

1 General comment

This is the produced BALANS land cover classification - as well as the generalised product based on the original land cover classification - for the Baltic Sea drainage basin. Please note that the product is not homogeneous, the approach has been to produce the best possible product pixelwise within the time and cost limits of the project. This means that the quality of the reference data sets the standard for the quality of the BALANS land cover classification. Since the land cover reference datasets primarily are country-based, the quality of the product is homogeneous within a country.

The land cover classification is based entirely on IRS 1C/1D WiFS data.

For further information concerning the classification work, please read the BALANS report *BALANS Land Cover and Land Use – Classification Methodology*, Balans-utv-24, ver 2.0, June 2001, written by Ulrihca Malmberg. The report is distributed with this dataset as a PDF file.

2 Differences between the generalised product and the original BALANS Land Cover and Land Use dataset

Both the original BALANS Land Cover and Land Use dataset and the generalised product has a pixel size of 150 x 150 metres. However, the generalised product has a minimum mapping unit of 10 hectares for the class *Water bodies* and 25 hectares for remaining classes, meaning that a single pixel can't have a different class than its neighbours.

The second difference from the original BALANS dataset is that the two classes *Coniferous forest* and *Deciduous forest* into the class *Forest*, the classes *Agricultural land* and *Seminatural areas* have been merged into *Open land*, and the class *Bare rock* has disappeared during the generalisation process.

2.1 Generalisation process

The generalisation process used is based on the generalisation performed in the swedish CORINE LC project. It consists of four different steps;

1. Filter the dataset to remove areas smaller than two pixels.
2. Creating a "bridge" between small areas of the same class that are of a maximum distance of one pixel from each other.
3. Merging small areas with classes of similar nature.

4. Let the surrounding classes expand into remaining small areas (the *Water* class is not allowed to expand).

3 Comments Agricultural land

Since the quality of the agricultural information that covered the entire drainage basin (*Baltic Sea Region GIS. Maps and Statistical Database* 10x10 km agricultural data and the *World Atlas of Agriculture*) was of too low quality, it was decided only to use agricultural information for areas where good agricultural reference data could be retrieved. This is the case for Sweden, Finland, Denmark and Poland. For these areas, the class *Open land* (code 2) has been subdivided into the classes *Agricultural land* (code 21) and *Seminatural areas* (code 22). For remaining areas, the class *Open land* (code 2) remains intact.

A separate agriculture image, *balticregiongis_agriculture_10km_resolution.img* has been created, giving the percentage of *Agricultural land* for 10x10 km squares. This information is derived from the pasture and arable datasets available on the web site *Baltic Sea Region GIS, Maps and Statistical Database*, <http://www.grida.no/baltic/index.htm>. The dataset is free of charge. For more detailed information about this dataset, please read the documentation on the web site.

4 Data format

The README.DOC file is a MS Word file.

The image files are Erdas Imagine IMG files and are compressed in ESRI run-length code. They are stored in the BALANS projection.

Projection parameters:

Projection: Transverse Mercator
 Ellipsoid (spheroid): WGS84
 Datum: WGS84
 Units: Metres
 Scale factor of central meridian: 0.997
 Longitude of central meridian: 21 00 00
 Latitude of origin: 0
 False easting: 1 000 000 (to avoid negative co-ordinates)
 False northing: 0

5 File description

File	No. Layers	Description
README.DOC	-	Word document.
BALANS_Landcover_Classification_Methodology_ver2.pdf	-	PDF file. Report about the classification methodology used to obtain the BALANS dataset.
Balans_landcover.img	1	Land cover classification. The class <i>Open land</i> has been divided into <i>Seminatural areas</i> or <i>Agricultural land</i> for areas with

		good agricultural reference data available. For areas with no agricultural information, the major class <i>Open land</i> (code 2) remains.
Balans_landcover_generalised.img	1	The generalised product of the original classification. This product should be available for free on the Baltic Environment Atlas web site. The forest classes have been merged into a single <i>Forest</i> class, the classes <i>Seminatural areas</i> and <i>Agricultural land</i> have been merged into <i>Open land</i> and the class <i>Bare rock</i> has disappeared during generalisation. The minimum unit for each area is 25 hectares for all land cover classes except for <i>Water bodies</i> that has a minimum mapping unit of 10 hectares.
balticregiongis_agriculture_10km_resolution.img	1	Agricultural information, showing the percentage of <i>Agricultural land</i> on 10x10 km squares for the entire Baltic Sea drainage basin, derived from the "Baltic Sea Region GIS, Maps and Statistical Database" dataset. Layer 1: Code 21 for <i>Agricultural land</i> , code 100 for <i>Not agricultural land</i> and code 254 for pixels with no agricultural information available. Layer 2: Percentage values.
Balans_scenes_per_pixel.img	3	IRS 1C/D WiFS and MODIS scenes used to classify each pixel. Three layer image, displaying the scene with the highest priority in layer 1 etc.
Balans_refdata_per_pixel.img	2	Reference data used to classify each pixel. Two band image listing land cover reference data in layer 1, and agricultural information in layer 2.

