PROBA-V CLOUD (C2)

Validation Report

333m Products

Version 1.0

Kerstin Stelzer, Michael Paperin, Uwe Lange

18.10.2019

Table of content

Change Log

Version	Date	Changes	Contributions
1.0	15.10.2019	First Version ProbaV-C2 Validation Report	Kerstin Stelzer, Michael Paperin, Uwe Lange

1 Introduction

In 2017/2018 a round robin exercise for different cloud screening algorithms had been performed, analysed by Brockmann Consult. A dedicated validation dataset comprising manually selected pixels (PixBox) had been collected and used within this exercise. This dataset was collected at 333m resolution data of Proba-V. One of the best performing algorithms was the method by Luis Gomez-Chova from University Valencia, and this algorithm is now implemented in the PROBA-V ground segment. It is foreseen to be implemented and applied to all three spatial resolutions, 1km, 333m and 100m.

Currently, the algorithm is implemented for PROBA-V 333m products and the scope of this document is to provide the validation results of the PROBA-V Cloud C2 for PROBA-V 333m products.

The validation data set and the validation methods are described. On the one hand, the validation is performed with a manually selected pixel collection and on the other hand by comparison of the different cloud flags in randomly selected PROBA-V images.

2 Validation data set

2.1 Pixel collection

The pixels are manually collected, classified and labelled by an expert user. The expert decides which of the pixels are to be considered and based on his experience he 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
  - Thick semi-transparent cloud
  - Average or medium dense semi-transparent cloud
  - Thin semi-transparent cloud
- 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 sub-classification of the semi-transparent cloud pixels as thin, medium or thick semi-transparent clouds, enables us to analyse the validation results in more detail and to understand which categories of semi-transparent clouds are captured by the cloud detection algorithm during the validation process.

Figure 2 shows a screenshot of a PROBA-V RGB image and the position and labelling of collected pixels.

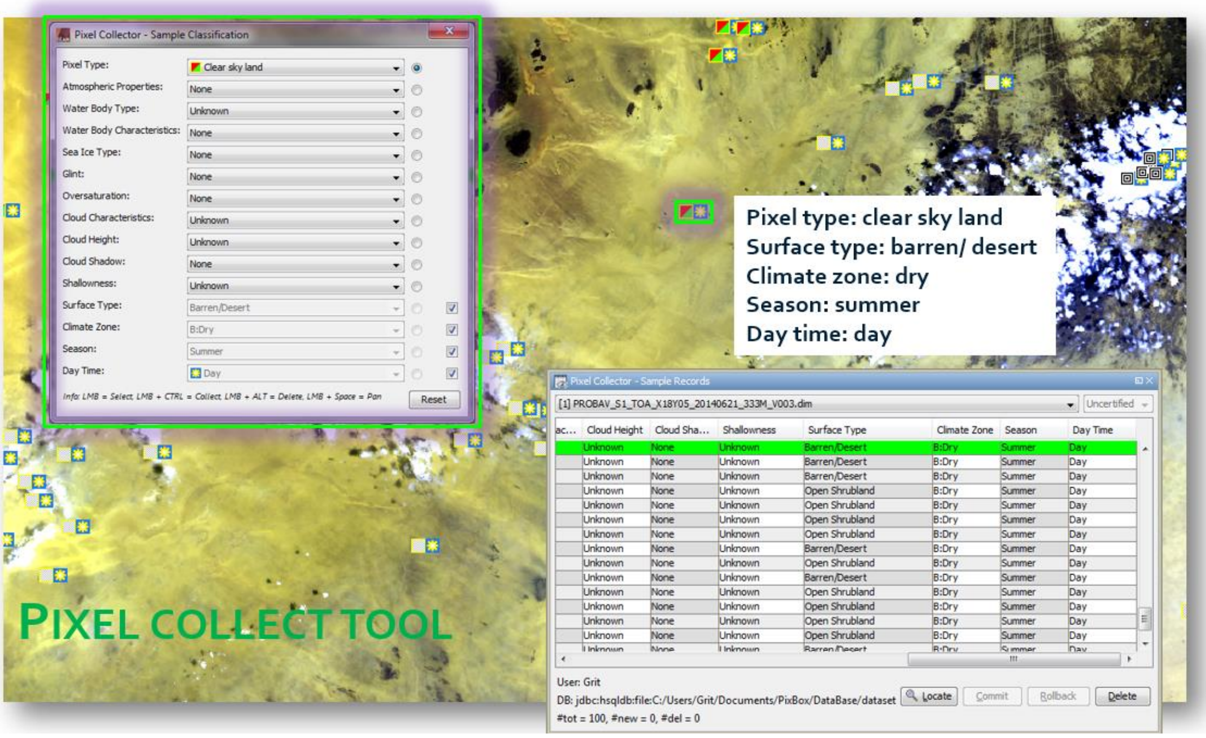


Figure 1: Pixel Collection tool with categories to be assigned to each pixel

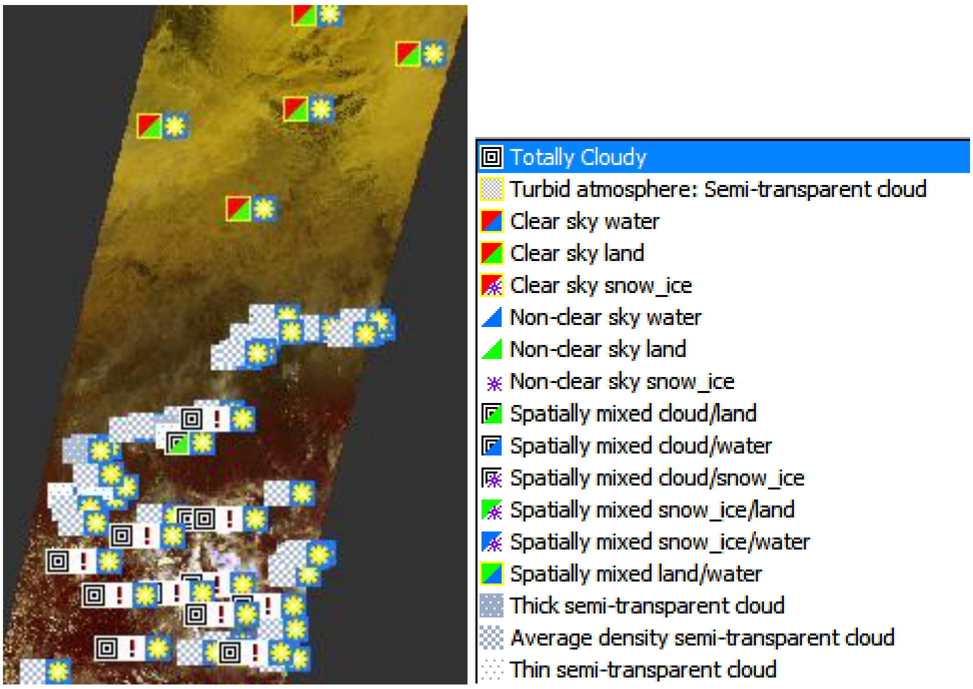


Figure 2: Example image showing the position of collected pixels and labelling.

2.2 Content of the validation data set

The validation data set originally contained 53000 entries collected from 61 different Proba-V 333m L2A images. During the development / improvement of the cloud detection algorithm, 10000 pixels were randomly extracted from this data set and provided University of Valencia for improving their training of NNs. Therefore, for the validation we only use the remaining 43000 reference points collected from 50 products. The input products cover the four days 21.03., 21.06, 21.09. and 21.12.2015. The pixels are collected based on the PROBA-V Level 2A products, processing version V001. The distribution of the different categories is shown in Figure 3.

