4) based on composites made from cloud- and cloud shadow free pixels in available scenes during predefined optimized parts of the growing season.

2.2.1.6 Quality of coastal and lake waters, use case 6

This use case aims at using cloud- and cloud shadow free water pixels from Sentinel-2 in order to investigate water quality and/or create water quality classifications. It is based on BOA reflectance values from coastal- and lake waters. The method is based on statistical analyses of images related to the engagement to the Water Frame Directive. For Sentinel-2 and Landsat 8 data this would be most relevant for particle (turbidity) and secchi disc depth (water transparency) investigations. The activity belongs to the *mapping* (Category 2) categories. Since Sentinel-2 are primarily a land sensor the radiometric sensitivity of Sentinel-2 are weaker than for Sentinel-3 and pixel averaging could be needed for some products.This would affect the geometric resolution of the final product. One could look for a combination of S2 and S3 for a better final product.

2.2.1.7 National archive for Level 1B, 1CN and 2AN data, use case 7

This use case, based on information from the above described use cases, outlines how the archive should function with different levels of products. The use case outlines the archiving of metadata for each product level and for the data used for atmospheric correction, geometric, and topographic correction. In order to ensure national users access to all the available data at any given time the archive must mirror all available Sentinel-2 and Landsat 8 data. This is particularly true for the Sentinel data as they are available online from the ESA-archive only within a limited timeframe.

Summarized this use case reveals:

1. The archive must have sufficient storage capacity to contain all Sentinel-2 level 1B, 1CN and 2AN data, and Landsat 8 data on the same levels, including metadata for all data stored.
2. Analyses aiming to present data from continuous areas need an archive supporting the production of cloud- and cloud shadow free composites from user specified geo-temporal frames (e.g. Norway in images or land cover maps (vegetation class maps)).
3. The archive must support bi-temporal analyses by supporting the production of cloud- and cloud shadow free composites from user specified and discreet geo-temporal frames.
4. The archive must enable the production of time series by delivering all picture data within a requested geo-temporal frame (e.g. the landscape greening process during the growth season or the changing snow coverage)

2.2.2 Discussion of use cases

The use cases covers four different categories for use of VIS, NIR and SWIR satellite images. In addition to the four categories covering the use of images, an extra use case describing the archive and its functionality is meant to clarify how the NSDC needs to be functioning.

We believe that our use cases cover the categories regarding the use of VIS, NIR and IR satellite images (e.g. Sentinel-2 and Landsat 8 data). This implies that other uses will be related to our use cases and fall into one of the four defined categories. As a consequence we do not expect other applications to put significant different demands on the preprocessing, storage and online accessibility to the data and metadata. The use cases do not however directly shed light on the need of an accurate national DEM.

The two most important functionalities of the archive/NSDC according to these use cases will be:

1. Delivery of level 1CN and 2AN datasets, produced within three hours after the data have been downloaded to a ground station.
2. Delivery of cloud/cloud shadow-free temporal composites within an area of interest defined by users.

Sentinel-2 alone will provide the user communities with an unprecedented amount of free data. The use of those data in concert with the free data from Landsat 8 will change the potential use of satellite image radically. Enhanced algorithms for identification of cloud and cloud shadows open up for the use of images with cloud cover in up to 50 % of the image (see page 67). Based on weather data from different geographical regions it will be possible to calculate optimal geo-temporal scales for areas of interest in different regions.

The use cases indicate that with Sentinel-2 and Landsat 8 the potential for cost efficient medium scale analyses, monitoring and mapping of sea and land, is possible in the near future. Background material and common knowledge on remote sensing and the use of advanced GIS analyses, makes clear that the realization of this leap forwards depends on high quality in other data sets. This means that ground reference data must be accurate, the national DEM must be as accurate as possible and the atmospheric correction must be of high quality.

The use cases are indicating a wide use of in situ data in projects utilizing satellite images. Strong arguments can be put to field supporting public availability to those data. On what terms they are made available, how they are technically made available and in which form they are stored and transferred must be studied. A large portion of these data will be important to other users, not only in the original sampling context, but also for new ways of utilizing satellite data. The study must also provide guidelines for how in situ data must be sampled regarding accuracy of positioning, measurement accuracy and other aspects of importance.

The use cases point out that use of remotely sensed data, particularly from satellite programs like Sentinel-2 and Landsat 8, involves important areas of common ground between users. Most projects will benefit from processing the data to Level 1CN and 2AN in a national archive. Availability of online free 1CN and 2AN data from a NSDC is instrumental for cost efficient continuity of already initiated use of satellite data. This analyze will be even more valid for exploratory use of Sentinel-2 or Landsat 8 data aiming to develop more cost efficient research and monitoring methods.

As an extended recommendation for optimizing functionality of the archive, we recommend a separate study on how the toolbox must be designed to optimize preprocessing data to produce the best BOA from L1AN data from Sentinel-2 and L1T from Landsat 8. At the time being we designate the following specifications for a toolbox linked to the archive as part of a NSDC:

- The toolbox must be optimal for Norwegian topographic and atmospheric conditions
- The toolbox must facilitate easy recalculation of data as better algorithms are developed
- The toolbox must be integrated in the architecture of the archive

2.3 Potential use of satellite data

Potential areas that could benefit from use of remotely sensed data are presented in Trollvik et al. (2012). In this report, selected use cases are illustrated in more depth to describe the potential usefulness of data from Sentinel-2.

The spatial and the temporal coverage of satellite data open for an elaborate use regarding monitoring of natural resources (ocean, coastal waters, land and lakes). The geometric revisiting time disregarding cloud cover and with the same viewing angle, will for Norway be less than 10 days for Sentinel-2A (which is the first of two satellites in the first of several Sentinel-2 generations) and less than 16 days for Landsat 8. Norway is situated far north and thus we will have a substantially shorter time lapse between revisits by the same polar orbiting satellite than areas closer to the equator. But due to high mountains and narrow fjords, there are some challenges for pixels above water in some coastal areas of Norway. The mountain shadow can for some fjords and valleys or parts of some fjords and valleys, cover most parts of the cross section, depending on the satellite viewing angle and the solar zenith angle. Future developments of DEMs could possibly correct for some of these challenges.

Large amounts of satellite data are already available and substantially more will be available in the future. Still, it is important to underline that satellite data cannot fully replace in situ sampling of data. The examples listed below show where data from satellites will contribute to further development of mapping and monitoring national resources:

- Mapping and monitoring of large river deltas. Changes in the use of land indicate planned or non- planned modifications.
- Monitoring of inland waters. Ice cover, lake water quality or water extent data for a lake from remote sensing.
- Mapping and monitoring of intervention free areas in Norway to protect biodiversity. Indication of changes in the use of land within the determined criteria for calculations of such areas.
- Monitoring of cultivated landscape through landscape changes.
- Monitoring of permafrost mires. Changes in vegetation in areas with permafrost mires can indicate degradations of such and could be caused by climatic changes.
- Loss of cultural monument due to non-planned changes.
- Secure biodiversity land through indications of land damage due to motorized traffic outside approved roads/areas.
- Potential harmful alga blooms (HAB). Forecasting and/or predictions for fisheries and aquaculture.

3 National satellite data centre

3.1 Arguments for a national satellite data centre

The use of satellite images requires a system for storage and retrieval of the image data and associated metadata. Today this can in principle be made on a desktop computer. Small scale handling of satellite images is however not an option when addressing the dataflow expected from Sentinel-2. It will provide many terabytes of data yearly over Norway. We expect that the use of remotely sensed data will increase with the introduction of Sentinel-2. All users will have to handle a large amount of data, including search and download from ESA's archive, orthorectification, atmospheric correction, data organization in a data storage system, search and access solutions, development of the methods and algorithms, purchase of necessary hardware and software, maintenance fees, and building up of expertise to handle the different processing steps, development, and data service.

The use cases (Chapter 2) demonstrate the current trend in image analysis of increased focus on change detection and use of time-series data to monitor development, e.g. multi-year phenology. Use cases not focused on change detection also show a need for images from several dates in order to make cloud-free or near cloud-free time composites. Thus, all use cases demonstrate the need for data from multiple dates. However, use of multi-temporal datasets requires precisely co-registered image data and atmospherically corrected data converted to surface reflectance. Precise co-registration can be obtained by using a detailed and accurate elevation model to resample the data in the orthorectification process.

Trollvik et al. (2012) concluded that ESA will not provide data fulfilling all these requirements. To resolve this dilemma two different solutions are possible:

1. The users will take care of all the necessary preprocessing, data storage, and data service.
2. The users will collaborate and establish a central processing unit that takes care of all necessary preprocessing, data storage, and data service.

Solution No. 2 will take the burden from each user to search for data in an international archive, preprocess the data to user ready form, and finally to archive the data in a data storage system. We have made an estimate of the costs of establishing a national solution in Chapter 3.7. Significantly higher cost with solution No. 1 is to be expected due to duplication of efforts, and the solution will also result in differing and often suboptimal procedures. Thus, we conclude that as a whole, the partners of Norway Digital are likely to save money and get a better solution and services if solution No. 2 is realized. A national satellite data centre for Sentinel-2 and Landsat 8 data is suggested by the technical committee. The main motivation is to facilitate easy use and access of data from Sentinel-2 and Landsat for Norway Digital partners, i.e. to realize plug-and-play functionality of the valuable datasets. The proposed service will also help to secure that Norway gets dividends from our share of the large investment in the space segment.

3.2 Definition of the main service principles

The analysis of the use cases and the necessary preprocessing steps reveals the need for a national satellite data centre capable of processing large volumes of satellite data. The satellite data centre must be capable of continuous processing and delivery of complete image- or pixel-based subsets to end users.

Quality issues related to for instance cloud cover pushes forwards increased use of products based on pixel-level information from several images. As a consequence, the archiving must support user defined geo-temporal requests, picking pixels of the best quality within a defined timespan. This view is strongly suggested by the use cases and by general trends in use of satellites data with short revisit time.

The suggested archive must meet the following principles:

- At any time online availability of original images
- Full temporal archiving allowing for retrieval of historical data
- Metadata must support pixel level use of the data
- High service reliability
- High service scalability
- Support standards relevant to distribution of spatial data support

In short we envision a system (Figure 4) where the information flows from satellites to the end user through a series of processes and storage steps making the data available both in image and pixel-based products.

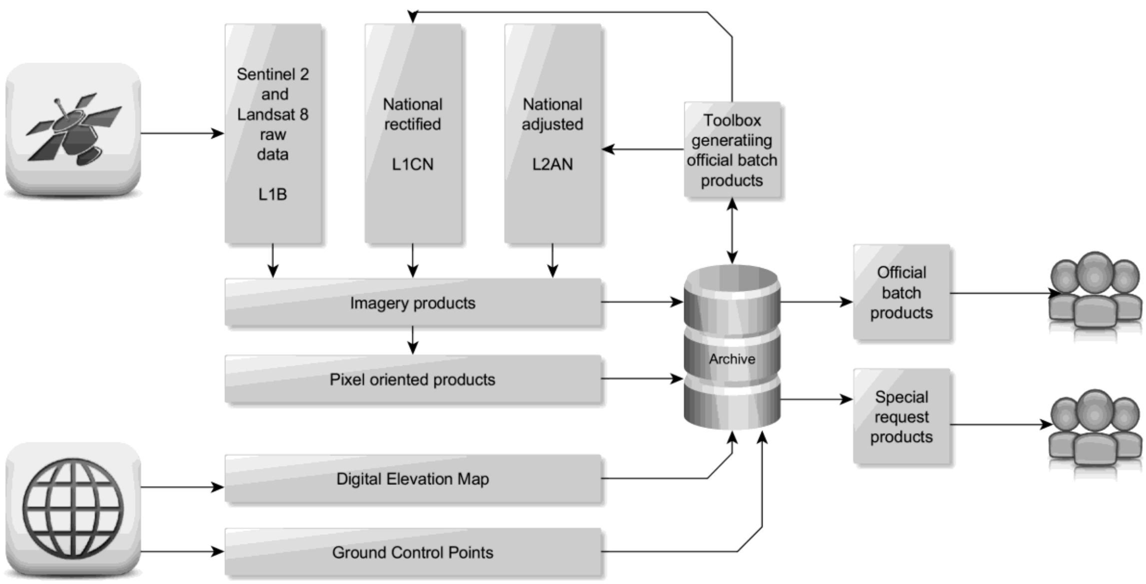


Figure 4 Overview of the path for data from satellites to the users.

In the following three sub chapters we will establish a context for the National Satellite Data Centre, briefly describe how we envision the process towards the data centre and consider some restraints and opportunities related to the future data centre.

3.3 Collaborative ground segment and the National Satellite Data Centre

ESA is offering member states a collaborative ground segment as a mechanism for improving on the product services provided by ESA's core ground segment. ESA has developed the Data Hub System (DHuS), software to allow partners to access Sentinel-2 data over a fast network through a dedicated hub. The partners can also use the software to manage their own mirror site. The DHuS is designed to enhance the speed, quality and capacity of data access (Figure 5).

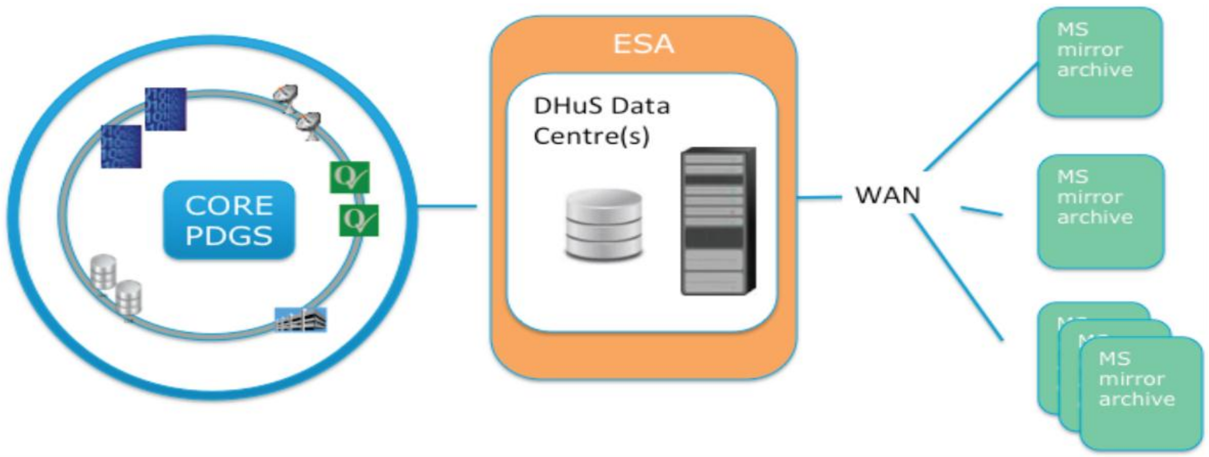


Figure 5 Collaborative ground segments are connected with ESA's core ground segment through a dedicated hub (DHuS) over a fast network (WAN), which gives near real time access to core products. Collaborative ground segments can establish mirror archives to secure fast data access for national end-users to both new and historical image datasets (figure from ESA, 2013).

A Sentinel-2 collaborative ground segment will be an important element of the proposed national satellite data centre. ESA plans to support the collaborative partners for the use of the data hub system, including possible customization, e.g., to set up scripts for systematic download (ESA, 2013).

The National Satellite Data Centre's (NSDC) main tasks will be to:

- download data from ESA's core ground segment
- pre-process data
- archive data
- provide data search interface
- disseminate data to national end users

A model based on the NSDC will:

- operate faster than ESA's core archive
- provide online data further back in time than ESA's core archive (more than 100 days)

- deliver data with higher geometric accuracy
- provide pixel oriented products
- provide user defined geo-temporal composites
- offer storage and preprocessing of optical satellite data from additional satellites
- consider a link to national in situ data

The main pre-processing tasks are orthorectification using an accurate national DEM and atmospheric correction to calculate surface reflectance. In addition, it is recommended that the NSDC also provides time composites of cloud free pixels (e.g. from data over a two to four week period) and multi-temporal datasets for time series analyses, e.g. phenology studies or change detection.

3.4 Process to establish the national satellite data centre

In this report we have so far shown the necessity of a centralized system to handle images towards operative products. Our use-cases provide examples outlining current and potential relevant products as well as toolbox systems for preparing some of the products.

Developing an archive with integrated toolboxes is a comprehensive task. The extent of the operation presupposes careful planning. It must include interaction with stakeholders, assessment of technical/economical restraints, development of a business model, integration in Norway Digital and more.

The process (Figure 6) should be funded as a research and development project led by a prospective technical facilitator.

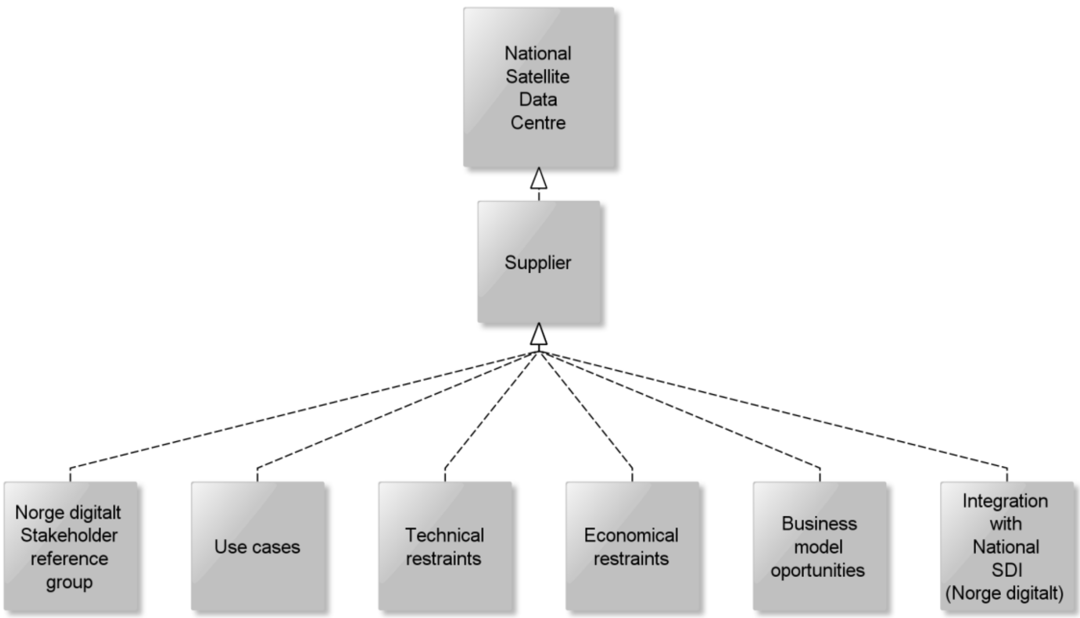


Figure 6 Overview of process to establish a National Satellite Data Centre.

The following process steps are suggested to carry the project forward:

- Dialogue with research and development operator to establish a partnership aiming at developing specifications for a National Satellite Data Centre
- Decision on stakeholders for reference group
- Develop use cases relevant for the process
- Funding of the process
- Timeframe considerations

3.5 User requirements

In the following chapters we aim to shed light on issues which should be brought into the process outlined above. The issues raised are considered relevant as basis for terms of reference to a development partner.

3.5.1 Data discovery and access

To serve its users and also to attract new users the data management solution should facilitate for easy access of the contained information. This can be done in two ways by

a) establishing a dedicated web-front end; and

b) supporting metadata standards which make data search and data retrieval possible across different platforms.

The retrieval of spatial data presupposes the operator's knowledge about the product and his/her ability to send a request to the data warehouse. A request to a warehouse would typically be through a web interface with map allowing the user to include spatial and temporal extent, spatial resolution, quality and more. The automated reply from the warehouse would usually contain more metadata making it possible for the user to evaluate the data and choose relevant image for purchase or direct download.

Current retrieval systems seem to be vendor centered, focusing on delivering images to clients from specific satellite solutions. This means that a user will have to visit several satellite data warehouses to retrieve all relevant images.

Some major satellite archives allowing for download of data worth mentioning are Earth Explorer from USGS[1] (Figure 7), Satellite Rapid Response System from Chelys[2]. A system from MapMart[3] allows for search across several commercial image providers.

[1] http://earthexplorer.usgs.gov/
[2] http://www.chelys.it/products/satellite-rapid-response-system/
[3] http://www.mapmart.com

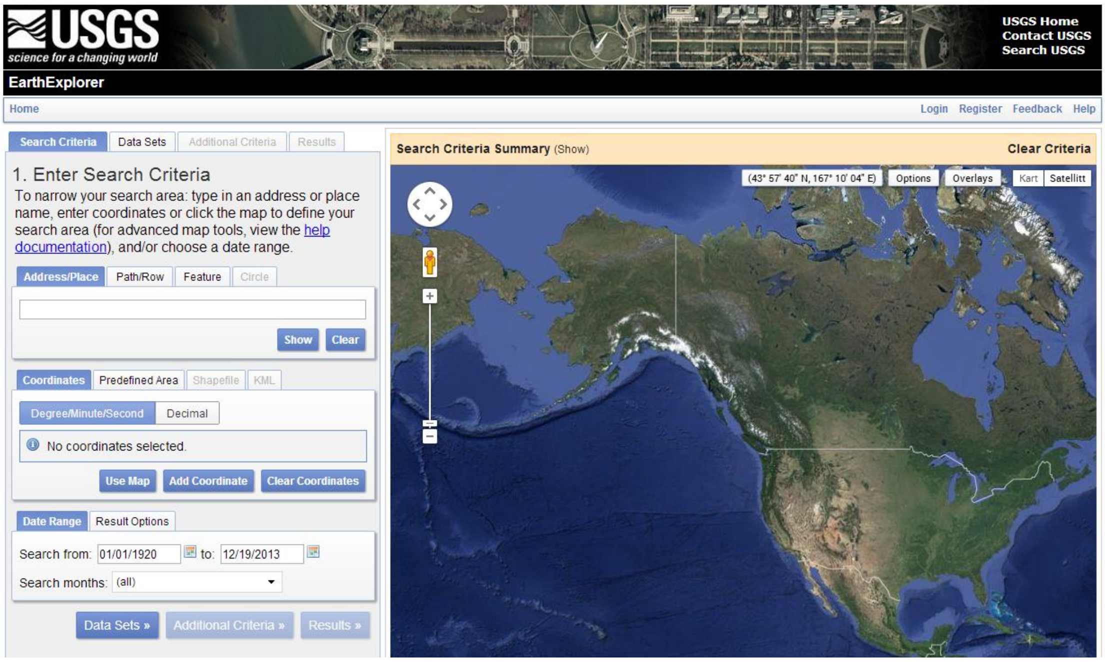


Figure 7 USGS Earth Explorer user interface.

Lastly some data are only available through manual request by phone or email to a sales representative thus increasing the consumption of resources for acquiring images.

For the product levels 1CN and 2AN, the archive system should be searchable through an OGC (Open Geospatial Consortium) catalogue service interface. Data access can be supported by NS-EN ISO 19128 OGC Web Map Service (WMS) or an OGC Web Coverage Service. WMS portrays geo-spatial data as static pictures on the client side (server renders the data) while WCS provides the actual data, and the client can render the data and input it to models for analysis. WCS supports basic spatial and temporal extractions and there are two basic subsetting operations that can be combined: *trimming*, which extracts a subarea (pixel or multi-pixel objects) of a dataset (delivers a coverage with all original dimensions inside a bounding box specified by a query), and *slicing,* which extracts a reduced dimension of a dataset (delivers a coverage with reduced dimension, e.g. only three bands from a seven band satellite image). Product level 3A is composed of pixels from more than one acquisition, and dynamic composition of level 3A products can be supported by the query and functional programming language XQuery. A service interface based on OGC WCS and XQuery can support extraction of arbitrary areas and time intervals (as pixels), as well as a standardized syntax for simple user-specific statistical analysis on-demand. This would require a database solution – a multi-dimensional data cube – where new image products will be ingested into seamless spatio-temporal regular grid storage.

A modern and effective data management platform to store and archive many thousands satellite images is needed. The platform must support the requested user functionalities (e.g. multi-dimensional subsetting). This includes full query support for

- searching for datasets in a spatial and temporal window
- searching for pixels that are cloud free inside a defined time window

- composing a time composite of cloud free pixels where available

The platform must support handling of metadata following the datasets, that is, the acquisition and processing history of each pixel is available and searchable.

Relational database systems are the most common today, but they do not directly support array data and cannot provide full query functionality. Binary Large Objects (BLOBS) database is another option, but it is equivalent to files with long strings of bytes and does not have query language functionality.

An array database management system (array DBMS) is developed to store and handle array data (raster data) that is composed of pixels sitting on a grid of one, two, three or more dimensions. Satellite images are a typical case, and are composed of pixels in two dimensions (x, y). Multiple images over the same area can be represented as triples, i.e. (x, y, t), where the "t" denotes the time dimension. Thus, a time series dataset is defined as a multi-dimensional subset of the array.

Raster data manager (Rasdaman) was the first implementation of an array DBMS to serve large multi-dimensional arrays. Rasdaman enables storage of multi-dimensional arrays in a conventional database and retrieval through its embedded SQL array query language. With Rasdaman it is possible to set up a 3D-cube with satellite image with full query functionality, e.g. such as subsetting of a 3D-segment defining a time-series dataset. Rasdaman is available as both an open source and commercially supported variants.

If all Sentinel-2 and Landsat datasets over Norway are stored in mirror archives, the data volume will be huge (Chapter 3.7), and it is uncertain if a database solution will be possible both from a technical and economical point of view. Will a raster database be able to cope with several hundred terabyte of data and will the costs to establish and run such a system be reasonable? We encourage further research into what metadata, interfaces and related processes should be supported by a national archive. Chapter 3.7 outlines several options for a data management platform.

3.5.2 Production line details

Acquiring a satellite image, processing it and finally storing it in a data store require large capacity storage and processing capacity. Figure 8 is used to illustrate a use case describing a national archive for level 1B, 1CN and 2AN data and metadata and represents a schematic view of the production line.

It reveals that not only must the data centre handle images, but also a number of other datasets important for preprocessing and for different uses of products derived from these data. Metadata regarding the different level products are important for declaring accuracy and quality of the data.

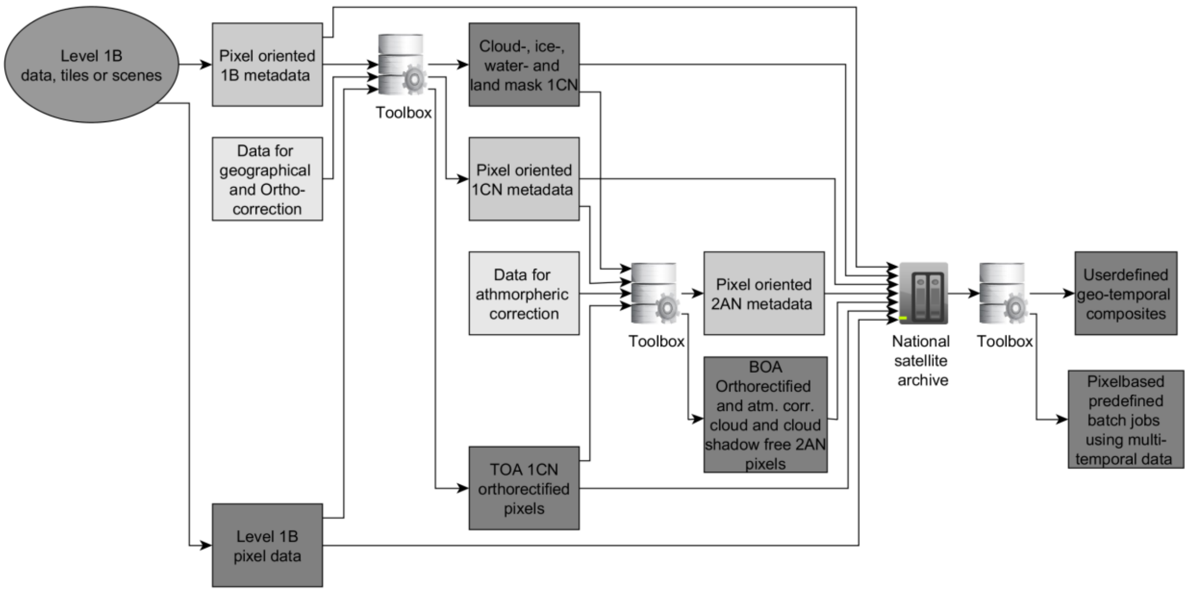


Figure 8 Satellite images from Sentinel-2 processed and stored in a national satellite data archive.

Sentinel-2 will potentially change how optical satellite data are used. High repetition frequency, wide swat and relatively high resolution open for more pixel based sampling, archiving and analyses. Pixel based use of the data will put different demands on the NSDC compared to traditional tile or scene based dataflow. In the user end we expect two sorts of products:

1) Predefined batch jobs according to the use cases mentioned in Chapter 2.2.1
2) Parameter based requests by users, e.g. geo-temporal composites as requested by users

An important challenge in such a product line is to balance the individual users pull on general system resources. It is not within the scope of this report to work out the details of how pixel based use of satellite data can be facilitated. A well designed pilot, engineered in close collaboration between users and a competent team of specialists should however clarify how this can be done.

3.5.3 Functionality and products

The main sources for satellite images the next five to ten years will be Sentinel-2 and Landsat 8. There is however reasons to expect a demand to allow the satellite centre to store and handle data from additional satellites. This will include both previously collected data and more recent images. We recommend that the first version of a National Satellite Data Centre is restricted to handle data from Sentinel-2 and Landsat 8 data.

The archive will be focused around image and pixel-level deliverables. Focusing on the pixel level is believed to be the only progressive way of handling the new Sentinel-2 datasets. In order to be able to do all and any repetitive analyses on the same area referring pixels to a standard national grid as part of a preprocessing is the only cost efficient approach.

Service oriented architecture helps to allow a transition from an application-centric view of the world to a process centric one. This means that clients (users) will be able to draw on resources without using separate applications but rather rely on standardized services or where standards are not developed provide well described functionality for known technical frameworks.

3.5.4 Integration with Norway Digital

There is a need for a national archive for storing 1B, 1C and 2A data and metadata from Sentinel-2 and Landsat 8. The archive architecture must support a cost-efficient use of these data by members of Norway Digital. The use cases reveal that a modern satellite data archive must have an architecture supporting a pixel-focused use rather than focusing on storing tiles or single scenes. This is supported by general international trends in viewing remote sensed data. As illustrated previously in this report, the pixel-focused use and the need for automated preprocessing calls for solutions with a scope beyond a classic archive notion.

Most uses of remote sensed optical data require in-situ data. In-situ data may be aerial photo, low altitude images, reference measurements, surface or substrate classifications and more. They may be linked to research, surveying, mapping or monitoring. Many in-situ data are collected in the course of publicly paid projects. It is advised that situ data should be sampled and reported according international and national standards.

Internationally this question has been addressed in the Copernicus Earth Observation program. The GMES in-situ coordination project was started in the European Environmental Agency in January 2010 and ended on October 31, 2013 (http://gisc.ew.eea.europa.eu/). The importance of easy access to standardized in-situ data has also been addressed by EuroGeographics in their briefing paper: "Proposal for a Regulation establishing the Copernicus Programme" [2013/0164 (COD)].

The technical committee has defined storage and extensive metadata bases related to in-situ data outside the scope of a NSDC. The technical committee suggests that the need for easy access to in-situ data relevant for image analysis is supported by partners of Norway Digital. This could be done by establishing a separate product describing relevant in-situ data either on institution level or as a collaborative between several Norway Digital partners. The product should then be available through the geoNorge metadata portal for Norway Digital spatial data. The technical committee considers this product as an important and relevant service for users of data from the NSDC.

3.5.5 Relevant standards

The importance of adhering to current standards has been mentioned several times in this document. We would like to stress the importance of doing a thorough review of current standards of relevance to the establishment of a National Satellite Data Centre.

The data content provided by the national data centre does probably not fall into any of the themes defined in the Norwegian Geodata Act (which is an implementation of the INSPIRE directive). However, since the Stakeholders (data owners) are primarily governmental, the principles stated in the Geodata Act should be investigated and implemented if applicable.

