GEOSPATIAL WORLD FORUM
18 - 21 January 2011, Hyderabad, India
Theme
Dimensions and Directions of Geospatial Industry

The Norwegian Approach to Land Cover Mapping Based on a Common European Nomenclature: Using Existing Data and Satellite Imagery

Anne Nilsen and Hanne Gro Wallin

INTRODUCTION

The European Environment Agency (EEA) is an agency of the European Union. The main task of EEA is to provide independent information on the environment.

The CORINE (Coordination of information on the environment) programme of EEA, dating back to 1985, is a European commission which compiles information on the state of the environment (Land cover, Costal erosion, biotopes etc.). CORINE Land Cover (CLC) is a compilation of national land cover data sets, which are integrated into a seamless land cover map of Europe. The data base is based on a standard methodology and nomenclature. On the basis of satellite data Europe was mapped with the help of computer assisted photo interpretation and a common nomenclature.

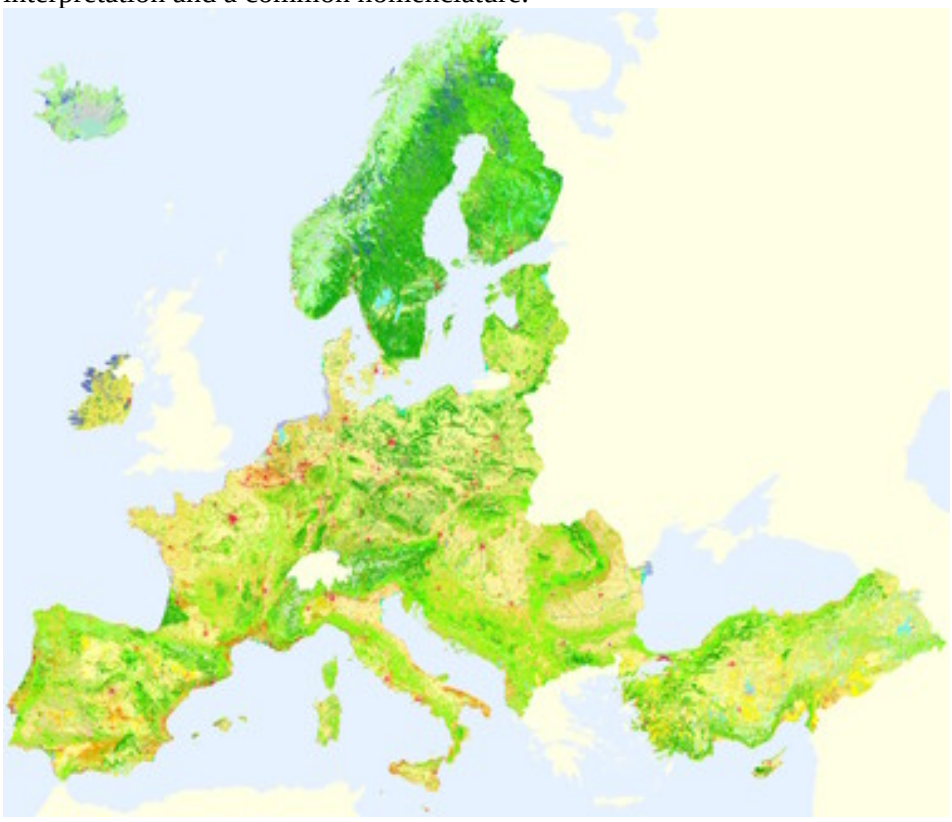


Figure 1. CLC2006 – a seamless land cover map of Europe. 36 of 39 countries have completed CLC2006 by February 2010. (www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster; Corine Land Cover 2006 raster data - version 13, 250 m)

CLC is not a detailed land cover data set. The smallest mapping unit is 25 ha. CLC has a standardized nomenclature on three different levels with 5, 15 or 44 classes. CLC is a small scale map showing build up areas, agriculture, forest and semi-natural areas, wetlands and water bodies. CLC2000 and CLC2006 are meant to represent the land cover situation close to year 2000 and 2006 respectively.

1. CLC2000

1.1. Data sources

The Norwegian approach in the making of CLC2000 utilized existing national land cover datasets wherever available.