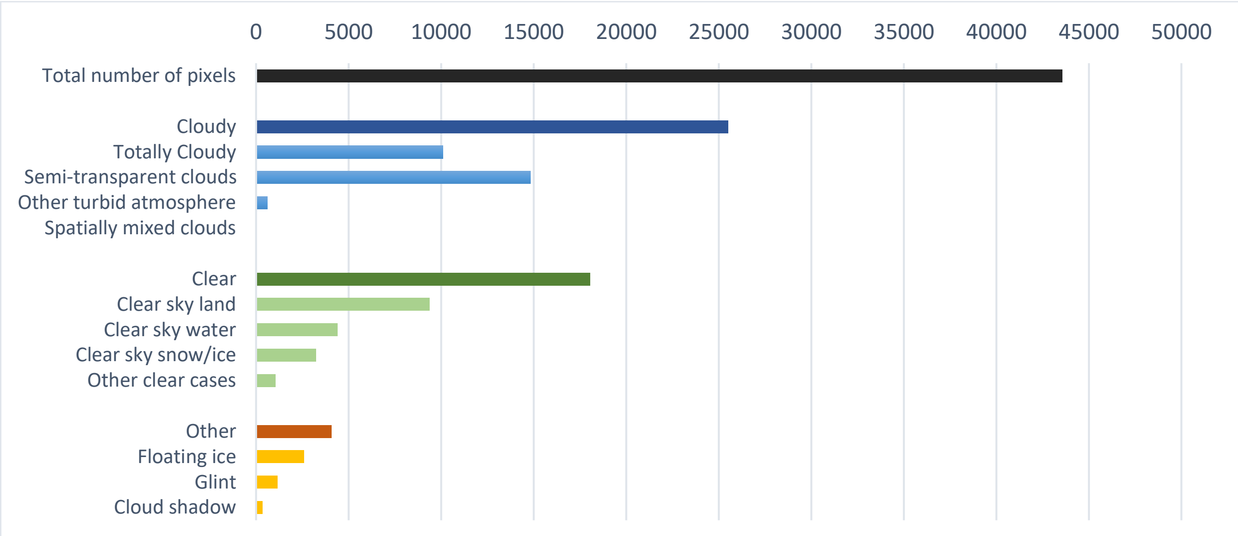


Figure 3: Distribution of surface types within the validation data set (numbers are the counts of pixels per category)

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 4 shows the global distribution of the validation data set pixels.

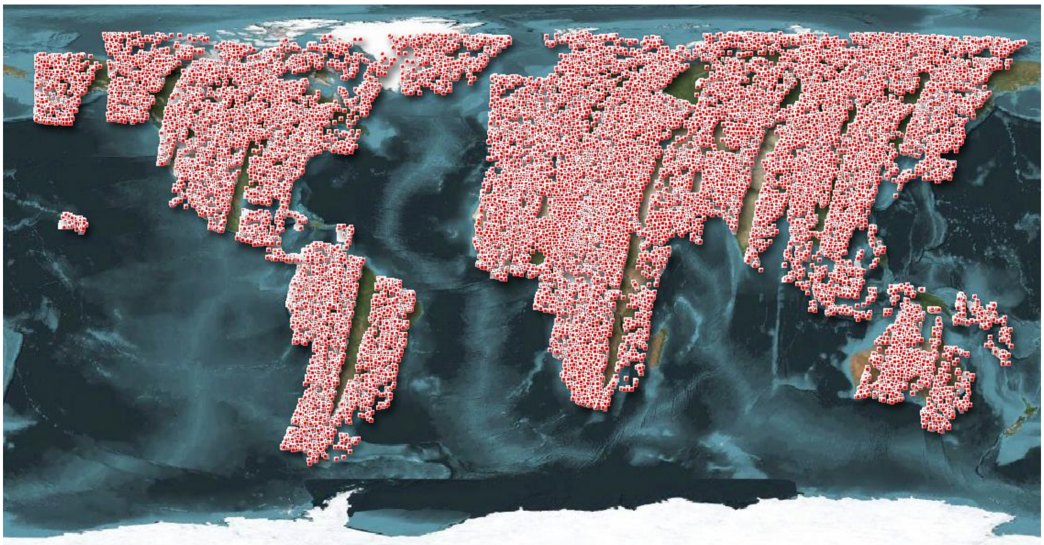


Figure 4: Position of globally collected pixels of the validation data set covering clear land, clear water, clear ice, totally cloud and semitransparent clouds

3 Validation Methods

Two methods were applied to verify the current implementation of the cloud detection algorithm:

1. Visual inspection of the cloud mask
2. Validation with reference data set

3.1 Validation by visual inspection

For the visual inspection, 22 randomly selected products have been investigated. Subsets of these products are investigated and displayed as RGB (NIR-RED-BLUE) with and without the overlay of relevant masks. The masks for CLOUD, SNOW and CLOUD SHADOW have been investigated. The different masks are displayed in changing colours, to highlight best the different clouds over a range of surfaces:

- A black CLOUD mask shows best if thin clouds and cloud borders are detected well (Figure 5 middle).
- A CLOUD mask in cyan is suited if the underlying surface is too dark to see the cloud flag in black.
- A semi-transparent overlay of the CLOUD mask is used to assess the detection of different densities of semi-transparent clouds (Figure 5 right)
- The SNOW ice mask is displayed in cyan or green to be separated well from the yellow CLOUD mask (see Figure 6)

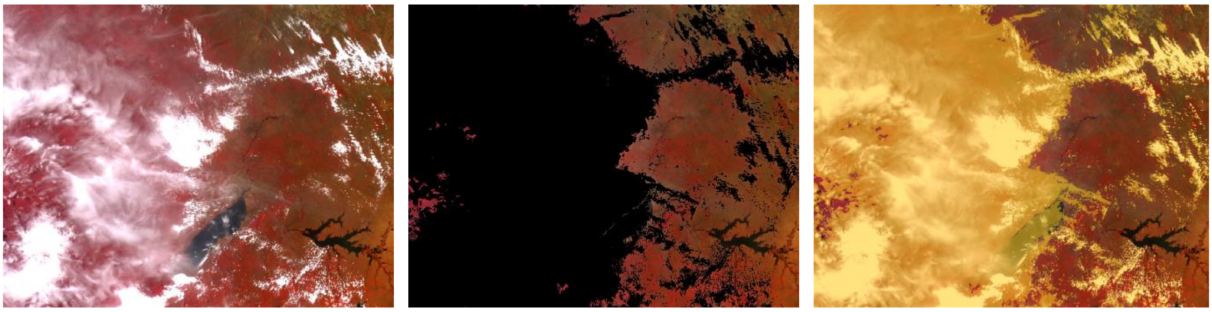


Figure 5: RGB image (left), RGB image with CLOUD mask overlay in black (middle) and in semi-transparent yellow (right) used for assessing the cloud mask