The Geodata Act states that the infrastructure provided should be available for stakeholders and themes described. For example, availability of metadata at the national geo portal will enhance the usage of these data and enable the technical interoperability. The following standards are candidates and should be fully investigated:

- INSPIRE Specification on Geographic Grid Systems (http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_Specification_GGS_v3.0.1.pdf.) This document describes a harmonized multi-resolution grid with a common point of origin and standardized location and size of grid cells. It covers quadrilateral grids used for the indirect geo-referencing of themes with typically coarse resolution and wide (pan-European) geographic extent. The grid is two-dimensional and mainly used for Spatial Analysis or reporting. In Norway a national grid system has been defined and is described in Appendix B.
- INSPIRE metadata http://inspire.jrc.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf. This document is a profile of ISO 19115 Metadata and ISO 19119 Services and provides technical guidelines for how metadata for data and services should be provided. If not the INSPIRE profile fulfills the requirement, the full ISO 19115 should be investigated. A revision of ISO 19115 as ISO 19115-1 is available at DIS (Draft International Standard), and should be considered.
- View service http://inspire.jrc.ec.europa.eu/documents/Network_Services/TechnicalGuidance_ViewServices_v3.11.pdf. This document specifies requirements and recommendations based on the European de jure standard ISO 19128 – Web Map Service (WMS). It defines an INSPIRE profile of ISO 19128 WMS.
- Discovery service.

  (http://inspire.jrc.ec.europa.eu/documents/Network_Services/TechnicalGuidance_DiscoveryServices_v3.1.pdf). This document specifies requirements and recommendation based on the OGC Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile.

The referenced INSPIRE specifications are technical guidelines that gives the technical details to ensure that the requirements in the directive / Geodata Act are fulfilled.

Other standards (like download, etc.) may be considered if required.

3.5.6 Previously acquired scenes

A number of satellite scenes have been acquired over the years by the partners in Norway Digital for various uses in Norway. The acquisition of those data is a result of tedious work put into finding scenes of sufficient quality, from the right area and within the right time span. The cost in the term of money when acquiring scenes from commercial satellite missions may also be significant. They represent valuable retrospective data. The archive should be able to store these data, according to the same standards as Sentinel-2- and Landsat 8 data. Development of a solution for storing them in the data centre is a necessary part of the pilot for the NSDC.

Previously acquired data hold some additional aspects not directly related to the NSDC, but never the less in need of attention. The partners in Norway Digital need to gather all previously acquired satellite images. Acquisitions of the data from commercial satellite missions may be for a restricted number of users. The license agreements should be renegotiated for those data and we propose that contingence plans are made for financing free public use of the data.

Needless to say, all this goes for optical satellite data acquired from other EO-programs than Landsat 8 and Sentinel-2 in the future.

3.5.7 Service quality and capacity

It is important that a National Satellite Data Centre has a high uptime and is able to absorb demanding requests. To understand what level of availability and data volumes are necessary, it is suggested that this issue is given focus in the planning of the centre. One should also consider user access levels to balance capacity with user resource needs.

3.6 Estimated data volume of image tiles

Trollvik et al. (2012) suggested archiving national levels 1CN and 2AN (Figure 9). In addition we have been recommended (C. Brockmann, pers. comm.) to store level 1B as well, in case reprocessing is needed at a later stage. ESA will only keep level 1B data in the on-line archive for a limited time (few months) and later it will be stored in an off-line, long term archive. Thus, historical level 1B data will be more difficult to procure.

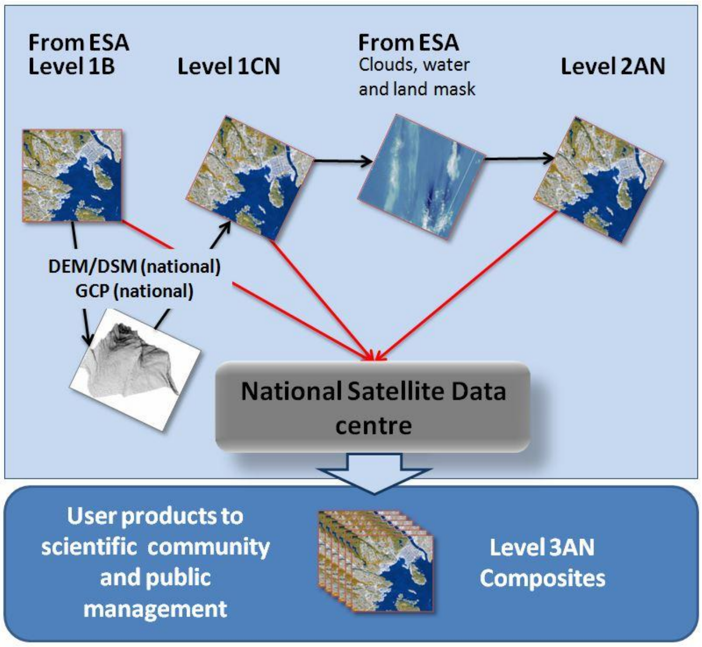


Figure 9 Different Sentinel-2 levels to be archived in the National Satellite Data Centre.

We suggest organizing orthorectified data products in image tiles using the SSBgrid system of spatial tessellations (Appendix B). SSB100KM is a tessellation with 100 x 100 km tiles and is shown in Figure 10. To fully cover the land area of Norway and coastal waters 73 SSB100KM tiles are needed.

Two advantages of using SSB100KM tiles are the following:

- pixels will be aligned on a common lattice and ready for time-series analysis and change detection, and

- data volume for each tile is not too big for easy transfer over a network.

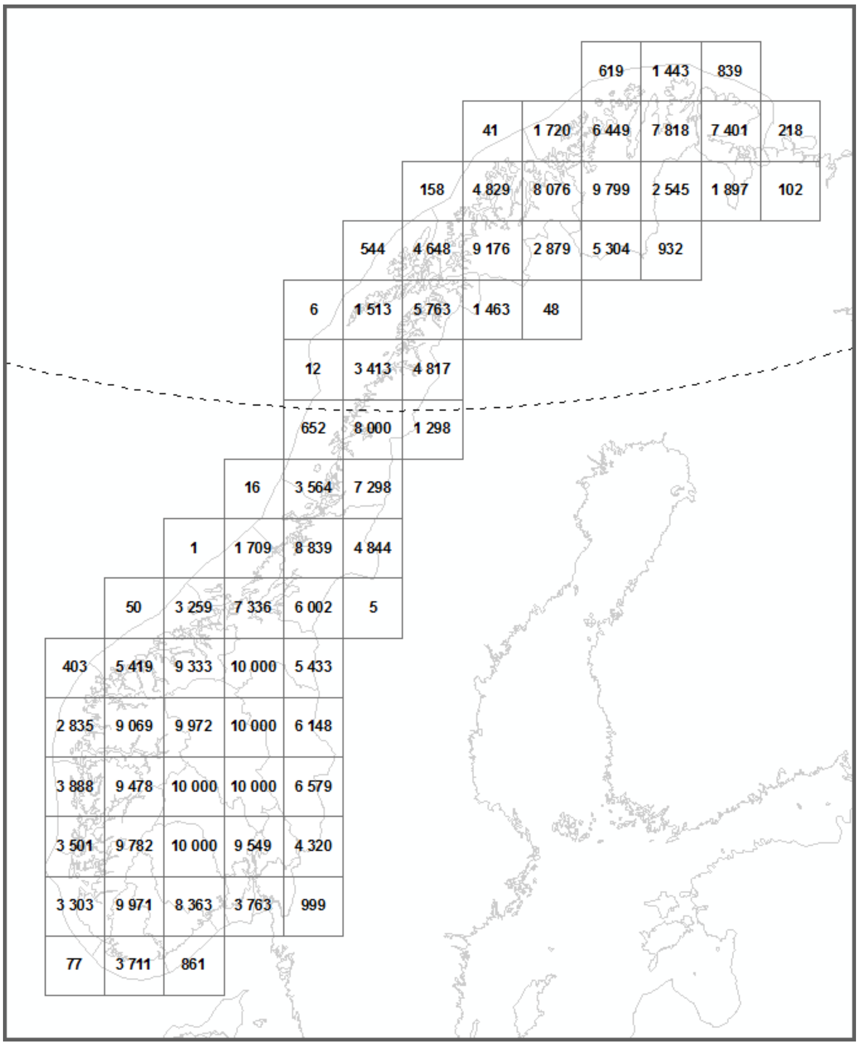


Figure 10 The 73 SSB100KM tiles that together contain the Norwegian mainland and coastal waters. The numbers inside the 100 km x 100 km tiles show the area [km^2] overlapping mainland Norway. The stippled line is the Arctic Circle. The area of Norway's mainland is 323 787 km^2 while the total area of the tiles is 730 000 km^2.

In Table 1 estimated data volumes for tiled Sentinel-2 and Landsat 8 data are presented. We have assumed 16-bit raster pixels, no compression, and all spectral bands in original pixel size (Landsat 8 thermal bands are assumed resampled from 100 m to 30 m). Some additional bands can be expected for product levels 1CN and 2AN, e.g. quality band, land/water mask, and cloud/cloud shadow mask, and including these would increase the volume estimates somewhat. We see from Table 1 that all tiles together will weigh about 82 GB with Sentinel-2 data and 23 GB with Landsat 8 data.

At the Equator, the revisit frequency is every 5 days for the Sentinel-2 mission (both satellites 2A and 2B operative) and every 16 days for Landsat 8. The Landsat mission consists of two satellites, Landsat 7 and Landsat 8. Landsat 7, launched in 1999, is still operative, but the scan line corrector stopped to work in 2003 leading to images with a "zigzag" appearance with nodata areas in both ends of the scan lines, and about 22 % of an image is therefore missing. We have therefore not included Landsat 7 in the volume estimates. A preliminary plan exists for Landsat 9, but no decision has yet been made to finance it. We therefor count on only one Landsat satellite for the next 5 to 10 years.

*Table 1 Estimated data volume of SSB100KM tiles (100 km x 100 km) for one product level. *denotes average number of effective revisits during a year at the geographic centre of Norway.*

	Sentinel-2 (2A + 2B)		Landsat 8	
Revisits	1	166.5*	1	52.0*
Data volume 1 tile	1.12 GB	186.01 GB	0.31 GB	16.18 GB
Data volume 73 tiles	81.54 GB	13.58 TB	22.70 GB	1.18 TB

During a year Sentinel-2 (with two satellites) will repeat a path 73 times and Landsat 8 22.8 times. At Equator the overlap between neighboring paths is small but the overlap increases with latitude L (see Figure 43 on p. 104) because Earth's circumference reduces with cos (L). In Figure 11 the average number of effective revisits during a year is shown. At 60°N the average number of revisits has doubled from 73 to 146 for Sentinel-2 and from 22.8 to 45.6 for Landsat 8; and at 64°N, which is the geographical center of Norway, the effective revisits are 166.5 and 52 respectively.

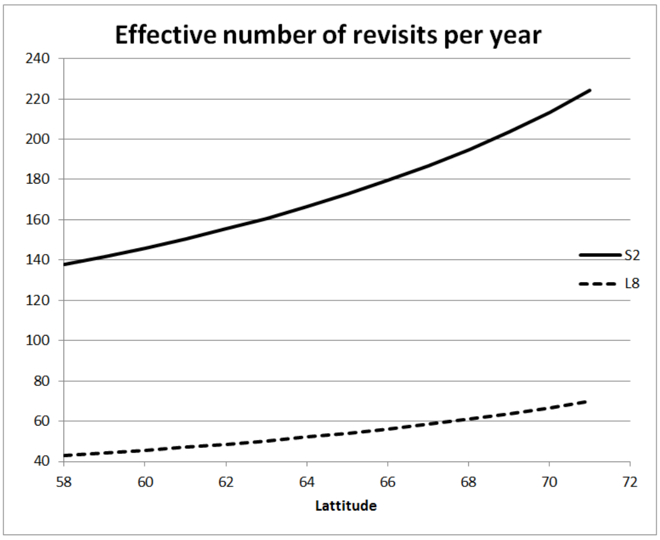


Figure 11 The average number of effective revisits during a year increases with latitude because the overlap between neighboring paths increases with latitude. S2 denotes Sentinel-2 (two satellites) and L8 Landsat 8.

The increase in data volume for one product level is approximately 13.6 TB/year for Sentinel-2 and 1.2 TB/year for Landsat 8 (Table 1). The total estimated data volume for the two missions together (two Sentinel-2 satellites and one Landsat satellite) is approximately 15 TB/year for one product level and 45 TB/year for three product levels (1B, 1CN, and 2AN).

In Appendix E similar estimates are made but based on a smaller area (400 000 km^2 corresponding to Norway's mainland plus a buffer), two Landsat and two Sentinel-2 satellites, and image file size (see p. 101). The data volume estimate is 8.5 TB/year for one product level and 25.5 TB/year for three product levels. An estimate for system storage requirement, which includes backup disks for a secure storage, is presented, and the physical storage requirement is 43 TB/year for one and 130 TB/year for three product levels.

Two facts which can lead to significant reduction of data volume are:

- Many acquisitions will have very high cloud coverage (see Figure 41) which prevents atmospheric correction (cloud cover < 50 % needed for atmospheric correction with ATCOR). We can assume at most 1/3 of the acquisitions will have cloud coverage less than 50 %; and
- Low light conditions in parts of Norway during wintertime because of low sun angle or polar night. We assume that about 10 % of the acquisitions will be too dark to process.

3.7 Outlines of technical data management solutions

3.7.1 Introduction

How do we build a system that enables specialists across Norway to access, integrate and analyse satellite images of the country in a timely manner? In particular, how do we select an appropriate system, that will fulfil user needs as much as possible, and which can be built quickly, at reasonable cost, with a high chance of project success?

This technical report attempts to answer these questions. The first parts of this text, Chapters 3.7.2 and 3.7.3, consider the data management problem as a black box problem – i.e. what are the user's requirements from such a system, from their own point of view? What technological problems must the system overcome?

In Chapters 3.7.3.2 and 3.7.4, a menu of potential technical solutions is presented with notes regarding benefits and costs, followed by a recommendation. A brief project schedule outlines actions to be taken subsequent to project funding. Further information and references are found in Appendix E.

3.7.2 Characterisation of the problem from the end-user perspective

Norwegian remote sensing specialists presently retrieve and process Landsat satellite sensor data from the USGS (United States Geological Survey) as they require it, when studying land status. Data is retrieved directly from USGS web servers. Typical processing might include applying filters to sensor data, cloud cover removal, and image compositing (spatial or temporal).

ESA will soon launch the new Sentinel-2 satellite system. This system consists of a pair of satellites which record data over a broader east-west area, at higher resolutions, covering more spectral bands, and with a much faster average revisit time compared to Landsat.

Currently specialists must retrieve and process data themselves on an adhoc basis from the published data products of the USGS. In the future, they will also need to access products from the ESA in order to make use of Sentinel-2. The two sets of data products will be used together and will likely undergo further processing. In addition to this, some specialists will be sharing processed data. This is illustrated in Figure 12.

There are four consequences resulting from the introduction of Sentinel-2 that will affect end-users. Firstly, there will soon be far greater amounts of data than ever before. Secondly, the data is arriving according to two schedules. Thirdly, the data is in different formats with different provisos and will require different post-processing. Fourthly, the data files must be gathered from different sources – which affect operations like downloading data, searching for data, comparing metadata, and so on. This, along with increasing interactions with other users, will require an increase in skills, effort and attention from all users in order to make good use of the data.

Since all of these consequences will affect the work of many specialist groups around Norway on a daily basis, and since the impact on each group may be quite similar, it would make sense to approach this data management problem systematically at a national level. The aim would be to simplify and standardise the process of data retrieval and processing for end-users of Landsat and Sentinel-2 data throughout Norway.

As the deployment of Sentinel-2 is taking place in 2015, there is an excellent motivation to take action now. It is the perfect opportunity to begin consolidating and organising the ways in which satellite images are gathered, processed, stored and used across Norway, before the data problem becomes unmanageable for remote sensing specialists and other users. This raises a question: what would be ideal at a national level, in terms of improving user productivity, standardising research data inputs and outputs, and reducing incrementally incurred costs of data use throughout Norway?

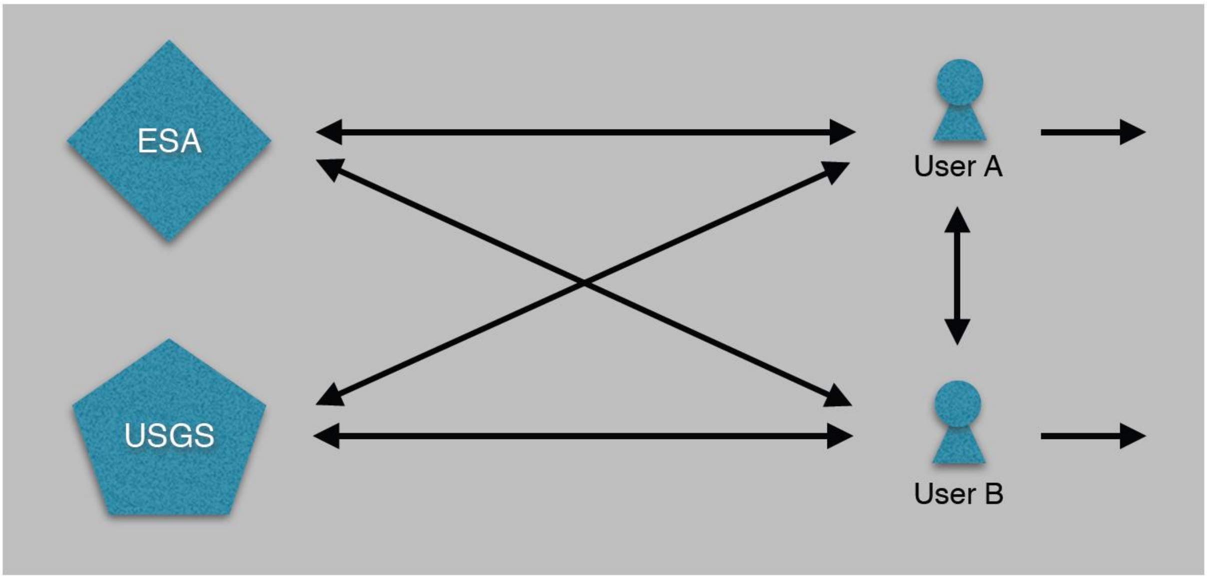


Figure 12 Presently, users will have to connect to USGS and ESA servers to locate and download sensor *data. They will process data by hand, and share materials with local and external users manually. The working environment will become increasingly complex for users throughout Norway as the number of data sources, products, processing workflows and total users increase. The distinct shapes used for ESA and USGS indicate distinct types of data and access systems.*

The most obvious answer is to place some type of data management platform between the user and the data sources. This platform would remove complexity, improve data access times, and assist the user with common data operations (Figure 13). A "Norwegian Satellite Data Platform" (incorporating a Norwegian Satellite Data Centre as the primary component) would simplify, standardise and aggregate information about all relevant satellite images available for Norway, and would automate prompt and accurate provision of data items as they are required.

Potentially, it might also be designed to perform other duties. These could include post-processing or analysis of data; spatial or temporal image compositing; long term archival; search facilities; notification of new data; sharing of results and processed data; warnings about missing data or poor quality data; and so on. It would also be straightforward to provide this system in a localised manner for Norwegian users – e.g. including Bokmål, Nynorsk and English – which might be more user-friendly for some users than an English-language-only based international download system.

From the user's perspective, the platform should seem like a useful black box. The user begins by accessing the platform and searching or browsing for the material they need for their current work. The platform then finds the correct data, helps the user confirm that it is correct (e.g. thumbnail preview, summary statistics, metadata), perhaps pre-processes it, and then delivers it as quickly and conveniently as possible to the user.

The national data platform might be designed in a variety of ways. It might have the form of user friendly desktop software for managing the data workflow locally, with plugins for commonly used desktop software. More likely though, the platform would be on the scale of a web-based national data cache or data archive with some processing features. Chapter 3.7.4 of this report outlines various potential system architectures that could be examined and tested through a prototyping study.

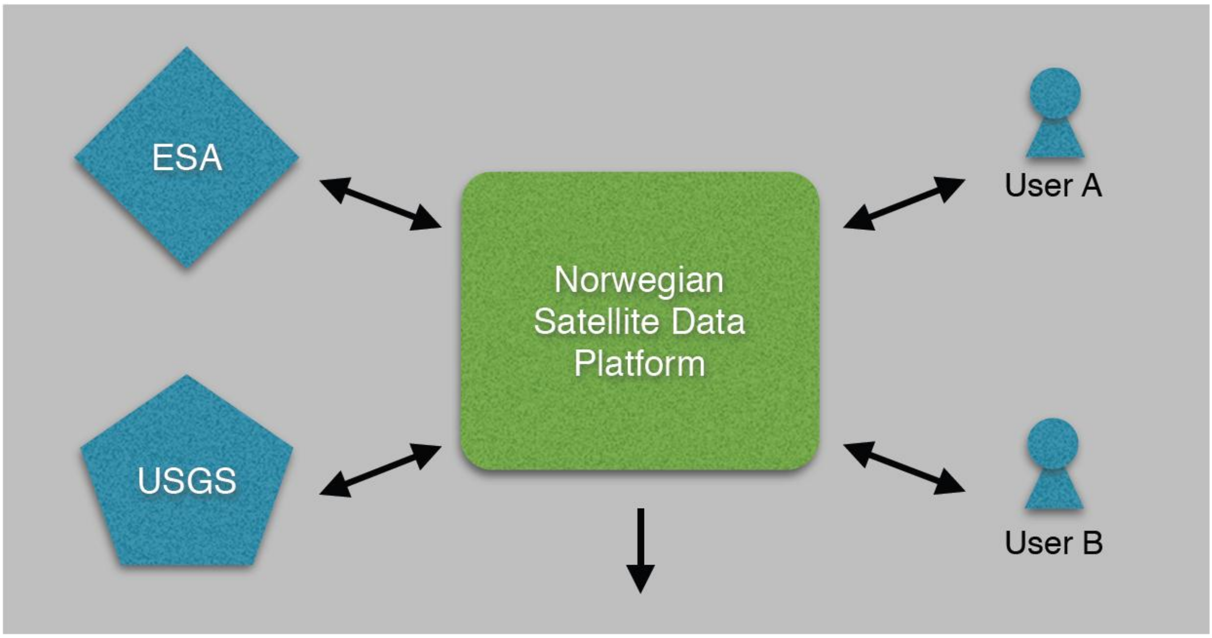


Figure 13 Proposed system. Users connect to a centralised national service in order to locate and download sensor data. Data processing can take place at the national data platform itself and/or locally for each user. Users might also share processed data products with each other and with external groups, through the data platform service. The complexity of accessing each satellite data source, and of data storage and management, are hidden from users by the data platform. The green box symbolises the role of the national data platform.

3.7.3 Characterisation of the problem from a technical perspective

3.7.3.1 Overview of primary system requirements

- The system should be capable of allowing the user to retrieve satellite images over Norway based on Landsat data and Sentinel-2 data.

  Landsat data is currently generated by two satellites, Landsat 7 and 8 (see references [1, 2, 3] in Appendix E), which work together to revisit points on the equator every 8 days. Landsat 8 has 11 spectral bands at varying resolutions, generally 30 m. The Sentinel-2 system is a new European remote sensing platform based on a pair of satellites which will operate together on a combined 5 day equatorial revisit schedule, with higher resolution sensors than Landsat (mostly 10/20 m), and with 13 bands.

  In both cases, revisits/re-imaging occurs more frequently at latitudes far from the equator and for areas with high typical cloud coverage, as is the case in Norway. These factors can result in a high degree of overlap in acquired images. The diagrams shown in Appendix D indicate the pattern of data acquisition.

- The system should enable the user to retrieve four types of sensor images. These are: 1) fast thumbnail previews, so that users may see the data they are requesting before download or analysis; 2) single images; 3) temporal composites and cloud-free composites; 4) a set of images representing an area over a timeframe.

- It must be possible to retrieve the stored data after clipping by an arbitrary geographical bounding box supplied by the user. There should be a convenient way for the user to specify this bounding box.

- The system must maintain a structured database of metadata information to keep track of the details of the images that are available.

- The system must allow users to browse or search for image data according to all of the stored metadata (e.g. producer, data, type of data, general statistics (e.g. cloud cover)).

- The system should work in the UTM33 system and should be based around the SSBgrid system used in Norway.

- The system should be capable of providing orthorectified data.

- It should be possible for the user to perform some common forms of re-processing or analysis on the server before an image is downloaded.

- The system should have a user-friendly interface that is suitable for a wide range of specialist users.

- Scientific journals increasingly mandate free access to raw data used by scientists in research papers. For example, from February 2014 onwards the largest journal in the world (PLoS One) will require all papers to provide data. Researchers are one of the classes of system user, so it would be helpful if the data platform facilitates this.

- If the system is implemented as a centralised server, the typical load on the system will be one user at a time. A 'high load' will be defined as 2–3 users. The system does not need to be designed to cope with more than 10 simultaneous users.
- The typical response time to a request for a single image will be 5–15 seconds, and should not be more than 60 seconds.

3.7.3.2 Technological constraints

Constraints upon data: an extremely large amount of data storage will be required if, for example, the project is intended to directly archive all of the relevant tiles from Landsat and Sentinel-2 over the next 10 years. Precise calculations are given for the data volumes in Appendix E.

Constraints upon data networking: In designing the system, it will be necessary to pay attention to network speeds and points of congestion between data storage and the researchers using the data. 100 GB of data takes around 1 day to transfer over a dedicated 10 Mbit/second connection. It may be worthwhile to sometimes allow certain types of approximate calculation or pre-preprocessing to take place on a central server nearby the data, rather than moving the data to the specialist's desktop machine.

Constraints upon data processing: Although many graphical operations can be easily parallelised in principle, many of the open source GIS-related libraries and programs available today do not offer parallelisation as an effective means of speedup of an individual task. It is usually available only as a way to enable multiple tasks to complete side-by-side. In the case of this system, generally there will be only 1 or 2 users running a task at the same time. Apart from single images or short temporal sequences, the datasets will not fit into RAM. Consequently, there will be limited opportunity to benefit from purchasing extra CPU/RAM hardware to improve the system after it is built. It is necessary to take care to develop a system which will not depend on 'upgrades' later.

Constraints upon hardware: Presently, and for many years, there has been only limited growth in hard disk capacity (~10–20 % per year) or in single threaded performance of CPUs (~5–10 % per year). When planning this work, we should not expect server technology in the near future to become much better or cheaper than today. The project should be planned around today's technology and prices.

Risks from maturity of software: There are several powerful open source software packages that might be suitable and interesting for this project, which are mature, well-studied and tested in large scale environments (UNIX, various file systems, PostgreSQL). There are also other very interesting and relevant programs that do not yet have many large scale production case studies or independent benchmarking (e.g. PostGIS Raster, rasdaman). There are also long-established projects that have yet to reach full standardisation and inter-compatibility (e.g. GDAL JPEG2000 drivers) which offer both opportunity and risk. The project will therefore benefit greatly from a prototyping phase, in order to explore opportunities and evaluate risks in full.

3.7.4 Outline of possible technical solutions

The aim of this project is to simplify, standardise and automate the tasks that surround analysis of remotely sensed data in Norway. This can be achieved in many ways; including client software and plugins; classical relational databases linked to file stores; database raster models; and real-time analytical raster databases. These models are now outlined with a brief explanation to highlight their unique features.

3.7.4.1 Desktop client workflow tool

Here, the idea is to focus on making existing technology better, by enhancing the GIS systems that are already used every day for data analysis on desktop machines (Figure 14). This could be through plugins to existing software that simplify tasks such as finding or obtaining data; or it could be a new standalone client that automates the workflows around obtaining and analysing data.

The advantages of this approach are that it does not require investment or the upkeep of centralised infrastructure; it can fit well with workflows that are already closely based around desktop applications; and the cost of development can be low, and the development time fast, if local specialists are interested in helping to improve and standardise what they already do.

The disadvantages include the problem of slow internet connections (when downloading large files from the USA); the difficulty of rolling out upgrades to all users as standards change over time; and the near-impossibility of building a centralised catalogue of thumbnails, metadata, statistics.

A desktop client can support other approaches – for example, a GIS plugin that connects and interacts with a national data service directly (in place of the normal web interface) can help to drive adoption of the new national data service.

Although this approach can help to hide some of the complexity of the data management problem from the user, it does not remove any of the complexity or necessary interconnections directly. It is unlikely to offer a solution to the problem of supporting interaction between users. This approach risks excluding users who use a less popular operating system or GIS/analysis tool. It could present problems in situations where a large national scale dataset is needed in order to correct the downloaded raw data (e.g. orthorectification).

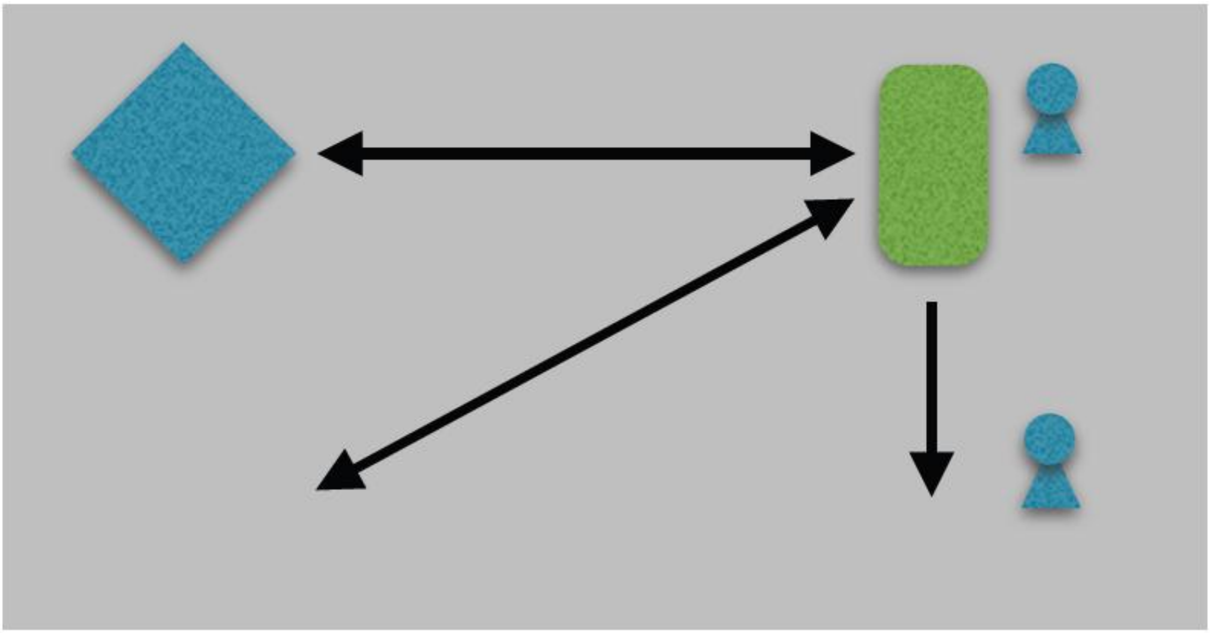


Figure 14 An outline of a desktop client based data management platform. The user accesses the client primarily to achieve a simple/standard view of diverse data services they wish to access. In the diagrams of Chapter 3.7, the blue diamonds represent satellite image archives such as the ESA and USGS, and the green box indicates the role played by the data platform technology.

3.7.4.2 Proxying data management platform

This web-based approach (Figure 15) may make the user believe they are accessing a large archive of curated data files. In fact, the system stores very little – only a relational database describing metadata, and perhaps low resolution thumbnails. When the user has selected their desired data, the proxy quickly goes to download the needed raw data over a high speed connection from the ESA or USGS, and processes it into the user's desired format. This system requires little storage, but requires good network connections to the real data stores and well written software to quickly process the data. It has the potential to be considerably faster than desktop client software. The user would have some control over the types of processing performed before data is returned.

The advantages of this system are the very low hardware purchase and maintenance costs; the ease of updating the management system compared to a desktop client; and the ease of extending the system to support new data sources or new data types. The disadvantage is that the system may not feel fast enough for the user, depending on the amount of data requested. This could be managed by providing the user with information while they select data, explaining how their data choices will affect the turnaround time for the system to provide them with the data.