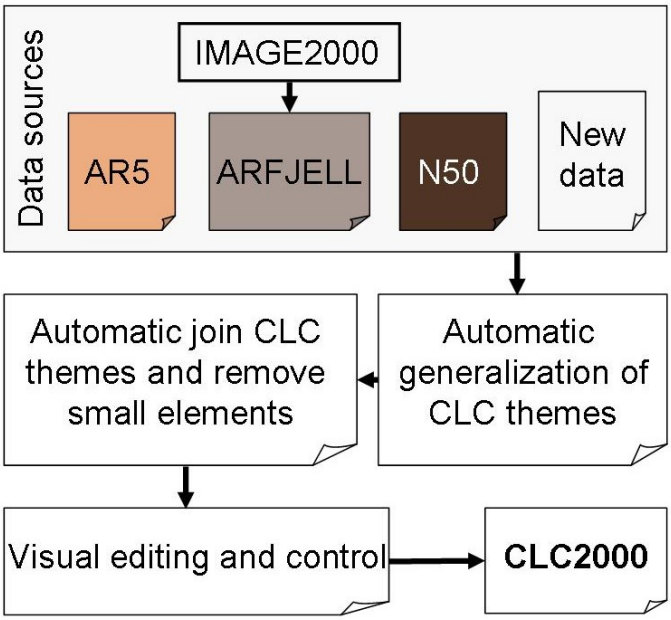


Figure 2. The data sources and steps involved in the production of CLC2000 (Heggem and Strand 2010).

1.1.1. AR5

AR5 is a Norwegian national land capability classification system and map dataset that describes land recourses, with emphasis on capability for agriculture and natural plant production. Minimum mapping unit is 0.2 ha. The dataset is produced and maintained by Skog og landskap. The following selection of data was used as input to CLC2000:

- Surface type: Arable land, pasture, forest, peat bogs, open land, urban area
- Forest cover type: Broad-leaved forest, coniferous forest, mixed forest
- Soil conditions: Organic, blocky or block field areas

AR5 covers the productive part of Norway. Areas close to, or above the tree line are not included in the dataset.

1.1.2. ARFJELL – AREA ABOVE THE TREE LINE

ARFJELL covers mountains and unproductive areas above the timberline (covering almost 45 % of Norway's land area) and includes a wide spectre of resource-types. It contains land

cover ranging from unproductive boulder fields to high-productive meadows. Five classes are produced: abiotic, sparse vegetation, lichen, intermediate vegetation and vigorous vegetation. Three classes were extracted from ARFJELL:

- Moors and heath land (composed of intermediate vegetation and vigorous vegetation)
- Sparsely vegetated areas (composed of scattered and sparse vegetation and lichen dominated areas)
- Bare rock

ARFJELL was produced using a semi-automatic method developed at Skog og landskap. The input datasets for ARFJELL were the IMAGE2000 images, mountain area mask generated from the N50 dataset and a snow and cloud mask. Training areas were selected through a qualified image interpretation for each of the five result classes. The training areas were carefully selected in areas of varying illumination and the normalized difference vegetation index (NVDI) is used as a supplement for the interpretation. Homogeneous polygons were aggregated. Mosaic classes may appear where the second largest class inside a polygon covers more than 20 % of the total area. The classified polygons were given to trained interpreters who manually accepted or corrected the proposed class for each polygon. In areas where suitable IMAGE2000 scenes were unavailable, other satellite images or aerial photos were used. As a result, the images used for ARFJELL are from the period 1994–2006.
The ARFJELL method was found to produce acceptable results for the mountain areas.

1.1.3. IMAGE2000

The delivery from EEA included Landsat 5 and 7 images (1999-2002).

A 50 m resolution digital elevation model is produced by Norwegian Mapping Authority. This model is interpolated from Norwegian 1:50 000 topographic maps which has an elevation accuracy of ±20 meters. The model was used for topographic normalization of satellite images.

1.1.4. TOPOGRAPHIC MAP (N50) AND CADASTRE

The digital topographic 1:50 000 map database (N50) is produced and maintained by the Norwegian Mapping Authority. The dataset consists of several themes which were used as input to CLC2000: Surface cover, transport, building and construction. Forest, peat bogs and agriculture was only used in areas without AR5 coverage. Road and rail network, buildings, camping ground, ports, airports and golf courses were edited before use. Urban fabric areas, industrial units, sport arenas, quarries/gravel pits, cemeteries and parks were used as a direct input.

The Cadastre information is the official ground property, address and building register in Norway. It contains all building points. To delineate classes under the artificial surfaces heading, a data set including all points older than 1 January 2001 created from this data base, was used.

