
 BROCKMANN CONSULT GMBH		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 1

PROBA-V CLOUD ROUND ROBIN

Description of the Test Data Set

Version 1.2

Kerstin Stelzer, Michael Paperin, Grit Kirches

13.10.2016





		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 2

Table of content

1	Introduction	3
2	Pixel collection	3
2.1	Collection criteria	3
2.2	Examples	5
3	Test data set	9
4	Issues	14

Change Log

Version	Date	Changes	Responsible
1.0	06.10.2016	Initial version to ESA	Kerstin Stelzer, Michael Paperin, Grit Kirches
1.1	07.10.2016	Updated version to ESA	Kerstin Stelzer, Michael Paperin
1.2	13.10.2016	Reviewed Version, provided to ESA	Rosario Quirino Iannone, Kerstin Stelzer

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 3

1 Introduction

The scope of this document is to describe the test data set that is provided to all participants of the Round Robin for a Proba-V Cloud Detection Algorithm. The document provides an orientation for the participants about the validation dataset, which will be employed for the validation of the participants' submission. In addition, pixel classification criteria are also explained as well as how different cloud types appear in the image. The test data set consists of 1350 pixels and is not meant to be used as training data set. The database of the test data set containing the labelled pixels is provided to the participants of the cloud detection Round Robin. The final validation data set, which will be used for the validation of all algorithms, will contain over 40.000 pixels and will not be available to the participants.

2 Pixel collection

2.1 Collection criteria

The pixels are manually collected, classified and labelled by an expert user. The expert decides which of the pixels are to be considered, and then, based on his experience, assigns pre-defined properties (e.g., "completely cloudy", "clear sky (land, water, snow/ice)", "semi-transparent clouds", "coastline") for each selected pixel. In a second level characterization it is specified if a turbid atmosphere comes from e.g. desert dust or fog, and water surfaces are further characterized as turbid water, floating vegetation or sun glint. The pixels are only collected if the expert has no doubt in the determination of its properties. The tool for pixel collection and labelling is called PixBox. The data is stored in a database.

The pixels have been collected and labelled for the following categories:

Clouds

- Totally cloudy (opaque clouds)
- Semi-transparent clouds
- Other turbid atmosphere (e.g. dust, smoke)



Clear surfaces

- Clear sky water
- Clear sky land
- Clear sky snow/ice
- Other clear cases

Spatially mixed pixels

- Spatially mixed cloud/land
- Spatially mixed cloud/water
- Spatially mixed cloud/ice

The semi-transparent clouds were further differentiated into three density classes. Those will be in the end taken as one semi-transparent cloud; but it enables to understand which categories of semi-transparent clouds are captured by the cloud detection algorithm during the validation process.

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 4

The classification of the semi-transparent clouds is the following:

- Thick semi-transparent cloud
- Average or medium dense semi-transparent cloud
- Thin semi-transparent cloud

Figure 1 shows a screenshot of a ProbaV RGB images and the position and labelling of collected pixels.