A major advantage of proxying systems over direct storage is that the user is guaranteed to get up-to-date information. As the ESA and USGS adjust their processes to improve the signal processing of the satellite telemetry, the resulting new and better data will automatically be fed to all users.

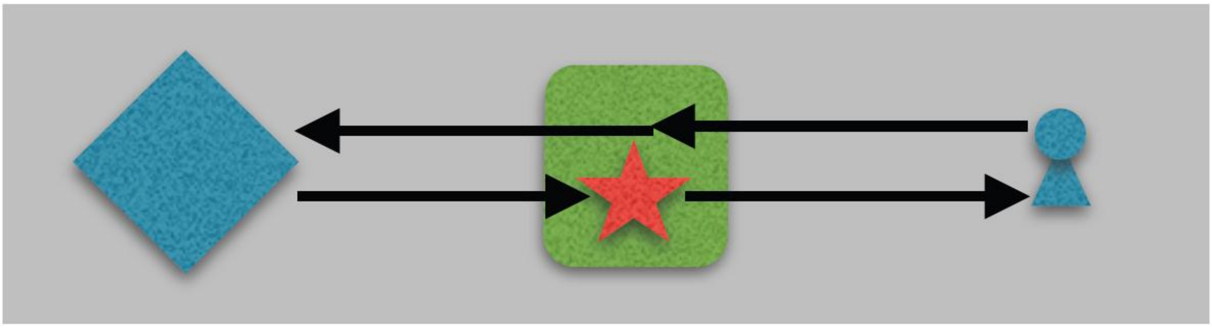


Figure 15 An outline of a proxying data management platform. Here, the user accesses a web based system which provides access to a relational database of metadata, perhaps with thumbnails. Once an image is selected, the system quickly returns the required image data by collecting and processing it dynamically from the ESA or USGS over a high speed link.

3.7.4.3 Caching proxy data management platform

This is similar to the simple proxy, but the proxy also has a cache, which is some storage of its own to keep files that are frequently requested (Figure 16), for example, popular files from particular projects, or very recent data.

Perhaps the user is able to request to the proxy that it uses the space to preload some important data. The cost of developing this system is similar to the simple proxy, because it must perform the same tasks in order to fill the cache with requested data. There are extra costs for storage, but this can be scaled to suit the budget and performance needs of users.

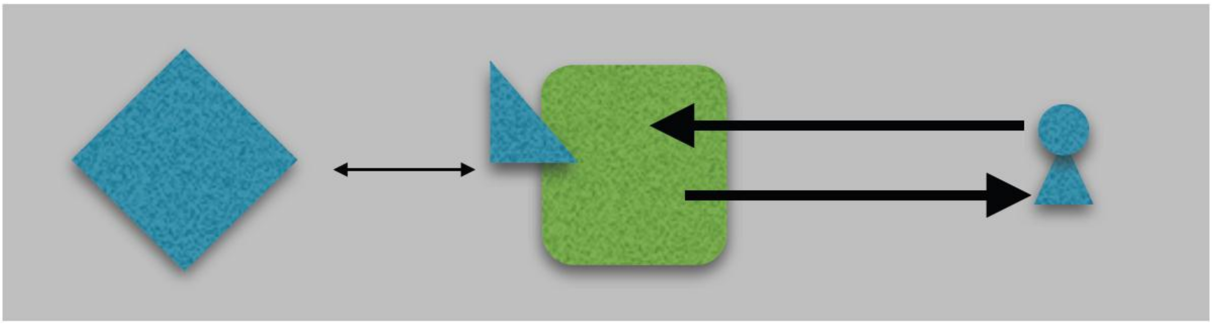


Figure 16 Outline of a caching proxying data management platform. The proxy stores a local copy of the most frequently used parts of the data from the original archives. Here, the user accesses a web based system similar to the proxy detailed above. Now, the proxy checks to see if it is already storing a prepared copy of the desired data, before contacting the ESA or USGS to obtain the raw data for processing. In this case, since the cache is storing a commonly-used fraction of ESA data, it is often able to answer the request directly. Occasionally, it must retrieve an image from the ESA/USGS archive dynamically (the thin arrow). However, even when it has the data, the proxy may still contact the ESA or USGS to check that the data is up to date.

3.7.4.4 Archive proxy data management platform

This may be an extension to the proxy caching model so that the cache is large enough to gradually accommodate every file that might be accessed; or it may be a deliberately designed system that stores all the original files after a one-time bulk download from the source archive. Alternatively, it may be that a set of special variations of the original data is stored locally, for example, an archive of composite images. The costs of this approach can become extremely high for a complete replicating archive (see Appendix E). Data must occasionally be refreshed from the source archive servers if it can become out of date. There are three main ways to build the archive proxy, based on two storage variations which are shown in Figure 17.

The first approach is a metadata database with file system storage for the raster files. This is a well-tested, classical approach. Here, image metadata is stored and queried from a metadata database (e.g. via a web interface), as a relational DB is effective for querying complex metadata. The data tiles themselves are stored in a traditional file system as ordinary files, as a file system is good for storing a large volume of bulk data objects. The advantages of this approach are that it is straightforward and fairly cheap to build. The main disadvantage is that it is not easy to analyse tile contents via database queries.

The second approach is a raster database which stores the raster data contents directly inside database tables (e.g. PostGIS raster) using a special raster database type using any lightweight compression or raw storage format that is available in the database (in-DB). This is usually not an efficient means of storage or computation but it can be easier to work with when advanced users want to program queries that relate to the detail inside the raster data. The advantages include being able to query the raster data directly, and perform large queries against many raster datasets at once. Also, web applications may be able to access graphics or other results directly through the DB interface without file management, which may be useful for programmers.

The third approach is a raster database that stores the tile metadata and also the raster metadata in the database, but keeps the raw raster data values within files in the file system (out-DB). It is a combination of both of the first two types of system (file system approach and raster database).

In PostGIS for example, storing to out-DB GeoTIFF raster files will allow you to store raster data directly in the database but also in the file system. The advantages are a potentially more compact representation (e.g. the availability of higher levels of compression); and the ability to re-use the same set of files for another purpose (e.g. direct web download). The disadvantages can come from incompatibility problems with some out-DB data formats and the risk of error (deleting a file which the database considers to be in use will effectively damage the database integrity).

In both the in-DB and out-DB models, the DB can be used directly as the backing for a proxy cache or an archive. In both cases, it is possible to perform analysis directly upon the data without the user having to transfer the raster data to their local machine. In practice, e.g. PostGIS Raster is not as fast as file-based approaches to processing data.

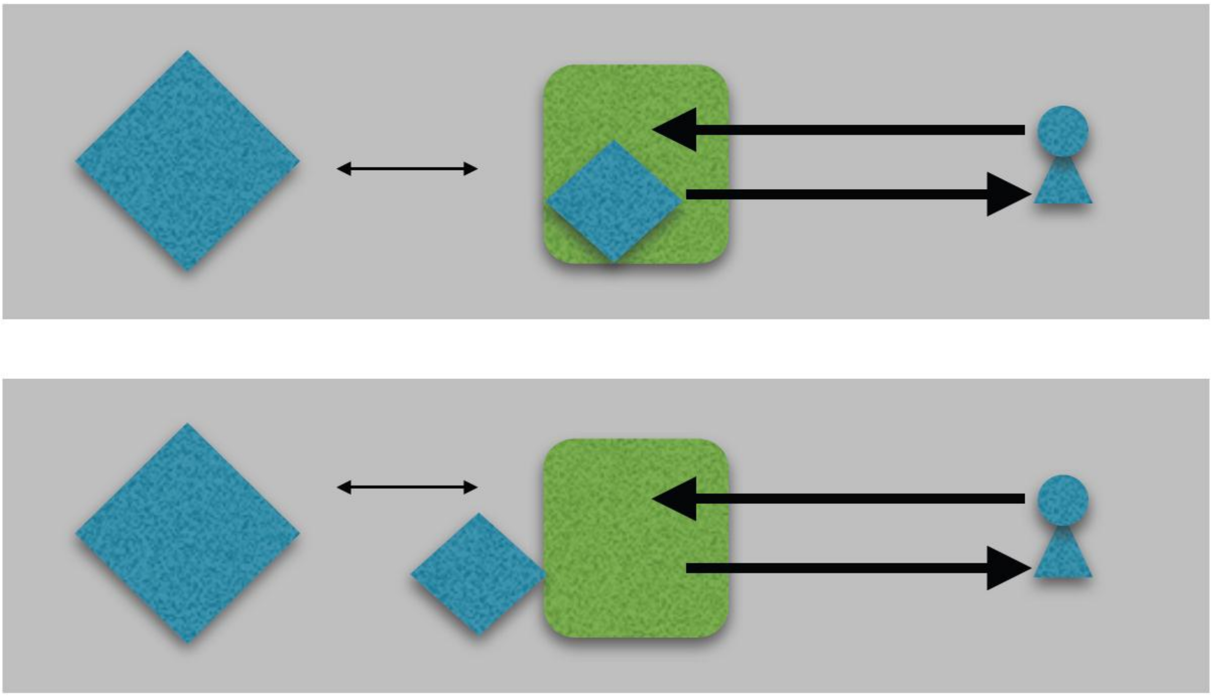


Figure 17 In the top diagram, the user accesses the proxy and the data is already held inside the database as database tables. In the lower diagram, the proxy must look up the data as files in a nearby file storage server; the database tables provide only metadata and a link to the data file, not the data itself. The small blue diamond indicates that a complete copy of the original archives is continually kept and occasionally updated.

3.7.4.5 Real-time analytical DB models

In this model (Figure 18), some or all of the raster data is stored in the database, similar to the previous example of an in-DB cached raster. Now, though, the data may be multiply-represented or stored in an unusual format in order to allow the various different access patterns during analysis take place very quickly. The idea is that the database should quickly process requests for slices of data such as ‘give me all data between time = 5 and time = 10 for area X1Y1X2Y2’.

Although the user is connecting over a network, he or she may have the feeling that they are working locally sometimes. The downside is that usually, far more data storage is required by the real-time analytical DB design; the range of available functions may be smaller and compatibility with other software may be lower. These systems are currently not as popular among developers in the open source GIS world as traditional DB systems, but have some potential to offer fast online raster processing.

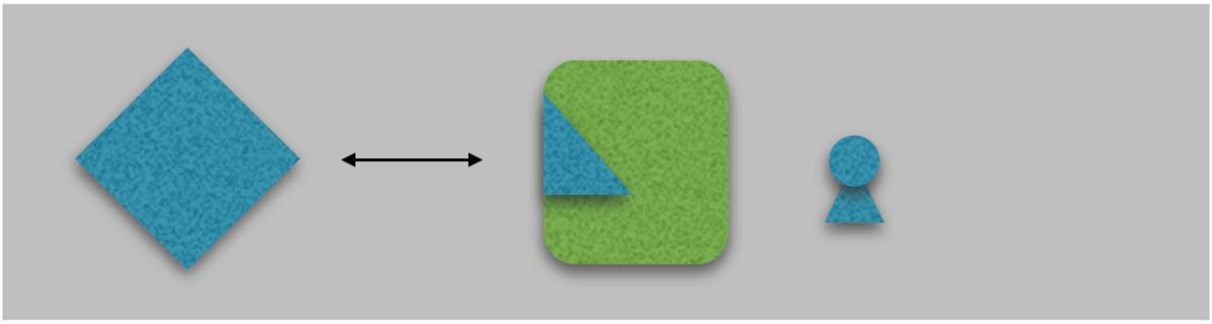


Figure 18 The user is shown here as though they are directly accessing the database – although in practice they are connected over a network. For some types of request (e.g. text answers to a calculation), they are able to receive results quickly and directly from the analytics DB. The purpose of the DB here is to try to return results from long tables as quickly as possible, and to centralise and record computation and access. However, such databases often have more restrictive data models, smaller development communities and costs such as higher space consumption per unit data. Here, the user is only able to perform analysis across the part of the data stored locally. It may be more difficult to fetch and integrate external data dynamically.

3.7.4.6 Hybrid models

It is not necessary to pick a single model when developing the system. Elements of the system can be combined. For example: Client desktop software might be used as an alternative to a web interface to access a proxy or archive, much as BEAM is used today (see Chapter 6.3). Alternatively, a complete archive might only be used to curate some particular type of unique data that cannot be served by proxy, for example, user-contributed data or data which is expensive to calculate.

It is also possible to extend many of these designs gradually. A direct proxy may have a cache added later. A caching proxy may grow in size as new funds become available until it begins to resemble an archive in practice. An archive may have an analytics database added later which gathers data from the main archive upon request for subsequent interactive analysis.

3.7.5 Recommendation and proposed next steps

All of these architectural possibilities should be examined within a prototyping phase to determine limits on how well each model can be made to serve the needs of the remote sensing community in Norway. It might be discovered, for example, that the community is extremely well served by plugins to popular existing software packages that simplify common work processes and are easy to use, instead of a web interface that interfaces less well with existing software.

If a centralised web-based model is strongly preferred from the outset, the author proposes that the following hybrid model would combine the best aspects of each of the outlined architectures, while avoiding high costs in terms of storage, processors, backups, system administration, power, hardware maintenance and replacement, and so on.

3.7.5.1 Proposed Hybrid Model

The data platform would be web based with independent subsystems enabling access to:

- Product metadata: e.g. date, cloud coverage statistics, other overview statistics, producer etc. via a web interface to a traditional relational database with web based search facilities.
- Previewing: A complete archive of miniature thumbnail versions of all available images, to enable instant previews of selected data whether it is stored locally or elsewhere.
- For instant access to large quantities of the most recent data and the most popular data in full high resolution formats, a caching proxy would be used with a cache size of several terabytes.
- For fast access to ‘hard to compute’ data – in particular, temporal composite and cloud-free composite image sets – pre-computed images would be built and stored in a long-term archive, with background tasks set to re-compute composites as new data becomes available.
- To support data sharing between users and in relation to scientific archiving, a temporary sharing area and a small long-term archive could be provided.
- For prompt access to older or less popular high resolution raw data: a web-based proxy or well written client software/plugin, that can search for the correct data at the ESA or USGS, download it, pre-process as appropriate and immediately forward it on to the user.
- Experimental real time analytics (developed later in the prototype phase). Here, the word “experimental” is chosen to highlight that this approach is relatively novel and probably higher risk to implement. A small-scale system implementation could be attempted using low spatial resolution resampled data (30x30m or less) and possibly with reduced temporal resolution (e.g. 1 sample per month) to explore feasibility, performance and costs.

  If the real time analytics database proves capable of handling large quantities of this lower resolution data, and if there is sufficient demand from users after they have tried the prototype, the product can be scaled up to operate directly on higher resolution data. Still, the overall aim here would be to undertake a long-term feasibility analysis into this approach, rather than to try to deliver a full resolution full archive analytics product as the primary deliverable for this project. Attempting a full product would probably represent a significant risk for the project and might not be possible with current technology. Generally, this type of technology is best suited to extending an existing data management platform that has already been established.

- It may be worth investigating the possibility of peer-to-peer storage, i.e. in the style of bittorrent. This might solve the problems of building centralised storage for satellite images. A problem may occur when the hosts who are meant to have the data are not available, or when the hosts have accidentally damaged the data.

3.7.5.2 Next steps and timetable

For this project, it would be most appropriate to begin with a prototyping stage. The aim would be to try scaled-down versions of the strategies outlined here, in order to establish accurate information about costs, benefits and limitations of each approach. This stage would also involve discussion with users across Norway in more detail about their patterns of usage, to ensure that the data platform provides the maximum benefit to the maximum number of users in Norway.

Year 1:

- Discussion of the patterns of usage for satellite sensor data across Norway, to identify which aspects of remotely sensed data analysis in Norway are most important when developing automated workflows, search facilities, faster access times.
- Short context and literature survey to examine whether new technology options have been developed in 2014–2015.
- Selection of technology options and development of prototypes to study each architectural approach.
- User trials, measurement of effectiveness leading to a report on cost/benefits for each potential approach.
- Development of a workflow to generate temporal and cloud-free composites.

Year 2:

- Development of a production system based upon the outcomes of the prototyping process and discussion process, perhaps similar to the recommended hybrid model outlined in this report.
- Simultaneously, development of a small-scale experimental online analytics system.

Year 3–10:

- Maintenance of the system, as use of the system increases, and increased quantities of data from new satellites become available.
- Further development of analytics according to success and popularity of the small scale system.

3.8 Cost estimates

It is difficult to estimate the cost of establishing and running the National Satellite Data Centre at this stage. As described in Chapter 4 a pilot project is needed in order to estimate total costs of the complete system. We have made a guesstimate of the total cost (see Table 2) of establishing the national centre including the pilot project. The cost is split up in the following parts: software development, data storage, maintenance solution, and web portal costs. We assume that open source software can be used to a wide extent.

We need a period of prototyping before we can give more precise calculations. The cost can be split in four components:

1) Pilot project
    a. Consultants
    b. Norway Digital technical committee (working group, travel expenses)
2) Purchase and development of software
    a. Purchase of software
    b. Development of maintenance solution
    c. Development of web portal
3) Purchase of hardware incl. installation costs
    a. Servers for running the software with a capacity to deliver all requested data
    b. Storage capacity
4) Maintenance agreement for keeping the system updated
    a. Software and hardware maintenance agreement
    b. Maintenance cost for keeping the database updated

Table 2 Estimated costs for establishment and maintenance of the satellite data centre

Costs to establish the data centre	*MNOK*
Pilot project (consultants and ND technical committee)	1.5
Purchase and development of software	2.0
High capacity data servers and storage	0.5
Total	**4.0**
Yearly maintenance costs	*MNOK*
Service agreement SW/HW	0.5
Operational service (i.e. adaption of new satellite data sets)	0.5
Development of system	0.2
Total/year	**1.2**

One obstacle of realizing the data centre is the initial cost related to purchasing of hardware, software and software development. It is not known to the technical committee if Norway Digital has the ability/mandate to finance the initial costs.

As mentioned earlier, our basic assumption is automatic adaption of new satellite images into the national satellite data archive and that only relatively few man-hours will be needed to keep the system running. In that respect, the maintenance cost will be relatively low and it is likely that users from Norway Digital partners will see the benefits of a centralized preprocessing and archiving service. It is therefore likely that the partners of Norway Digital are willing to pay the running costs of a national satellite data centre.

3.9 Time schedule

We have made a rough time schedule for the pilot project and the establishment of the centre (Table 3). Landsat 8 is already operative, since February 2013, and Sentinel-2 is planned to be launched the first half of 2015. In the planning of the project, we have taken this into consideration and aim to have the centre up and running when Sentinel-2 starts to deliver data.

Table 3 Project time schedule

Tasks	2014/Q2	2014/Q3	2014/Q4	2015/Q1	2015/Q2	2015/Q3
Pilot project						
Establishment of the centre						
Purchase of SW and HW						
Development period						
System in service						

4 Preparations for implementation – a pilot project

In this report, we have described and analyzed selected use cases in order to document how optical satellite data are used today and what is the expected use in the near future. The study has documented the needs of the different users and the service and functionality the users expect from a national satellite data centre for Sentinel-2 and Landsat data.

The report has identified three areas for further focus:

- Automated orthorectification
- Automated atmospheric correction
- A system for data archiving, searching, and dissemination

The technical committee recommends that Norway establish a national satellite data centre for Sentinel-2 and Landsat 8 data to facilitate the use of the datasets in the different operational services based on satellite data. The use of satellite data will increasingly be focused on multi-temporal analyses where a huge amount of datasets will be involved. The users will need data that represent true surface reflectance in order to compare observations obtained from different days and possibly different sensors. ESA plans to use a global DEM that we expect will lead to inaccuracies in the orthorectified product covering Norway. With increasing focus on time series analyses and change detection, accurate geometric rectification is crucial, and we therefore recommend to orthorectify Sentinel-2 and Landsat images using a national DEM of better accuracy than the global DEM ESA and USGS are using. Another reason to perform this preprocessing step is the possibility to align the rectified images to SSBgrid defined by Statistics Norway. SSBgrid is a standardized grid to facilitate combined use of datasets from different data sources and data providers. When data are aligned to the same grid, overlay analyses, time series analyses, and comparisons are very easy. It is ideal to interpolate satellite images as few times as possible, and it is therefore best to align the orthorectified images to SSBgrid when they are processed from level 1B (Sentinel-2) or level L0Rp (Landsat). Thus, we propose a fixed grid structure for all orthorectified satellite images based on SSBgrid.

4.1 Orthorectification to a national level 1C

More than 1000 images from Sentinel-2 and Landsat 8 will be recorded over Norway each year and it will therefore be necessary to automate the orthorectification process. This will require a very accurate reference image that will be used in automatic image-to-image matching to obtain control points (CP) between a given band of the image to refine and the reference image. The CPs will be used to enhance the geometric model that links the coordinate of a point in the focal plane with a point on the Earth surface. Once the geometric model of the image is enhanced, the orthorectification process can be performed with a resampling algorithm (Baillarin et al., 2012). This process will convert L1B/L0Rp datasets to national level 1C datasets (1CN). In a pilot project, methods to perform the geometric rectification have to be investigated and tested.

Some questions need to be examined:

- What elevation model is the best available for orthorectification?
    - DTM, a national DEM from the Norwegian Mapping Authority, generated from contour lines from big scale maps (1: 1 000 and 1: 5 000) in developed/urban areas and scale 1:50 000 in rural/mountainous areas. The vertical accuracy fluctuates from two to three meters in urban areas to five to twenty meters in mountainous areas. This is a terrain model the follows the ground surface over forested areas. Over dense forest, the contour lines often lay above the terrain surface.
    - DEM based on a combined dataset from matching of aerial images and aerial laser scanning. Norway Digital has taken an initiative to establish detailed terrain and surface models based on aerial laser scanning. Above the tree line, matching of aerial images is an alternative.
    - WorldDEM based on data from the TanDEM-X mission. This is a surface model and follows the tree canopy over forested areas.
    - ASTER GDEM V2 based on data from the ASTER radiometer on-board the Terra satellite. This is a surface model and follows the tree canopy over forested areas.
- What is the accuracy of the elevation models?
- How is the reference image going to be constructed? Will the reference image made by ESA be available for collaborative ground segments?
- What software is available for geometric enhancement and correction?
    - Sentinel-2: Will ESA provide the software they will be using to enhance the geometric sensor model of level 1B (dataset in sensor geometry) and orthorectify to level 1CN? As far as we know, ESA is still discussing if they shall provide collaborative ground segments with a program to process from level 1B granules to level 1C tiles (B. Hoersch, 12/12/2013).
    - Landsat 8: What software is available to enhance the geometric sensor model of level 0Rp (sensor geometry) and orthorectify to level 1CN?
    - What parts of the geometric enhancement can commercial image processing software perform?
- What is the estimated cost of an operational service?

4.2 Atmospheric correction to a national level 2A

ESA has commissioned DLR to develop a module for BEAM to perform atmospheric correction on Sentinel-2 datasets. The module will be called SEN2COR and the license will be restricted for use on Sentinel-2 data. This means that another solution has to be found for processing Landsat 8 datasets. ATCOR is also available as stand-alone software from ReSe Applications and this version will process both Sentinel-2 and Landsat 8 datasets. ATCOR can be used both as an interactive tool and in batch mode where images are processed automatically.

MACCS is an alternative to ATCOR. It has been developed to do atmospheric corrections of the images acquired by the Venµs satellite that was launched in 2010. The satellite has a 2-days repeat cycle and MACCS takes advantage of this short time duration to estimate the aerosol content in the atmosphere: the algorithm assumes that surface reflectance vary slowly with time and aerosol vary quickly with time. Thus an observed change in radiance at the top of the atmosphere is likely caused by change in the atmosphere. Hagolle et al. (2008) state that it will be possible to use short term variations over 10 to 15 days to retrieve surface reflectance. This means that MACCS can be used to do atmospheric corrections on Sentinel-2 datasets, but it is unsure if it is suitable for Landsat 8 datasets with a 16 days repeat cycle.

Cloud and cloud-shadow detection are important preprocessing steps. ATCOR uses spectral rules with thresholds to detect haze, clouds and cloud shadows and can remove the effect of haze, thin cirrus clouds and cloud shadow. MACCS uses multi-temporal information to detect clouds and use the contrast of the relative stability of the surface reflectance of unclouded pixels with the quick variation of reflectance of pixels affected by clouds and cloud shadows.

Some questions need to be examined:

- What processing steps are necessary to employ ATCOR (MACCS) on a dataset and to what degree is it possible to automate the steps?
- Can ATCOR (MACCS) be used to bulk process many datasets in batch mode with minimal user input?
- How well do the cloud and cloud-shadow detection methods work? Can this process be executed in batch mode with minimal user input?
- What is the estimated cost of an operational service?
- Will a user license be available for MACCS?

4.3 Data management systems

Data management of large raster data volumes comprising thousands of datasets is challenging. In recent years new database management system technology has emerged that may be able to meet this challenge and deliver full query functionality based on SQL query language. However, with the estimated storage requirement of 130 TB/year (Chapter 3.6) and 1 300 TB after 10 years, it can be problematic to use database technology to store the raster data.

Some questions need to be examined:

- What alternatives exist to handle large volumes of satellite data and with the functionality that is requested for the proposed national satellite data centre?
- What is the right level of service from a centralized system? What processing should be made on the server side and on the client side? More processing on the server side will save large volumes of data traffic, but it can be challenging to develop a very flexible system.

4.4 Organization of further work

How to follow up this report can be summarized by the following points:

- Discussions of the proposals in the report within each institution and the Norway Digital reference group.
- Organize a meeting with the Norwegian Space Centre, Kongsberg Satellite Services AS, and the technical committee of Norway Digital to discuss the proposals in the report. If we conclude to go further with a pilot project the following questions must be answered:
    - How should the pilot project be organized?
    - Who should participate in a project group?
    - How is the project going to be financed?
    - What should be the time schedule for the project?

Part II: Preprocessing of image data – tools and elevation model

5 Discussion of elevation models for orthorectification

5.1 Digital elevation models

When working in a pixel oriented environment, positional accuracy is important. Better quality of the elevation model used for orthorectification will contribute to final results containing fewer errors.

ESA plans to use a surface model (DSM) mainly based on data from the Shuttle Radar Topography Mission (SRTM), which took place in year 2000, to orthorectify Sentinel-2 data. For areas north of 60 degrees the elevation model is based on other datasets used to fill in gaps in SRTM. What datasets are used to fill in the voids is unknown to us, and we are therefore not able to assess the quality of these.

Our assessment is that a national elevation model with higher resolution and better accuracy will be a much better basis for orthorectification because it will have no voids (we suppose) and significantly fewer large errors in the elevation values. In the following we discuss the quality of the SRTM DSM compared to the current national terrain model. We are aware that we are comparing a surface model (DSM) from SRTM and the existing national terrain model (DTM). A DSM represents the terrain surface and over forest it lays on top of the vegetation canopy and, in contrast, a DTM represents the ground surface. In our assessment we have taken this difference into consideration.

5.2 The Shuttle Radar Topography Mission

The Shuttle Radar Topography Mission (SRTM) was an international research effort to obtain a digital elevation model on a near-global scale from 56° S to 60° N. Its origin was a specially modified radar system (X-band) that flew on board the Space Shuttle Endeavour during the 11-day STS-99 mission in February 2000. Pulses from X-band radars do not penetrate dense forest canopies and the echoes come from a surface high up in the canopy close to the tree tops. The elevations that are measured therefore represent the ground on open land and the vegetation canopy on forested land. For this reason, the resulting elevation model is called a digital surface model (DSM) to distinguish it from a digital terrain model (DTM) that represents the ground elevation.

The elevation model is arranged into tiles, each covering one degree of latitude and one degree of longitude. The resolution of the raw data is one arcsecond (30 m), but this has only been released over United States territory. For the rest of the world, the best accuracy available is three arc seconds (90 m N-S, and 45 m E-W). The vertical accuracy for these data is stated to be better than 30 m for a 90 % linear error, with a horizontal accuracy better than 50 m for circular error (90 % probability). This DSM is however not available above 60 degrees north (Figure 19), and supplementary data are of unknown quality and may be encumbered by even worse errors.

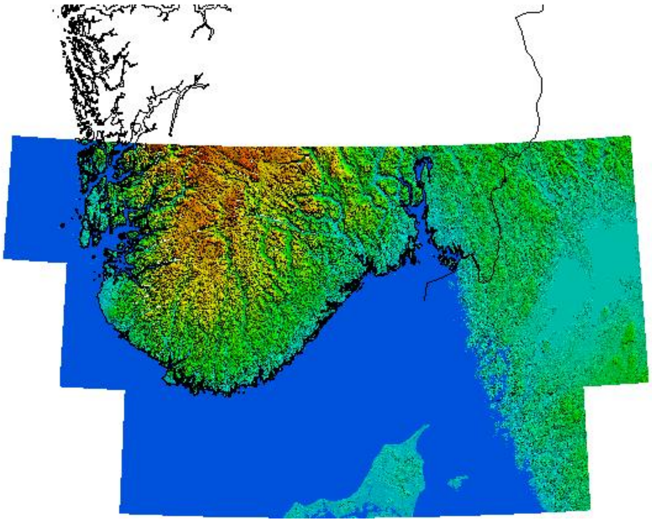


Figure 19 Coverage of DEM from The Shuttle Radar Topography Mission (SRTM) in Norway.

5.3 The national terrain model

The national digital terrain model (DTM) from the Norwegian Mapping Authority is represented by a 10 meter grid and is generated from contour intervals from big scale maps (1: 1 000 and 1: 5 000) in developed/urban areas and scale 1:50,000 in rural/mountainous areas. The vertical accuracy fluctuates from two to three meters in urban areas to five to twenty meters in mountainous areas.

Norway Digital has taken an initiative to establish a new national DTM from airborne laser scanning. If this initiative is realized, Norway will obtain a detailed national DTM with grid size of one meter or better and accuracy (both horizontal and vertical) better than one meter by year 2020.

5.4 Geometric errors caused by errors in an elevation model

At the outer edge of a Sentinel-2 acquisition, errors in a DEM used for orthorectification will lead to a horizontal error a little less than 1/5 of the vertical error (Δh) as shown in Figure 20. A vertical error of 50 m will cause about 10 m horizontal error. For Landsat 8, which has a more narrow field of view, the horizontal error will be less and about 1/8 of Δh.

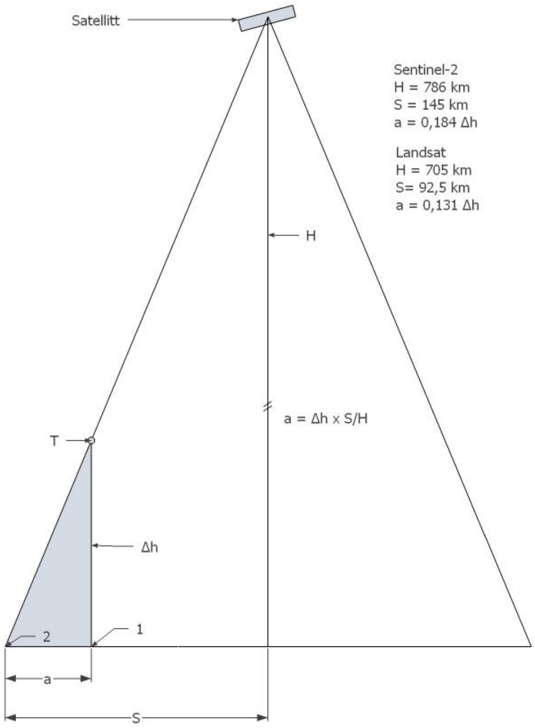


Figure 20 At the outer edge of a Sentinel-2 coverage, an error Δh in the DEM used for orthorectification of the image, will cause a horizontal error, a, that is about 1/5 of Δh. An object, T, will erroneously be located at point 2 instead of at point 1.

5.5 Comparison of SRTM DSM and the national DTM

A test was performed by the photogrammetric department at the Norwegian Mapping Authority. The SRTM DSM was compared with the national DTM up to 60° north. An inland area was chosen to avoid contribution from ocean level that is set to zero meters in the SRTM DSM (Figure 21).