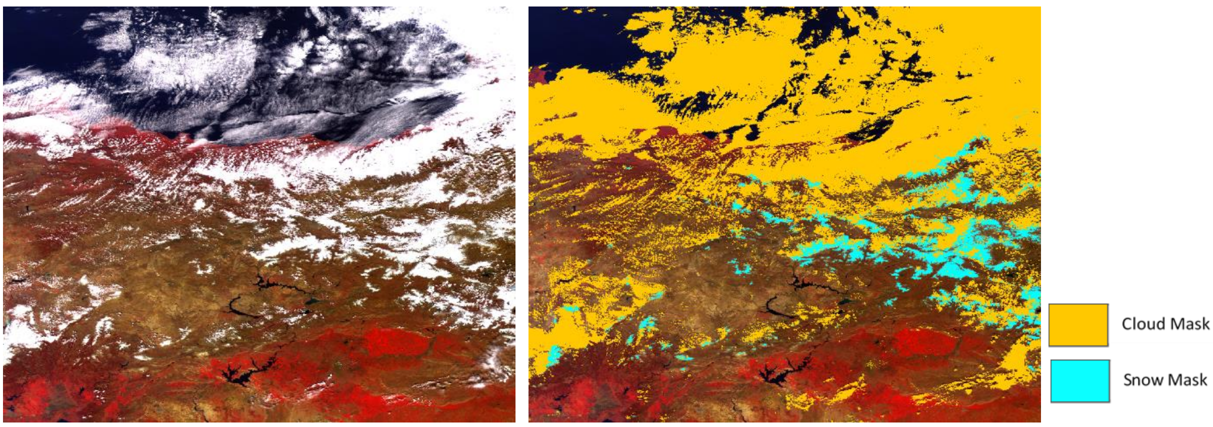


Figure 6: RGB image (1) and RGB image with CLOUD mask (yellow) and SNOW mask (cyan) overlays (2).

112 subsets have been assessed in total. From these subsets, an overall assessment for different surfaces was performed and summarized in 3 quality categories which are further explained in chapter 4.1.6.

3.2 Validation with reference data set

The second validation method is based on the reference dataset containing 43000 pixels. For the statistical analysis, confusion matrices and related statistics were used, namely the User's Accuracy (UA), Producer's Accuracy (PA), Overall Accuracy (OOA) and respective errors (E) (see Figure 7). Further, the values for Krippendorf's alpha is given (Krippendorf 2004[1]). A high User's Accuracy for the CLOUD flag means that the pixel under the CLOUD mask is most probably a cloud. If the Producer's Accuracy is high for the reference cloud it means that a reference cloud is most probably classified as CLOUD. A low Producers' Accuracy for clouds indicates that not all clouds are classified as CLOUD (error of omission or high number of false negative), while a low User's Accuracy for the CLOUD flag accuracy indicates that the CLOUD flag has classified also clear surfaces (error of commission or high number of false positive).

PixBox reference

cloud mask	0	1	sum	UA	UE
0	TN	FN			
1	FP	TP			
sum			Σ		
PA				OAA	
PE					

Figure 7: Confusion matrix showing the agreement between cloud mask and validation data set for clear surfaces and clouds. UA = User's Accuracy, E = Error, PA = Producer's Accuracy, OAA: Overall Accuracy, TN: True Negative, TP: True Positive; FN: False Negative, FP: False Positive

The different accuracy measures are available for assessing the quality of a classification (cloud/no cloud) and calculated as follows:

$$OAA = \frac{TN + TP}{TN + FN + FP + TP} * 100$$

$$UA_{TP} = \frac{TP}{TP + FP} * 100 \quad and \quad UA_{TN} = \frac{TN}{TN + FN} * 100$$

$$PA_{TP} = \frac{TP}{TP + FN} * 100 \quad and \quad PA_{TN} = \frac{TN}{TN + FP} * 100$$

4 Results

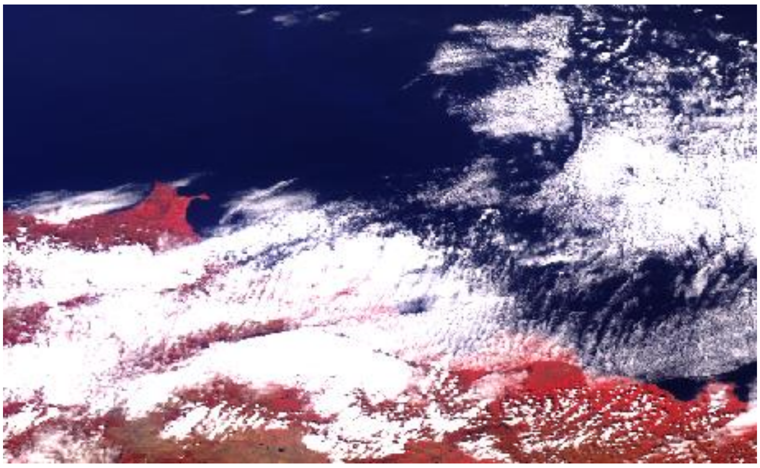
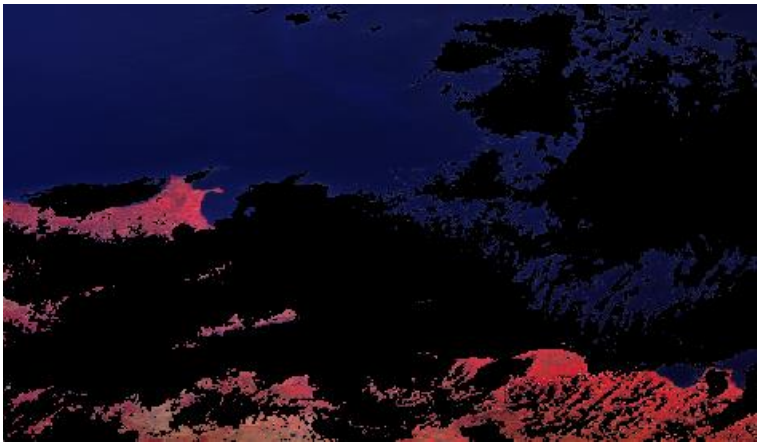
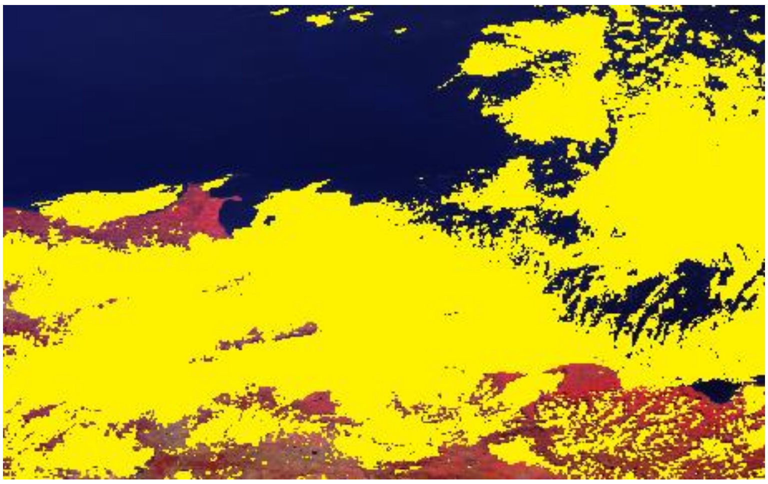
4.1 Visual Cloud Mask Assessment