1.1.5. DATASETS ESTABLISHED OR MODIFIED – NEW DATA

Some land use elements could not be retrieved completely from existing digital map datasets:

- Ports
- Airports

- Golf courses
- Industry and construction sites
- Mineral extraction sites
- Dump sites
- Camping grounds
- Sport and leisure facilities
- Beaches
- Inland marshes
- Intertidal flats

Different datasets, either delivered from other institutes or prepared at Skog og landskap, were used to locate potentially large enough areas (> 25 ha) with the classes mentioned above. This was done by displaying the datasets together with aerial photos from Norge-i-bilder (lit. "Norway in pictures", NiB). NiB is a national database of orthorectified aerial photos and satellite images. During the project period approximately 50 % of Norway was covered with images with a resolution better than one meter. NiB was available through WMS.

In addition a "young forest dataset" was extracted from IMAGE2000 and contains areas dominated by young forest and clear cuts. This was the input data source used as the basis for the transitional woodland/shrub class in CLC2000.
The first version of the young forest dataset was produced by Skog og landskap during the CLC2000 project. The input data sets were the IMAGE2000 images and training areas (young forest/not young forest) were selected manually for each scene.

1.2. The generating of CLC2000

1.2.1. INTEGRATION OF THE DATASETS

When all datasets were established and assembled and the satellite image interpretation was completed, the separate datasets could be integrated and generalized into CLC2000. AR5 was the main data source for the natural land cover classes below the timberline and ARFJELL above the timberline. N50 data was used in areas where neither AR5 nor ARFJELL was available. N50 is also used to separate lakes, water courses and marine water. N50 and the cadastre datasets were the main data sources in built-up areas.
Together, the data sources gave a complete dataset covering the whole of Norway including most of the required classes. As some of the classes could not be derived directly from any of the data sources, special analyses were in some cases needed to extract the necessary information.
All datasets collected were used as input in the generalization process. As some of the datasets were related to the same CLC class, or the same areas, a priority list was made according to accuracy and importance of the datasets. The input features were ranked according to priority and the feature with the highest priority would be selected in cases where two or more features were present in the same area.

1.2.2. THE GENERALIZATION PROCESS

As the input datasets are more detailed than the CLC specification for minimum mapping area and minimum width, the datasets needed to be simplified. An automatic generalization method was developed in order to produce the CLC2000 dataset. The conceptual steps in the generalization process are described in 0.

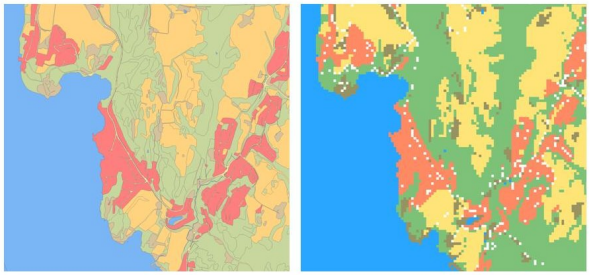
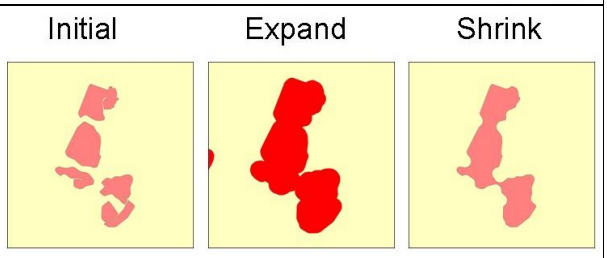
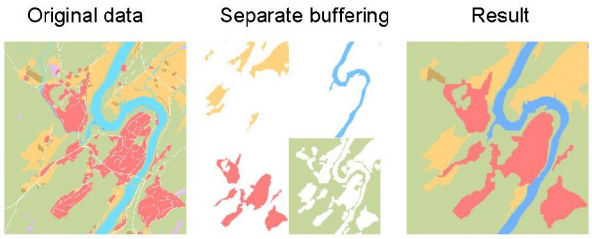
All available input datasets was converted to raster and recoded into CLC classes. All generalizing was carried out on raster datasets. Very small pixel groups were removed from the dataset and the geometry were slightly smoothened.	
Larger units were obtained by buffering: • All pixel groups were expanded 50 m. As a result they merge with adjacent objects. • All the expanded pixel groups were then shrunken 50 m. The buffering was done for each class separately. Areas closer than 100 m were then joined and narrow parts eliminated. Small (< 25 ha) pixel groups were finally removed from the dataset.	
All buffered classes were joined. A priority table was used to determine the correct class in areas where two or more buffered classes did overlap The dataset was then cleaned and remaining small areas were moved into neighbour areas. In the final step the dataset was converted into vector data.	