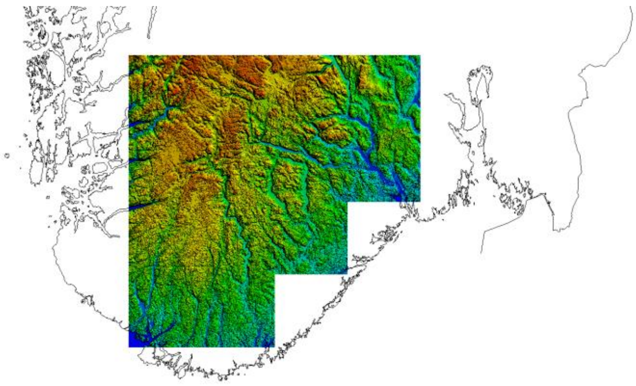


Figure 21 Area from SRTM DSM used in comparison with the national DTM.

The SRTM DSM has a drawback as the model has gaps as shown in Figure 22. Areas with missing values typically occur in deep valleys where terrain restricts the view of the instrument during acquisition.

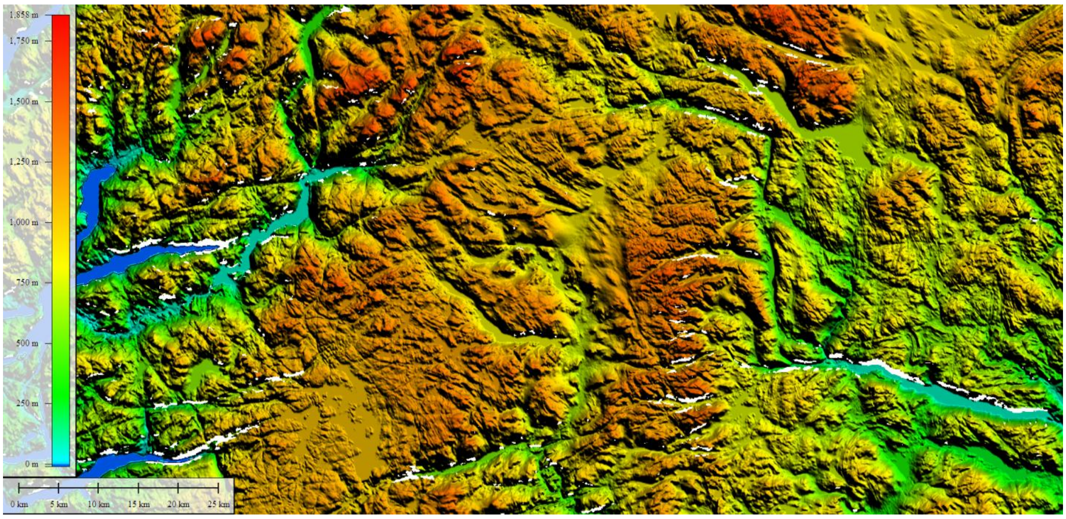


Figure 22 A section from SRTM DSM used for comparison with the national DTM. White areas show missing values.

A total of 7,692,943 elevation points from SRTM DSM was compared with data from the National DTM. About 30 % of the points from SRTM DSM falls within two meter deviation, 75 % of the pixels are inside seven meters, and 91 % falls within 11 meters, compared with the national DTM (Figure 23). The average deviation is 5 meters from 0; the median is -2 meters and the standard deviation 7 meters.

Table 4 Comparison between SRTM DSM and the national DTM

Statistic	**SRTM DSM compared with the National DTM**
Standard deviation	7.257 m
Average deviation from 0	5.018 m
Average	–2.162 m
Median	–1.949 m
Min. deviation	–566.405 m
Max. deviation	610.323 m
Number of points	7,692,943
Number of points with a deviation more than 20 m	107,202
Percent deviation > 20 m	1.39 %

Table 4 shows that the biggest vertical error is 610 meters. That will lead to a horizontal displacement of 112 meters in the outermost pixel if the error was present in a Sentinel-2 acquisition, and 80 meters in a Landsat 8 acquisition. These figures apply in an ideal situation, when the orientation of the images is perfect. In a real situation, errors in orientation (aero-triangulation) will contribute to a larger vertical deviation, and the displacements of pixels will increase. This will especially apply in steep areas, where a small movement in x, y causes a large deviation in elevation.

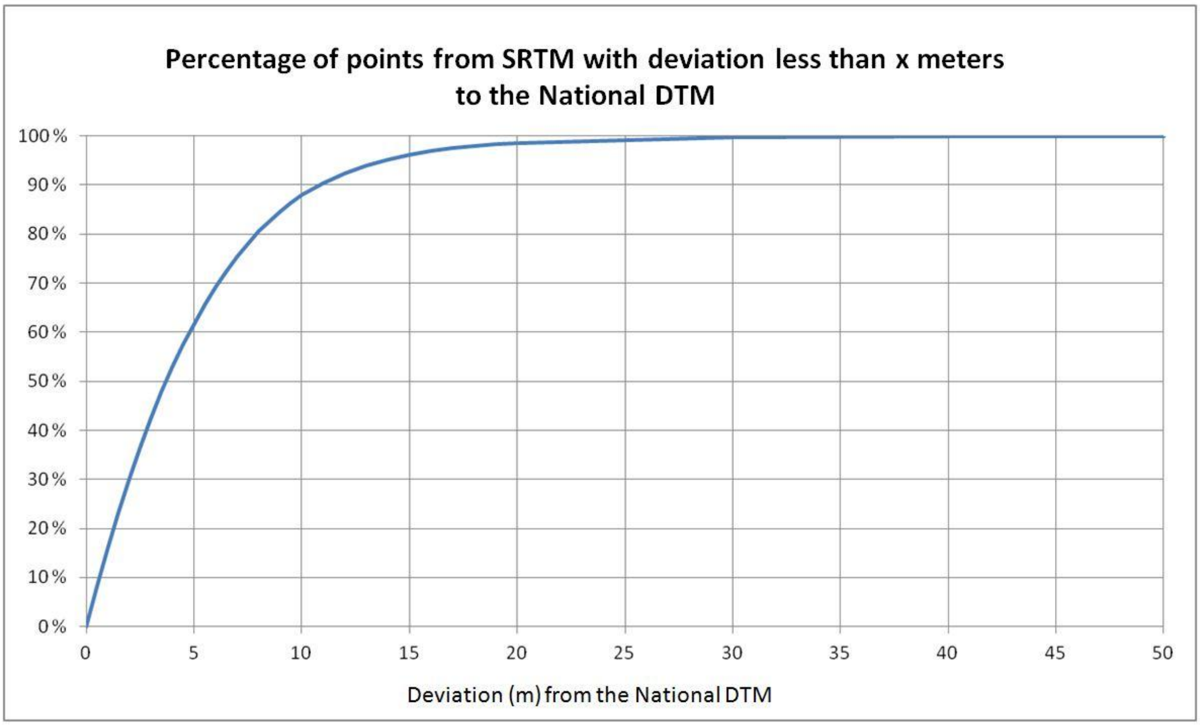


Figure 23 The diagram shows a comparison between the SRTM DSM and the national DTM. About 1.4 % of the points have a deviation larger than 20 meters.

5.6 Conclusion

Low resolution and poor coverage render the SRTM DSM unacceptable as a basis for orthorectification of satellite images over Norway. The test of SRTM DSM also reveals serious errors. It has too many areas with zero values and a significant number of the points with a vertical deviation of more than 20 meters. This will in turn contribute to significant geometric errors in the orthorectified images.

We believe that use of SRTM DSM, a coarse and inhomogeneous elevation model, will severely undermine the realization of cost efficient return from the national investments made in the Sentinel-2 project, not to speak of the usefulness of the data obtained.

A national elevation model must have high resolution in order to meet the challenges mounted by the country's rugged terrain. However, the current national DTM, partly based on contour lines, does not have the quality that makes it ideal for geometric rectification of images that are used in pixel-oriented analyses. Multi-temporal analysis requires precisely aligned pixels, and to obtain this, a very accurate and uniform elevation model is necessary. The technical committee supports the initiative from Norway Digital to establish detailed digital terrain and surface models based on aerial laser scanning.

6 Processing of Sentinel-2 and Landsat data

6.1 Geometric correction and orthorectification

Sentinel-2 data will be provided as product levels 1B and 1C. Level 1C will be orthorectified by using a global DEM with 90 m resolution, which is partly derived from the Shuttle Radar Topographic Mission (SRTM). As SRTM did not cover areas north of 60° N other datasets have been used to fill in the no-data voids. In areas with high relief no-data voids also occur and these are either filled by data from auxiliary DEMs or by interpolation. In mountainous terrain, topographic displacement can be significant in satellite images, especially when the sensor has a very wide swath such as the 290 km of Sentinel-2. At the edges of a Sentinel-2 dataset, an elevation error Δh will lead to a horizontal error of 1/5 of Δh (Gjertsen and Trollvik, 2013). To avoid geometric errors that can be detrimental to multi-temporal analyses, a more accurate national DEM has to be used in the orthorectification process. Townshend et al. (1992) concluded that a misregistration of one pixel resulted in an error equivalent to greater than 50 % of the actual differences in the NDVI. To achieve an error of only 10 %, registration accuracies of 0.2 pixels or less are required. They state that high levels of registration accuracy are needed for reliable monitoring of change and further that manual methods are unpractical. High accuracy can be achieved in part with accurate satellite navigation information and by subsequent use of image-matching methods.

With several hundred acquisitions each year from Sentinel-2 over Norway, it will be necessary to automate this preprocessing step. Image co-registration involves locating and matching similar regions in two images to be registered. In manual registration, humans carry out the registration task visually using interactive software to extract many control points. In automatic registration, an autonomous algorithm performs these tasks. The CPs are used to compute parameters of a geometric transformation. Manual selection of CPs is repetitive, laborious, and time consuming task that becomes prohibitive with a large amount of data. Location of CPs can often be difficult and the process can often lead to inaccurate points and large registration errors. Research in image registration has as a goal to improve the accuracy, robustness and efficiency of fully automatic algorithmic approach to the problem (Le Moigne et al. 2011). To achieve a fully automatic process, a cloud free reference image with high geometric accuracy will be needed to provide control points with known map coordinates. Several image-to-image matching algorithms are developed to match images with reference images (Le Moigne et al., 2011). The matching will provide new control points that can be used to enhance the geometric sensor model that describes the relationship between the image coordinates and the ground coordinates:

$$(x, y) = f(\text{latitude}, \text{longitude}, \text{elevation}).$$

By determining where the satellite is pointing, approximate ground coordinates can be given to an image. With systematic and random errors, the navigation model does not accurately report where the satellite is pointing. Recent navigation models use information from Global Positioning System (GPS) satellites and are usually accurate within a few pixels. Sub-pixel accuracy is crucial for applications such as change detection. Precision corrections therefore correct for these errors by registering an image to known ground features, e.g. via a precise reference image.

To achieve accurate geometric correction the following steps are important:

1. producing and updating a reference image dataset (involving simultaneous adjustment of the sensor models with large space triangulation adjustment);
2. using ground control points obtained from matching against the reference image dataset to enhance the level 1B sensor model of new acquisitions; and
3. using a detailed and accurate DEM to orthorectify to level 1CN.

ESA will use their enhanced global SRTM DEM in all steps. The first two steps probably need special software that might be provided to the collaborative ground segments by ESA (still unclear); the third step is much easier and can also be performed with commercial software. To bring in a national DEM in all three steps, the collaborative ground segment needs access to special software for steps one and two. Centre National d'Etudes Spatiales (CNES) is developing on behalf of ESA a prototype processor for geometric correction (Baillarin et al., 2012).

6.2 Atmospheric correction and processing tools

ESA provides a toolbox called BEAM, an open-source, modular software for analysing optical satellite images. BEAM was originally developed for the users to process data from ENVISAT, and ESA is developing a Sentinel-2 toolbox building on BEAM. A new module has been developed to do atmospheric correction of level 1C datasets and produce level 2A. Thus, ESA will not produce level 2A but will instead provide this toolbox for the users so that they can do the processing themselves. It is however not likely that the geometric rectification included in BEAM provides the required accuracy (C. Brockmann, pers. comm.) and thus another software solution must be implemented. Several commercial software solutions are possible candidates.

Removing the atmospheric effect from satellite images to obtain true surface reflectance is an important pre-processing step before analysing the data, particularly important when multi-temporal datasets are to be compared and analysed. The pixel values of level 1C datasets represent radiance at the top of the atmosphere (TOA). As the atmosphere is highly variable in space and time and its influence on the radiance values can be very high, correcting for this effect is important for many uses. The TOA radiance can change as much as 100 % in the visual bands over areas with very low reflectance such as conifer forests, as a function of variance in optical thickness of the atmosphere (Hagolle et al., 2008). ESA has given the German national space center, DLR, the commission to develop an atmospheric correction for Sentinel-2. The module is a special version of ATCOR, a software tool for atmospheric and topographic correction owned by DLR (Rolf Richter, pers. comm.). ATCOR is based on MODTRAN, a radiative transfer code modelling the propagation of radiation through the atmosphere. One advantage of ATCOR compared with similar software is its capability to do corrections in rugged terrain. There are several methods and software tools to perform atmospheric correction of satellite images. In this report, we have focused on ATCOR because it has been selected by ESA. Another option does however exist which has been developed for the French Land Thematic Data Center. It is named MACCS and is, in contrast to ATCOR, based on analysis of bi-temporal or multi-temporal datasets to determine the aerosol content and optical thickness of the atmosphere. MACCS is based on the Successive Order of Scattering (SOS) radiative transfer code (Hagolle et al. 2008). We include a description of MACCS because it is an operational method that is used by a national satellite data center in France.

In the following sub-chapters we present brief descriptions of BEAM, ATCOR and MACCS.

6.3 BEAM

The BEAM (Basic ENVISAT Toolbox for (A) ATSR and MERIS) Earth Observation Toolbox and Development Platform is open source software developed by Brockmann Consult GmbH on behalf of ESA. It consists of a collection of executable tools and Application Programming Interfaces (APIs) for utilization, viewing, processing and analyzing satellite data. BEAM was primarily developed as an ENVISAT/MERIS toolbox, but has since evolved to a toolbox processing Earth observation data in general. BEAM supports raster data formats, such as GeoTIFF and NetCDF and data format from other Earth Observation (EO) sensors than Envisat's optical instrument MERIS (MEdium Resolution Imaging Spectrometer) such as MODIS L2 on Aqua and Terra and TM on Landsat 5 (see further details in Table 5).

Table 5 Overview of formats from various satellite instruments accepted in BEAM

Instrument	Platform	Formats
MERIS L1b/L2	Envisat	Envisat N1
MERIS L3	Envisat	NetCDF
AATSR L1b/L2	Envisat	Envisat N1
ASAR	Envisat	Envisat N1
ATSR L1b/L2	ERS	Envisat N1, ERS
SAR	ERS	Envisat N1
OLCI[1)]	Sentinel-3	NetCDF/SAFE
SLSTR[1)]	Sentinel-3	NetCDF/SAFE
MSI[1)]	Sentinel-2	JPEG2000/SAFE
CHRIS L1	Proba	HDF4
AVNIR-2 L1/L2	ALOS	CEOS
PRISM L1/L2	ALOS	CEOS
MODIS L2	Aqua, Terra	HDF4
AVHRR/3 L1b	NOAA-KLM	NOAA, METOP
TM	Landsat 4	GeoTIFF
TM	Landsat 5	GeoTIFF, FAST
ETM+	Landsat 7	GeoTIFF
OLI, TIRS	Landsat 8	GeoTIFF
SPOT VEGETATION	SPOT	HDF

[1)] In development - available in 2014

BEAM consists of the main components:

- VISAT, an intuitive desktop application used for EO data visualization, analyzing and processing
- A set of scientific data processors running either from the command-line or invoked by VISAT

- A command line tool, graph processing tool, to execute processing graphs (Java API, Application Programming Interface)
- A Java API with ready to use components for remote sensing application development and plug-in points for new BEAM extension modules

BEAM consists of different functional tools:

- Import and export tools
- Imaging tools
- Analysis tools
- Product generation tools
- Time series tools

BEAM also includes other features as the OPeNDAP Client, Graph Processing Framework (GPF) and Data Processors. Typical data processors are the ICOL, the FuB WeW Water processor (MERIS) and the Collocation. Many of these are plug-ins which must be downloaded after installation of BEAM through the Module Manager on the VISAT interface. The standard format of input/output format of BEAM is BEAM-DiMAP which is a profile of the DIMAP format developed by SPOT-Image, France. BEAM can be used for batch processing through command lines and additional modules can be developed through Java BEAM API and integrated in BEAM VISAT[4].

The future toolbox for processing the Sentinel series data, including S2 and S3, will be based on the BEAM software. How it will be implemented and its final modules and processing tools is not decided and will be further developed in the future.

6.3.1 Cloud handling in BEAM

Cloud detection

Cloud detection is a pixel by pixel classification starting with detecting bright pixels (Santer and Ramon, 2011). This is based on using reflectance threshold at 443 nm and threshold values on the spectral slope between the blue and the NIR of the Rayleigh corrected reflectance. Unless the bright pixels are classified as snow using the MERIS Normalized Difference Snow Index (MNDSI) with the 865 and the 890 nm channels, the pixels are classified as clouds.

Cloud probability processor (plug-in BEAM)

A cloud probability algorithm has been developed and implemented in BEAM by the Free University and Brockmann Consult GmbH. Nine spectral bands of MERIS (Refl1-Refl6, Refl9, Refl10 and Refl13), the ratio of band 11 and 10 (an indication of the absorption due to O_2), the ECMWF surface pressure and the exact wavelength of band 11 is used as input to the algorithm.

Simulated cloud and non-cloud free top of atmosphere radiance is produced by using the radiative transfer model MOMO, which calculate the polarized light field in an atmospheric ocean system (AOS) (Hollstein et al., 2010). A neural network is fed with reflectances, the pressure and the output

[4] http://www.brockmann-consult.de/beam-wiki/display/BEAM/Wiki+Home.

of the radiative transfer model. After a post-processing a probability value (0 to 1) is given as output indicating if a pixel can be regarded as a cloud or not.

6.3.2 Atmospheric correction

Atmospheric correction is a necessary step in order to obtain surface reflectance. When the photon reaches the remote sensor its path and origin is not plain and may have reflected off numerous sources (Figure 24), exemplified by:

- Atmospheric scattering due to different types of aerosols
- Specular reflection at the land
- Inland lakes and sea surfaces
- Reflection from surfaces before and after atmospheric corrections

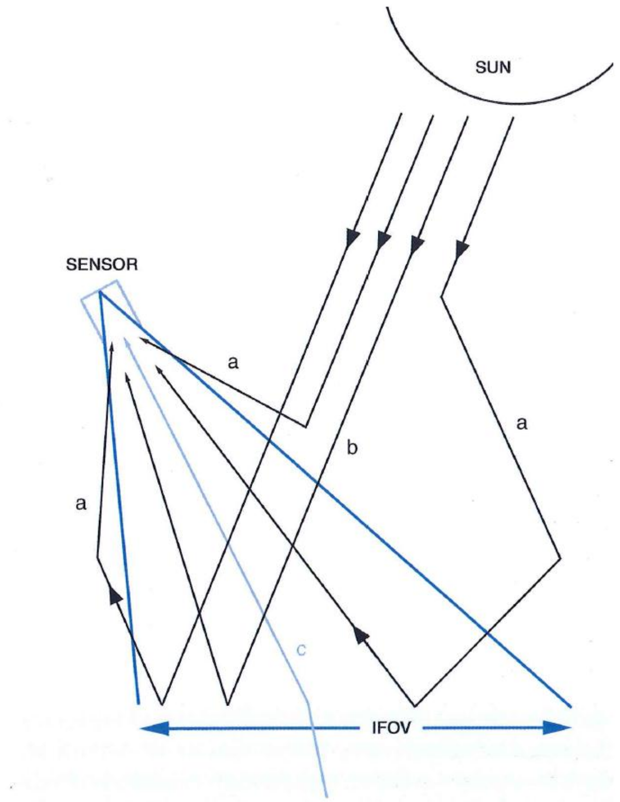


Figure 24 The path of light before it reaches the remote sensor (adapted from IOCCG (2000) and Sathyendranath (1986)). (a) Light scattered by the atmosphere. (b) Specular reflection of direct sunlight at the sea surface. (c) Upwelling light leaving the water surface and travelling in the direction of the sensor. The field of view (IFOV) or the pixel size of the sensor on the surface of the water body is determined by the geometry of the sensor and its altitude and viewing angle.

6.3.3 Atmospheric corrections over land

The initial technique used for atmospheric correction over land is detection of Dark Dense Vegetation (DDV) using a threshold on a spectral index. For MERIS the Atmospheric Resistant Vegetation Index (ARVI, Kaufman et al., 1992) was chosen as a tool for detecting DDV. Here

$ARVI = \frac{\rho_{NIR} - \rho_{rb}}{\rho_{NIR} + \rho_{rb}}$ and

$$\rho_{rb} = \rho_r - \gamma(\rho_b - \rho_r)$$

where ρ_b, ρ_r, and ρ_{NIR} are reflectances corrected for molecular scattering and gaseous absorption observed in the blue, red and the near-infrared channels (443, 670 and 865 nm). ϒ was set to 1.3, a value that reduces sensitivity of ARVI to the amount of aerosol (Santer et al., 2005). The spatial coverage of DDV (on average ~1 % of land surfaces) and the aerosol retrieval has been extended to brighter surfaces. This is denoted LARS (Land Aerosol Remote Sensing, Borde et al., 2003). The LARS reflectance can be viewed as a mixture of bare soil reflectance and dark vegetation reflectance.

When the surface reflectance is predicted, two-wavelength aerosol optical depth retrieval is done over each MERIS pixel. The best couple (AOT and Ångstrøm exponent) is selected.

No more than two aerosol parameters can be retrieved with the MERIS spectral bands (Santer and Ramon, 2011). The focus was put on the Aerosol Optical Thickness (AOT) at one wavelength and an Ångstrøm exponent describing the spectral dependence of AOT in order to correct easily all 15 MERIS bands from aerosol contamination (Santer et al., 1999). The level 2 product processing is done by a unique processor and stored in a unique file with the same spatial resolution as the level 1B data.

6.3.4 Atmospheric corrections over marine and lake waters

Atmospheric correction over water is the procedure used to extract the marine signal from the signals measured by the remote sensor in various spectral bands (Antoine & Morel, 2011). Water leaving radiances, defined as the upwelling radiance just above the sea-surface, should be retrieved at each pixel, from the total radiance received by the instrument. The “marine radiances” at the TOA level are those radiances which originate from photons that have crossed the atmosphere down to the ocean and then have twice crossed the air-sea interface before reaching the sensor after a second atmospheric travel. These radiances contain information about the optical properties of the oceanic upper layers (Figure 25, see below).

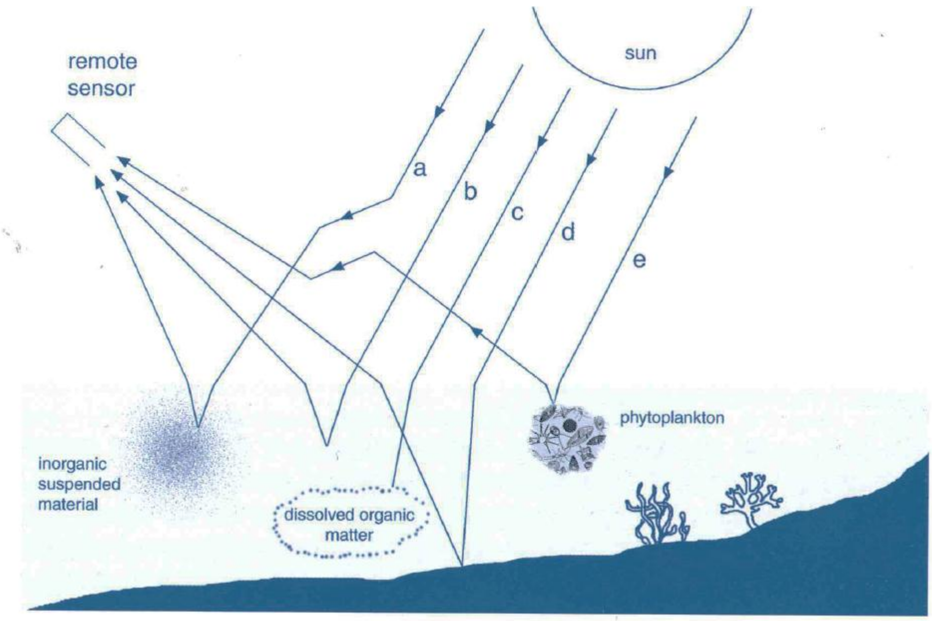


Figure 25 Graphical view of the factors influencing the upwelling light reaching the water surface (adapted from IOCCG (2000)). (a) Upward scattering by inorganic suspended material; (b) upward scattering

from water molecules; (c) absorption by the yellow-substances component; (d) reflection off the bottom; (e) upward scattering from the phytoplankton component.

The radiance backscatter of the ocean-atmosphere system measured by an ocean-colour sensor is a signal in the visible part of the spectrum largely dominated by the “atmospheric path radiance”. These photons have never penetrated the water column, but are scattered by air molecules and/or aerosols. The water-leaving radiance represents only about 10 % of the total radiance in the most favourable conditions (clear atmosphere, small sun-zenith angle and favourable viewing angle) and has to be properly extracted.

The determination of the radiances origination from molecular scattering (Rayleigh scattering) is easily handled as only the illumination, atmospheric condition and wind speed at the sea surface have to be known. The estimation of the radiances due to aerosol scattering is however a more challenging task. The aerosol distribution (both along the vertical and through the field of view) and the aerosol optical properties are not known when processing an ocean colour scene.

Open ocean, case 1 waters

Atmospheric correction schemes have been developed for open ocean case 1 waters since the era of the Coastal Zone Colour Scanner (CZCS) s based on following assumptions (Bovet et al., 2008):

- Atmospheric path radiance can be split in a molecular scattering component (Rayleigh scattering) and an aerosol scattering component.
- The water leaving radiance in the near- infrared spectral bands is neglectably small (due to water absorption) so that the radiance at top of atmosphere, after subtracting the contribution by molecular scattering, is only influenced by aerosols.
- The spectral extinction of aerosol can be described by an exponential function, which allows an extrapolation from the near infrared to the blue-green spectral range.

This procedure has been developed in several versions since the launch of Coastal Zone Color Scanner (CZCS) in November 1978 by NASA and their principles are summarized in an overview by Gordon and Morel (1983). New generation of ocean color sensors required however a more sophisticated procedure to retrieve water substances.

The radiative transfer equation (RTE) is used to describe the radiant field within the ocean-atmospheric system. For a passive medium (i.e. with no internal source of radiation), the equation accounts for the loss of radiance due to scattering and absorption in the direction of propagation, and for the gain of radiance due to scattering from other directions. The Monte Carlo technique is used where photons can be sorted out with respect to their history (number and type of scattering events).

To generate TOA total radiances for molecular or compound (molecules plus aerosols) atmospheres, the radiative transfer was simulated using Monte Carlo (Morel and Gentili, 1191; 1993). Concerning Rayleigh scattering, a 50-layer profile is taken from Elterman (1986).

The total radiance measured in near infrared bands is used first by the algorithm, which is equal to the path radiance since there is no near infrared (NIR) oceanic signal over case 1 waters. Radiances at the TOA level are converted into reflectances through

$$\rho(\lambda, \theta_s, \theta_v, \Delta\phi) = \frac{\pi L(\lambda, \theta_s, \theta_v, \Delta\phi)}{E_s(\lambda)\mu_s}$$

where $E_s(\lambda)$ is the extraterrestrial irradiance (the solar irradiance which reaches the Earth after passage through the atmosphere) for the wavelength in question, L is the radiance, μ_s is the cosine of the solar zenith angle θ_s, θ_v is the satellite viewing angle and $\Delta\phi$ is the azimuth difference between the sun-pixel and the pixel-sensor half vertical planes (Antoine & Morel, 2011). “TOA total reflectance” means the sum of path reflectance (the contribution of the sky to the total reflectance) simulated above a black ocean plus the water leaving reflectance independently calculated, transmitted to the atmosphere.

After the correction of atmospheric effects, the water-leaving reflectances ρ_w are obtained. These are related to the bio-optical state of the water body, through the ratio of the backscattering coefficient (b_b) to the absorption coefficient (a); b_b/a (Morel and Gentili, 1991; 1993; 1996).

Turbid, case 2 waters

The standard level 2 MERIS products from ESA (MEGS, MERIS ground segment processor) do not provide optimal results for all ocean areas and adapted algorithms are required in order to retrieve concentrations of water constituents or to provide the inherent optical properties. This accounts in particular for case 2 water areas (mainly coastal and inland waters). Processors have been developed as plug-ins for BEAM. They consist of procedures for atmospheric corrections and for determining the inherent optical properties and concentrations of water constituents for coastal and different lake waters (case 2 waters).

Atmospheric correction deals not only with turbid waters (where also the NIR spectral bands are influenced by scattering of suspended particles) but also scattering by thin or sub visible cirrus clouds including aged jet trails and specularly reflected sun light present in the nadir radiances and a combination of these problems (Doerffer & Schiller, 2008).

The atmospheric correction of case 2 waters is meant to solve all of the problems mentioned above. The atmospheric correction is defined as the determination of the water leaving radiance reflectance spectrum ($RL_w(\lambda)$) from TOA radiance reflectance spectrum ($RL_{TOA}(\lambda)$). This requires determination of the radiance backscattered from all targets above the water surface including air molecules, aerosols, thin clouds, foam on the water surface, as well as all radiance which is specularly reflected at the water surface, in particular the sun glint. In addition the transmission of the solar radiance through the atmosphere to the water surface and of the radiance from the water surface to the sensor has to be computed.

Case2 Regional (Plug-in)

The Case2 Regional (C2R) is implemented as a plug-in in BEAM. This module consists of three different algorithms which deal with atmospheric corrected reflectances and water quality constituents for MERIS data. “It requires determination of the radiance backscattered from all targets above the water surface including air molecules, aerosols, thin clouds and foam on the water as well as all radiance which is speculary reflected at the water surface, in particular sun glint. Furthermore, the transmission of the solar radiance through the atmosphere to the water surface and of the radiance from the water surface to the sensor has to be computed." (Doerffer and Schiller, 2008)

This atmospheric correction is based on simulations of radiative transfers. A neural network (NN) is trained to use this simulated radiance reflectance, and the NN is used for the parameterization of the relationship between the TOA radiance reflectance and the water leaving radiance reflectance.

The water leaving radiance reflectance is simulated for all bands. The water leaving radiance reflectance RL_w is determined directly from the corresponding RL_{tosa}[5] values by the NN. The disadvantage of this is that a full bio-optical model for the water has to be included, but the advantage is that the extrapolation from the NIR bands is avoided.

For the bio-optical model, two NNs is used, a forward (forwNN) and a backward (invNN). Input is water leaving radiance reflectance from 8 MERIS bands which are outputs from the atmospheric correction described above. Inherent optical properties such as total scattering of particles (suspended matter, TSM) b_tsm, the absorption coefficient of phytoplankton pigments a_pig, and the absorption of dissolved organic matter a_gelb (gelbstoff) are derived at 443 nm (band 2). The algorithm relates the bi-directional water leaving radiance reflectance with the concentration variables. An inverse model is used to get an estimate of the concentrations by reflectance and geometry ($\theta_s, \theta_v, \Delta\phi$) data. As an optional optimisation step these concentration are used in a minimization procedure with a forward model iteratively to minimize the difference between the calculated reflectance (using concentrations and geometry) and the measured ones.

The FUB/WeW Water Processor (Plug-in)

The FUB/WeW Water Processor is an additional plug-in. This processor consists of different algorithms, and one is for atmospheric correction of MERIS data. The algorithm makes use of MERIS Level 1TOA radiances in the bands 1-7, 9-10 and 12-14. By using these bands, water leaving reflectance and case II water properties along with AOT bands are retrieved. The combination masks GLINT_RISK, BRIGHT, INVALID and SUSPECT are additional masks applied before the retrieval of the data.

The retrieval procedure of this processor is based on four separate artificial neural networks. These neural networks are trained based on the results from the radiative transfer simulations by the MOMO model (Figure 26). The MOMO code takes different atmospheric and oceanic conditions into account and generates a look up table (Hollstein et al., 2010).

[5] tosa - top of standard atmosphere.