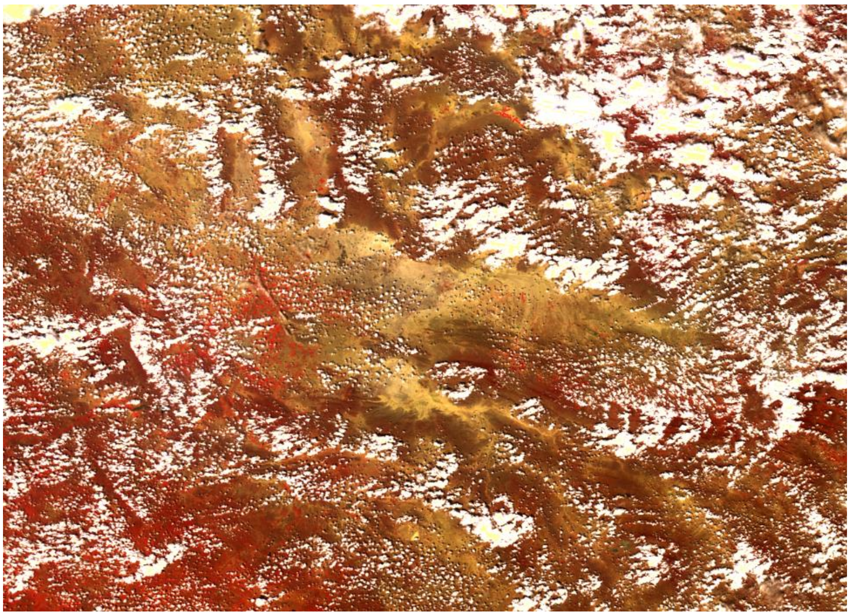
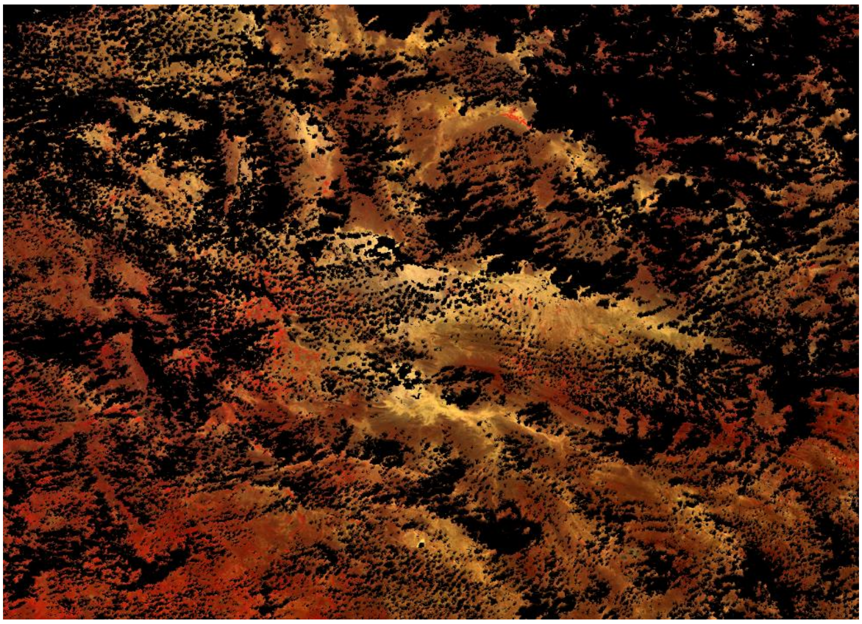
Selected examples of the investigated subsets are provided in the following sections. They cover different cloud and surface types. A short description of the subset as well a short assessment of the performance

[1] Krippendorff, K: (2004): Reliability in Content Analysis: Some Common Misconceptions and Recommendations, Human Communication Research 30(3) (2004), pp.: 411-433

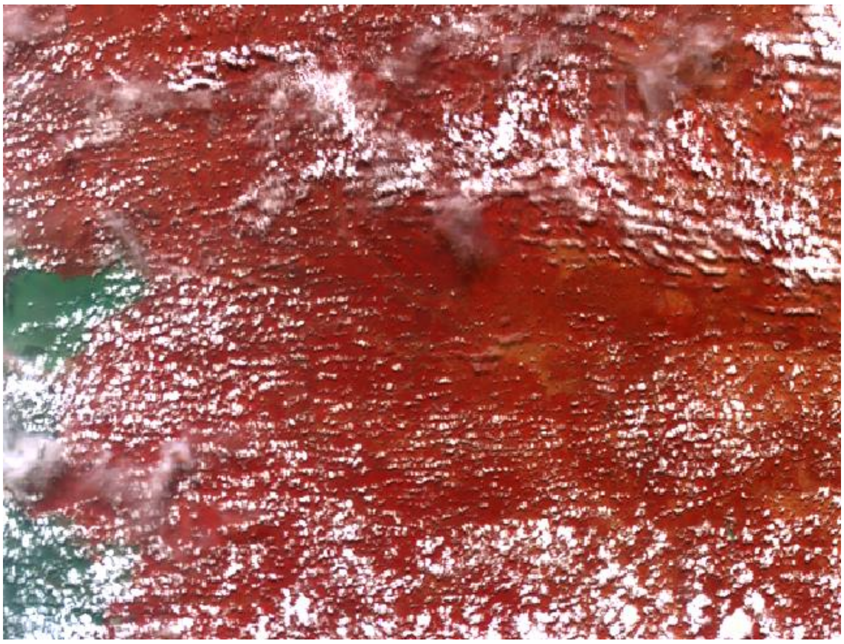
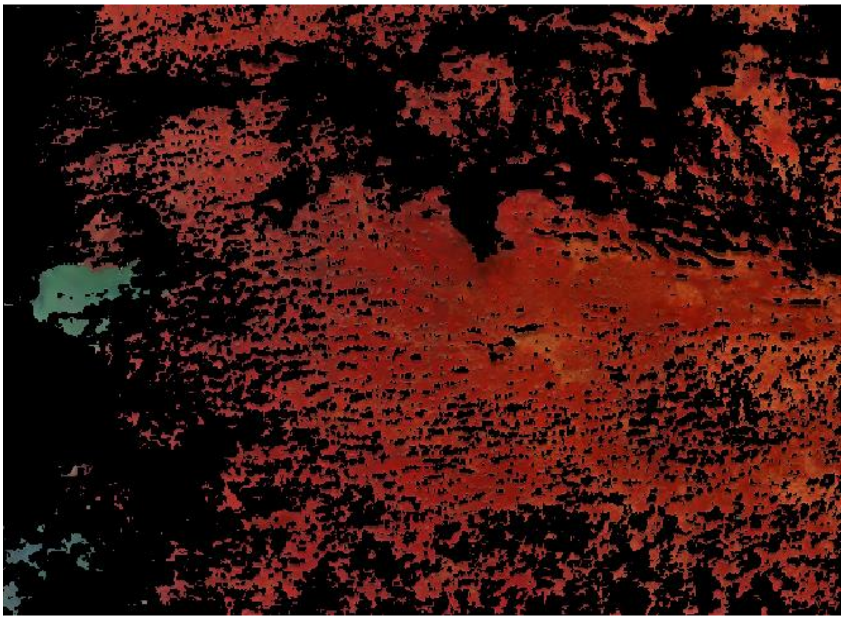
of the cloud flag are given. Dedicated sections show the results for CLOUD detection, SNOW detection and CLOUD SHADOW detection.