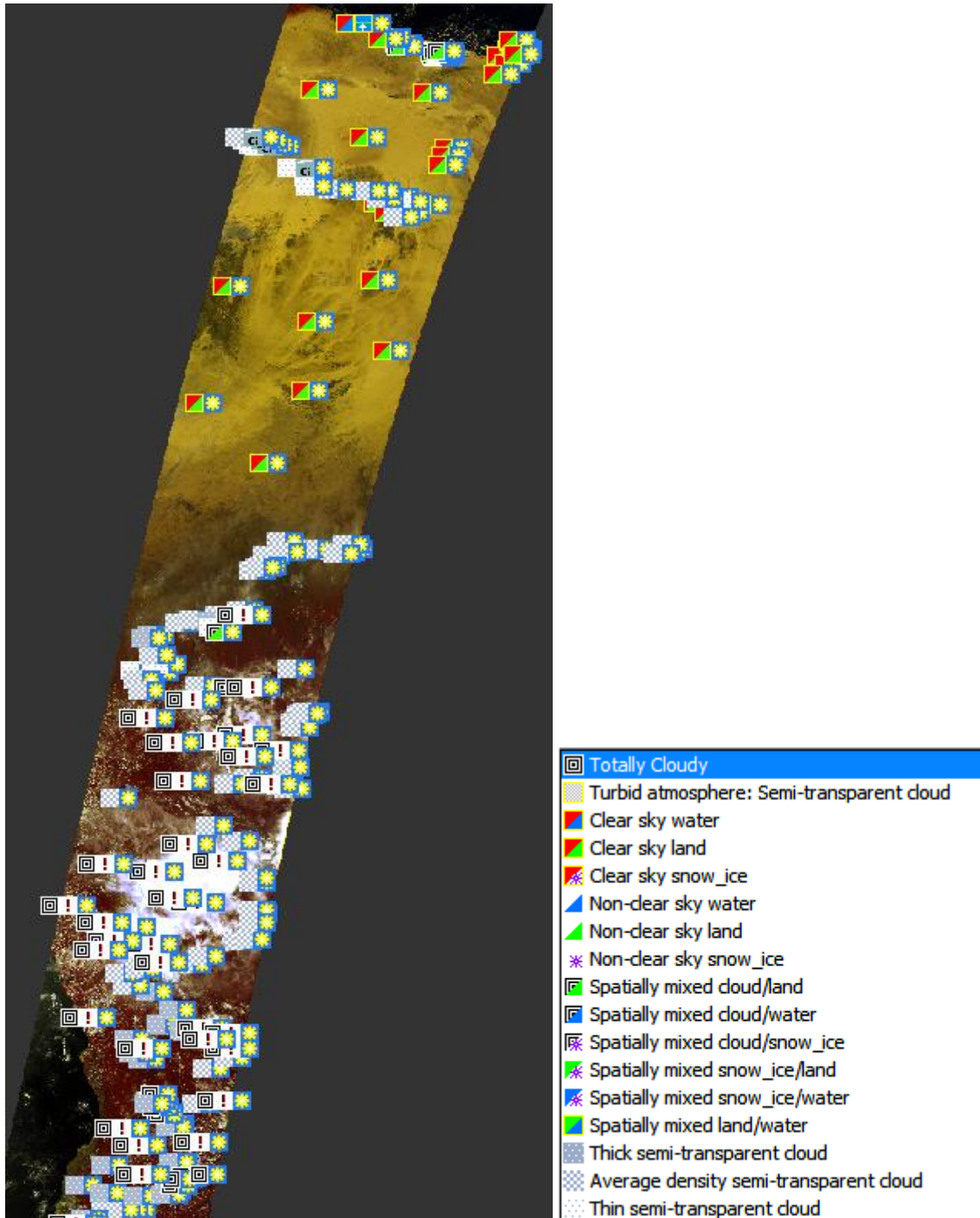




Figure 1: Example image showing the position of collected pixels and labelling.

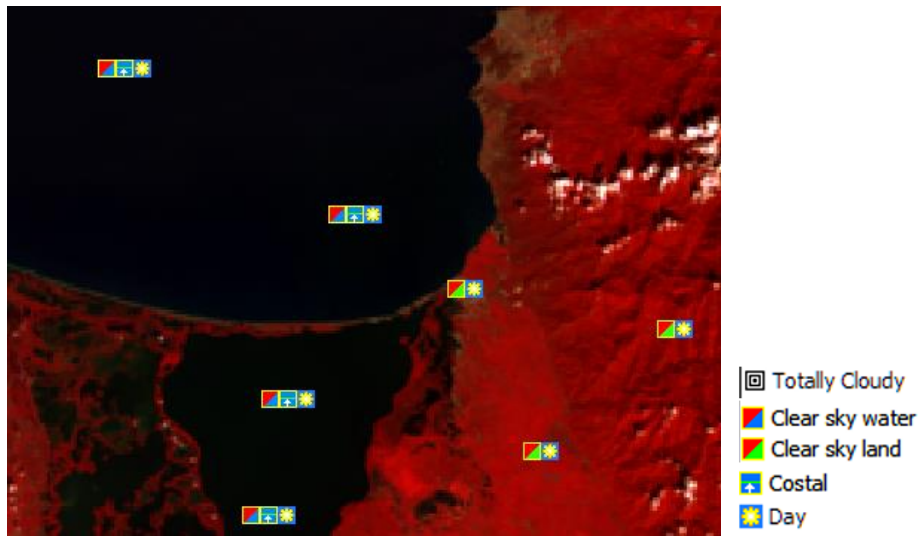
		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 5

2.2 Examples

The following images provide some examples of the collected pixels and the underlying surface types. It shall provide an orientation for the validation data set.