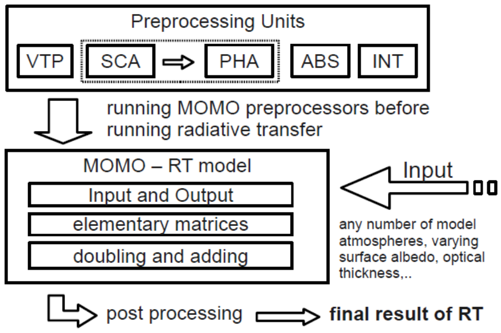


Figure 26 Schematic overview of the radiative transfer model MOMO (Inherited from Hollstein et al. (2010)).

Processing chain

A typical processing chain for atmospheric correction over water from coastal areas is listed below:

1. Downloading L1 products (TOA radiance or reflectance data) from available archive
2. Using toolbox for subsetting to area of interest (AOI) if this is not already done by the archive service
3. Processing the satellite image
4. ICOL processing (optional)
5. Atmospheric processing, with and without ICOL
6. Extracting pixel data at stations through pixel extraction tools

It is important to mention that this would be a substitute for the L2 MEGS products. It will still require proper validation with in situ measurements after processing and the processing is done to allow for adjacency effects (reflections from land) and turbid waters (coastal and inland waters).

6.3.5 BEAM and functionality for Sentinel-2 data

The focus of the multi-spectral imaging mission decides the width and the resolution of the spectral bands. A conservative approach with a basic set of bands is satisfactorily for detecting and mapping land cover/change, while for an improved surface monitoring mission with an adequate quantification of geo-biophysical variables, a high amount of spectral bands with narrow widths would be needed. Therefore with a larger number of bands, a more accurate classification, land use change mapping as well as vegetation status monitoring for agriculture and forestry is feasible. See Drusch et al. (2010) for further description of heritage on the MSI-sensor on Sentinel-2.

Landsat 5 TM will satisfy the current minimum service requirements in areas with extensive regional in-situ information. Still, for more improved land cover/change classification, atmospheric correction, cloud/snow separation and the quantitative assessment of vegetation status in the future, additional spectral channels are needed. For the Sentinel-2 sensor the bands described in Table 6 are required to satisfy the mission objectives for both the basic and the advanced services. Several of the channels are heritage from other missions such as MODIS (Moderate Resolution Imaging Spectrometer), ALI (Advanced Land Imager), LDCM (Landsat Data Continuity Mission) for channel 1 and MERIS among others for channel 2, 3, 4, 5, 6, 7, 8a and 9. See Senintel-2 MRD.pdf for further description of heritage on the MSI-sensor on Sentinel-2.

Landsat 8 provides measurements of the Earth's terrestrial and Polar Regions in the visible, near-infrared, short wave infrared and thermal infrared light. Land use planning and monitoring in regional to local scales, support of disaster responses and evaluations and monitoring of water use is the routine use of these measurements. It will also supply measurements for climate, carbon cycle, ecosystems, water cycle, biogeochemistry and Earth surface/interior research (http://landsat.gsfc.nasa.gov/?page_id=4091). Operational Land Imager (OLI) and Thermal InfraRed Sensor (TIRS) band data can be converted to TOA spectral radiance and/or TOA planetary reflectance using the radiance rescaling factors and the reflectance rescaling coefficients provided in the metadata file (http://landsat.usgs.gov/Landsat8_Using_Product.php).

In Table 6 an overview of the different channels in the MERIS, MSI and OLI/TIRS sensors is given and compared to show partly the overlapping spectral bands. Sentinel-2 has heritage from both the MERIS sensor and Landsat sensors and have therefore good correspondence with both the MERIS and the OLI/TIRS sensors.

Table 6 An overview over the channels of the MERIS, MSI, and OLI/TIRS sensors

	Channel	Band center (nm)	Sensor (satellite)	Chanel	Band center (nm)	Sensor (satellite)	Channel	Band center (nm)	Sensor (satellite)
Visible	1	412.5	MERIS (Envisat)						
	2	442.5	MERIS (Envisat)	1	443	MSI (Sentinel-2)	1	433	OLI/TIRS (Landsat 8)
	3	490	MERIS (Envisat)	2	490	MSI (Sentinel-2)	2	482	OLI (Landsat 8)
	4	510	MERIS (Envisat)						
	5	560	MERIS (Envisat)	3	550	MSI (Sentinel-2)	3	562	OLI (Landsat 8)
	6	620	MERIS (Envisat)						
	7	665	MERIS (Envisat)	4	665	MSI (Sentinel-2)	4	655	OLI (Landsat 8)
	8	681.25	MERIS (Envisat)						
	9	708.75	MERIS (Envisat)	5	705	MSI (Sentinel-2)	8	590	OLI (Landsat 8)
				6	740	MSI (Sentinel-2)			
Infrared	10	753.75	MERIS (Envisat)						
	11	760.625	MERIS (Envisat)						
	12	778.75	MERIS (Envisat)	7	783	MSI (Sentinel-2)			
				8	842	MSI (Sentinel-2)			
	13	865	MERIS (Envisat)	8a	865	MSI (Sentinel-2)	5	865	OLI (Landsat 8)
	14	885	MERIS (Envisat)						
	15	900	MERIS (Envisat)						
				9	945	MSI (Sentinel-2)			
				10	1375	MSI (Sentinel-2)	9	1375	OLI (Landsat 8)

		11	1610	MSI (Sentinel-2)	6	1610	OLI (Landsat 8)
		12	2190	MSI (Sentinel-2)	7	2200	OLI (Landsat 8)
					10	10800	TIRS (Landsat 8)
					11	12000	TIRS (Landsat 8)

BEAM was originally developed to facilitate the utilization of image data from the optical sensors MERIS and AATSR aboard Envisat. BEAM now supports a growing number of other optical instruments as well as general raster data formats (GeoTIFF, HDF, and netCDF).

For Sentinel-2 the processing toolbox will be based on BEAM. The development started early 2014 and will last for 30 months and the first version will be released in autumn 2014 (C. Brockmann, pers. comm.). The ICOL processor in BEAM deals with the adjacency effect in coastal areas, especially in the near infrared bands. This effect is due to photons reflected and scattered towards the sensor from nearby land surfaces. BEAM currently supports input products such as MERIS 1b, Landsat 5 TM and Landsat 7 ETM+. BEAM was also operational for reading Landsat 8 OLI/TIRS data in GeoTIFF formats in September 2013. As a consequence, it is realistic to anticipate an ICOL processor handling Landsat 8 data as well.

6.4 ATCOR

Atmospheric and Topographical Correction (ATCOR) is a tool for correcting remotely sensed data for atmospheric effects. It is based on a radiative transfer model that calculates the attenuation of the solar energy traveling from the top of the atmosphere to the ground target and back to the satellite sensor. This tool developed by Richter at Deutschen Zentrums für Luft- und Raumfahrt (DLR) (Richter and Schläpfer, 2013) is a widely used tool. ATCOR is distributed as an add-on module to ERDAS Imagine and PCI Geomatics, two widely used image analysis software packages, and as a stand-alone tool from ReSe Applications. ATCOR can do an all-automatic processing of images and calculate BOA values from the TOA radiance values measured.

Remotely sensed data are used for mapping surface properties on Earth from space. Scattering and absorption by air molecules and aerosol particles does however affect signals recorded by the satellite sensors. The influence typically varies from day to day, from hour to hour and over distances of a few kilometers. Figure 27 shows a basic sketch of the three radiation components in the solar region in the wavelength spectrum from 0.35–2.5 µm:

- The path radiance L_1, composed of sunrays or photons scattered by the atmosphere into the sensor's instantaneous field-of-view (IFOV).
- The reflected radiance L_2 from a target area covered by the IFOV (pixel on the ground). Direct and diffuse solar radiation hit the target and a fraction is reflected towards the IFOV of the sensor. The sum of diffuse and direct solar radiation hitting the surface is called global flux E_g.
- The reflected radiance L_3 from the neighbourhood of the target carries no information of the target pixel. The adjacency radiation is scattered by the air molecules and aerosols into the sensor's IFOV.

The total radiance L as measured by the sensor is the sum of the three components:

$$L = L_1 + L_2 + L_3$$

ATCOR reads the radiometric calibration coefficients from the metadata of the image datasets and calculates radiance with:

$$L = c_0 + c_1 DN$$

The calibration coefficients, the offset c_0 and gain c_1, are measured in the laboratory before launch and in flight to correct for drift in the values. DN represents a digital number, usually an unsigned integer with values in the interval $[0, 2^B - 1]$, where B is the number of bits. The coefficient values are provided in a metadata file that follows each dataset (see also Chapter 6.6).

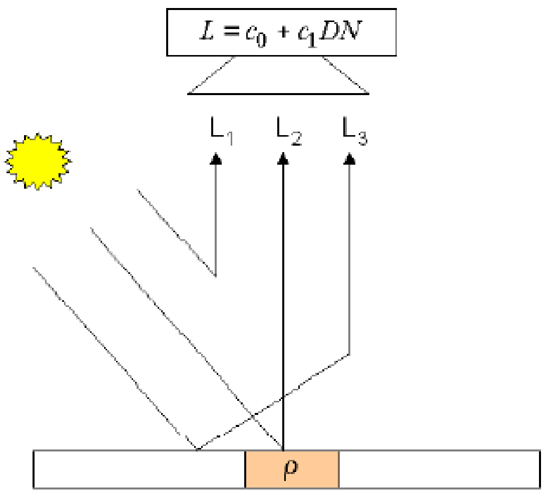


Figure 27 Radiance measured by the sensor on a remote sensing platform (Richter and Schläpfer, 2013). Only one component, L_2, carries information of the target pixel.

A very simple reflectance model can be outlined as follows if the adjacency effect is ignored:

$$L = L_1 + L_2 = L_1 + \frac{\tau \rho E_g}{\pi} = c_0 + c_1 DN$$

where τ is the transmittance of the atmosphere, ρ is the reflectance of the target and E_g is the global flux incident on the target. Solving for the target reflectance ρ we get:

$$\rho = \frac{\pi L - L_1}{\tau E_g}$$

The path radiance L_1, global flux E_g, and transmittance τ are unknown and must be calculated in order to calculate the target reflectance.

All the unknowns are influenced by the visibility or optical thickness of the atmosphere. The optical thickness is a number determined by three components:

- Molecular scattering
- Aerosol scattering
- Molecular absorption.

This sum up to total optical thickness δ. Visibility, a quantity used in meteorology, is related to optical thickness. Transmittance is defined as

$$\tau = e^{-\delta}.$$

Transmittance decreases with increasing optical thickness (or decreasing visibility). ATCOR uses MODTRAN 5, a radiative transfer code, to calculate the fraction of radiation that will pass a certain path length of the atmosphere. The calculations are based on the optical properties of the atmosphere. Reference areas of known reflectance are necessary to estimate the optical properties, and ATCOR uses the dark dense vegetation (DDV) method to define reference pixels. If a SWIR band exists (as in Sentinel-2 and Landsat 8 datasets), the image is searched for dark pixels in this band. A visibility of 23 km is assumed and the SWIR reflectance is calculated. Only pixels with reflectance above 1 % are used to exclude water pixels and the algorithm searches until it defines 1 % of the pixels in the image as dark reference pixels. A correlation with the red band is used to estimate the red reflectance of the reference pixels:

$$\rho_{0.650} = 0.5\rho_{2.5}$$

where 0.650 (µm) denotes the red band and 2.5 (µm) the SWIR band. The value of the red reflectance is used to estimate the optical properties of the atmosphere using the MODTRAN 5 code in the form of a look up table (LUT). In the LUT, the radiance at the top of the atmosphere is modeled as a function of the visibility. From the intersection with the actually measured radiance by the sensor the actual visibility can be determined. This is illustrated in Figure 28, where the stippled line represents the measured radiance value of a reference pixel and the curved line the MODTRAN 5 model. The unknown parameters describing atmospheric properties necessary to calculate surface reflectance are determined by using the estimated visibility.

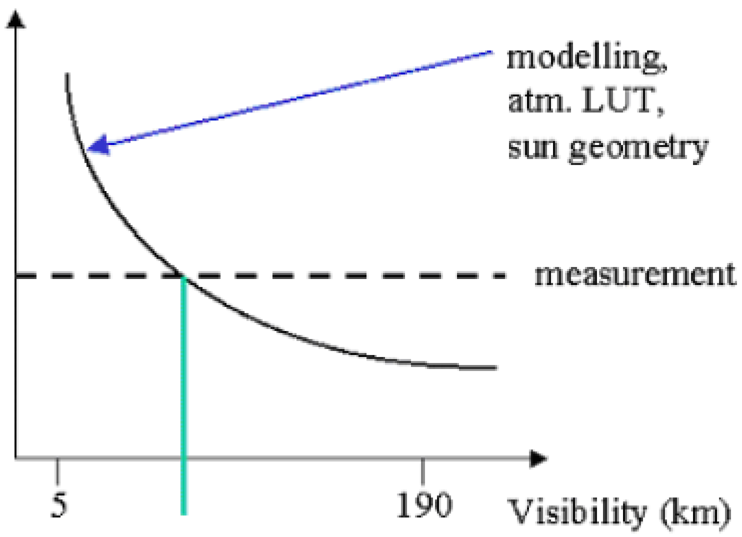


Figure 28 Estimation of visibility with a reference pixel with known reflectance ρ in the red band (Richter and Schläpfer, 2013).

ATCOR uses a reflectance model that includes the adjacency effect, L_3 and therefore has to make several iterations to calculate the reflectance of a target pixel. The strength of the adjacency effect depends on the difference in brightness between the actual pixel and its surroundings and thus several iterations are necessary. For rugged terrain a fourth component L_4 is introduced. It takes into account reflected terrain radiance from slopes in the neighborhood hitting the target pixel. A terrain view factor $V_{terrain}$ defined as $1 - V_{sky}$, is calculated and used to scale the effect of terrain radiance. Several iterations are necessary because the effect depend on the brightness of the terrain.

Empirical bidirectional reflectance distribution function (BRDF) corrections are available as a post-processing step following the reflectance calculations. There is large variation in the direct irradiance on slopes in rugged terrain as slopes facing the sun receive more irradiance than slopes facing away from the sun's direction. The reflectance model of ATCOR assumes Lambertian (isotropic) reflectance, which is unrealistic for many surface categories. Natural cover types, e.g. forests and meadows, usually exhibit bidirectional brightness variations, and typically show increasing reflectance with increasing solar zenith and view angle. This leads to overcorrection of poorly illuminated terrain slopes in the reflectance model ATCOR uses. ATCOR correct for this effect by using a geometric function in a post-processing step:

$$\rho_g = \rho_L G$$

where $\rho_L$is the reflectance assuming Lambertian reflectance. G is determined automatically by ATCOR and has values in the interval [0.2, 1.0].

ATCOR does some preprocessing steps before calculating reflectance. These consist of masking areas covered with haze, cloud, water and shadow. After masking the areas ATCOR can run an automatic process to remove haze, cirrus clouds and shadows. The main processing steps of ATCOR are illustrated in Figure 29. The processing steps necessary in searching for and finding dark reference pixels (DDV pixels), calculating surface reflectance in the red band of the reference pixels, and finally in estimating the visibility of the reference pixels, are illustrated in Figure 30.

ATCOR needs some cloud free areas in order to work. It will not lead to useful results unless the cloud cover is lower than 50 %. The code can be fully automated in a batch process (D. Schläpfer, pers. comm.). DLR, which is the owner of the code, has granted ESA an ATCOR license for a special ATCOR version tailored to Sentinel-2. ESA has a non-exclusive, non-transferable right to use the code for the Sentinel-2 project. The special code version can be used in BEAM (described in Chapter 6.3). The code name for the special version of ATCOR will be SEN2COR (R. Richter, pers. comm.). It will only work with Sentinel-2 data and cannot be used for Landsat 8 data (D. Schläpfer, pers. comm.), and an ordinary ATCOR license will therefore be necessary to process Landsat 8 data.

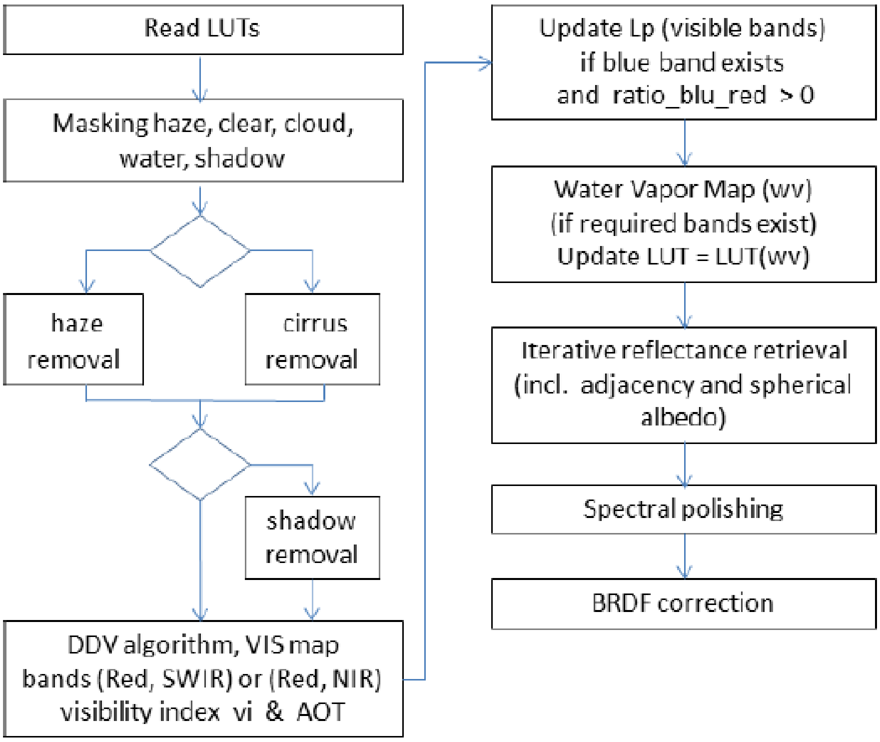


Figure 29 The main processing steps to estimate surface reflectance (Richter and Schläpfter 2013).

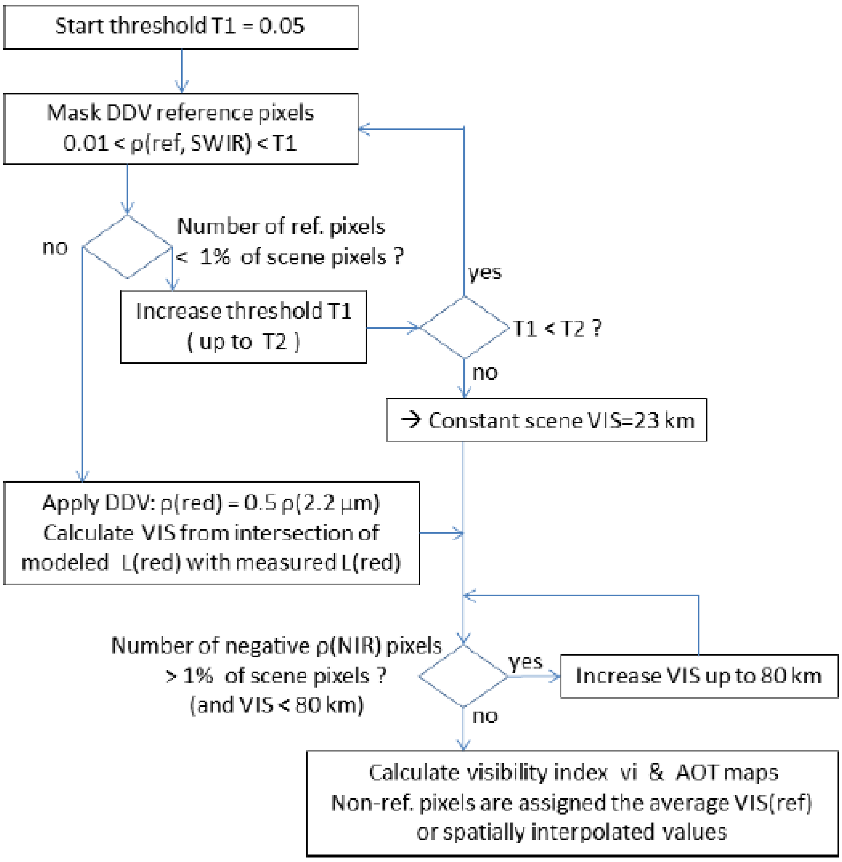


Figure 30 Sketch of the process to estimate visibility and aerosol optical thickness. DDV is dark dense vegetation, often coniferous forest (Richter and Schläpfter 2013).

6.5 MACCS

MACCS is a Multisensory Atmospheric Correction and Cloud Screening processor. The processor yields three levels of products:

1. Top of the atmosphere (TOA) reflectance and orthorectified data
2. Bottom of the atmosphere (BOA) surface reflectance and orthorectified data (incl. cloud and cloud shadow masking)
3. A weekly composite from a 15 day time span from surface reflectance. This means reusing recordings made during the last eight days of the last composite.

The idea behind this processor is to provide scientific users and public policy actors with datasets that are ready to use with a “plug and play” setup for time series and change analyses.

MACCS uses multi-temporal and multi-spectral information and criteria to:

- detect clouds and cloud shadows
- estimate aerosol content and perform corrections

Very few users have ordered high resolution images with high cloud cover and few users have used time series with more than 10 images until recently (Hagolle et al., 2010). Time series based on relatively few images made it possible to manually delineate clouds and cloud shadows. Sentinel-2 will however in the near future provide time series with large numbers of acquisitions, e.g. 50 to 100 or more. Thus, the number of images to process will make it impractical and impossible to manually detect and screen out clouds and cloud shadows.

Screening out clouds and cloud shadows and making a cloud mask is a necessary preliminary step before further image analysis. This includes computing spectral indices, atmospheric corrections, image compositing for cloud free datasets and time series analysis. This is costly and time consuming if done manually on individual recordings. Hagolle et al. (2010) describe a method for producing cloud masks called Multi-Temporal Cloud Detection (MTCD). It is based on the fact that the surface reflectance of a land pixel usually varies very slowly with time, especially in the wavelength range 400–500 nm. Thus, a very large increase in reflectance in this range over a time span of few days is very likely caused by the appearance of a cloud. The classical approach to detect clouds is based on thresholds in spectral bands.

When available the thermal band is used for detecting clouds colder than the land surface. Thin and very low clouds are rarely significantly colder than the land surface and it is better to base detection of those on reflectance in the blue band. This may, however, fail in the event of very bright land surfaces.

Bands in the short wave infrared (SWIR) domain are often used to discriminate snow from clouds. This method is based on the fact that SWIR reflectance is much lower for snow than for clouds. The multispectral sensor MSI on Sentinel-2 will have an infrared band (1365–1395 nm), and this band is ideal for detecting cirrus clouds, which are high-altitude clouds. Water vapor absorption is very strong in this band and land surface will be invisible. This property will allow cirrus clouds to be detected against a very dark background. The availability of different spectral bands on the different satellites varies and thus the classical approach to detect clouds varies with the different satellites. The MTCD method circumvent this by utilizing multi-temporal datasets and relies on sudden increases in reflectance in the range 400–500 nm, a range present in most satellite missions. Hagolle et al. (2010) argue that the approach should perform better than single-date thresholds in the blue band.

6.5.1 Cloud detection

The Multi-Temporal Cloud Detection method is based on the following steps (Hagolle et al. 2010):

- A threshold on the difference in the blue band is computed. For an image on day D, a cloud free reference image is needed, to compute the variation. The reference image D_r is a composite of the most recent cloud free pixels. The threshold depends on the number of days between D and D_r and when dates are very close the threshold is usually around 0.03, but the value doubles when the dates are separated with e.g. 30 days to allow for change in surface reflectance.
- The algorithm does not work well above inland and coastal waters because waters are prone to sudden changes in reflectance due to turbidity, sun glint or foam. Thus water pixels must be discarded before computing the cloud mask.
- Sudden changes of the land surface may occur: over agricultural areas due to cropping and ploughing and over natural areas due to snow cover, fires, or quick drying of the vegetation. To cope with such changes two extra tests are added to check if the change is really caused by cloud cover:
    - Test if the change in red reflectance is much greater than the change in blue reflectance. This usually happens when a field is ploughed or cropped or vegetation dries between two dates.
    - Test the neighborhood to see if the pixels are well correlated with the images taken before date D. It is very unlikely that a cloud stays at the same place and with the same shape over many days, so a high correlation means a transparent atmosphere and no clouds.
- SWIR bands are used to discriminate snow and clouds and the method is based on the Normalized Difference Snow Index (NDSI) that is defined with $(\rho_{Green} - \rho_{SWIR}) / (\rho_{Green} + \rho_{SWIR})$
- The cloud mask is dilated to cover thin clouds that usually surround thicker clouds. A width of two pixels is used.

6.5.2 Atmospheric correction

A satellite borne sensor records radiance at the top of the atmosphere (TOA) and the effect of the aerosols in the atmosphere is included in the measurements. The abundance and nature of the aerosols are very variable with time and space and therefore difficult to model. Over dark surfaces, like dense conifer forest, the TOA reflectance can change as much as 100 % caused by variations in the atmospheric condition. Thus, correcting for atmospheric effects prior to time series analyses and change detection is necessary.

Atmospheric correction is no easy task which this simple equation, a first order radiative transfer function, explains (Hagolle et al., 2008):

$$\rho_{\mathrm{TOA}} = t_{\mathrm{g}}[\rho_{\mathrm{surf}} \cdot T_{\mathrm{atm}}(\mathrm{AOP}) + \rho_{\mathrm{atm}}(\mathrm{AOP})],$$

where ρ_{TOA} is the measured TOA reflectance, t_{g} is the transmittance due to atmospheric absorption, ρ_{surf} is the unknown surface reflectance, T_{atm} is the scattering (Rayleigh and aerosol), AOP is the Atmospheric Optical Properties, and ρ_{atm} is the atmospheric path reflectance. The equation has two unknowns: ρ_{surf} and AOP.

In MACCS, two assumptions are used to solve the problem:

- The aerosol optical properties vary quickly with time and slowly with location
- The surface reflectance vary slowly with time and quickly with location

Thus, according to the two assumptions, a variation in the TOA reflectance over a short time period is likely due to variation in aerosol optical properties. This offers a possibility to estimate the AOP and is the basis for the MACCS method for atmospheric correction. The method uses two images, one from day D and another from day D + h, where h is the number of days after day D. It is assumed that the images are acquired within a few days, at the same local time (sun synchronous orbit), and at the same viewing angle. This assures that the sun angle and viewing angle are nearly constant and the reflectance is therefore not significantly affected by these factors.

The Successive Orders of Scattering code is a central component in the method. It provides look-up tables (LUT) for TOA reflectance as a function of surface reflectance for different values of the following parameters:

- Aerosol optical thickness
- Viewing and solar angles
- Wavelength
- Surface altitude

An inversion of the look-up tables is made to calculate surface reflectance (BOA reflectance) as a function of TOA reflectance and the parameters given above. Minimization of a cost function is used to estimate a value for AOP. Over a small neighborhood around pixel coordinates (i, j) a value for AOP is searched that minimize the sum of squares of differences between the acquisitions of day D and day D + h (Figure 31). Over a short duration of a few days, the surface reflectance is normally stable and change little. Observed changes in TOA reflectance are mainly caused by variation in aerosol optical thickness.

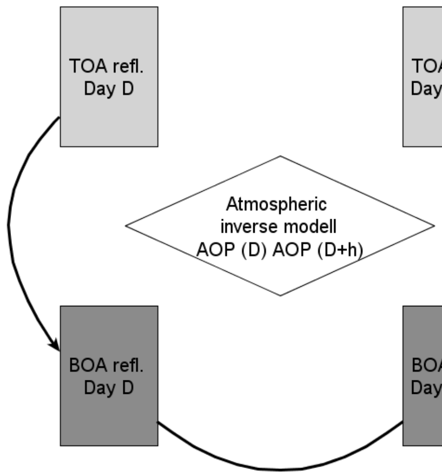


Figure 31 The illustration shows a scheme of the cost function and how it is used to find the AOP that minimizes the difference between surface reflectance at the bottom of the atmosphere (BOA) of day D and D + h (after Hagolle et al., 2008).

A problem with the method occurs when the TOA reflectances for days D and D + h are identical, because then any constant value for AOP or optical thickness will produce the same surface reflectance. This problem is solved by introducing a second cost function that adds a term that compares the surface reflectance of day D and D + h with the calculated surface reflectance of day D from a previous iteration using day D and day D – h. The rationale with the expanded cost function is that it is unlikely that the optical thickness is stable over all three days.

Surface reflectance variations will cause a problem for the method. Human activities such as ploughing and harvesting, darkening of the soil after rain, sudden increase in vegetation in the spring and drying of the vegetation in the summer will cause sudden change in surface reflectance. Since the method assumes stable reflectance, this will cause a problem. An algorithm that detects pixels with sudden change in reflectance is therefore needed, and the aerosol inversion algorithm cannot be applied for these pixels. Spectral bands in the near and middle infrared bands are ideal to detect such reflectance changes since these bands are little affected by aerosols. Other limitations of the method are when the surface is very uniform (e.g. deserts) or very bright. In these situations the inversion of the optical thickness is not stable. The method uses a continental aerosol model to extrapolate aerosol values estimated in the blue and green bands, and this can sometimes lead to inaccuracies when the situation deviates from the continental model. Undetected thin cirrus clouds will also affect the inversion of optical thickness.

The MACCS method can be used with Sentinel-2 data, even if the revisit time is 5 days (Hagolle et al., 2008). The SWIR bands of the Sentinel-2 sensor, centred on 1.6 µm and 2.2 µm, are ideal for detecting sudden surface reflectance variations, and the SWIR band centred on 1.38 µm is ideal for detecting thin cirrus clouds.

6.6 Radiometry and radiometric correction

Radiance is the basic unit measured by an optical remote sensing instrument. It is a physical quantity that corresponds to brightness of an object and is measured as watts per square meter per steradian [W · m^{-2} · sr^{-1}]. The radiance from perfectly diffuse objects (Lambertian reflectors) is the same in all viewing directions. Many natural objects have anisotropic reflectance and the radiance varies with viewing direction, e.g. forested areas have a hotspot in the direction of the sun (tree shadows are not visible). With the wide swath (290 km) of Sentinel-2, the view angle will vary relatively much across an image, and this could lead to significant variations in the measured radiance values from non-Lambertian targets. Bi-directional reflectance distribution functions (BRDF) model such variations.

The optical image sensor is a converter that measures light intensity and converts it into an analogue signal that can be read by an instrument. The sensor is composed of photosensitive detectors that are monochromatic light meters, i.e. they don't separate different colours or spectral bands. A spectral filter with a precisely defined bandwidth is therefore placed in front of each detector to block out photons with wavelengths (frequencies) outside the upper and lower cutoff frequencies of the filter. Each detector is a capacitor which can store electric charge. When an optical image is projected through the lens onto the detectors, an electric charge is created that is proportional to the light intensity at the location. The charge is accumulated over the exposure time.

After exposure, a charge amplifier converts the electric charge of each detector to voltage and amplifies this electrical signal before the voltage is converted to a digital number (DN) in an analogue-to-digital (A/D) converter. The resolution of the converter determines the number of discrete values it can produce from the range of the continuous signal. Since the discrete values are stored in binary form, the resolution is expressed as a power of two. The converters of the sensors aboard Landsat 8 and Sentinel-2 have 12 bits resolution and can thus represent 2^{12} = 4096 values in the interval [0, 4095]. The converter of Landsat 7 has 8 bits resolution and can represent 2^8 = 256 values in the interval [0, 255]. A/D-conversion introduces errors that need to be minimized, and quantization error (rounding error) and non-linearity are two common sources. The sensor sensitivity indicates how much the output changes with a change in the measured physical quantity. The sensitivity is defined as the ratio between the output signal and the measured quantity. For an optical sensor that converts light intensity to an analogue voltage signal, the sensitivity can be expressed as a ratio with the unit [V/W]; i.e. output voltage per unit radiant power. For a linear sensor, this ratio is constant over the range of measurements. If the output signal is not zero when the input is zero, the sensor has an offset or bias. The range of values of the output signal is limited, and when the input exceeds the limits the detector will saturate.