4.1.1 Cloud Mask

Opaque clouds - PROBAV_L2A_20140321_080337_3_333M_V101	
	**RGB** The subset shows different cloud types over water and land surfaces. Opaque clouds, semi-transparent clouds and small cumulus clouds are present.
	**Assessment** The cloud mask in black shows that almost no white spots remain after flagging the clouds with the cloud mask, indicating a low number of false positives. This applies true for all cloud types (opaque, semi-transparent and cumulus). Cloud Mask
	**Assessment** The cloud mask in yellow shows that the cloud detection over the dark surfaces (water) is working well for the different cloud types. Cloud Mask

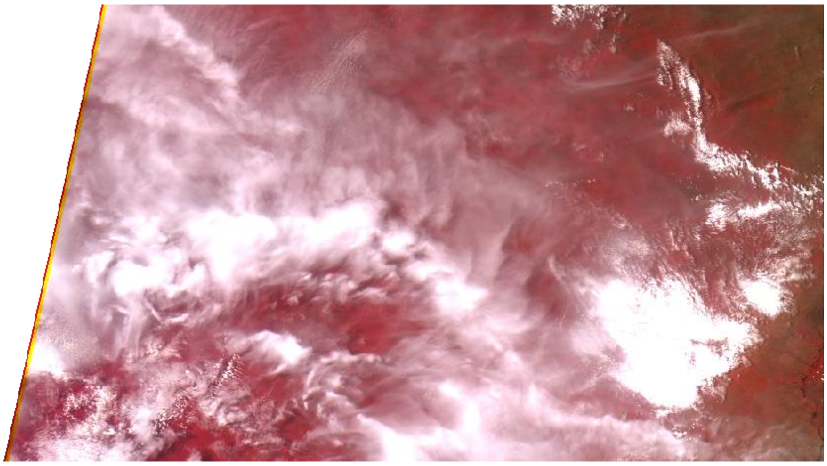
Cumulus clouds over land PROBAV_L2A_20140921_172401_3_333M_V101	
	**RGB** The subset shows cumulus clouds in different sizes covering mountainous areas.
	**Assessment** The cloud mask in black covers the clouds and no cloud border is left out. Cloud Mask

The following two examples show the performance of the cloud mask for small cumulus clouds over land and water.

Cumulus clouds over land and water PROBAV_L2A_20140321_080337_3_333M_V101	
	**RGB** The subset shows cumulus clouds over land and water surfaces. They are overlaid by higher semi-transparent clouds.
	**Assessment** The cloud mask is capturing all small cumulus and overlaying semi-transparent clouds over land and water; no white spots remain. Cloud Mask

The following examples show the performance of the cloud mask with respect to semi-transparent clouds. Several examples are shown because semi-transparent clouds can show many different structures and opacity and the more transparent the clouds are, the more difficult they are to differentiate from the underlying surfaces. However, it depends on the follow-up analysis (e.g. atmospheric correction, land use classification), if thin semi-transparent clouds influence the L2 products.

Semi-transparent clouds over land PROBAV_L2A_20140321_080337_3_333M_V101



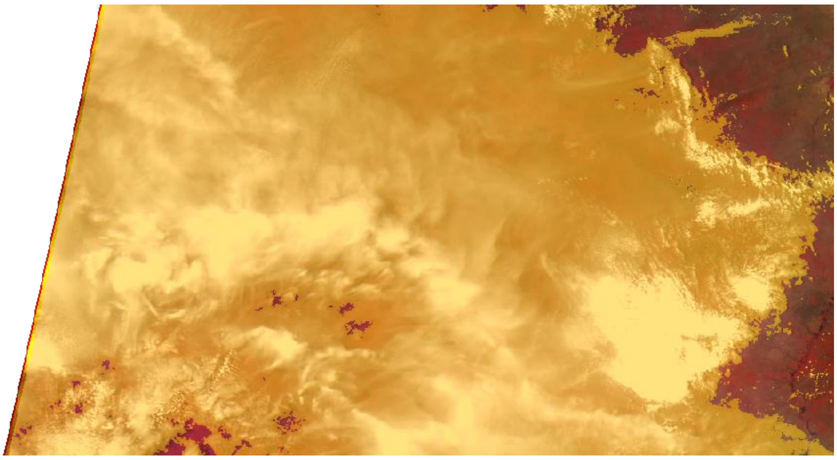
RGB

The subset shows opaque and semi-transparent clouds over land with different opacities.

Assessment

The cloud mask overlaid in black shows that there is no semi-transparent cloud remains outside of the cloud mask. Also, very thin semi-transparent clouds are detected as cloud.

Cloud Mask



Assessment

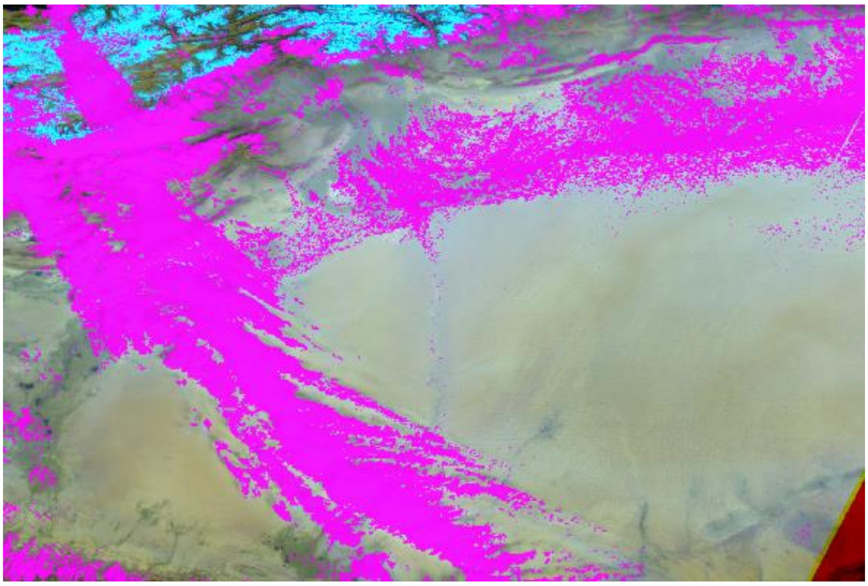
The transparent cloud mask overlay underlines that the cloud detection works for all levels of semi-transparent clouds, also the thin semi-transparent clouds.

Cloud Mask

Semi-transparent clouds over bright land - PROBAV_L2A_20140321_062703_1_333M_V101

RGB

This subset shows semi-transparent clouds over bright surfaces in high altitude with mountains in the northwest part. The mountains are partly covered by semi-transparent clouds.