Clear land, clear water and totally cloudy:

land and water:



Snow:



totally cloudy:

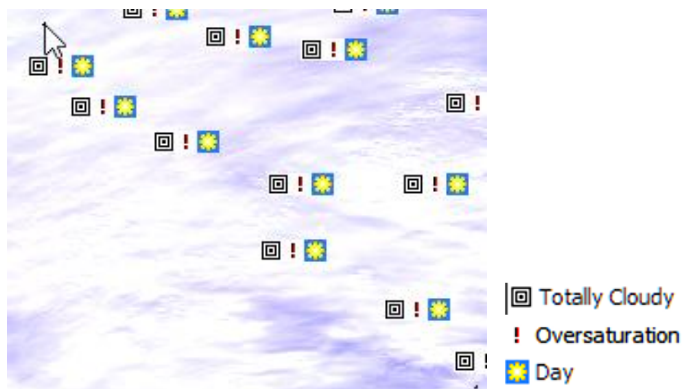




Figure 2: examples of clear surface (land and water and ice) and totally cloudy pixels (below).

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 6

Semi-Transparent cases over land:

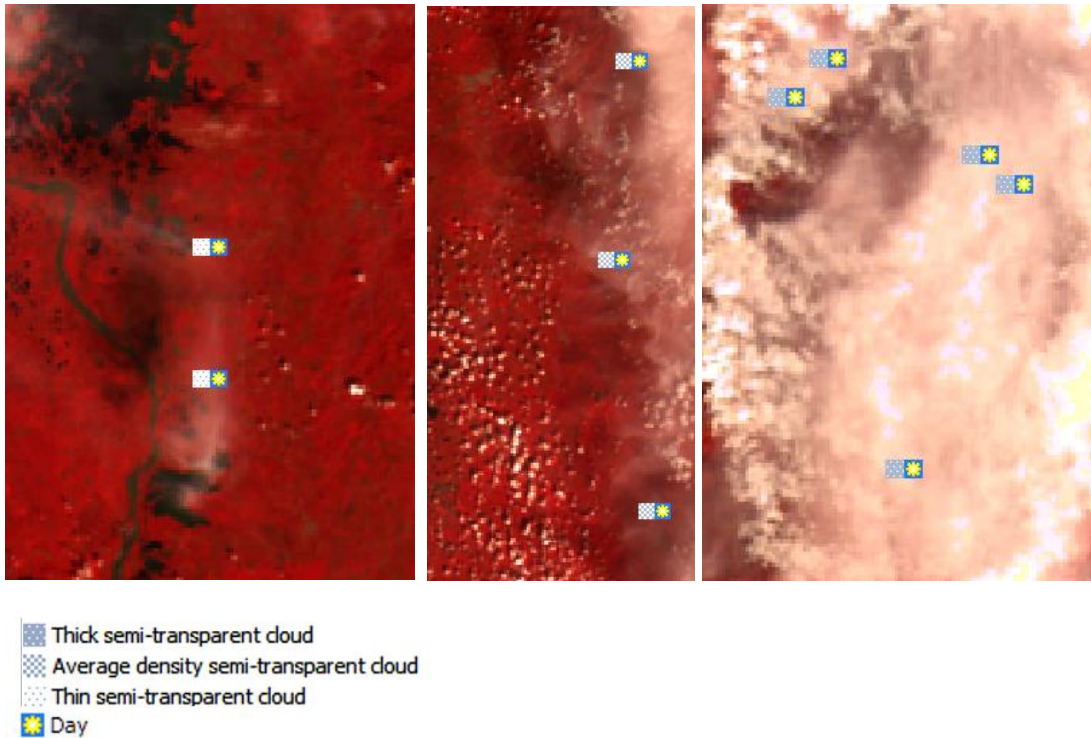


Figure 3: examples for semi-transparent clouds over land: left: thin; middle: medium; right: dense

Semi-Transparent cases over water:

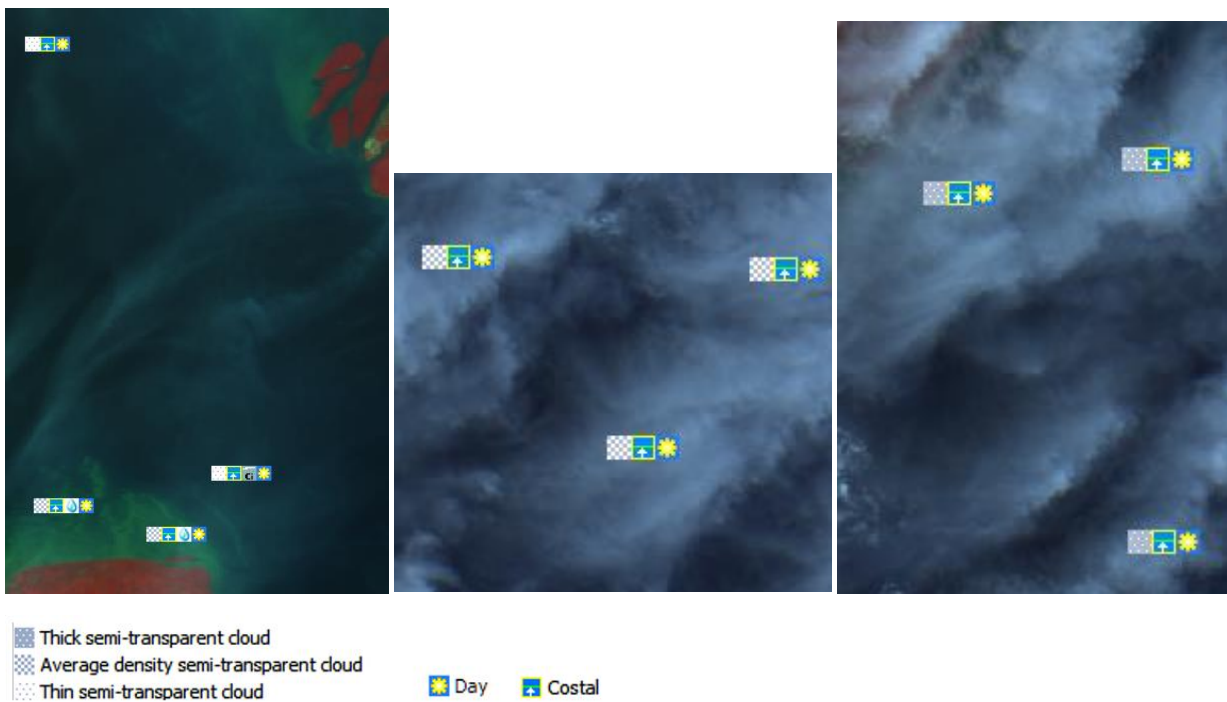


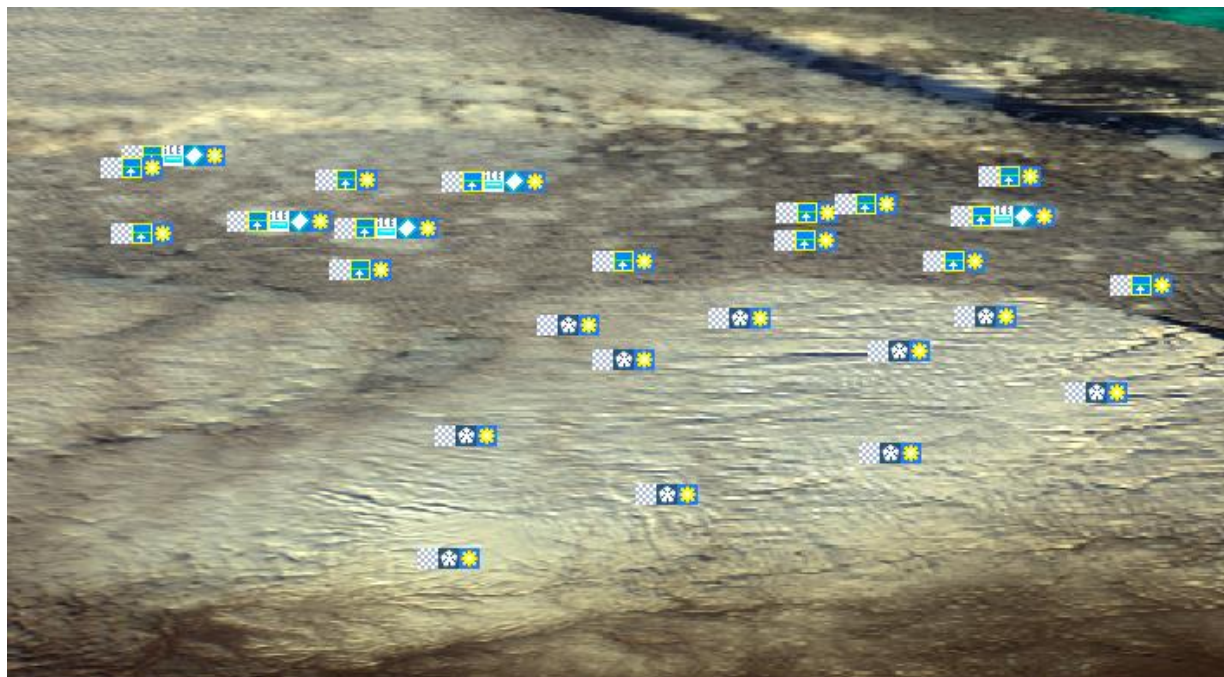


Figure 4: examples for semi-transparent clouds over water: left: thin; middle: medium; right: dense

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 7

Semi-transparent clouds over ice:








-  Average density semi-transparent cloud