The following formula gives the relation between the digital number V_{DN} representing quantized voltage and spectral radiance:

$$L_\lambda = \frac{kV_{DN}}{\beta^2 A_p \cos\theta} = \frac{\varphi_\lambda}{\beta^2 A_p \cos\theta}$$

where L_λ is spectral radiance [W · m^{-2} · sr^{-1} · nm^{-1}], *k* is the calibration constant to convert the voltage signal to spectral flux φ_λ [W · nm^{-1}], β^2 [sr] is the instantaneous field of view (IFOV), A_p [m^2] is the aperture of the telescope, and θ is the viewing angle. Each image dataset is supplied with a metadata file that contains coefficients that allow the user or a dedicated program to calculate spectral radiance values from the digital numbers (see p. 65).

Sentinel-2 level 1B data will be delivered as spectral radiance values and level 1C as spectral reflectance values at the top of the atmosphere (Gjertsen and Trollvik, 2013). Spectral reflectance at the top of the atmosphere (i.e. at the satellite) $\rho(\lambda)_{sat}$ can be computed with

$$\rho(\lambda)_{sat} = \frac{\pi L_\lambda}{(E_{\lambda,s} \cos\theta_z)/d^2}$$

where *d* is Earth-Sun distance (in astronomical units), $E_{\lambda,s}$ spectral irradiance from the sun, and θ_z solar zenith angle. For level 2A, the reflectance at the bottom of the atmosphere $\rho(\lambda)$ is computed. A general formula for this calculation is

$$\rho(\lambda) = \frac{\pi(L_\lambda - L_{\lambda,\mathrm{d}})d^2}{T_{\lambda,v}(E_{\lambda,s} \cos\theta_z T_{\lambda,z} + E_{\lambda,\mathrm{a}})}$$

where $T_{\lambda,v}$ is transmittance from the ground to the satellite, $T_{\lambda,z}$ transmittance from top of the atmosphere to the ground, and $E_{\lambda,\mathrm{a}}$ diffuse irradiance from the atmosphere onto the ground. ATCOR and MACCS described in Chapters 6.4 and 6.5 are examples of tools that can perform atmospheric correction.

The core ground segment will perform radiometric correction when processing raw data from level 1A to 1B. To get physical values such as radiance and reflectance from the digital numbers, the sensor sensitivity needs to be known very precisely. The ground station will, after receiving the data, do radiometric corrections to deal with among other things: dark signals, detector relative sensitivity variations, defective detectors, crosstalk, and de-noising (Baillarin et al., 2012).

For time-series analysis it is important that the radiometric calibration of the detectors is updated, so that the physical values are comparable over time. The preflight values of detector sensitivity are accurately measured in the laboratory, but after launch, in-flight radiometric calibrations are necessary because the detector sensitivity will degrade over time.

In addition, to fully exploit the synergy of the Landsat and Sentinel-2 missions, the physical values of the two sensors need to be inter-calibrated. In a summary report from the Sentinel-2 preparatory symposium in 2012 (Arino et al., 2012), it was pointed out that ESA and USGS should cooperate and make sure the data are processed to a common standard and that the sensors, MSI and OLI, are inter-calibrated to enable data fusion. After launch, vicarious radiometric calibration and validation (Cal/Val) processes should be coordinated between the Landsat and Sentinel-2 teams: e.g. agree on and share methods and datasets (Arino et al., 2012). We support the recommendations in Arino et al. (2012).

6.7 Summary

Several important preprocessing tasks have been described in this chapter. ESA will produce and deliver two public products: level 1B and 1C. Level 1B will be delivered with an enhanced sensor model describing the position and viewing direction for each detector element in the focal planes. The data will not be projected and resampled and will thus be delivered in sensor geometry. The sensor model of new acquisitions is enhanced by a complex process where an accurate reference image database is used in an image matching process to find control points with known geographic coordinates. The control points are used to refine the sensor model. Level 1C is an orthorectified product where the image has been projected to a cartographic projection (UTM/WGS84) and where the pixels in the process have been resampled to fit a predefined output grid defined in the map coordinate system.

For level 1C, the land-water mask and cloud mask are computed based on spectral criteria (Baillarin et al., 2012). Low- and middle-altitude clouds have high reflectance in blue and SWIR bands; high-altitude clouds (cirrus) have low reflectance in the blue bands and high reflectance at 1380 nm (corresponds to band 9 of the Landsat 8 OLI sensor and band 10 of the Sentinel-2 MSI sensor); and water has low reflectance in the near infrared and SWIR bands. As far as we know, the cloud mask attached to level 1C products will not include cloud shadows. However, cloud shadows can be detected with ATCOR and MACCS, but robust detection of cloud shadows is probably more difficult than robust detection of clouds and water.

Orthorectification errors can be high if the elevation model used has low resolution and accuracy. On behalf of ESA, CNES is developing a prototype processor for geometric corrections (Baillarin et al., 2012). This processor will generate a global reference image dataset, where the sensor models have been simultaneously enhanced based on a large space-triangulation adjustment. ESA has not decided (at the time of writing) if the processor will be made available for collaborative ground segments. It is still unclear to us what alternative software could be used to orthorectify level 1B data using a detailed and accurate national DEM. Our understanding is that ESA will use the enhanced SRTM DEM in two steps: 1) in the enhancement of the geometric sensor model of the level 1B datasets, and 2) in the orthorectification process to level 1C. Special software is probably needed to perform the first step, but for the second step it is likely that commercial image processing systems will provide a solution. Thus, to improve the geometric accuracy by using a national DEM can be a highly complex task if both steps have to be made, but probably much simpler if only step two is needed to obtain satisfactory geometric accuracy.

Atmospheric correction will not be made by ESA, but a toolbox will be available to the users to perform this preprocessing step. A module for BEAM based on ATCOR has been made. A limitation is that this toolbox will only be licensed to work on Sentinel-2 data. Therefore, another solution is needed for Landsat data. ATCOR can also be delivered as stand-alone software that can process data from both satellite missions. MACCS can be an alternative, but its limitation is that the method needs very frequent acquisitions (every three to five days), which can be difficult over Norway with many cloudy days. A highly robust and automatic processor is needed to mass produce atmospherically corrected datasets. We need to test the robustness and to what degree the process can be automated.

In Table 7 we present a summary of the preprocessing steps together with an assessment of the expected degree of difficulty to perform the steps by a collaborative ground segment.

*Table 7 Summary of the preprocessing steps and assessed degree of difficulty. * is fairly easy, ** is somewhat difficult, and *** is difficult or complex.*

Preprocessing step / Satellite	Sentinel-2	Landsat 8
Land-water mask	*	*
Cloud mask (low and middle-altitude clouds)	*	*
Cirrus mask (high-altitude clouds)	*	*
Cloud-shadow mask	**	**
Geometric enhancement of level 1B sensor model	***	***
Orthorectification to level 1CN	*	*
Atmospheric correction to level 2AN	**	**

References

Antoine, D., and Morel, A., 2011. MERIS ATBD 2.7. Atmospheric Correction of the MERIS observations Over Ocean Case 1 waters. Issue 5, revision 1.

Arino, O., Paganini, M., Koetz, B., Seifert, F. M., and Pinnock, S., 2012. Sentinel-2 Preparatory Symposium: session summaries, 14 May 2012, Issue 0, Rev. 2, 32 p.

Baillarin, S., Meygret, A., Dechoz, C., Petrucci, B., Lacherade, S., Tremas, T., Isola, C., Martimort, P., and Spoto, F., 2012. Sentinel-2 level 1 products and image processing performances, IGARSS, pp. 7003–7006.

Borde, R., Ramon, D., Schmechtig, C., and Santer, R., 2003. Extension of the DDV concept to retrieve aerosol properties over land from the Modular Optoelectronic Scanner Sensor, International Journal of Remote Sensing, Vol. 24(7), pp. 1439–1467.

Bouvet, M., Fisher, J. and Brockmann, C., 2008. Atmosphere and glint correction Algorithm, AGC ATBD, GKSS-KOF-AGC-ATBD01, 45 p.

Doerffer, R. and Schiller, H., 2008. MERIS Regional Coastal and Lake Case 2 Water Project Atmospheric Correction ATBD.

Drusch, M., Gascon, F., and Berger, M., 2010. GMES Sentinel-2 Mission Requirements Document, Issue 2, Revision 1-8/03/2010, EOP-SM/1163/MR-dr.

ESA, 2013. Collaborative Ground Segment Proposals status – Overview of ESA support activities, GOCG Workshop #8, 4 December 2013, ESA HQ, Paris.

Elterman, L., 1968. UV, visible and IR attenuation for altitudes to 50 km, Air Force Cambridge Research Laboratories, Environmental Research papers, No. 285, AFCRL-68-153, 49 p.

Falldorf, T., Strand, O., Panzacchi, M., and Tømmervik, H., 2014. Estimating lichen volume and reindeer winter pasture quality from Landsat imagery, Remote Sensing of Environment, Vol. 140, pp. 573–579.

Gjertsen, A. K., and Trollvik, J. A., 2013. Europa og USA kommer med nye jordobservasjonssatellitter: Hva gjør Norge for å forberede mottak og anvendelse? Kart og Plan, No. 2, pp. 113–127.

Gordon, H. R., and Morel, A., 1983. Remote assessment of ocean colour for interpretation of satellite visible imagery, a review. In Lecture notes on Coastal and Estuarine Studies, 4, Springer Verlag, New York.

Hagolle, O., Dedieu, G, Mougenot, B., Debaecker, V., Duchemin, B., and Meygret, A., 2008. Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: Application to Formosat-2 images, Remote Sensing of Environment, Vol. 112, pp. 1689–1701.

Hagolle, O., Huc, M., Villa Pascual, D., and Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENµS, LANDSAT and SENTINEL-2 images, Remote Sensing of Environment, 114, pp. 1747–1755.

Hollstein, A., Bismarck, von J., Fishcer, J., and Preusker, R., 2010. Description of the Vector Radiative Transfer Model MOMO, http://www.brockmann-consult.de/beam-wiki/download/attachments/17563679/momo_model_description.pdf?version=1&modificationDate=1291371266113.

IOCCG, 2000. Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters. Sathyendranath, S. (ed), Reports of International Ocean-Colour Coordinating Group, No. 3, IOCCG, Darthmouth, Cananda.

Kaufman, Y.J., and Tanré, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS MODIS. IEEE Transactions on Geoscience and Remote Sensing, 30, pp. 261–270.

Le Moigne, J., Netanyahu, N. S., and Eastman (eds.), R. D. 2011. Image Registration for Remote Sensing, Cambridge University Press, New York, 484 p.

Morel, A., and Gentili, B., 1991. Diffuse reflectance of oceanic waters: its dependence on sun angle as influenced by the molecular scattering contribution, Appl. Opt. 30, pp. 4427–4438.

Morel, A. and Gentili, B., 1996. Diffuse reflectance of oceanic waters. III. Implication of the bidirectional of the remote sensing problem, Appl. Opt. 35, pp. 4850-4862.

Morel, A., and Gentili, B., 1993. Diffuse reflectance of oceanic waters. II. Directional aspects, App. Opt. 32, pp. 6864–6879.

Pedersen, A., Borgersen, G., Folkestad, A., Johnsen, T. M., Norling, K., and Sørensen, K., 2012. Basisovervåkning av kystvann-Trøndelag, foreløpige resultater etter ett års undersøkelser i 2011. Rapport TA-2958/2012, 139 p.

Santer, R., and Ramon, D., 2011. MERIS aerosol remote sensing over land, ATBD 2.15, 31 p.

Santer, R., Carrere, V., Dubuisson, P., and Roger, J. C., 1999. Atmospheric correction over land for MERIS. International Journal of Remote Sensing, Vol. 20, pp. 1819–1840.

Santer, R., Zagolski, F., Ramon, D., Fischer, J., and Dubuisson, P., 2005. Uncertainties in Radiative Transfer Computations, consequences on the MERIS products over land, International Journal of Remote Sensing, Vol. 26(20), pp. 4597–4626.

Sathyendranath, S., 1986. Remote sensing of phytoplankton: A review, with special reference to picoplankton. In: Photosynthtetic Picoplankton (Canadian Bulleting of Fisheries and Aquatic Sciences, Volume 214), T. Platt and W. K. W. Li (eds.), Canadian Government Publishing Centre, Ottawa, pp. 561–583.

Strand G-H., and Bloch, V. V. H., 2009. Statistical grids for Norway: Documentation of national grids for analysis and visualization of spatial data in Norway, Statistics Norway, document 2009/9, 39 p.

Townshend, R. G., Justice, C. O., Gurney, C., and McManus, J., 1992. The Impact of Misregistration on Change Detection, IEEE Trans. on Geoscience and Rem. Sens., Vol. 20(5), pp. 1054–1060.

Trollvik, J. A., Gjertsen, A. K., Sørensen, K., and Larsen, R., 2012. Forberede mottak og anvendelse av optiske satellittdata for Norge digitalt, Report ANB11-05, Norwegian Mapping Authority, June 2012, 63 p.

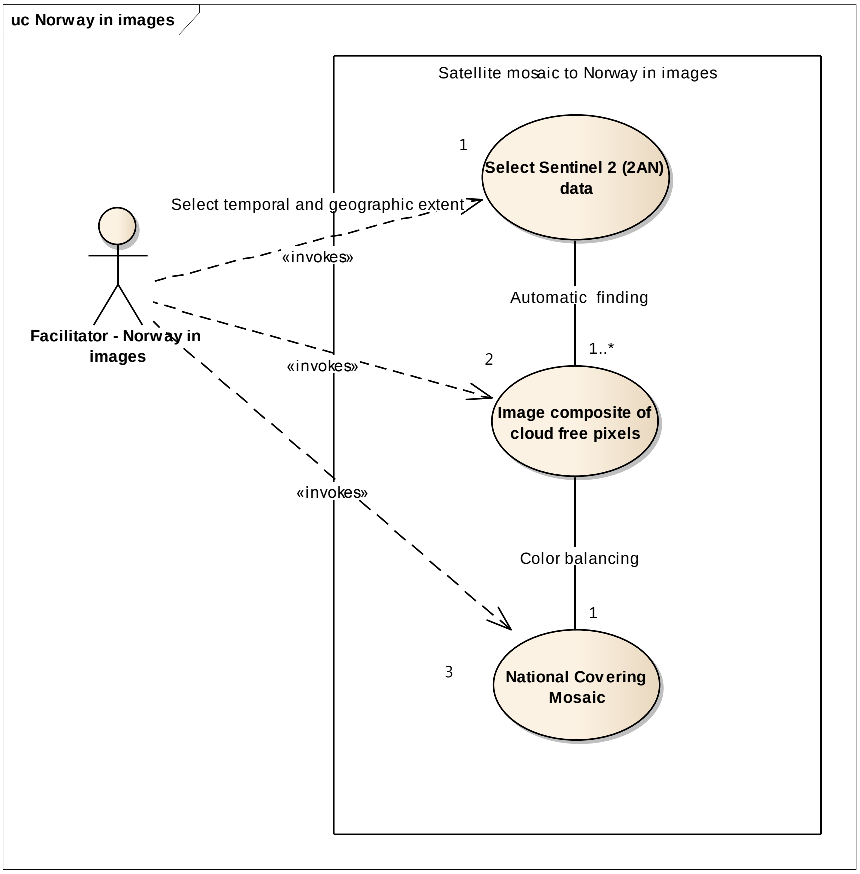
Appendix A – Use cases

Use case – Satellite mosaic for Norway in images

Ref CEN/TC287 TR15449 part 4 Service centric view annex B.

This template is an extended version of the original template defined by Cockburn [6], in particular extended with a possibility to describe Requested Information Resources found suitable in an SDI setting.

Table A 1 Description of use case

Use Case	Description
Use Case Name	Routine for periodic production of national mosaics from Sentinel-2 to Norway in Images
Use Case ID	1
Revision and Reference	V01
Use Case Diagram	
Status	In progress
Priority of accomplishment (optional)	Must have: The system must implement this goal/ assumption to be accepted.
Goal	A tool which can select from a stack of available Sentinel-2 data, find pixels free of clouds and cloud shadows, and put them together in a mosaic.

Use Case	Description
Summary	A web based tool that makes it possible to present Sentinel-2 level 2AN data available in the Norwegian Digital Satellite Archive in selected time intervals. Cloud mask is used to locate cloud free pixels and cloud shadows. Resulting pixels are put together in a composite (Level 3AN). Color balancing (adjustments) are performed. The final result is a more or less national covering mosaic to be put into Norway in Images. The use case should be performed automatically in periodic intervals (10-12 mosaic per year) The resulting mosaic can be used as a backdrop for other thematic information, change detection over years, change detection within a year, snow coverage, wood cutting and more.
Category	*Categorization of use cases according to overall reference architecture.*
Actor	General public via Norway in Images (Viewing and download services)
Primary Actor (initiates)	Administrator of Norway in Images
Stakeholder (optional)	The Norwegian Mapping Authority (Kartverket)
Requested Information Resources (optional)	• Sentinel-2 level AN or 3AN(Depending on what Levels archived in NSDC)
Preconditions	Sentinel-2AN with accompanying metadata must be accessible in NSDC.
Triggers (optional)	Ordering of national Satellite coverage from the Norwegian Mapping Authority (Can be an automated (batch) process)
Main success scenario	1. The user opens the Web portal (NSDC) which starts with a national index map. 2. To reach the download service it is needed to log on to the system via BAAT (TBD). 3. The user chooses area by setting up a geographic filter (eg. the whole country) 4. The user chooses source data (Sentinel-2AN) and time span (from date – to date) for the composition. 5. The system returns a preliminary view which indicates the coverage to be delivered 6. The user confirms the order, chooses Composition (RGB, and CIR), pixel size, file format, Tiling etc. 7. The result is a national mosaic from a composite of Sentinel-2 pixels accompanied by metadata at pixel level 8. Post processing to smoothen out sharp edges in the mosaic. (Or color adjustments earlier in the process) 9. Transfer of the dataset to Norway in Images archive
Extensions	5 a) If the coverage does not match the user requirement, go back to step 3

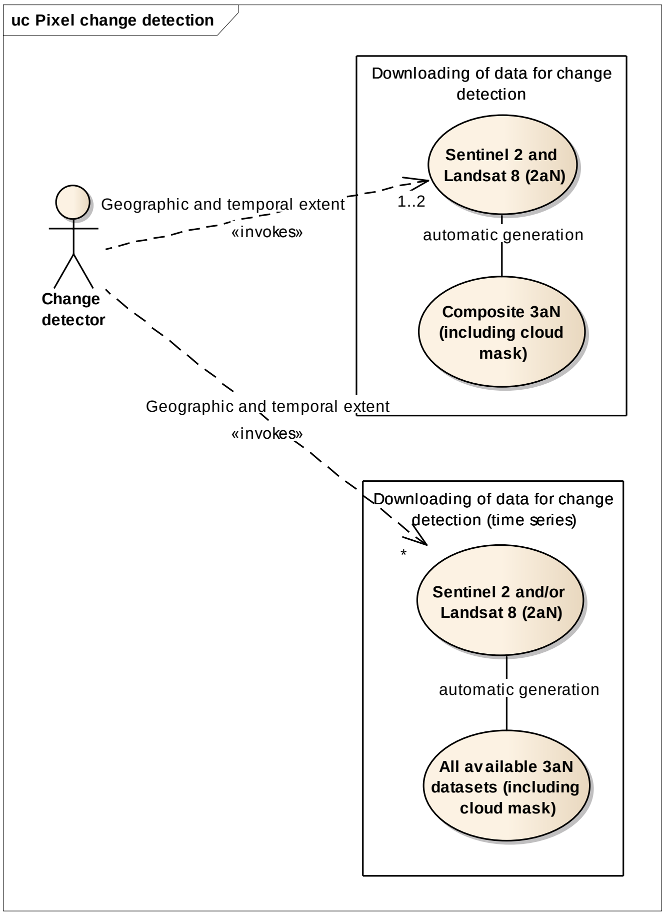
Use Case	Description
Alternative paths (optional)	
Post conditions	
Non-functional requirements	6a) The system must return the expected data volume and time for delivery. 7a) The user receives an e-mail with an URL to where the data can be downloaded.
Validation statement	*List of statements that indicate how to validate the successful realization of the use case.*
Notes	*Additional notes or comments (also by other users).*
Author and date	Jon Arne Trollvik 08.02.2013

Use case – Pixel change detection

Ref CEN/TC287 TR15449 part 4 Service centric view annex B.

This template is an extended version of the original template defined by Cockburn [6], in particular extended with a possibility to describe Requested Information Resources found suitable in an SDI setting.

Table A 2 Description of use case

<table>
<tr><th>Use Case</th><th>Description</th></tr>
<tr><td>Use Case Name</td><td>Pixel change detection</td></tr>
<tr><td>Use Case ID</td><td>pixel_change</td></tr>
<tr><td>Revision and Reference</td><td>0.1</td></tr>
<tr><td>Use Case Diagram</td><td>


In addition Draft only.
Visual reference:
use_case_pixel_change.png
Draft::
use_case_ pixel_change.graphml
</td></tr>
<tr><td>Status</td><td>Planned</td></tr>
</table>

Use Case	Description
Priority of accomplishment (optional)	
Goal	System provides common baseline semi-automated information about changes in areas based on between periodic changes in pixels/cluster of pixels for the stakeholders. Common analyses are done once and common tools for analyses are developed once for members of Norway Digital.
Summary	Users define areas of interest and maintain a database of current status for these areas on a pixel level. Changes in their status are indicated and the operator informed.
Category	*Change detection*
Actor	DN, SSB, S&L, Kartverket
Primary Actor (initiates)	Norway Digital
Stakeholder (optional)	Members of Norway Digital, public
Requested Information Resources (optional)	Pixel database (read) Aerial photo (read)
Preconditions	The users chose to initiate production of change pixel based change detection database
Triggers (optional)	
Main success scenario	1. Level 1C Sentinel-2 data (top of atmosphere images) is downloaded to National Collaborative Ground Segment (NCGS) every third day and Landsat 8 every 16 day from May 15 through September 15. 2. Quality enhanced by the use of better data for geographically and geometrically correcting and rectification by using a more precise national DEM and national ground control points (GCP's) => level 1CN product 3. Toolbox doing best available atmospheric correction and correcting for slope and neighboring effects, cloud and cloud shadow detection and correcting for geophysical variables as fAPAR (*Fraction of Absorbed Photosynthetically Active Radiation)* leading to a level 2AN product 4. Archive containing corrected 100x100 km tiles as described in 1, 2, and 3 showing cloud and cloud shadow free pixels. 5. Archive-design allows ready making of composites from user defined time frames – level 3 products 6. Archive (WMS) service for announcing and distributing level 3 products for common use amongst Norway Digital partners
Extensions	1a. Establishing a NCGS is required in order to establish use case 2a. New and more precise national DEM is instrumental for cost efficient utilization of Remote sensed data as is an adequate number and distribution of GCP's 3a. The NCGS must have access to the best possible toolbox featuring the most updated algorithms available at any given time 4a. Archive for satellite images and relevant byproducts must be established and sufficient resources must be allocated to planning, establishing and implementation of optimal and cost effective archive architecture by 31.12.2014 5a. The NCGS must host an effective tool (computer software) for producing composites, this must be purchased or developed by 31.12.2014 6a. There must be developed an extension of the archive functions facilitating exchange of relevant level 3AN products (composites) for reuse

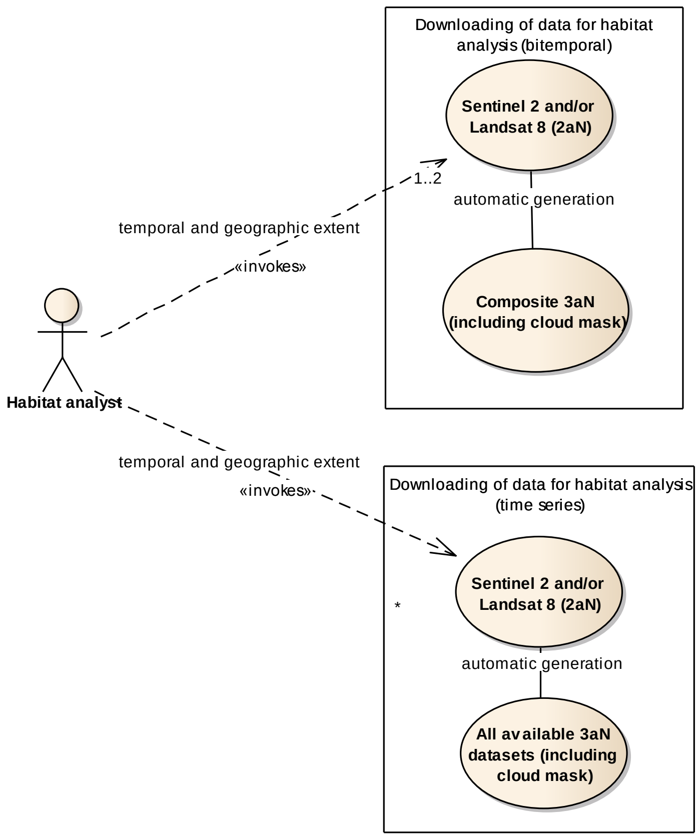
Use Case	Description
Alternative paths (optional)	
Post conditions	Norway has an archive making level 2AN 100x100 km tiles potentially from every third day (Sentinel-2) or every sixteenth day in the May 15. – September 16. time frame. The same archive also contains byproducts and composites generated by the users. The contents of the archive are available to stakeholders and the general public
Non-functional requirements	The archive should be accessible on-line for Norway Digital members. The main investments must be made through governmental funding. This is an consequence of ongoing investment in ESA's activities, including the Sentinel program (147 mill EUR), cost of upgrading and maintenance should be the only cost that will be divided between the Norway Digital members
Validation statement	An archive with level 2AN and level 3 Sentinel-2 and Landsat 8 products is established Contents of archive are available on line for users Members of Norway digital has supplemented or replaced in situ data sets with data from this archive Analyses of changes in land cover and land use are more cost efficient and frequently updated
Notes	
Author and date	Johan Danielsen & Ragnvald Larsen, Trondheim, April 12. 2013

Use case – Habitats

Ref CEN/TC287 TR15449 part 4 Service centric view annex B.

This template is an extended version of the original template defined by Cockburn [6], in particular extended with a possibility to describe Requested Information Resources found suitable in an SDI setting.

Table A 3 Description of use case

Use Case	Description
Use Case Name	DN habitat modelling
Use Case ID	dn_habitat_modelling
Revision and Reference	0.2
Use Case Diagram	 Includes also drafts: Visual reference: • use_case_habitats.png Draft:: • use_case_habitats.graphml
Status	Planned
Priority of accomplishment (optional)	
Goal	System generates models of species and habitats distribution

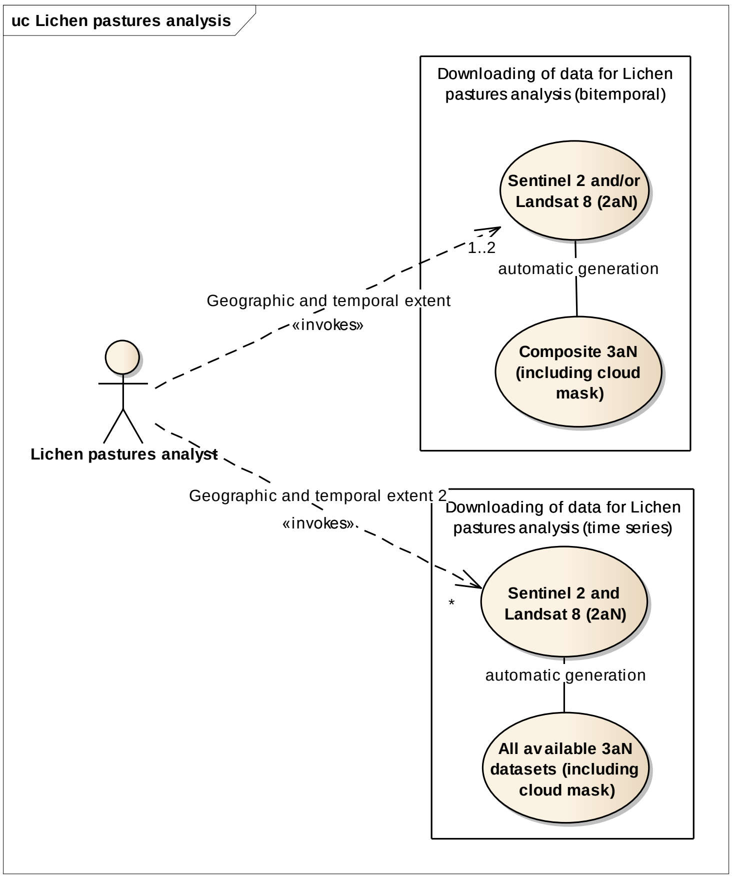
Use Case	Description
Summary	DN regularly publishes spatial representations of modelled habitats and species distribution. The models are based on current base knowledge. Used as a basis for area and species management. Will also be indicative for further research and mapping.
Category	*Context dependent*
Actor	DN, local authorities (political/professional), general public, land owners, proponents
Primary Actor (initiates)	System owner (DN)
Stakeholder (optional)	County governors, municipalities, universities, etc…
Requested Information Resources (optional)	Pixel database level 2AN, Composite summer Geo-referred well documented habitats
Preconditions	DN (The Directorate for environment) initiates development of habitat models as management tools
Triggers (optional)	
Main success scenario	1. See Change detection use case template 1. through 5. 2. Reviews of rapports on habitat modeling based on satellite images 3. New habitat models based on satellite images and knowledge extracted from reviews adapted to Norwegian needs
Extensions	7a. See Change detection use case template 1a. through 5a. 8a. Resources must be allocated to production of review 9a. Operative capable of developing relevant habitat models based on available in situ information combined with satellite images
Alternative paths (optional)	
Post conditions	Relevant habitat models are developed for a number of important species and habitats with an accuracy level at more than 80 %
Non-functional requirements	DN establishes WMS-services for public use of relevant habitat models for planning and further research
Validation statement	Habitat models based on satellite images and relevant in situ data are developed for relevant species and habitat Habitat models based on satellite images and relevant in situ data for relevant species and habitat is used when feasible in order to make in situ mapping of habitats and species more cost efficient Habitat models based on satellite images and relevant in situ data are developed for relevant species and habitat are used in planning when this is considered as the best available knowledge
Notes	
Author and date	Johan Danielsen & Ragnvald Larsen, Trondheim April 16. 2013

Use case – Lichen pastures analysis

Ref CEN/TC287 TR15449 part 4 Service centric view annex B.

This template is an extended version of the original template defined by Cockburn [6], in particular extended with a possibility to describe Requested Information Resources found suitable in an SDI setting.