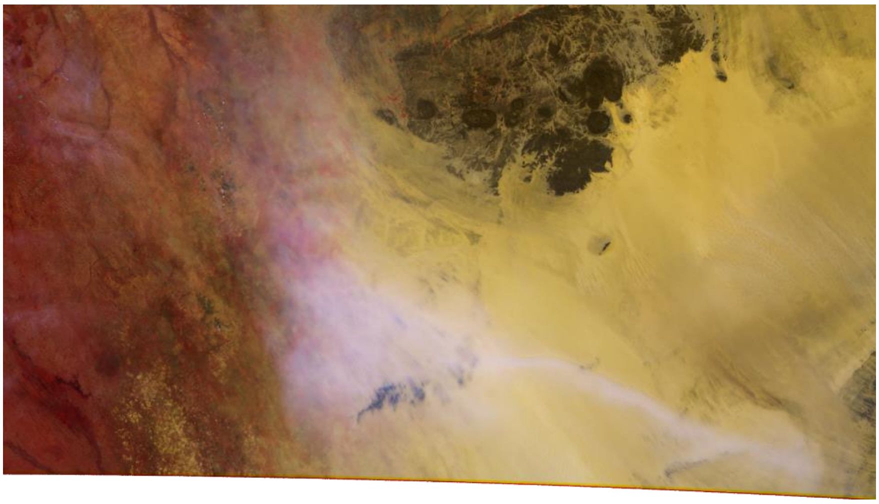


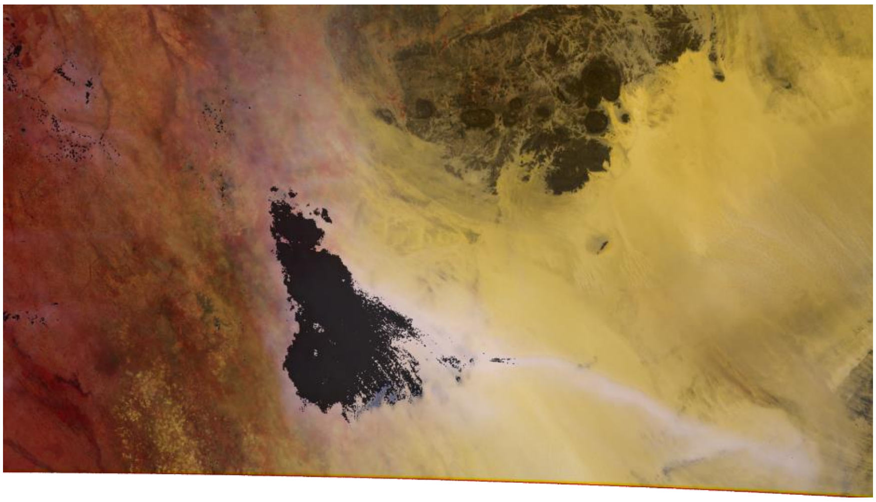
Assessment

The cloud mask in pink is covering the semi-transparent clouds in the west but is also masking turbid/dizzy atmosphere (from west to east). The differentiation between snow (cyan) and clouds is working well, some areas along the edge of the snow-covered areas are wrongly detected as cloud.

Cloud Mask

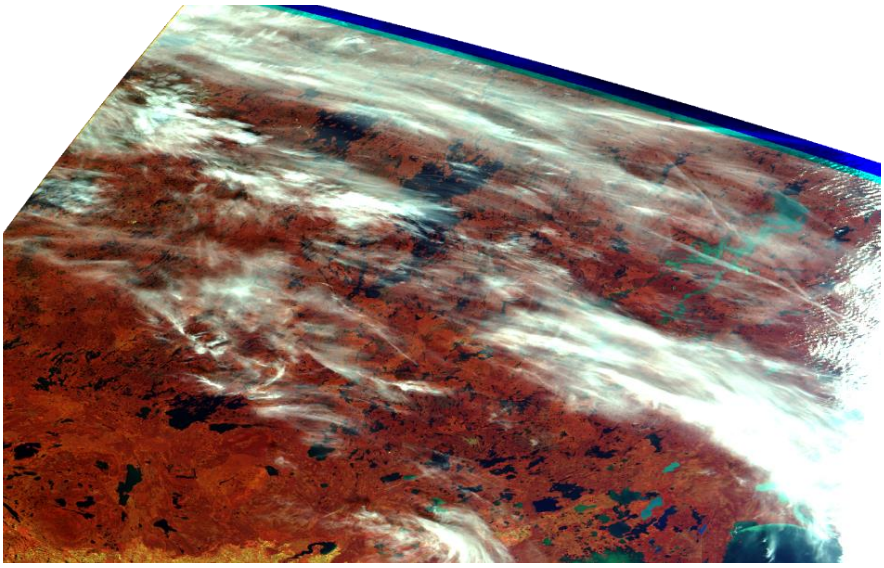
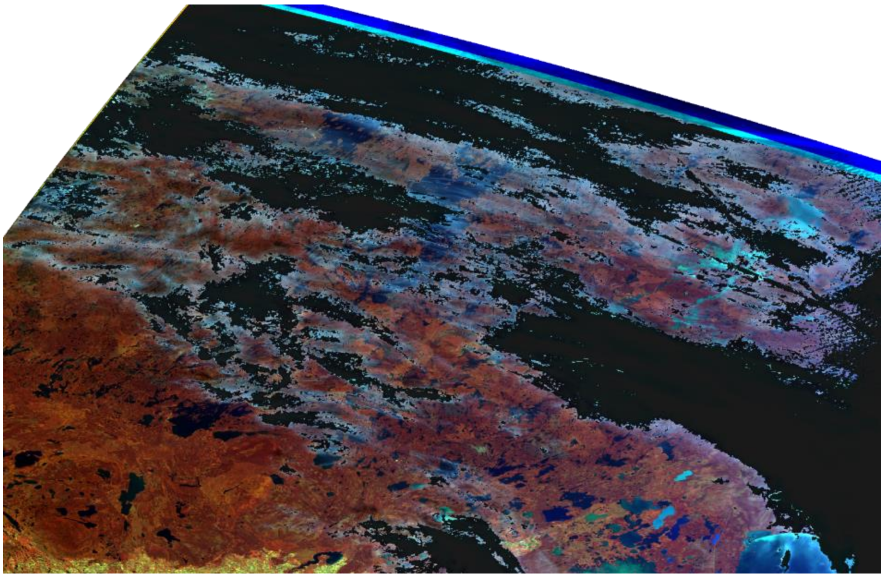
Snow Mask

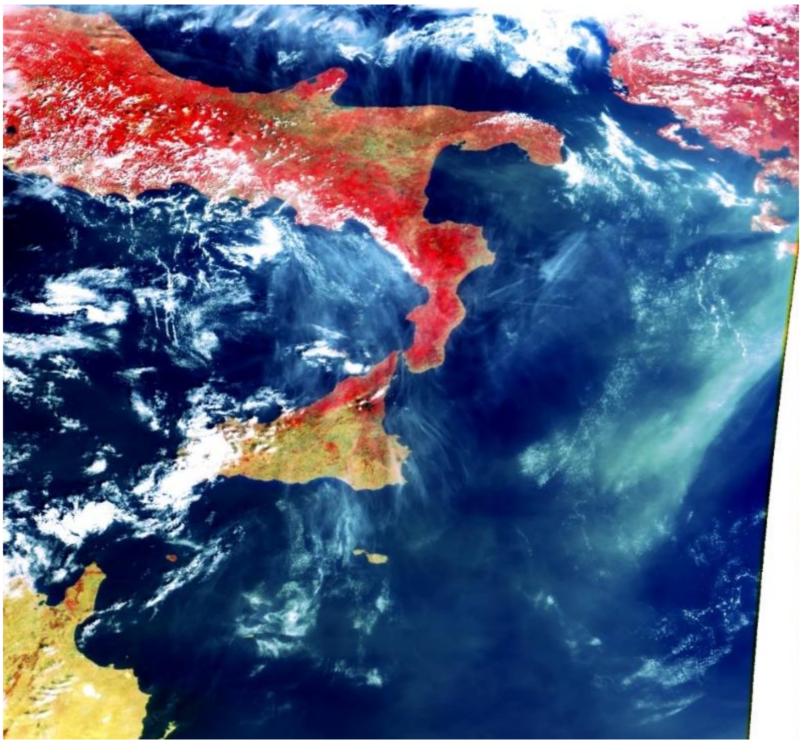
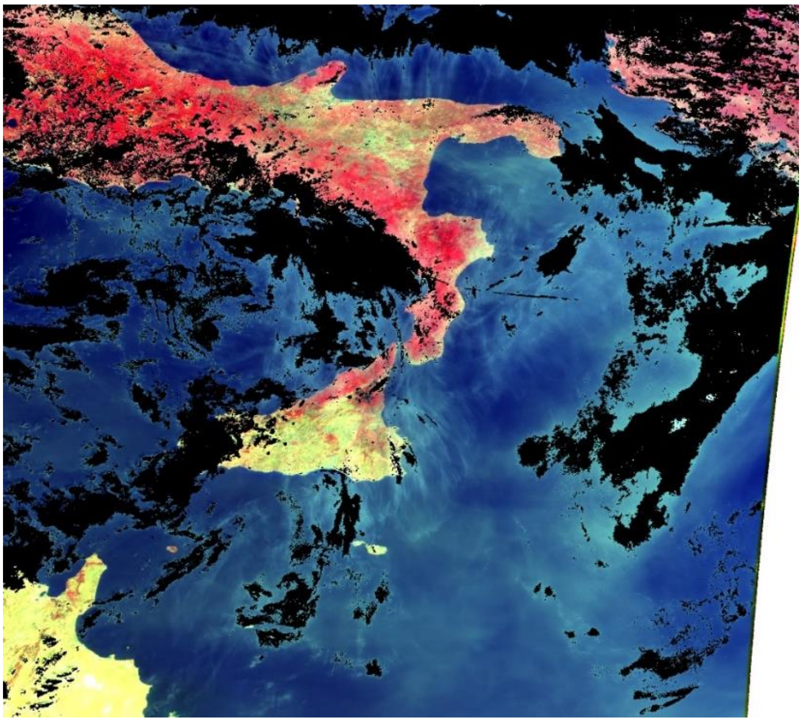
<table>
<tr><th colspan="2">Semi-transparent clouds over bright land - PROBAV_L2A_20140921_104339_1_333M_V101</th></tr>
<tr><td></td><td>RGB