Figure 3. The conceptual steps in the generalization process (Heggem and Strand 2010).

1.2.3. INTEGRATION OF GENERALIZED CLC2000 AND MANAGEMENT OF SMALL AREAS

All CLC2000 themes were generalized and then joined into one dataset using raster overlay. During the generalization process, buffering and merging lead to larger, more continuous areas.

The new, overlay dataset contained gaps with no assigned class and areas with an extension less than 25 ha area. These were removed in a two step process:

1) Merging the empty areas with neighbour polygons based on the thematic content of the gaps (found in the input data) based on a simplified version of the CLC standard priority table.
2) Remaining parcels less than 25 ha were allowed to split and transferred into neighbour areas and the automatic generalization was completed.

1.2.4. POST PROCESSING OF AUTOMATIC GENERALIZED CLC2000

The dataset resulting from the automatic generalization process required post processing in order to meet the CLC2000 specifications. Step one converted the raster datasets into vector files. The second step was a visual control and manual editing. The final step was to assemble the regional tiles into one CLC2000 dataset.

The Norwegian CLC2000 was verified by EEA in April 2009 (Heggem and Strand 2010).

1.3. The generating of revised CLC2000

The revised CLC2000 were generated by merging only the technical changes in the CLC-Changes database (chapter 2.2) together with the existing CLC2000.

Revised CLC2000 = CLC2000 + CLC-Changes$_{technical}$

- CLC2000 = the original CLC database for 2000 (polygon larger than 25 ha and wider than 100 m)
- CLC-Changes$_{technical}$ = technical changes (code errors in original CLC2000 were corrected)

2. CLC2006

CLC2006 is based on CLC2000.

2.1. Data sources

2.1.1. IMAGE2000

The delivery from EEA included Landsat 7 and 5 scenes (1999-2002).

2.1.2. IMAGE2006

The delivery from EEA included SPOT4 and IRS P6 scenes (2005-2006).

2.1.3. TOPOGRAPHIC MAPS

Topographic maps were available in 1:5 000, 1:50 000 and 1:250 000. They were available as seamless, georeferenced digital raster images through WMS.

2.1.4. DATASETS ESTABLISHED OR MODIFIED – NEW DATA

A cadastre dataset with all new buildings in the period between 31 December 2000 and 1 January 2007 was established to easier detect new land use. The data included a classification of the buildings and information about the size. These were used to mark areas where CLC change was expected.

NiB was linked to the GIS software through WMS and thus available to the interpreters to verify the actual changes and/or the sizes of the areas.

A “forest mask” showing changes in the forest between 2000 and 2006 was made for the areas with most extensive forestry. The objective was to automatically classify forest land into 1) clear cut or recently clear cut areas (open areas and young forest) and 2) tree covered areas. The first category is temporary open areas where regrowth has started or is expected to start soon. The second category has a cover of trees (is wooded) and is anything from young thinning stand to mature and old forest.
An overlay analysis was used to detect the following change categories: 1) forest in 2000 and forest in 2006, 2) forest in 2000 and clear cut in 2006, 3) clear cut in 2000 and clear cut in 2006, and finally 4) clear cut in 2000 and forest in 2006. Category 2 and 4 correspond to change of forest state.