6 Land Cover class codes – Full resolution

Valid for file;

- Balans_landcover.img

Code	Land Cover class
0	Outside of Baltic Sea drainage basin
1	Artificial surfaces
2	Open land
21	Agricultural land
22	Seminatural land
3	<i>Forest</i>
31	Coniferous forest
32	Deciduous forest
4	Wetland
5	Water bodies
6	Glaciers and perpetual snow
7	Bare rock

99	Cloud
254	No data

7 Land Cover class codes – Generalised product

Valid for file;

- Balans_landcover_generalised.img

Code	Land Cover class
0	Outside of Baltic Sea drainage basin
1	Artificial surfaces
2	Open land
3	Forest
4	Wetland
5	Water bodies
6	Glaciers and perpetual snow
99	Cloud
254	No data

8 Agricultural codes

Valid for files:

- balticregiongis_agriculture_10km_resolution.img

DESCRIPTION BALANS_AGRICULTURE.IMG	
0-100	Indicating the percentage of Agricultural land for a 10x10 km area. Based on Baltic Sea Region GIS, Maps and Statistical Database agricultural data (Arable land and Pasture).

9 Scene codes

Valid for file

- Balans_scenes_per_pixel.img

The image is a three layer image, displaying all the scenes (cloudfree areas) used in the classification of each pixel. The value for the scene with the highest priority is stored in layer 1, the scene with the second highest priority in layer 2 etc. A zero value in any layer indicates that no scene has been used for that priority level.

These codes indicates what scenes have been used in the classification of each pixel. The actual scene numbers have not been used, since it would create a 16-bit image, that would need a large amount of disk space.

Code	Scene	Registration date
1	21/30	1999-07-11
2	22/21	1999-08-02
3	23/25	1999-07-30
4	26/20	1999-06-26
5	26/28	1999-05-18
6	30/20	1997-07-11
7	30/32	1999-05-20
8	32/28	1999-07-11
9	33/25	1997-08-19
10	36/22	1999-06-13
11	36/25	1997-09-03
12	38/22	1997-06-09
13	38/27	1999-06-15
14	39/25	1999-06-12
15	44/22	1999-06-29
16	45/25	1999-05-17
17	MODIS 26 july 2000	2000-07-26
18	MODIS 27 august 2000	2000-08-27
19	MODIS 19 september 2000	2000-09-19

10 Reference data codes

Valid for file

- Balans_refdata_per_pixel.img.

The image indicates what reference data has been used to classify each pixel. The image consists of two layers, the first indicating from what reference data the land cover information has been collected, while the second layer informs what reference data has been used to separate the class *Open land into Agricultural land* and *Seminatural areas* (this is done by using an agricultural mask).

The original resolution is also given for each reference data set, indicating the quality of the reference data.

Code	Reference dataset	Original resolution
0	Outside of Baltic Sea Drainage Basin	-
1	Corine LC	100 m. Generalised product, minimum mapping unit 25 ha
2	Swedish Terrain Type Classification	25 m
3	Finnish Land Use and Forest Classification	25 m, resampled to 150 m
4	Agricultural class from IAKS 99 – Sweden	25 m

5	Agricultural class from Finnish Land Use and Forest Classification	25 m, resampled to 150 m
6	Agriculture class from AIS (Areal Informations Systemet) Denmark	originally vector data, based on data in scale range 1:20 000 - 1:50 000
7	Agriculture class from Poland national Corine dataset	Originally 25 m, resampled to 150 m
254	No reference data exists	

11 Validation results

11.1 Terminology

Accuracy variable	Description
Target accuracy	The probability that a randomly selected point for a given class in the classification has its correspondance in reality.
Interpretation accuracy	The probability that a randomly selected point for a given class in reality has its correspondance in the classification.
Areal difference	How many percentages in the classification that have been over- or underclassified (-) compared to reality.
Average accuracy	The probability that a randomly selected point for a given class in the classification has its correspondance in reality <u>and</u> that a randomly selected point for a given class in reality has its correspondance in the classification.

11.2 Results Sweden

Two areas were validated, one in the northern region with sparse growing coniferous forest and large wetland areas (terrain type map 26KNO, 26KNV, 26KSO and 26KSV), and one north of Stockholm with mixed landscape (terrain type map 11INV).