This subset shows semi-transparent clouds over differently coloured surfaces (Sahara).</td></tr>
<tr><td></td><td>Assessment

The semi-transparent cloud is detected in the region where the underlying surface is darker. The continuation of the semi-transparent cloud over bright surfaces in the south-eastern part is not detected.

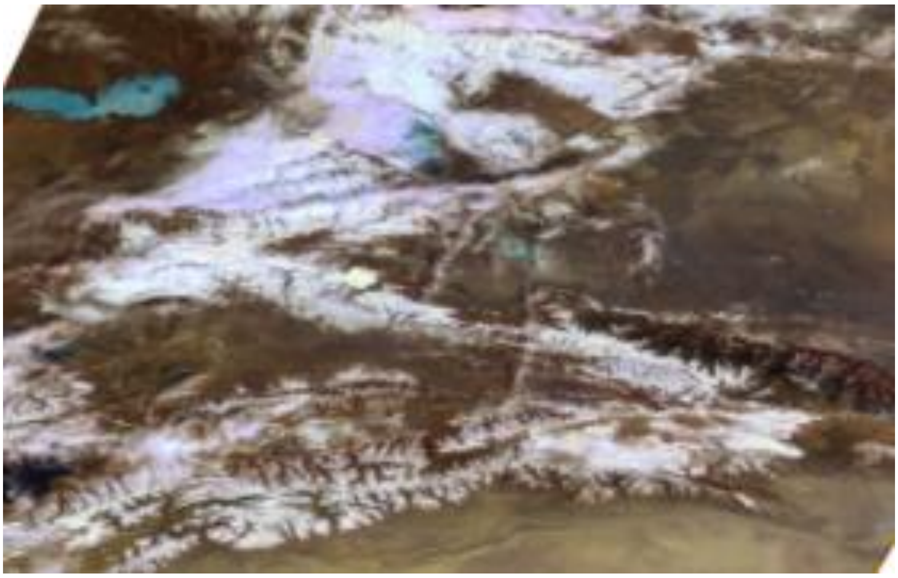
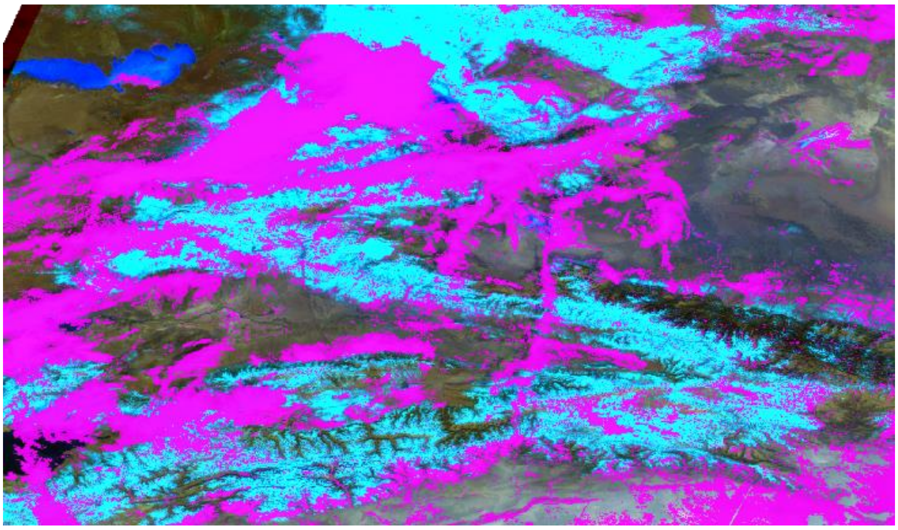
Cloud Mask</td></tr>
</table>

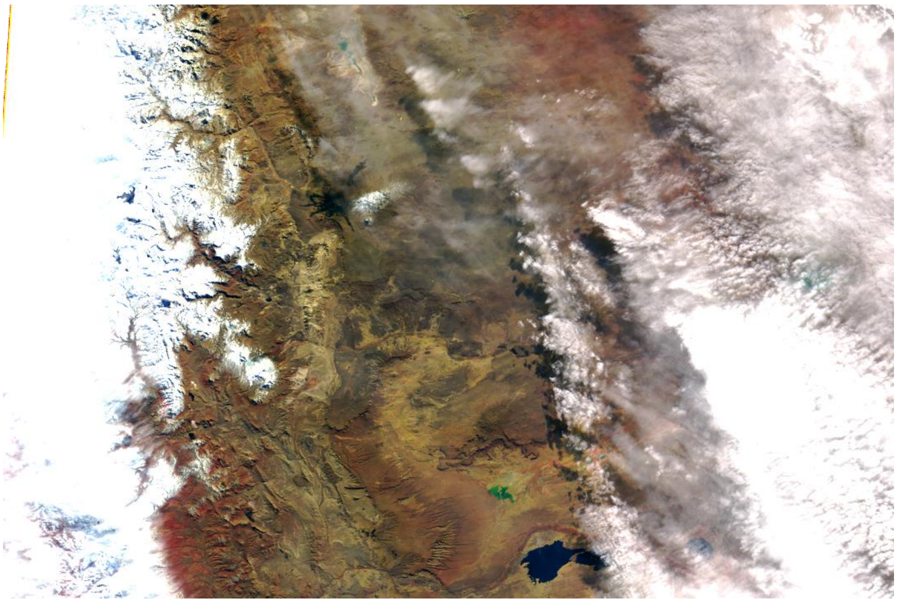
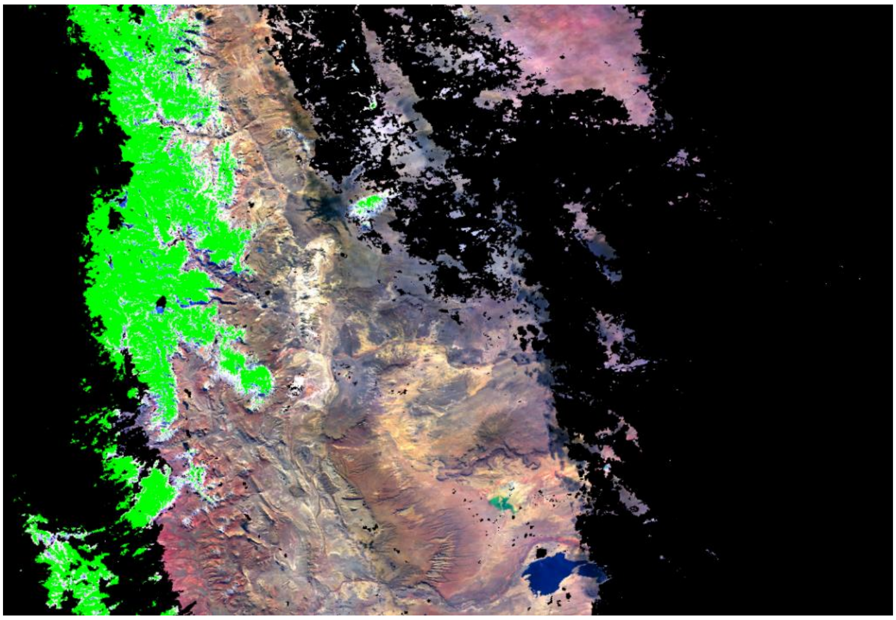
Semi-transparent clouds over land and different types of inland water bodies – PROBAV_L2A_20140921_190959_1_333M_V101	
	**RGB** The subset shows semi-transparent clouds with different opacities over land and inland waters with different water types (bright and dark water).
	**Assessment** Some semi-transparent clouds are well captured and some thin semi-transparent clouds are left out by the cloud mask over both surface types - land as well as over water. Bright water is (correctly) not flagged as cloud. Cloud Mask