-  Coastal
-  Snow
-  Ice
-  Day

Figure 5: examples for semi-transparent clouds over ice (medium)

Spatially mixed cases:

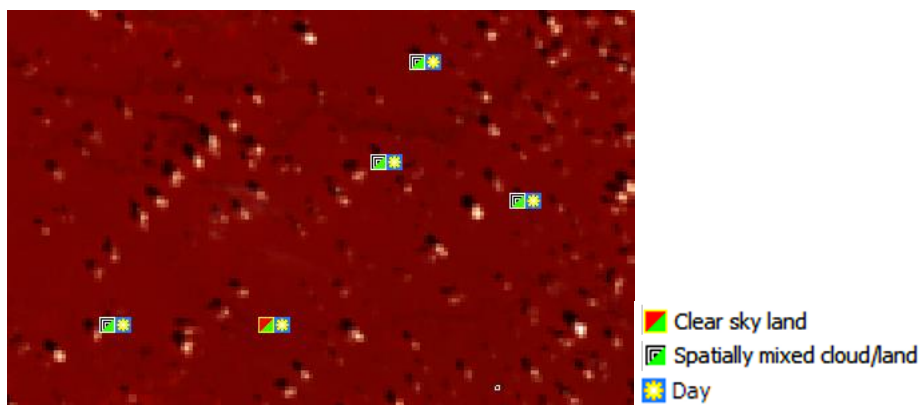




Figure 6: examples of spatially mixed pixels cloud/land

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 8

And some more examples over land of different cases:

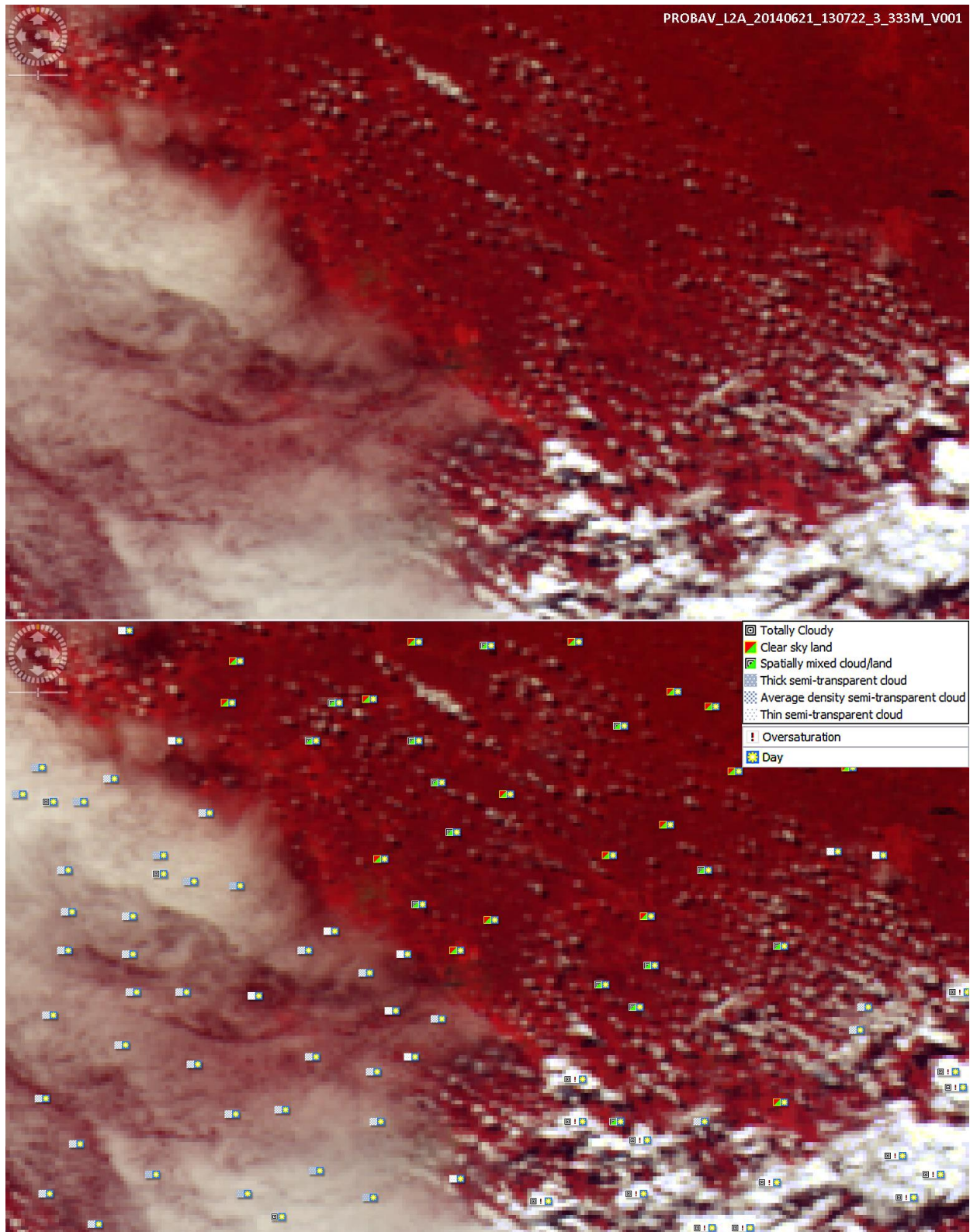




Figure 7: different cloud categories over land

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 9

3 Test data set

The test data set contains 1350 entries collected from 4 different images. The following distribution of categories is available in the test data set:

Category	No. of pixels	Percentage
Total	1350	100%
Totally Cloudy	400	30%
Semi-transparent cloud	438	32%
Other turbid atmosphere cases	23	2%
Clear sky water	192	14%
Clear sky land	205	15%
Clear sky snow_ice	67	5%
Other clear cases	25	2%
Total clear cases	489	36%
Floating ice	67	5%
Glint	59	4%
Cloud shadow	46	3%

The requirement was to collect 30% totally cloudy, 30% semi-transparent and 40% clear cases. The relationship between land and water pixel was requested to be 70:30 (land:water).

Figure 8 shows the global distribution of the test data set pixels. The pixels are collected on the basis of the ProbaV Level 2A products.

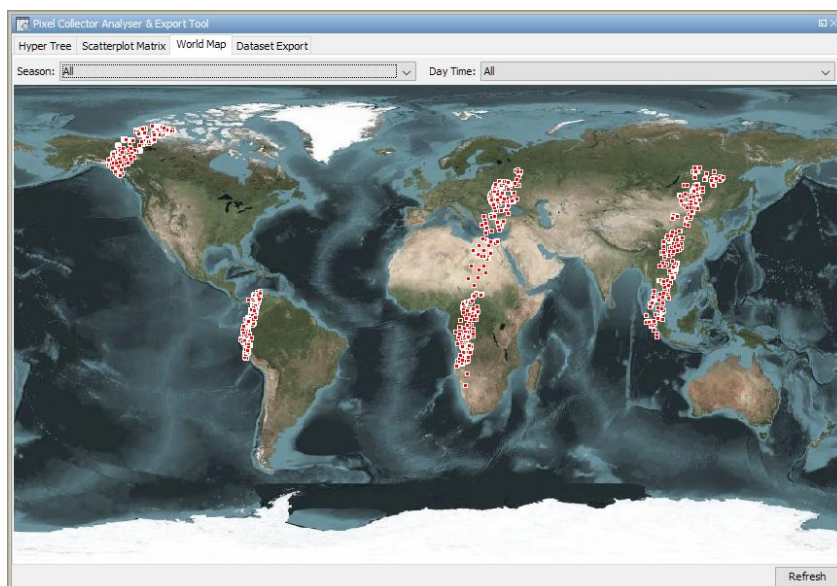




Figure 8: Position of globally collected pixels of the test data set covering clear land, clear water, clear ice, totally cloud and semitransparent clouds

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page

The test data with the 1350 entries with the attributes for each pixel are stored in the test data set in csv or excel format.