Table A 4 Description of use case

Use Case	Description
Use Case Name	Lichen mapping
Use Case ID	dn_lichen
Revision and Reference	0.2
Use Case Diagram	 Includes also: Visual reference: • use_case_habitats.png Draft:: • use_case_habitats.graphml
Status	Planned
Priority of accomplishment (optional)	

Use Case	Description
Goal	Establish mapping, monitoring and measuring of extent and volume of lichen pastures for reindeer based on composites from Sentinel-2 and Landsat images. Initiate mapping and monitoring of lichen heath as part of NiN mapping
Summary	Annual semi-automated mapping and monitoring of lichen heath in wild reindeer habitat in Norway based on cloud-, cloud shadow free and snow free composites based on Sentinel-2 and Landsat 8 pictures in the timeframe from June 1. to September 15.
Category	*Mapping and monitoring*
Actor	DN, The Species Databank, regional managers, Land owners
Primary Actor (initiates)	System owner (DN)
Stakeholder (optional)	DN, regional managers (political/professional), general public, land owners
Requested Information Resources (optional)	• user observation (read) — Satellite data, — Optimised DEM (Better DEM will improve quality of product and increase cost-efficiency) — Annual composites compiled from data collected 01. June – 15. September — Aerial photos — Ground truthing / field observations made systematically by trained personell
Preconditions	DN initiates mapping and monitoring of lichen heath for two purposes: 1. Winter pasture for wild reindeer 2. Nature type vulnerable to overgrazing and climatic change
Triggers (optional)	
Main success scenario	7. Level 1C Sentinel-2 data (top of atmosphere images) is downloaded to National Collaborative Ground Segment (NCGS) every third day and Landsat 8 every 16 day from May 15. through September 15. 8. Quality enhanced by the use of better data for geographically and geometrically correcting and rectification by using a more precise national DEM and national ground control points (GCP's) => level 1CN product 9. Toolbox doing best available atmospheric correction and correcting for slope and neighboring effects, cloud and cloud shadow detection and correcting for geophysical variables as FAPAR (*Fraction of Absorbed Photosynthetically Active Radiation)* leading to a level 2AN product 10. Archive containing corrected 100x100 km tiles as described in 1., 2. and 3. showing cloud and cloud shadow free pixels in addition to pixels with zero snow-cover. 11. Archive based tool for making composites from user defined time frames – level 3 products 12. Methodology for production of maps of lichen cover and volume is developed, repeatability assured by the production of automatized production method which is published

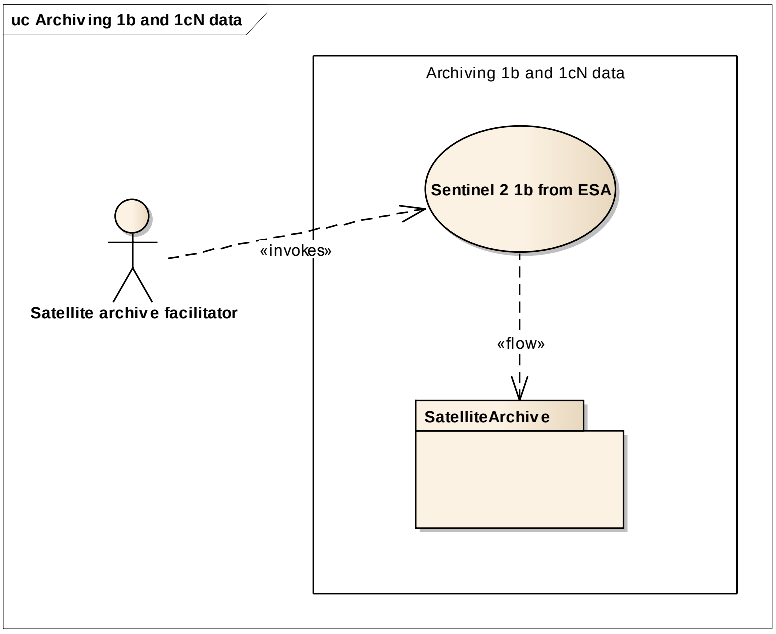
Use Case	Description
Extensions	10a. Establishing a NCGS is required in order to establish use case 11a. New and more precise national DEM is instrumental for cost efficient utilization of Remote sensed data as is an adequate number and distribution of GCP's 12a. The NCGS must have access to the best possible toolbox featuring the most updated algorithms available at any given time 13a. Archive for satellite images and relevant byproducts must be established and sufficient resources must be allocated to planning, establishing and implementation of optimal and cost effective archive architecture by 31.12.2014 14a. The NCGS must host an effective tool (computer software) for producing composites, this must be purchased or developed by 31.12.2014 15a. Operative doing the actual mapping based on level 3 products
Alternative paths (optional)	
Post conditions	Norway has an archive making level 2AN 100x100 km tiles potentially from every third day (Sentinel-2) or every sixteenth day (Landsat 8) in the June. – September 15. time frame.
Non-functional requirements	Users allocate resources to mapping of lichen heath and facilitates storage and availability of data online for public use
Validation statement	By 2016 maps of lichen heaths in all Norwegian wild reindeer areas are available online to the regional managers and the public through the Website of the Environmental Directorate
Notes	
Author and date	Johan Danielsen, Trondheim, April 15. 2013

Use case – Archiving 1B and 1CN data

Ref CEN/TC287 TR15449 part 4 Service centric view annex B.

This template is an extended version of the original template defined by Cockburn [6], in particular extended with a possibility to describe Requested Information Resources found suitable in an SDI setting.

Table A 5 Description of use case

Use Case	Description
Use Case Name	The need of a specular archive for 1B images from Sentinel-2 (S2).
Use Case ID	1
Revision and Reference	V01
Use Case Diagram	
Status	In progress
Priority of accomplishment (optional)	Must have: The system must implement this goal/assumption to be accepted.
Goal	1B from S2 will be downloaded from the ESA archive consecutively and archived in a national specular archive.
Summary	Level 1B will be archived consecutively in a national specular archive. These data will also be archived one month in the short term archive at ESA, and moved afterwards to a long term archive. A national satellite archive will give easier access to the user who needs 1B data. 1CN data is developed after ortho rectification with the use of a national DEM (Digital Elevation Model) with higher accuracy than the one used by ESA. Both 1B and 1CN data will be archived in the national satellite archive. In the case of further development of the national DEM, reprocessing of 1B is necessary to develop new 1CN data and 1B data therefore need to be easy accessed and therefore archived in the national satellite archive.
Category	
Actor	The user of 1B data. The user of 1CN data. The facilitator of the processing chain of the satellite data.

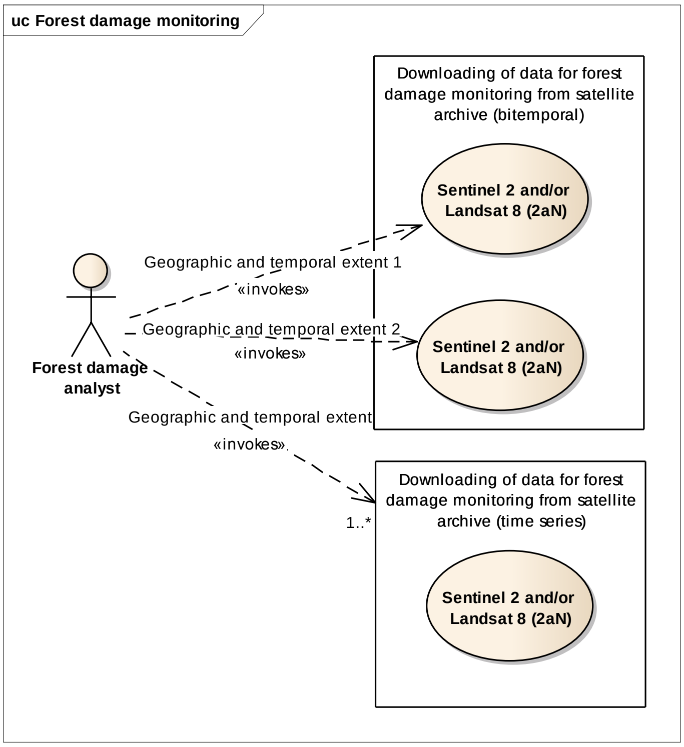
Use Case	Description
Primary Actor (initiates)	
Stakeholder (optional)	
Requested Information Resources (optional)	ESA archive, a national satellite archive, toolbox (BEAM).
Preconditions	
Triggers (optional)	A new DEM.
Main success scenario	1. Downloading 1B from ESA archive to the national satellite archive. 2. Develop 1CN data and archive in the national satellite archive.
Extensions	2 If a new DEM is developed, 1B need to be reprocessed.
Alternative paths (optional)	1. User downloading 1B from ESA archive. 2. User downloading 1C from ESA archive.
Post conditions	Retrieve the products through a web based portal or through a URL link received on e-mail.
Non-functional requirements	
Validation statement	List of statements that indicate how to validate the successful realization of the use case.
Notes	Additional notes or comments (also by other users).
Author and date	Anna Birgitta Ledang 3.5.2013

Use case – Forest damage monitoring

Ref CEN/TC287 TR15449 part 4 Service centric view annex B.

This template is an extended version of the original template defined by Cockburn [6], in particular extended with a possibility to describe Requested Information Resources found suitable in an SDI setting.

Table A 6 Description of use case

Use Case	Description
Use Case Name	Forest Damage Monitoring
Use Case ID	1
Revision and Reference	V01
Use Case Diagram	
Status	In progress
Priority of accomplishment (optional)	
Goal	A monitoring system that as early as possible gives information about change in state of the forests so that actions to mitigate damage can be planned and carried out.

Use Case	Description
Summary	Damage of productive forests costs forest owners large sums each year, and introduction of alien pest infestation and climate change will probably result in even more damage. Worldwide, large episodes of forest damage are either caused by extreme weather or introduced alien pests. In order to decrease the potential damage and economic loss, the state of the forest areas needs to be monitored so that necessary actions can be planned and implemented in due time. Today, Norwegian Forest and Landscape Institute is hosting a forest damage database. Damages are monitored as field observations or in a few cases from aircraft when extensive areas are affected. This monitoring service needs improvement. The detection of forest damage will be based on methods developed in our own research projects, available methods from the literature, and refinements of these. Forest damage events will be detected as a sudden change from bi-temporal image pairs. In a further development of an operational service we also intend to use detection as a deviation from normal phenological variation during the season, where the normal is identified from several years of time-series of data (Eklundh et al. 2009). The detection algorithms will be based on either a sudden decrease in a vegetation index, or a sudden increase in brightness due to less shadowing after a forest damage event, e.g. after defoliation and wind-throw. In an operational service, we will use all available acquisitions from Sentinel-2 to establish a time series of all pixels. Thus, a damage event covered by a pixel will be detected as early as possible as a sudden change in spectral values (change in indexes). The cloud mask delivered with the Sentinel-2 datasets will be used to remove pixels with cloud cover. A map showing detected damage events will be continuously updated from automatic processing of Sentinel-2 data (level 2AN from NSDC), enabling early detection.
Category	*Categorization of use cases according to overall reference architecture.*
Actor	Public Norway through the Ministry of Food and Agriculture
Primary Actor (initiates)	Norwegian Forest and Landscape Institute operates a program financed by the Ministry of Food and Agriculture
Stakeholder (optional)	Forestry sector, forest owners, forest owners associations, public foresters, insurance companies, general public
Requested Information Resources (optional)	
Preconditions	Sentinel-2 level 2AN with accompanying metadata is accessible from NSDC.
Triggers (optional)	

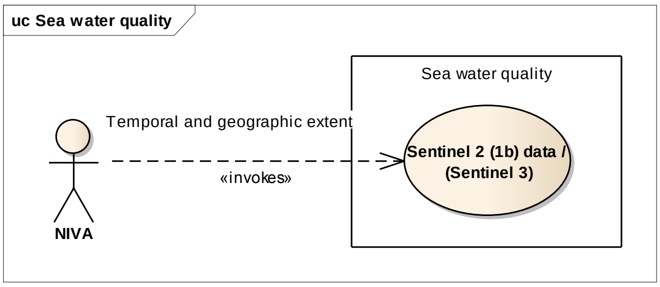
Use Case	Description
Main success scenario	1. A few hours after download to a station the Sentinel-2 data are processed to level 2AN and archived in NSDC. 2. The data are accompanied with a cloud and shadow mask. 3. Data are combined with earlier acquisitions and a pixel-based time series is updated with the newest data. 4. The updated time series for each pixel is analyzed. 5. Pixel variations that deviate from an expected normal variation are flagged as unexpected incidents that may indicate a change and/or damage event. 6. The detected events are interpreted either manually by experts at Norwegian Forest and Landscape Institute or automatically by use of a reference dataset that contains examples of several damage types. 7. The forest resource theme of Kilden, the web map portal at Norwegian Forest and Landscape Institute, is updated with the new information of the forest state. Stakeholders are informed if of new possible damage events that are a threat to forest health and economic value.
Extensions	Landsat 8 data are used in combination with Sentinel-2 data. This requires that also Landsat 8 data can routinely be processed to level 2A, i.e. surface reflectance values, and that a cloud/shadow mask can be automatically produced. An accurate cloud mask is essential for atmospheric corrections and time series analysis.
Alternative paths (optional)	
Post conditions	
Non-functional requirements	
Validation statement	*List of statements that indicate how to validate the successful realization of the use case.*
Notes	*Additional notes or comments (also by other users).* Eklundh, L., Johansson, T., Solberg, S. 2009. Mapping insect defoliation in Scots pine with MODIS time-series data, Remote Sensing of Environment, Vol. 113(7), pp.1566-1573.
Author and date	Arnt Kristian Gjertsen, 12 February 2013.

Use case – Sea water quality

Ref CEN/TC287 TR15449 part 4 Service centric view annex B.

This template is an extended version of the original template defined by Cockburn [6], in particular extended with a possibility to describe Requested Information Resources found suitable in an SDI setting.

Table A 7 Description of use case

Use Case	Description
Use Case Name	Routine for production of temporal composites for determining sea water quality through classification.
Use Case ID	1
Revision and Reference	V01
Use Case Diagram	 In addition, figure 1 and figure 2 in Appendix B present examples of composites for the current use case, based on MERIS data.
Status	In progress
Priority of accomplishment (optional)	Must have: The system must implement this goal/assumption to be accepted.
Goal	Calculate monthly, seasonally or periodically averages of chlorophyll –a and Secchi disc depth products based on available Sentinel-2 products. Use cloud free pixels to produce temporal composites.
Summary	A web based tool that makes it possible to download L1B products from the National satellite Archive for developing ocean colour products such as chlorophyll–a and Secchi disc depth. Before downloading, a timestamp and an area of interest (AOI) must be chosen. Marine satellite data need careful processing such as atmospheric correction since 90 % of the upward directed radiance comes from atmosphere, including sunglint and skylight, specularly reflected at the sea surface (Doerffer and Schiller, 2008). Additional, photons reflected towards the sensor, from land for instance, can give large errors in derived geophysical quantities and a processing taking this into account is also needed. Temporal composites will be created defining filters using information from bands and masks such as cloud_ice, water, sunglint and on specific reflectance values. Both classification of water quality and time series could be developed using these products.
Category	
Actor	Miljødirektoratet (Norwegian Environmental Agency), Vannregioner
Primary Actor (initiates)	NIVA
Stakeholder (optional)	The Water Directive, Vann-nett, Vannmiljø

Use Case	Description
Requested Information Resources (optional)	Sentinel-2 level 1B. Combination with Sentinel 3 could be explored. In situ data from Ferrybox or similar.
Preconditions	Sentinel-2 1B and corresponding Landsat products with accompanying metadata must be available in NSDC.
Triggers (optional)	The need of these products, which could be from once a month to once a year.
Main success scenario	3. Access the Web Portal, a national index map. 4. User need to have an account to access download of necessary products. 5. The source of the products needs to be chosen (Sentinel-2) and product level (1B). 6. Define a time stamp (start and end date). 7. Define a regional AOI (Area Of Interest). 8. The portal returns a preliminary view of datasets. 9. Mark requested datasets 10. Choose format of the files. 11. Confirm the requested order. 12. Download products
Extensions	8a. Shall be on N1 or dim formats where pixel values can be extracted through an external toolbox or a dataset with pixel values. Anyhow, both need the metadata for further work on the processing chain.
Alternative paths (optional)	3. Download L2AN from NSDC. 4. Download L1B through other sources or media than NSDC.
Post conditions	Retrieve the products through a web based portal or through a URL link received on e-mail.
Non-functional requirements	
Validation statement	List of statements that indicate how to validate the successful realization of the use case.
Notes	Additional notes or comments (also by other users).
Author and date	Anna Birgitta Ledang 2.5.2013

Appendix B – National grids for Norway

A fixed national grid defined by Statistics Norway, SSBgrid, is an open ended definition of a family of spatial tessellation models for Norway (Strand and Bloch, 2009). All SSBgrids are defined in UTM33/WGS84 (EUREF89) and are aligned such that the northwest corner coordinates (*Xc, Yc*) of a grid cell containing the position (*X, Y*) in a SSBgrid with cell size *K* are

$$(X_C, Y_C) = \left(T\left[\frac{X+f}{K}\right] \times K - f, T\left[\frac{Y}{K}\right] \times K\right),$$

where *T* is a truncation operation and *f* is a "false easting" set to 2,000,000 meters to avoid negative numbers. Each raster cell has a unique ID calculated with

$$ID = 2 \times 10^{13} + X_c \times 10^7 + Y_c.$$

The ID makes it possible to distribute SSBgrid data as tables instead of raster data. Several SSBgrid datasets can easily be joined together using the ID as join attribute. Figure 32 to Figure 35 illustrate the SSBgrid system.

For change detection and time series analyses it is important that the image datasets involved are accurately aligned so that the pixels involved represent the same area. By using a standardized grid structure, such as SSBgrid, for the orthorectified datasets, pixels from all datasets will be perfectly aligned and ready for analysis. Norway is covered by the UTM zones 32 to 36. Zone 33, centered on the 15°W meridian, is therefore the natural choice for a dataset that covers the whole country.

UTM is a conformal projection that preserves angles, approximates shape of small areas but distorts distance and area. The distortion of scale increases towards the ends of each zone, which is a six-degree band of longitude, but is held below 1 part in 1,000 in each zone (at Equator). The distortion will of course be greater for areas outside the zone borders, e.g. when a UTM projection is used outside its defined band of longitude. At high latitudes the scale errors will however be relatively small.

The scale factor of the Mercator projection increases with the secant of the latitude $\sec\varphi$, i.e. it is 1 at the equator, 2 at 60°N and infinite at the poles. In contrast for the UTM projection, that is a transverse Mercator projection, the scale factor increases with the distance from the central meridian. We have

$$d = \arcsin(\sin(\lambda - \mu) \times \cos\varphi$$

$$\text{scale factor } k = 0.9996 \times \sec d$$

where d is the distance (in radians) from a point P at longitude λ and latitude φ and a point Q on the central meridian at longitude μ and the same latitude as P. The number 0.9996 is due to the scale reduction factor of the UTM projection at the central meridian. From the equations, we can see that the scale factor at a point P is a function of the angular distance from the central meridian and the latitude.

The mainland of Norway is located between 4.95 and 31.06 degrees east and 57.97 and 71.13 degrees north. Using the above equations, we find that at Kirkenes, located at 30.05 degrees east and 69.725 degrees north, the scale factor $k = 1.003671$. Thus the true scale is less than 0.4 % larger than the nominal scale. The distortion of an area will be $k^2 = 1.007356$ and thus less than 1 % larger than the true area. A pixel with a nominal area of 100 m^2 in the SSBgrid system will represent an actual area of$100 \times 1/k^2 = 99.2698\ \mathrm{m}^2$ when located at Kirkenes. As can be seen from Figure 32, areas far east are also far north, thus the distance to the 15°E central meridian will therefore be moderate even for the most eastward areas.

We conclude that this small distortion of scale in the SSBgrid system is not a problem for the use with satellite data. What is important is that the pixels from different acquisitions are accurately aligned and therefore ready for change detection and time series analysis.

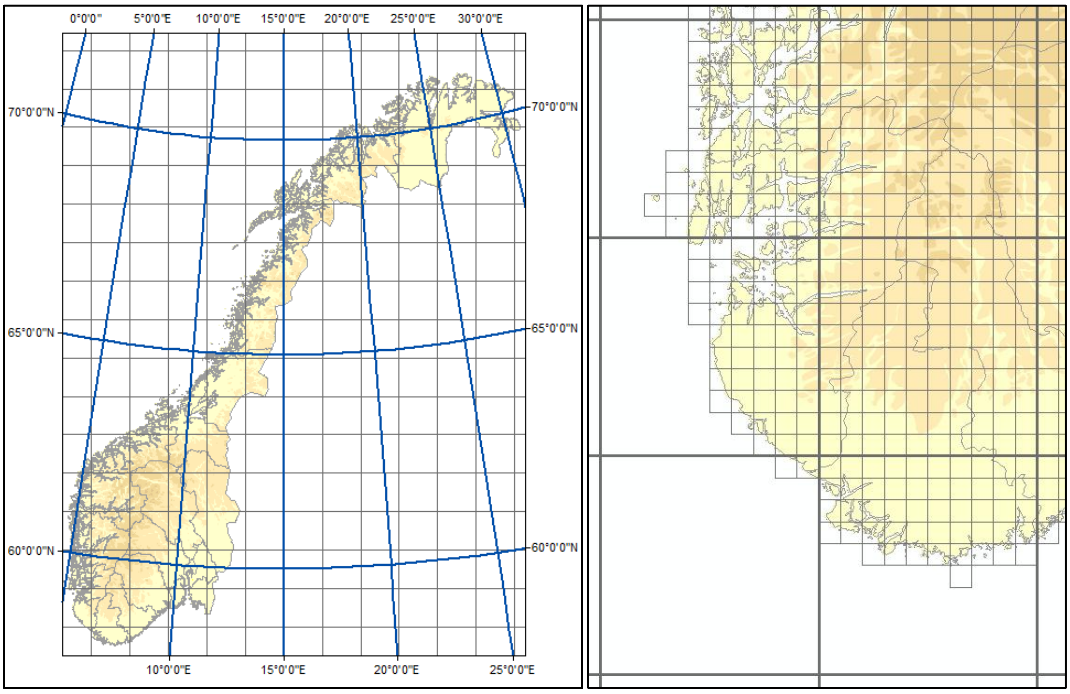


Figure 32 The official SSBgrid defined by Statistics Norway. SSBgrid is defined in UTM zone 33/WGS 84 and consists of a family of spatial tessellation models for Norway. The SSBgrids SSB100KM and SSB10KM are shown. In all, 73 100Km tiles cover a part of the landmass of mainland Norway.

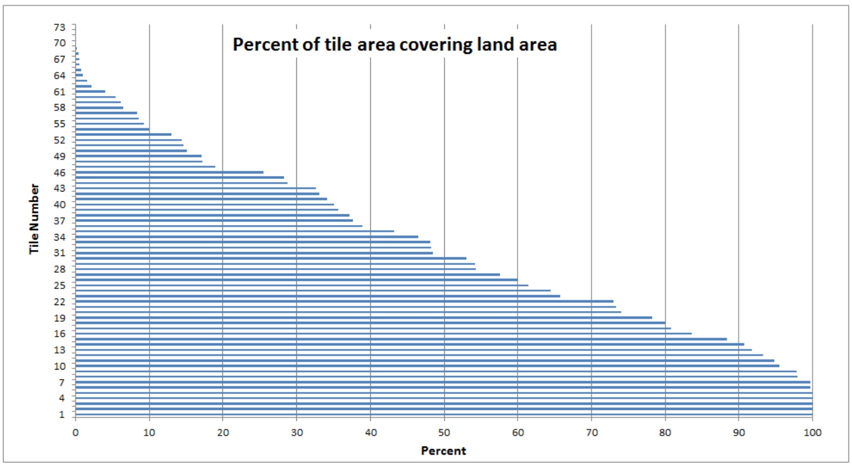


Figure 33 In all, 73 100KM tiles cover a part of the landmass of mainland Norway. Many tiles are only partly covering land.

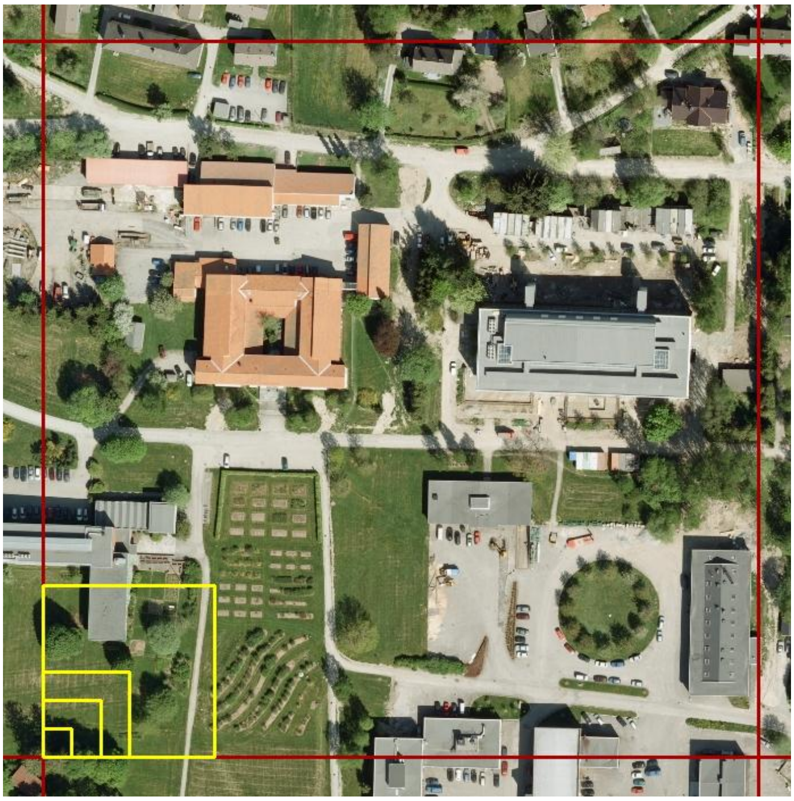


Figure 34 SSBgrid with 250 m large cells in red (SSB250M) and pixels in sizes 10 m, 20 m, 30 m, and 60 m in yellow. When pixels are aligned to a common fixed grid, combined use of data from different acquisitions and satellite missions is easy.

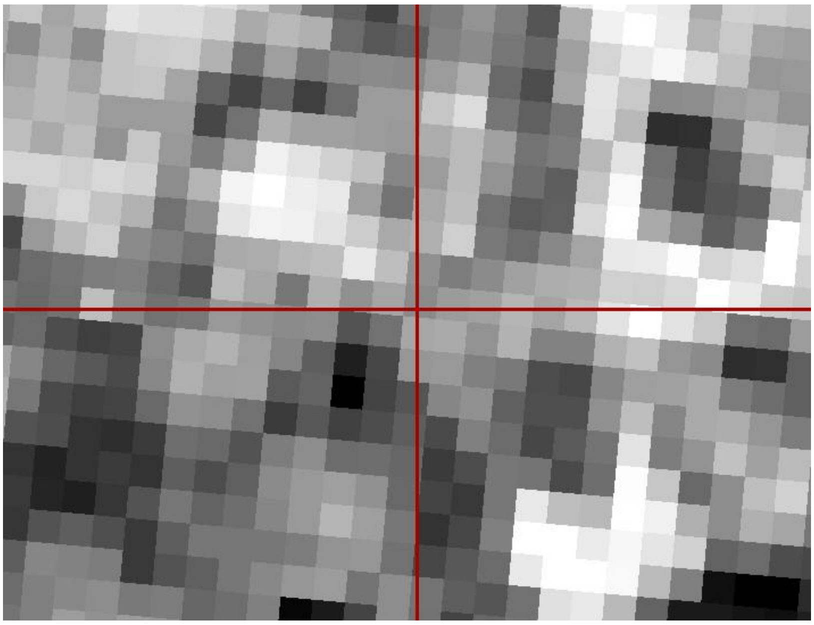


Figure 35 Orthorectified Landsat 8 dataset (product level L1T from USGS) and SSB5KM (red). The pixels are not aligned with the grid.

Appendix C – MERIS composites of six scenes in July 2011

From several single scenes, a temporal composite can be created with the arithmetical mean of the pixel values choosing the cloud free pixels. The same principle can be used with monthly mean products to a seasonal product. This is exemplified below with scenes from MERIS covering the Trondheim fjord in 2011 (Figure 36 and Figure 37).

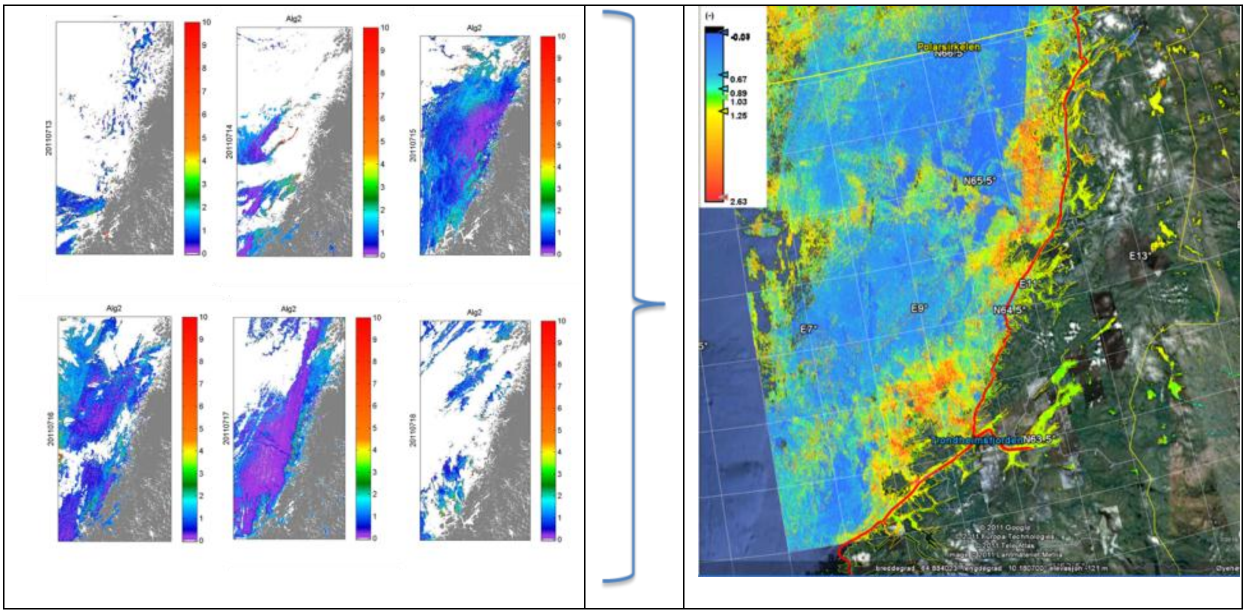


Figure 36 From single scenes to pixel based data to monthly means. Chlorophyll–a 13 of July to 18 of July 2011 from the Trondheim fjord (pers.comm. K. Sørensen).

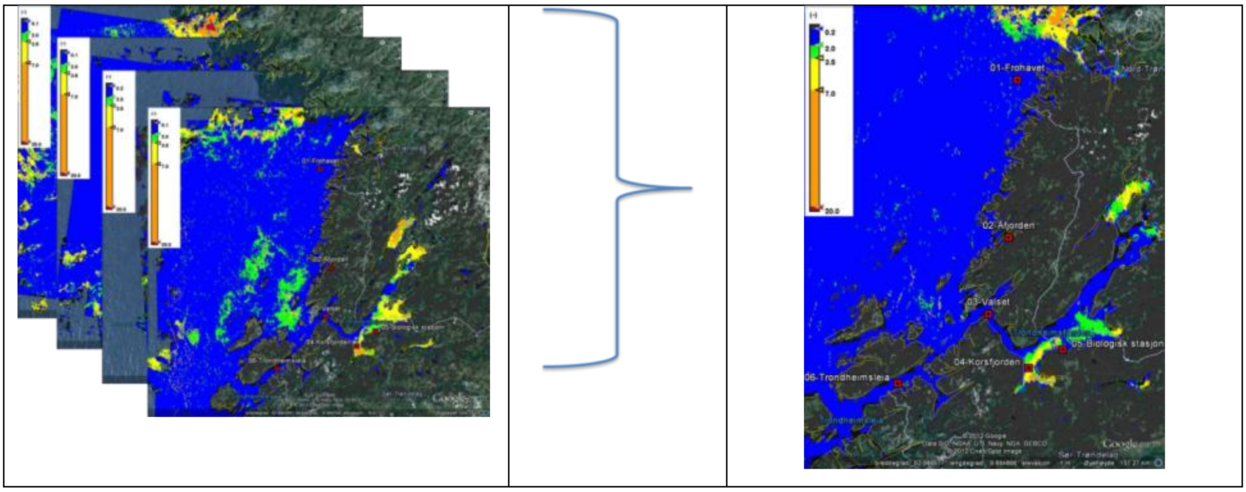


Figure 37 Chlorophyll–a for June to September 2011 from the Trondheim fjord: monthly means to seasonal means. Adapted from Pedersen et al., (2011).

Appendix D – Landsat 8 acquisitions

Landsat 8 was launched in February 2013 and started acquiring images over Norway in April 2013. Images from four selected scenes from the Landsat WRS-2 system have been selected and illustrated in Figure 38. In the following figures (Figure 39 to Figure 42) all acquisitions from April to September are shown. We can see a clear difference in the amount of cloud cover. ATCOR can do atmospheric correction if the cloud cover is less than 50 %.