Number of areas	Validation								
	1	4	5	7	21	22	31	32	Total
1	91	0	0	0	4	3	1	0	99
4	0	57	0	0	4	6	5	12	84
5	0	0	89		4	0	1	1	95
7	0	0	0	0	0	1	0	0	1
21	1	2	0	1	87	3	1	1	96
22	9	15	1	1	28	82	8	25	169
31	1	23	10	47	30	43	99	25	278
32	2	3	0	1	14	1	3	31	55
Total	104	100	100	50	171	139	118	95	877

Class	Target	Interpretation	Areal diff	Average
1	91,9	87,5	-5,1	89,7
4	67,9	57,0	-19,0	62,0
5	93,7	89,0	-5,3	91,3
7	0,0	0,0	-4900,0	0,0
21	90,6	50,9	-78,1	65,2
22	48,5	59,0	17,8	53,2
31	35,6	83,9	57,6	50,0
32	56,4	32,6	-72,7	41,3

11.2.1 Comments

Each validation point corresponds to 36 pixels (6x6) in the classification. The number of validation points selected for each class is not representative compared to the actual land cover distribution for the area.

The tables below shows the area representation for each land cover class in the classification (in percentage) and the area representation for each class for the validation points used.

Area 11INV (north of Stockholm)		
Class	Classification distribution (%)	Validation point distribution %
1	9,9	20,3
4	0,7	10,8
5	3,2	10,8
7	0,0	10,8
21	17,6	21,6
22	27,2	8,6
31	38,6	14,0
32	2,8	3,0

Area 26K (northern Sweden)		
Class	Classification distribution (%)	Validation point distribution %
1	0,1	2,3
4	19,9	11,3
5	3,0	11,3
7	0,1	0,0
21	3,6	22,4
22	12,1	22,4
31	53,0	12,0
32	8,2	18,3

Bare rock (only existing in area north of Stockholm) has primarily been classified as *Coniferous forest*. This is not surprising, since the bare rock areas are very small and scattered (normally smaller than the resolution of 150 m for the classification), hence not possible to derive a "pure" bare rock pixel for this area.

Coniferous forest has been highly overclassified and mixed with most of the other classes. Primarily this is because mixed pixels (that has a large representation, especially in the area north of Stockholm) have been classified as *Coniferous forest*.

Deciduous forest has been highly underestimated. This areas are very small and scattered and are basically "mixed" pixels in the satellite data. The wrongly classified pixels for this class have in first hand been classified as *Coniferous forest* or *Seminatural areas*.

Agricultural land has been highly underestimated. This class has been derived from a mask, and no particular conclusion can be made in this matter, other than that the quality of the mask. It has to be noted that the mask is originally a raster data set of 25 m resolution, and that it has been generalised to 150 metres. A quality reduction from this generalisation is to be expected.

11.3 Results Poland

Number of validation areas	Validation points						
	1	4	5	21	31	32	Totalt
1	125	1	1	2	0	1	130
4	0	3	0	0	0	0	3
5	1	1	82	0	0	0	84
21	6	12	5	131	3	3	160
31	1	4	12	1	129	11	158
32	0	1	1	0	0	78	80
Total	133	22	101	134	132	93	615

Class	Target	Interpretation	Areal diff.	Average
1	96,2	94,0	-2,3	95,1
4	100,0	13,6	-633,3	24,0
5	97,6	81,2	-20,2	88,6
21	81,9	97,8	16,3	89,1
31	81,6	97,7	16,5	89,0
32	97,5	83,9	-16,3	90,2

11.3.1 Comments

Wetlands have been highly underestimated, primarily classified as *Agricultural land* instead. The *Wetlands* have primarily being derived directly from reference

data, and this dataset for the polish area is the european CORINE LC that is a highly generalised product. The *Agricultural land* has been derived from a mask, hence no comments can be made concerning any "misclassification" in this matter.

11.4 Results Finland

No of areas Classification	Validation							
	1	4	5	21	22	31	32	Totalt
1	64	0	0	2	0	0	0	66
4	0	3	0	0	0	0	0	3
5	1	1	110	0	0	1	3	116
21	10	0	0	98	1	1	11	121
22	22	9	0	7	43	8	8	97
31	24	102	1	14	101	105	74	421
32	0	2	0	1	1	0	1	5
	121	117	111	122	146	115	97	829

Class	Target	Interpretation	Areal diff	Average
1	97,0	52,9	-83,3	68,4
4	100,0	2,6	-3800,0	5,0
5	94,8	99,1	4,3	96,9
21	81,0	80,3	-0,8	80,7
22	44,3	29,5	-50,5	35,4
31	24,9	91,3	72,7	39,2
32	20,0	1,0	-1840,0	2,0

11.4.1 Comments

Wetlands have been highly underestimated, and most areas have been classified as *Coniferous forest*. *Artificial surfaces* have also been underestimated (mixed with various other classes) and the same goes for *Seminatural areas* and *Deciduous forest*. As for Sweden, deciduous forest in Finland is a class normally with small stands and often mixed with coniferous forest. *Coniferous forest* has been overrepresented, and the confusion matrix indicates that mixed pixels (especially mixed *Seminatural areas/Coniferous forest* and *Deciduous forest/Coniferous forest* has been classified as *Coniferous forest*.