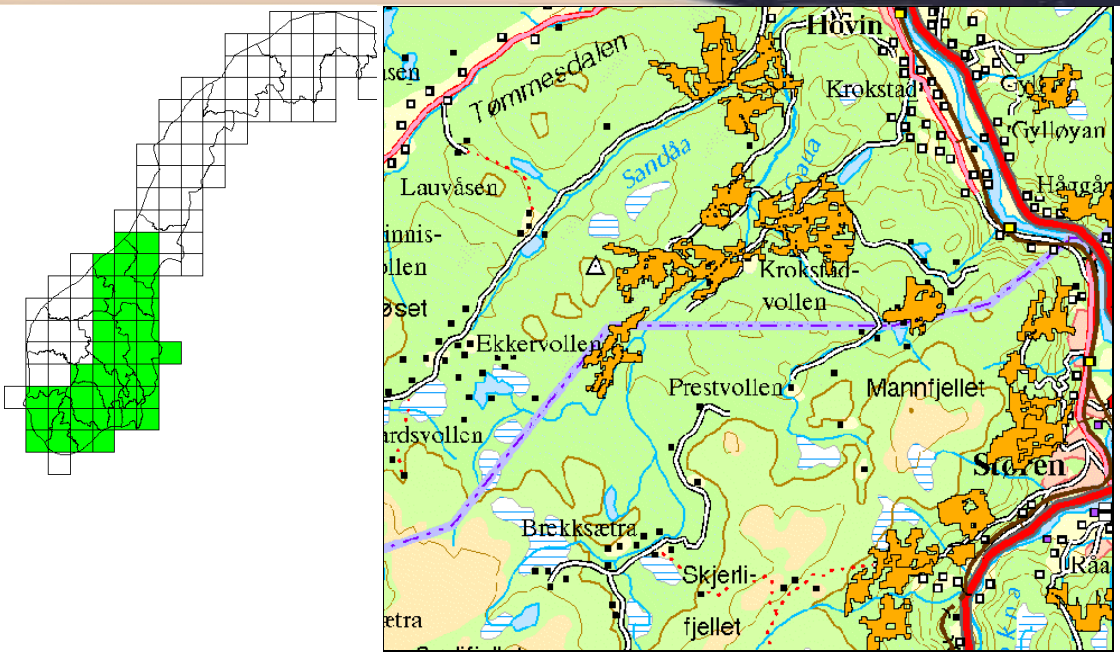


Figure 4. The figure shows areas where forest probably has been logged between 2000 and 2006. The clusters (in orange) are here drawn on top of a digital topographic map N250. © Norway digital. (Aune-Lundberg and Strand 2010a)

2.2. The mapping of CLC-Changes

The changes in the CLC2000 dataset between year 2000 and year 2006 were mapped manually following the instructions in the CLC2006 technical guidelines (EEA 2007). The changes were stored as a new dataset called CLC-Changes. IMAGE2006 was the basis for change mapping, as described in the technical guidelines, but the operators also extensively used the online access to NiB through a WMS integrated into the CLC-Changes working environment. Topographic maps and other ancillary data were also available online. For mapping the changes, Norway was divided into 112 tiles, each measuring 75 km x 75 km (see Figure 4).

Two different changes in the land use were mapped:

1. Real changes; differences in land use between 2000 and 2006
2. Technical changes; corrections of the CLC2000 database.

2.3. The generating of CLC2006

The production of CLC2006 was done by combining CLC2000 and CLC-Changes:

CLC2006 = CLC2000 + CLC-Changes

- CLC2006 = CLC database for 2006 (polygon larger than 25 ha and wider than 100 m)
- CLC2000 = the original CLC database for 2000 (polygon larger than 25 ha and wider than 100 m)
- CLC-Changes = changes (larger than 5 ha and wider than 100 m) between year 2000 and 2006 and corrections of the CLC2000 database
- + indicates a GIS process, including automatic generalization and some actions taken by a photo interpreter

The Norwegian CLC2006 was verified by EEA in February 2010 (Aune-Lundberg and Strand 2010a).

3. EXAMINATION OF THE CLC-QUALITY

The aims of the analyses are to evaluate the classification correctness and accuracy of CLC, compared with detailed vegetation maps. We look at how well the CLC classes fit the Norwegian vegetation zones, and finely we want to see the strength and weaknesses using highly generalised CLC classes.

For the analyses we used land resource maps (scale 1:5000 and 1:50 000) in the overlay (intersect) with the CLC2000 map (Aune-Lundberg and Strand 2010b).

3.1. Results and discussion

The accuracy of the CLC2000 classes based on vegetation contents was quite good.
Due to the generalization process and the large scale with minimum polygon size of 25 ha, is it inevitable with a certain amount of other vegetation groups than expected and intentioned.