The single pixels are identified by a unique pixel ID. The source product of the pixel is identified by a product ID:

PRODUCT_ID	NAME
1005212284	PROBAV_L2A_20140321_094935_1_333M_V001.hdf
1007605388	PROBAV_L2A_20140621_042255_1_333M_V001. hdf
1007605536	PROBAV_L2A_20140921_204654_2_333M_V001. hdf
1007605798	PROBAV_L2A_20141221_155703_2_333M_V001. hdf

The column names and description are listed in Table 1 and the coding of the respective category is listed in the tables thereafter. The main classification is done in the PIXEL_SURFACE_TYPE_ID, while the following categories further characterize the individual surface types.

Table 1: Description of the columns of the test data set

Column Name	Description
ID	Unique ID of the pixel
PRODUCT_ID	ID of the ProbaV product
PIXEL_X	X Position of the pixel in PRODUCT_ID
PIXEL_Y	Y Position of the pixel in PRODUCT_ID
LATITUDE	latitude of the pixel in PRODUCT_ID (decimal degrees, WGS84)
LONGITUDE	longitude of the pixel in PRODUCT_ID (decimal degrees, WGS84)
PIXEL_SURFACE_TYPE_ID	Main class of pixel surface type
ATMOSPHERIC_PROPERTIES_ID	Description of atmospheric properties if not atmosphere is not clear (e.g. smoke, sand storm)
WATER_BODY_TYPE_ID	Water type differentiated between Ocean and inland water
WATER_BODY_CHARACTERISTICS_ID	Specifies if special conditions are in or on the water (e.g. ice, turbid)
SEA_ICE_TYPE_ID	Further specification if Sea Ice is identified
GLINT_ID	Specifies if water body is in glint
OVERSATURATION_ID	Specifies if bright surfaces in saturation in at least one band
CLOUD_CHARACTERISTICS_ID	Specifies if the cloud is a cirrus cloud
CLOUD_HEIGHT_ID	Not applied
CLOUD_SHADOW_ID	Specifies if the pixel is under cloud shadow
SHALLOWNESS_ID	Not applied
SURFACE_TYPE_ID	Coarse surface type specification derived from IGBP map (International Geosphere Biosphere Programme) (IGBPa_1198.nc)
CLIMATE_ZONE_ID	Climate zone derived from Köppen Climate Zone Map
SEASON_ID	The season of a pixel is derived from the latitude (Northern hemisphere/Southern Hemisphere) and date
DAY_TIME_ID	Not applied (all day scenes)

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page



PIXEL_SURFACE_TYPE_ID	NAME
0	Totally Cloudy
1	Non clear atmosphere (none cloud)
2	Clear sky water
3	Clear sky land
4	Clear sky snow_ice
8	Spatially mixed cloud/land
11	Spatially mixed snow_ice/land
12	Spatially mixed snow_ice/water
14	Thick semi-transparent cloud
15	Average density semi-transparent cloud
16	Thin semi-transparent cloud

ATMOSPHERIC_PROPERTIES_ID	NAME
0	None
1	Turbid atmosphere: Desert dust
2	Turbid atmosphere: Smoke
3	Turbid atmosphere: Volcanic eruption
4	Turbid atmosphere: Other

WATER_BODY_TYPE_ID	NAME
0	undefined
1	Lake
2	River
4	Coastal

WATER_BODY_CHARACTERISTICS_ID	NAME
0	None
1	Snow
2	Ice
3	Bright turbid water (blue or brown)

SEA_ICE_TYPE_ID	NAME
0	None
1	Floating sea ice
2	Brash sea ice

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page



GLINT_ID	NAME
0	No Glint
1	Glint

OVERSATURATION_ID	NAME
0	None
1	Saturation in at least one band

CLOUD_CHARACTERISTICS_ID	NAME
0	Unknown
4	Cirrus



CLOUD_SHADOW_ID	NAME
0	None
1	Cloud shadow

SURFACE_TYPE_ID	NAME
1	Evergreen Needleleaf Forest
2	Evergreen Broadleaf Forest
3	Deciduous Needleleaf Forest
4	Deciduous Broadleaf Forest
5	Mixed Deciduous Forest
6	Closed Shrubland
7	Open Shrubland
8	Woody Savanna
9	Savanna
10	Grassland
11	Permanent Wetland
12	Cropland
13	Urban
14	Crop/Natural Veg, Mosaic
15	Permanent Snow/Ice
16	Barren/Desert
17	Water Bodies
18	Tundra