The WRS-2 paths 196, 197, 198, and 199 are shown over Europe in Figure 43. The overlap between scenes is increasing from Equator, where it is at its minimum, and towards the poles. The overlap over Norway is large and all locations are covered by at least two passes during a 16 days cycle. Many locations are covered with three or even four passes (in the north of the country). In Figure 44 the set of 70 Landsat scenes covering mainland Norway is shown together with the subset of 47 scenes where each scene covers at least 25 % of land area. Table 8 gives an overview of the paths and day classes of the 70 scenes and Table 9 shows all the 47 scenes in the subset. The percentage of the scene area that covers land is shown. The days (dayclass) during a revisit cycle of 16 days are numbered from 1 to 16. Only day classes 4 and 11 are not represented in the set of 70 scenes, and if, for example, day class one is June 1, then there will be acquisitions somewhere over Norway all days from June 1 to June 16 except for June 4 and June 11.

USGS delivers an orthorectified image (level 1T), which covers an area of 183 x 170 km^2, in a file that is approximately 1.8 GB (including 11 spectral bands and one quality band). During orthorectification, the image is rotated in the resampling process to align it to a map projection and as a result large areas around the image are filled with a background value (usually the number zero). In a typical orthorectified product (level 1T) over Norway only 53 % of the pixels contain image data, the rest is filled with the background value (see Figure 39 to Figure 42). Thus, the volume of image data constitutes approximately 1 GB of a 1.8 GB file.

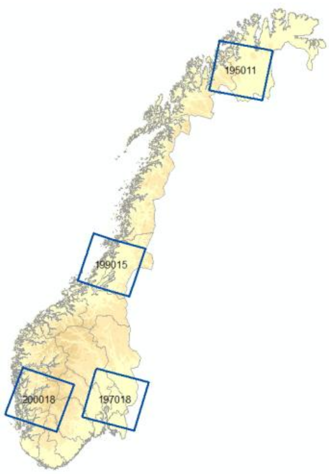


Figure 38 Four Landsat 8 WRS-2 scene outlines over Norway. All Landsat 8 acquisitions from April 2013, when the first recordings were made, to middle of September 2013 are illustrated in the four following figures. One scene is 183 km x 170 km.

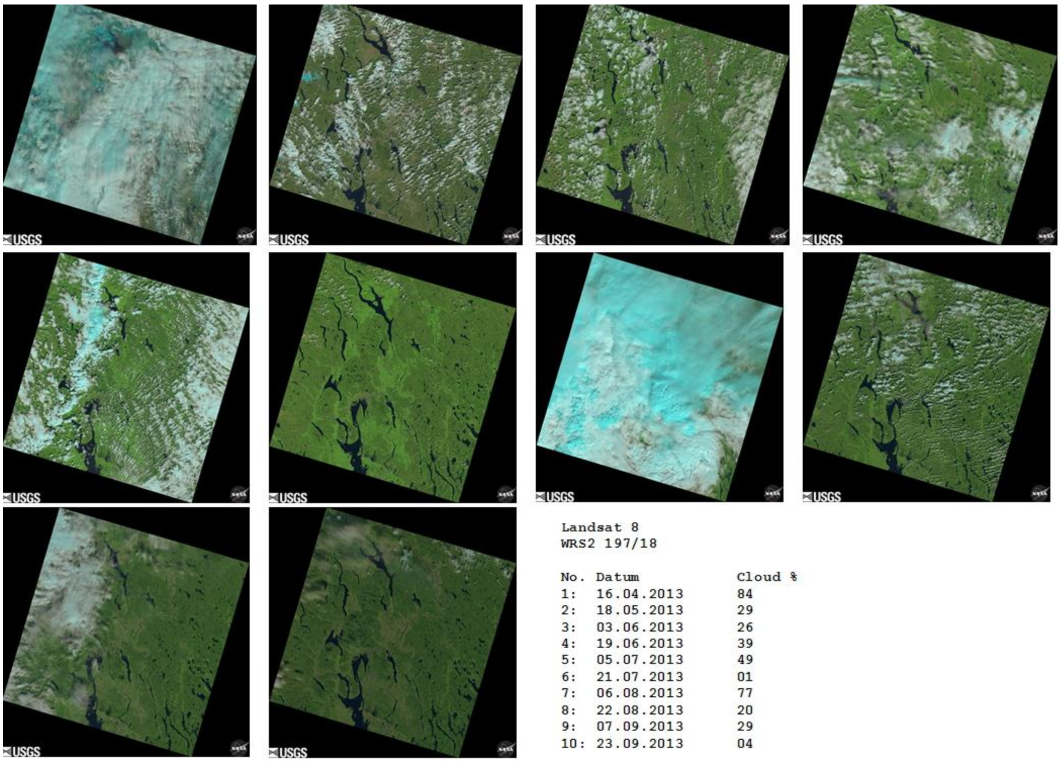


Figure 39 Quick looks of Landsat 8 acquisitions for the scene WRS-2 197/18 dated from 16.04.2013 to 23.09.2013. The cloud cover is less than 50 % in 8 of 10 acquisitions (80 %). The background value with no image data is black.

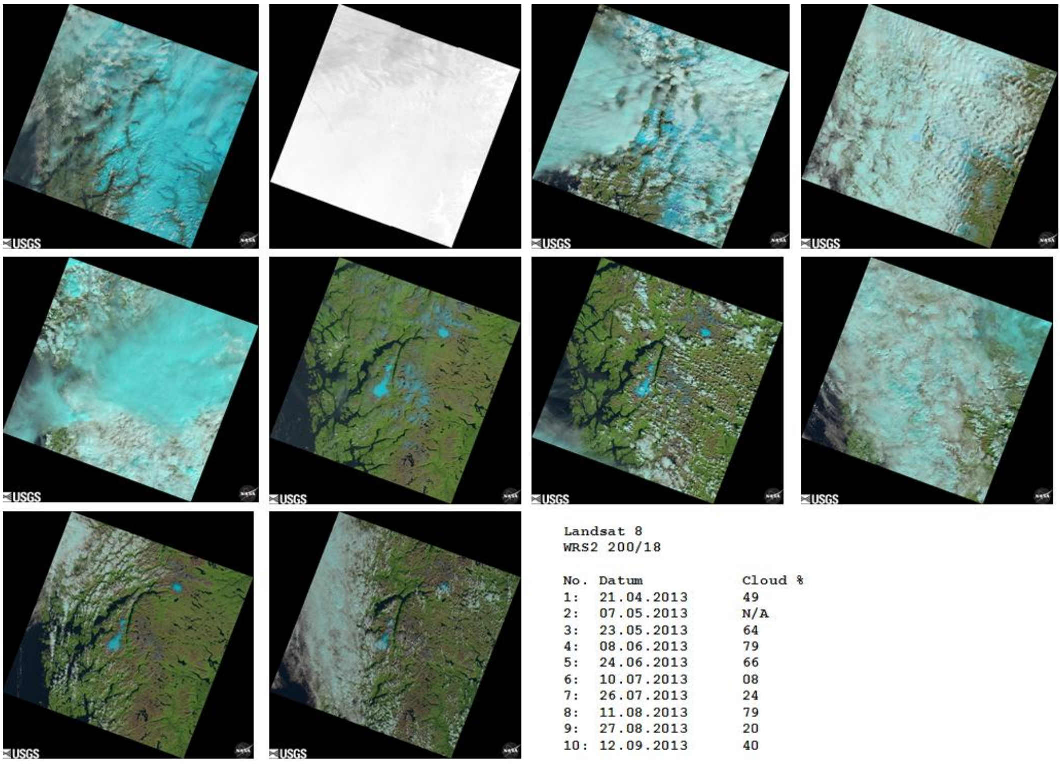


Figure 40 Quick looks of Landsat 8 acquisitions for the scene WRS-2 200/18 dated from 21.04.2013 to 12.09.2013. The cloud cover is less than 50 % in 5 of 10 acquisitions (50 %). The background value with no image data is black.

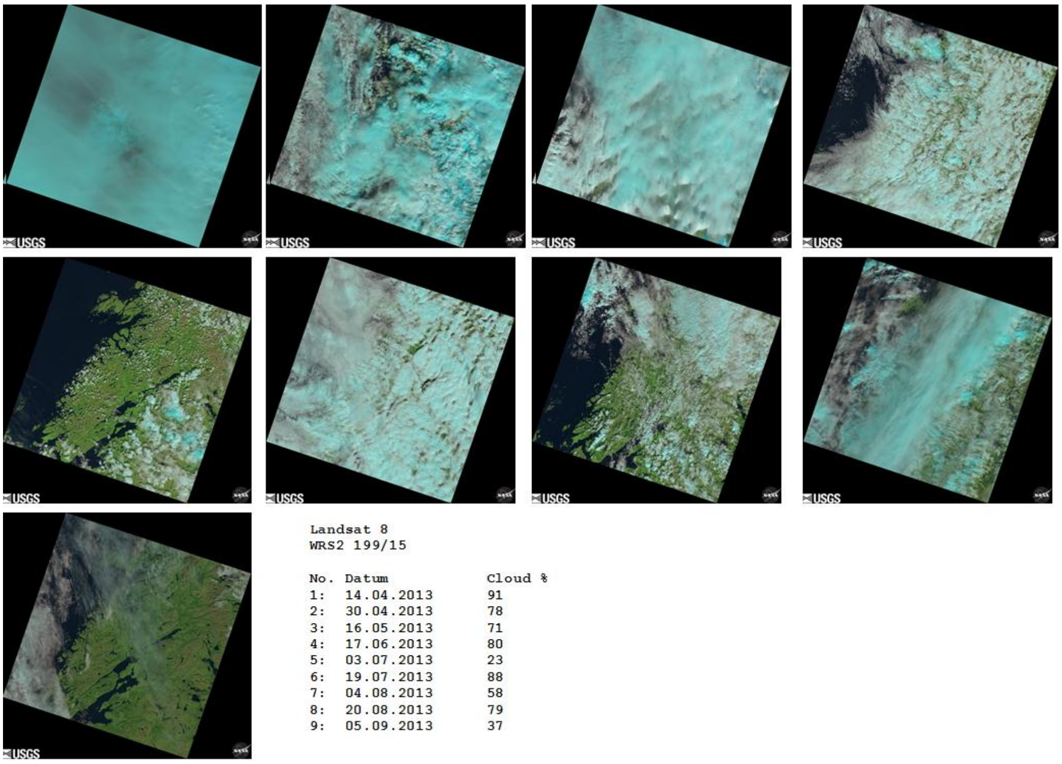


Figure 41 Quick looks of Landsat 8 acquisitions for the scene WRS-2 199/15 dated from 14.04.2013 to 05.09.2013. The cloud cover is less than 50 % in 2 of 9 acquisitions (22 %). The background value with no image data is black.

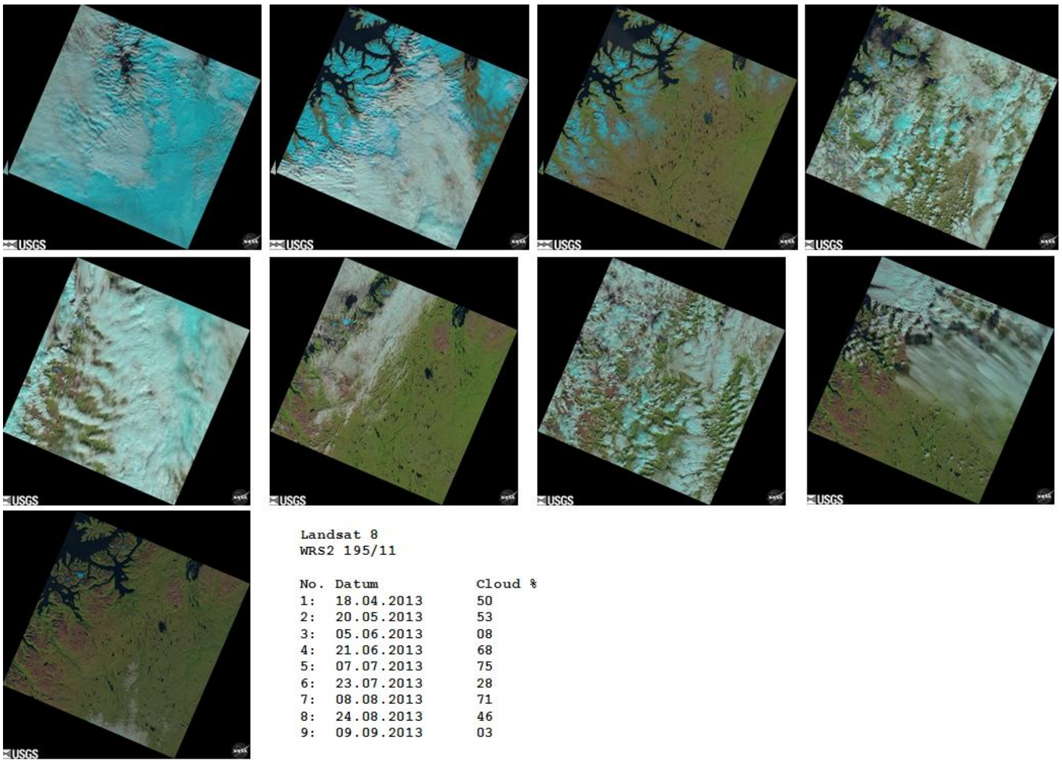


Figure 42 Quick looks of Landsat 8 acquisitions for the scene WRS-2 195/11 dated from 18.04.2013 to 09.09.2013. The cloud cover is less than 50 % in 4 of 9 acquisitions (44 %). The background value with no image data is black.

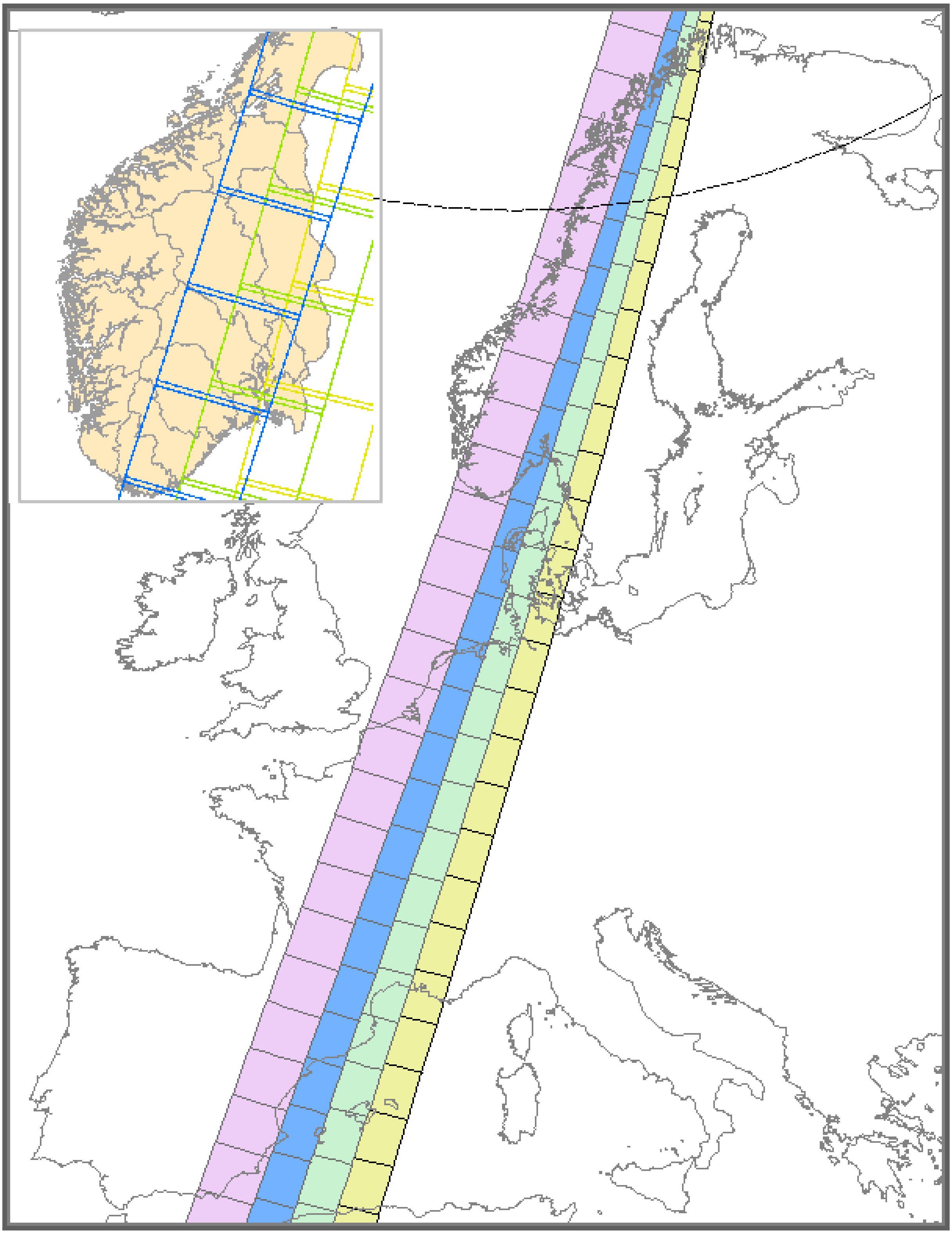


Figure 43 The Landsat WRS-2 scene reference system illustrated over Europe. We clearly see that the overlap between paths 196, 197,198, and 199 increases from south to north: the overlap is at its minimum at equator. The inset of South Norway shows paths 196, 197, and 198, and we can see how much they overlap. Many locations in Norway are covered with three or even four passes (North Norway).

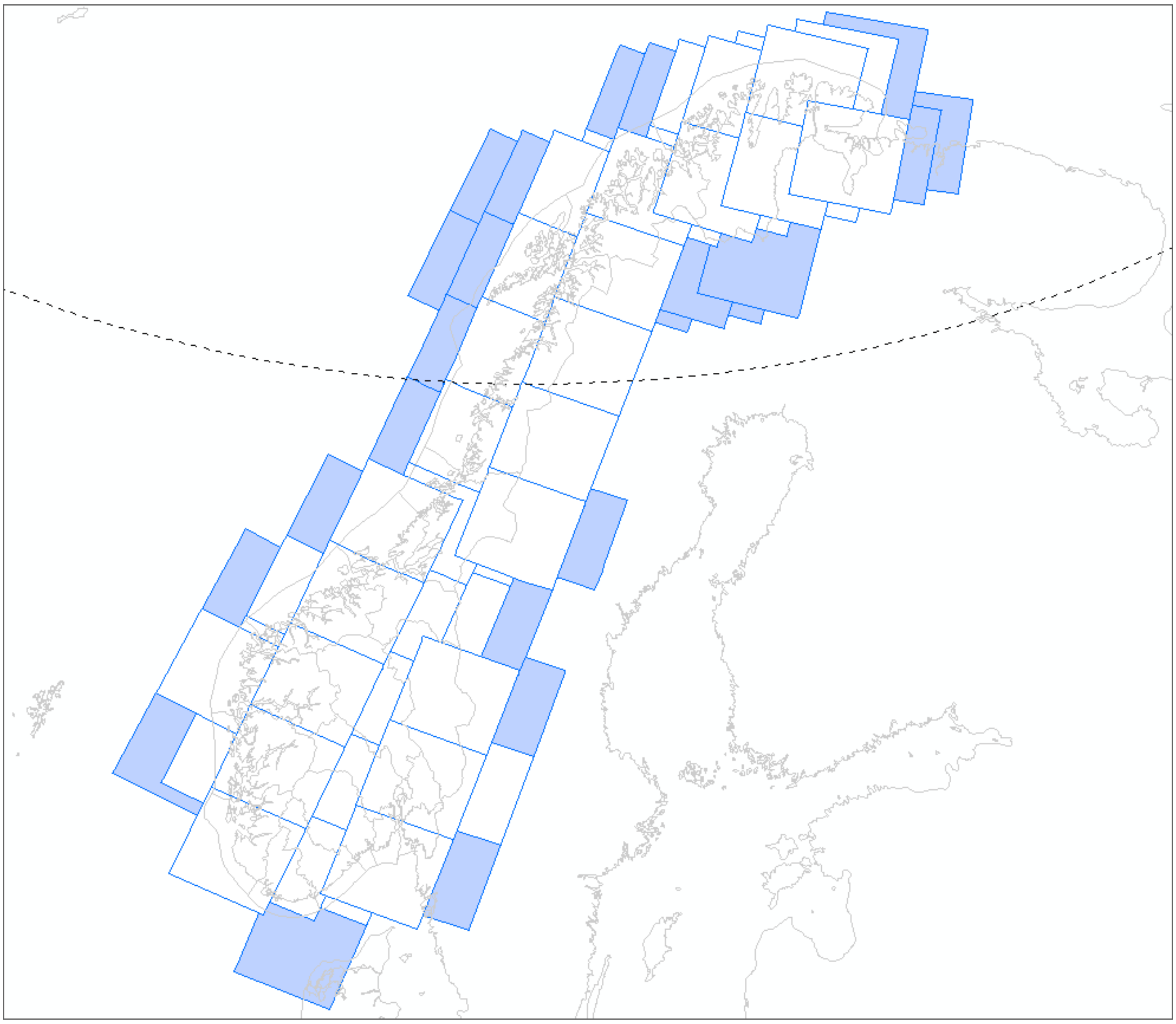


Figure 44 The set of 70 Landsat scenes that together covers Norway's mainland and coastal waters. The subset of 47 scenes with 25 % or more of the scene area inside the mainland is shown in white and the other in blue. Paths from 189 in the east to 201 in the west are represented. Landsat 8 records data over Norway on 14 days of the 16 days in the revisit period (12 of 16 if only the scenes in the subset are counted).

Table 8 Overview of the 70 Landsat WRS-2 paths that together cover Norway's mainland. Number of scenes of each path is shown together with the day class, i.e. the day number in the 16 days revisit cycle.

PATH	189	190	191	192	193	194	195	196	197	198	199	200	201	202
COUNT	1	1	2	2	3	3	3	7	10	11	9	9	6	3
DAYCLASS	2	9	16	7	14	5	12	3	10	1	8	15	6	13

Table 9 Overview of the group of 47 Landsat WRS-2 scenes where each scene includes at least 25 % land of mainland Norway.

PATH	ROW	DAYCLASS	AREA %		PATH	ROW	DAYCLASS	AREA %
191	11	16	0,29		198	15	1	0,60
192	10	7	0,27		198	16	1	0,63
192	11	7	0,41		198	17	1	0,97
193	10	14	0,38		198	18	1	0,98
193	11	14	0,54		198	19	1	0,64
194	10	5	0,40		199	11	8	0,29
194	11	5	0,75		199	12	8	0,43
195	10	12	0,37		199	13	8	0,45
195	11	12	0,87		199	14	8	0,51
196	10	3	0,26		199	15	8	0,61
196	11	3	0,72		199	16	8	0,92
196	18	3	0,35		199	17	8	1,00
197	11	10	0,61		199	18	8	1,00
197	12	10	0,41		199	19	8	0,81
197	13	10	0,27		200	15	15	0,31
197	14	10	0,26		200	16	15	0,76
197	15	10	0,29		200	17	15	0,97
197	17	10	0,59		200	18	15	0,86
197	18	10	0,80		200	19	15	0,38
197	19	10	0,31		201	16	6	0,44
198	11	1	0,51		201	17	6	0,71
198	12	1	0,48		201	18	6	0,37
198	13	1	0,44		202	17	13	0,29
198	14	1	0,50					

Appendix E – Information relating to the report on technical data management solutions

E1. Relevant open-source technology options for evaluation during prototyping

For storage of metadata with temporal and spatial information: *PostgreSQL, PostGIS*

For image previews: *GDAL, PostGIS Raster, Imagemagick*

For archived & cached data storage:

- File system based storage (GeoTIFF, multiple JPEG2000 GDAL drivers, other GDAL drivers).
- Fully in-DB storage (PostGIS Raster internal format (TOAST), rasdaman internal formats).
- Out-DB raster DB with filesystem storage (PostGIS Raster with GeoTIFF).
- Varnish cache.

For experimental analytics: PostGIS Raster, rasdaman / datacube approach.

Some of these technology choices may impact the overall development of the system.

JPEG2000 (and wavelet compression, and similar): The JPEG2000 system is likely to enable up to a 50% saving in terms of space consumption compared to raw data, but it is not recommended by the PostGIS Raster development team for out-DB lossless storage of raster data. Also, costly extra data-processing is required to compress and decompress the data whenever it is accessed.

Datacube/analytics: In the case of 'datacube' style database raster analytics or array stores, there is often a trade-off whereby blocks of data can be accessed more quickly than traditional databases, but extra space or complex preprocessing is required in return.

There is also the question of technology maturity. There is no question that file systems exist which are capable of storing thousands of petabytes of files safely and reliably with good data access times. The PostgreSQL database system has also proven capable of storing up to 2000 terabytes of data [5], though this required approximately 1000 PC servers to achieve. In contrast, open source technologies such as PostGIS Raster and rasdaman have never been tested at that scale. Rasdaman is currently developed with the aim to have it operating at the 100TB scale [6], but it is hard to find evidence of it being used in long-running industrial systems at the many-terabyte scale, or independent benchmarks.

E2. Estimates of space requirements for a tile archive for Landsat and Sentinel-2 data

When considering the possibility of a full data archive here, it is quite possible to estimate the hardware storage requirements reasonably well. The first assumption made here is that data might not be compressed (or not effectively compressed), as some storage/analysis systems will not support allow this (e.g. in-DB rasters or datacubes).

USGS Landsat tile data storage requirements for all of Norway:

(1) Area of Norway ~= 400,000 km^2
(2) File size of 1 raw Landsat tile ~= 2 GB
(3) Area coverage of 1 raw Landsat tile ~= 170 km x 183 km = 31,110 km^2
(4) Horizontal overlap between neighbouring orbital passes at 64°N = 1/cos(64) = 2.28x
(5) Equatorial revisit period = 8 days
(6) Equatorial revisits per year = 365.26 / 8 = 45.7x
(7) Data requirements per year = 2.7 TB/year

ESA Sentinel-2 tile data storage requirement estimates for all of Norway:

(8) Area of Norway ~= 400,000km^2
(9) File size of 1 raw Sentinel-2 tile >= 7 GB.
(10) Area coverage of 1 raw Sentinel-2 tile ~= 290 km x 276 km = 80,040 km^2
(11) Horizontal overlap between neighbouring orbital passes at 64°N = 1/cos(64) = 2.28x
(12) Equatorial revisit period = 5 days
(13) Equatorial revisits per year = 365.26 / 5 = 73.1x
(14) Data requirements per year = 5.8 TB/year

System storage requirements for this quantity of data:

(15) 10 year project lifespan = 10 x yearly amount
(16) RAID10, local backup, offsite backup = 4 x basic storage amount
(17) 3% active disk failure rate per year over (10 years less 3 years warranty) = 1/0.79 = 1.27 x

Total physical storage requirements for 10 year tile archive*: 432 TB.

It might be better to consider a 'tile archive' approach later in the prototyping process after other options have been explored. The result of the calculation might be modified further as follows:

- An 8-day revisit period is assumed for Landsat data, which implies use of the present Landsat 7 satellite data. Landsat 7 has some problems, and it may be replaced in the next 10 years by a new Landsat 9 satellite. If the archive will not store Landsat 7 data, the Landsat data estimates should be halved for the period until a potential Landsat 9 satellite might be available.
- In-database storage usually requires extra space for deleted rows, vacuuming, unused rows, performance features (indexes, de-normalisation), file system management, and so on. This increases space requirements by perhaps 20 %.
- The assumptions are for 10 years of future data only. If historical data is also to be included, the estimates should be revised substantially upwards to account for this.
- If compressed tiles may be used instead of raw data, a 10–50 % reduction space is possible depending on the type of compression available (e.g. gzip, JPEG2000, TOAST) [2].
- The project could be specified for only 5 years, not 10, reducing space requirements by 50 %.

- It might be acceptable to remove tiles with high cloud cover (estimated ~66 %) or low light (estimated ~10 %), however, this carries a related research and programming cost and may not be suitable for all users of the system.
- These assumptions are for only one level of data. If multiple versions of the data must be stored (e.g. level 1B, 1C …) then storage estimates should be multiplied by an appropriate integer.

E3. Notes to Appendix E2

(1, 8) 400,000 km^2 = 323,787 km^2 Norwegian mainland + ~20 % buffer accounting for tiling around the shape of Norway. This is a compromise between the official area of 323,787 km^2 and the recommended SSBgrid coverage extents given in [7] of 1,800,000 km^2.

(2) Via USGS factsheet [2]. We cannot assume that tiles can be stored in a more heavily compressed format if they might be stored and accessed directly in-database. Once the implementation system is known, the USGS figure of 1 GB may perhaps be substituted.

(3) Via [1, 3].

(4) Geographical center of Norway is at 63°59′26″N, 12°18′28″E. The circumference at latitude L is cos(L) of the circumference of the equator (approximately, i.e. using a spherical model).

(5) Via [2]. This is affected by the choice to include Landsat 7 data (or potential Landsat 9 data).

(7, 14) Size requirement per year = (area / tile area) * tile file size * revisits per year * orbit overlap.

(9, 10) See p. 126 in reference [8]. Level 2A products will generally have a slightly different raw file size per unit area when uncompressed, and a different shape (100km x 100km) [10]. Quoted figures for level 2 data may include the effects of compression and/or binning.

(9) File size of 1 raw Sentinel-2 tile >= 7 GB.

(10) Area coverage of 1 raw Sentinel-2 tile ~= 290 km x 276 km = 80,040 km^2

(12) Equatorial revisit period = 5 days.

(17) Via 10000+ disk study at http://blog.backblaze.com/2014/01/21/what-hard-drive-should-i-buy/.

E4. Estimates of space requirements for a map archive based on integrated tiles

An alternative approach might be to write software which integrates tiles together to remove overlap and multiply-represented data. Bands would be represented in a compact manner with only 1 sample of sensed data being stored per 8 or 5 day cycle respectively.

The data requirements for this approach can be calculated using the stated band resolutions for Landsat products and Sentinel-2.

Assuming 16 bit raster pixels, which is the only size commonly available in software that is big enough for the band data, 40 x 10,000 km^2 tiles would require 12.44 GB per Landsat equatorial revisit period and 44.68 GB per Sentinel-2 equatorial revisit period [9].

This yields 568 GB per year for Landsat and 3264 GB per year for Sentinel-2.

Again, these numbers would be multiplied by 10 years (38.3 TB) and by (x4, x1.26) to represent a real world project length and real world physical storage medium requirements including RAID, disk failures, and backups.

The total would thus be 193 TB for a 10 year project, further subject to the provisos above about historical data, database space requirements, compression etc. An additional project cost is introduced by this approach in developing software capable of integrating the tiles in this manner.

E5. References

[1] http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7_Handbook.pdf

[2] http://landsat.usgs.gov/landsat8.php

[3] http://landsat.usgs.gov/band_designations_landsat_satellites.php

[4] Landsat and Beyond: Sustaining and Enhancing the Nation's Land Imaging Program. By Committee on Implementation of a Sustained Land Imaging Program, Space Studies Board, Division on Engineering and Physical Sciences, National Research Council

[5] http://www.computerworld.com/s/article/9087918/Size_matters_Yahoo_claims_2_petabyte_database_is_world_s_biggest_busiest?taxonomyId=18&intsrc=hm_topic

[6] Rasdaman example. http://www.earthserver.eu

[7] http://www.ssb.no/a/english/publikasjoner/pdf/doc_200909_en/doc_200909_en.pdf

[8] Sentinel-2 Level 1 Products and Image Processing Performances, Towards Horizon 2020. Baillarin et al., International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B1, 2012

[9] Figures calculated by Arnt Kristian Gjertsen and checked by Graeme Bell. They assume bands 10 and 11 of Landsat are resampled at 30m per product specification for compatibility with measurements in other bands

[10] https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a

The Norwegian Mapping Authority
NO-3507 HØNEFOSS
Tel: +47-32 11 83 00
Email: post@kartverket.no
www.kartverket.no

Norwegian Forest and Landscape institute
P.O.Box 115
NO-1431 ÅS
Tel: +47-64 94 80 00
Email: post@skogoglandskap.no
www.skogoglandskap.no

Norwegian Environment Agency
P.O.Box 5672
NO-7485 TRONDHEIM
Tel: +47-73 58 05 00
Email: post@miljodir.no
www.miljodirektoratet.no

Norwegian Institute for Water Research
NO-0349 OSLO
Tel: +47-22185100
Email: niva@niva.no
www.niva.no