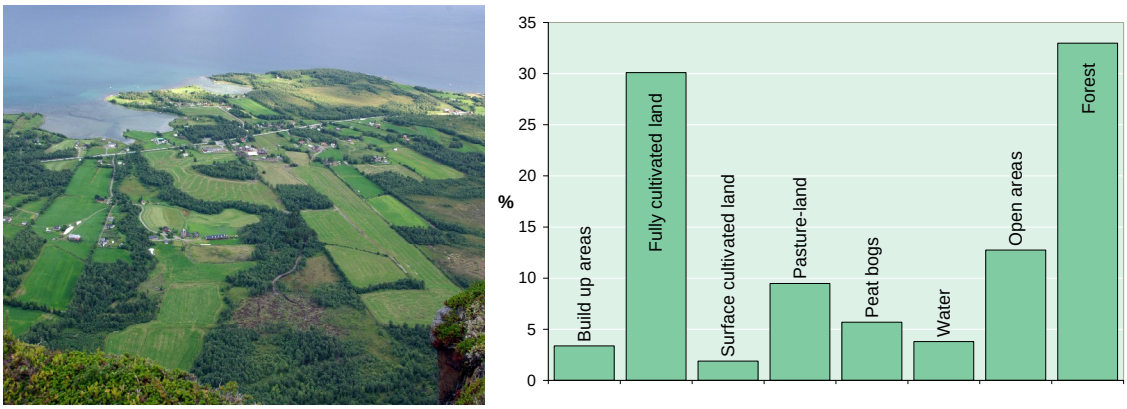


Figure 5. Land principally occupied by agriculture, with significant areas of natural vegetation. Photo: Per Bjørklund, Skog og landskap. (Aune-Lundberg and Strand 2010b)

The CLC class “Land principally occupied by agriculture, with significant areas of natural vegetation” reflects the land use in agricultural areas in a good way. The agricultural areas are small in extent and goes often into a mosaic with forest, semi natural areas and wetlands. According to previous investigations, the class should have 25-75 % arable land and pasture.

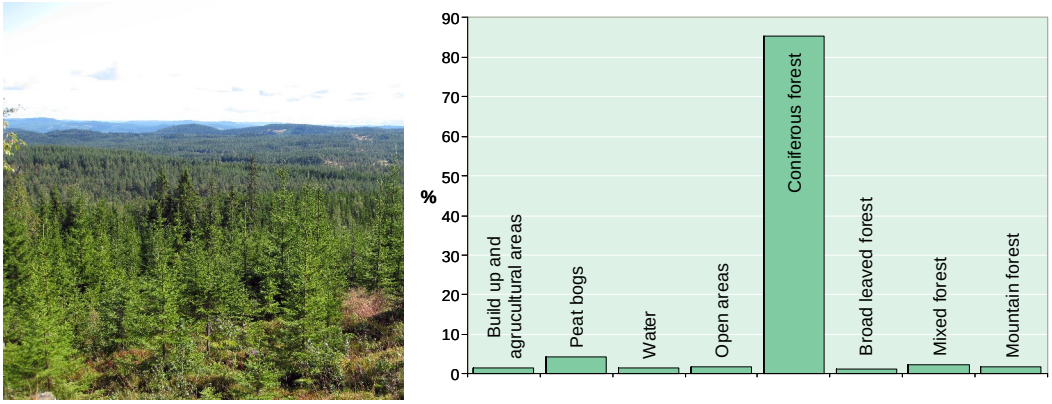


Figure 6. Coniferous forest. Photo: Linda Aune-Lundberg, Skog og landskap. (Aune-Lundberg and Strand 2010b)

As we can see from 0, the CLC class “Coniferous forest” is rather homogenous. There are only a few occurrences of other land types within this class. The histogram, showing that almost 85 % of the area is forest, largely reflects the actual land cover.

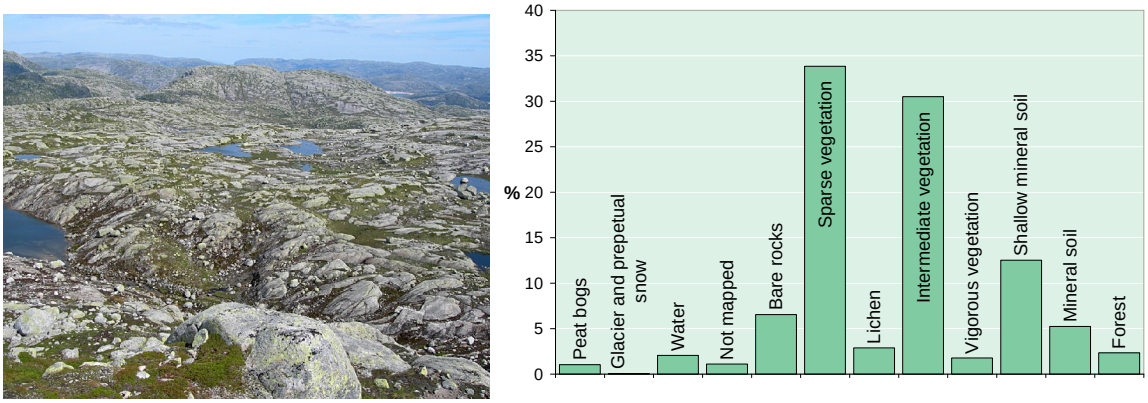


Figure 7. Sparsely vegetated areas. Photo: Yngve Rekdal, Skog og landskap. (Aune-Lundberg and Strand 2010b)