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 13

CLIMATE_ZONE_ID	NAME
0	Unknown
1	A: Tropical
2	B: Dry
3	C: Temperate
4	D: Cold
5	E: Polar

SEASON_ID	NAME
1	Spring
2	Summer
3	Autumn
4	Winter

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page

4 Issues

The collection does not contain pixels which are not fully understood. This was the case for pixels flagged by all SM_FLAGS.GOOD (thus valid pixels) but which seem to have all the same or very similar value (around 0.5, not always exact the same value). Because those pixels have a fixed value, the cloud or snow covered area look “flat” in those parts (Figure 9). Figure 10 shows the values of the blue band along such an area, interrupted by saturated pixels. Those pixels have not been collected because it is not understood why they are flagged as valid, but look unrealistic.

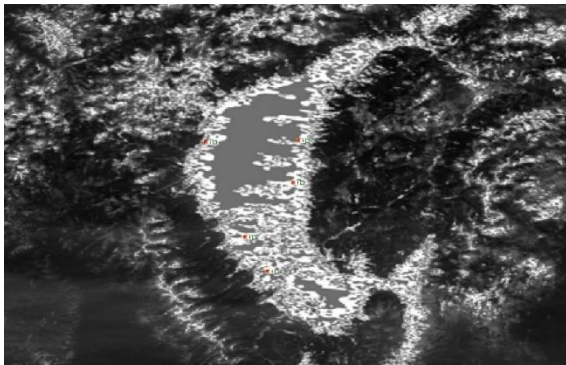


Figure 9: Cloud or snow pixels (in this case snow) set to fixed values and are not flagged as GOOD in bright clouds

In some cases, the pixels in between those flat areas have realistic values, but are not collected because of the artificial behavior.

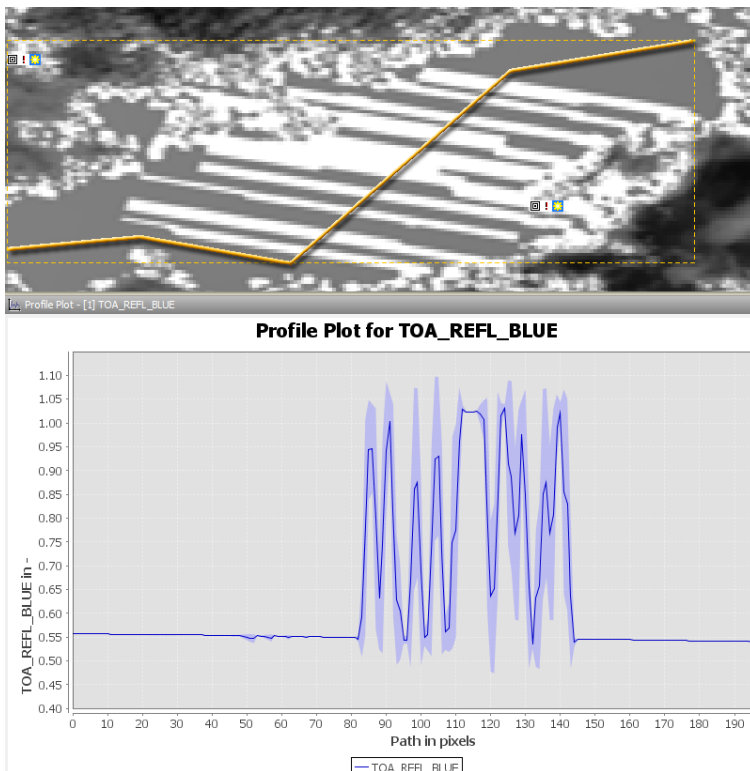




Figure 10: Cloud pixels set to fixed values; not flagged with GOOD flag (above); Blue band reflectance along a transect showing the effect of the fixed values

		Document:	ProbaV_RoundRobin_Testdataset_1.2.docx	
		Date:	13.10.2016	
		Version:	1.2	Page 15

On the other hand, pixels are sometimes not flagged as GOOD in at least one band due to saturation. For bright clouds and snow, this occurs quite often and it was decided after consultation of ESA to include those pixels although they are not fully flagged by the GOOD flags.