For large parts of the mountainous areas, were the CLC class “Sparsely vegetated areas” are highly presented, sparse and intermediate vegetation dominate. We also have elements of bare rock and areas with thin soil cover as fragments within this CLC class. This is actually how the Norwegian mountain areas look.

Based on this analysis it looks like CLC gives a good visual impression of the general land cover patterns in Norway. But the heavily generalized CLC classes contain a wide spectre of vegetation classes, and gradients such as poor to rich vegetation will not be seen in this small scale, low detailed map.

4. CONCLUSIONS

Examination of the thematic content and variation inside the CLC classes in Norway shows that the results are reasonable with respect to the class definitions.
The Norwegian methods have furthermore turned out to be highly cost-efficient. As an example, the man-hours used for compilation of CLC2006 in Norway were only 50 % of the initial estimate made by EEA. The reason is not that Norwegian analysts are more competent than their European colleagues, but rather that the systematic utilization of auxiliary data allowed the analysts to use their time more efficient.

5. REFERENCES

Aune-Lundberg, L. and Strand G.-H. 2010a. “CORINE LAND COVER 2006. The Norwegian CLC2006 project”. Report 11/2010 from Norwegian Forest and Landscape Institute, Aas.

Heggem, E. S. F. and Strand G.-H. 2010. “CORINE LAND COVER 2000. The Norwegian CLC2000 project”. Report 10/2010 from Norwegian Forest and Landscape Institute, Aas.

Aune-Lundberg, L. and Strand G.-H. 2010b. "CORINE LAND COVER CLASSES. Examination of the content of CLC classes in Norway". Report 5/2010 from Norwegian Forest and Landscape Institute, Aas.

EEA. 2007. CLC2006 technical guidelines, EEA Technical report 17/2007. European Environmental Agency, Copenhagen.

Paper reference number: PN-122

Title: "The Norwegian approach to land cover mapping based on a common European nomenclature: Using existing data and satellite imagery"

Presenters: Anne Nilsen and Hanne Gro Wallin

Authors: Eva S. F. Heggem, Linda Aune-Lundberg, Geir-Harald Strand, Anne Nilsen and Hanne Gro Wallin at The Norwegian Forest and Landscape Institute (Skog og landskap).

Mail: PO Box 115, 1431 Aas, Norway

E-mail: ban@skogoglandskap.no and hgw@skogoglandskap.no

Phone: +47 64948000

Fax: +47 64948001

Presenters' images: Attached as two jpg-files.

Presenters' biography:

Anne Nilsen
Master of Science in Geomatics, The Norwegian University of Life Sciences (UMB), 1994-1997
Research assistant in remote sensing, UMB, 1997-2003
Scientific Adviser remote sensing/GIS, Skog og landskap, 2003-present

Nilsen was the administrative CORINE coordinator of both the CLC2000 and CLC2006 projects. She's responsible for producing statistics from the landscape monitoring and land use statistics describing the state of land resources in Norway. In addition, Nilsen is in charge of the map production (paper and internet) at Skog og landskap. Since 2005 she has been involved in different projects on the Balkans and in Africa teaching GIS and remote sensing.

Hanne Gro Wallin
Scientific adviser in photogrammetry and image interpretation, Skog og landskap, 1995-2002
Head of remote sensing department, Skog og landskap, 2002-present

Wallin was the administrative CORINE leader and in charge of the high resolution data control in CLC2000. She coordinates the aerial image interpretation for landscape monitoring and represents Skog og landskap in the “Aerial photographs of Norway”-project which ensures aerial images covering the entire country at intervals of 5-10 years. In addition, Wallin is strongly involved in “Norwegian landscapes in retrospect”. Since 2007 she has been engaged in different projects abroad; image interpretation of forest reserves in Serbia, teaching remote sensing in Bosnia and Sudan.