Semit-transparent clouds over water - PROBAV_L2A_20140921_104339_1_333M_V101	
	**RGB** The subset shows mixed clouds over land and water.
	**Assessment** Almost all clouds are masked as cloud, some thin clouds are not masked, but the colour stretching of the image is quite bright. 

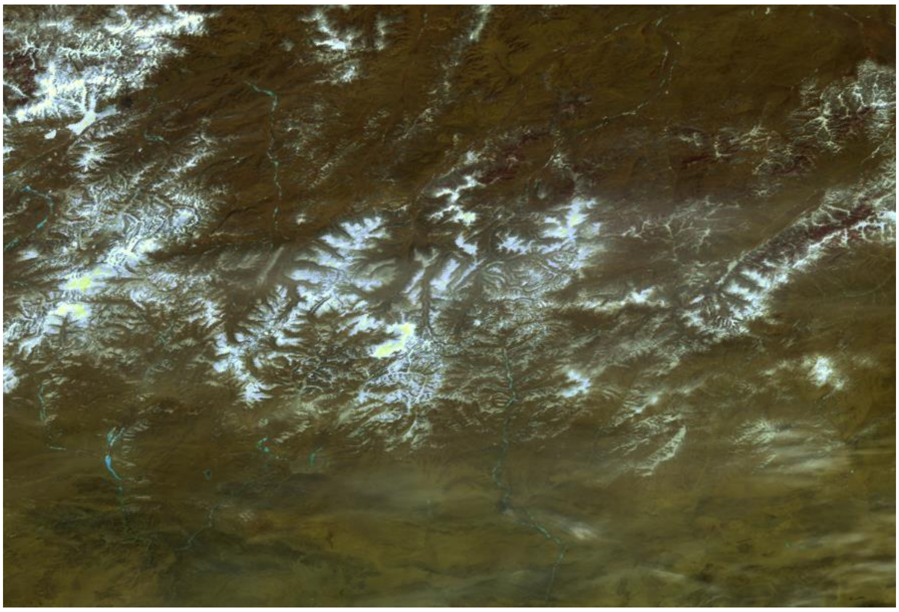
4.1.2 Combination Cloud and Snow masks

Clouds and snow surfaces are difficult to distinguish. Therefore, the two masks need to be assessed in common. The following examples show different snow areas (mountains, ice on water, high latitude areas).

Clouds and snow over land - PROBAV_L2A_20140321_080337_3_333M_V101 (check)	
	**RGB** Mixed clouds over land and mountains covered by snow. High altitudes.
	**Assessment** Snow mask and cloud mask are well distinguished. Snow borders at the edge of the mountains are classified as cloud. Cloud Mask Snow Mask

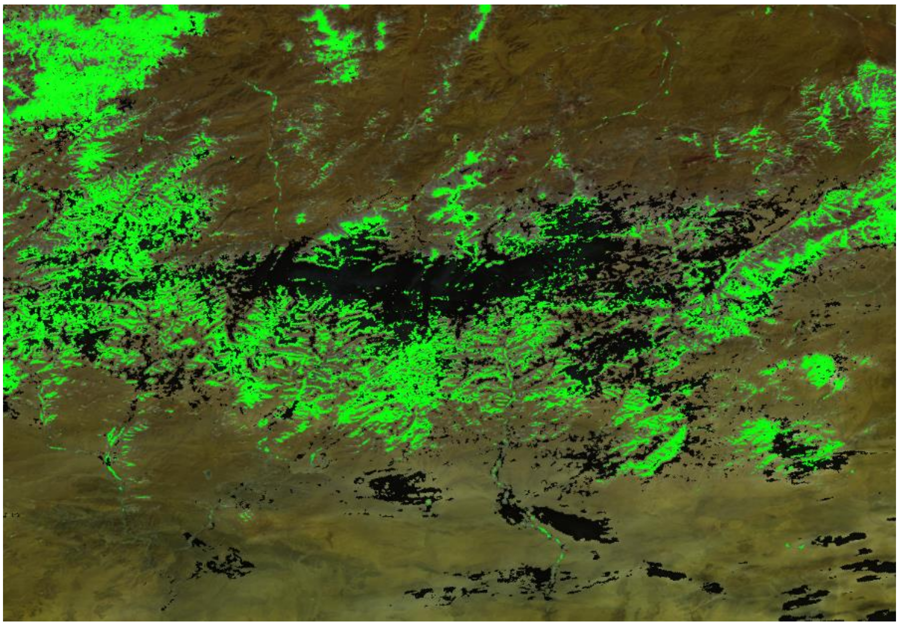
Clouds and snow over land - PROBAV_L2A_20140621_144525_3_333M_V101 (check)	
	**RGB** Mixed clouds over land and mountains covered by snow.
	**Assessment** Snow mask and cloud mask are well distinguished, no cloud mask at the border of mountains/snow in this example 

Semi-transparent clouds and snow-covered mountains - PROBAV_L2A_20140321_044122_1_333M_V101



RGB

Thin clouds over snow-covered mountains



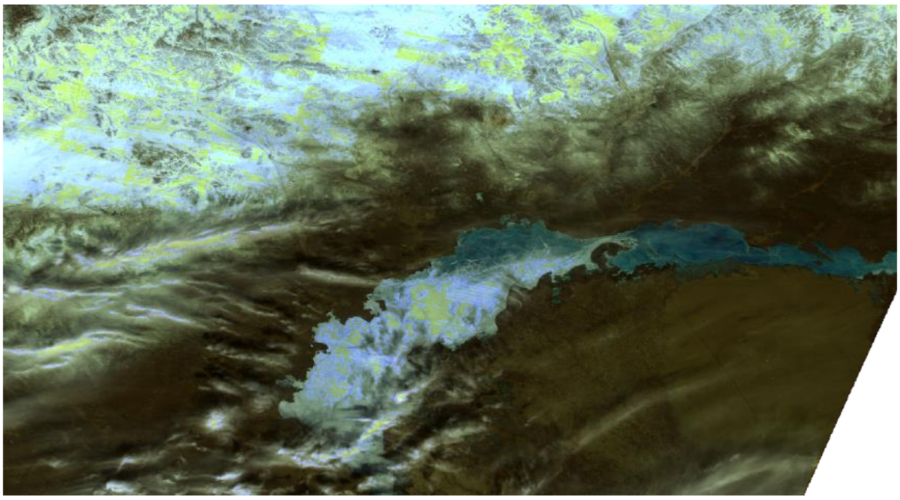
Assessment

Very thin clouds above mountains covered with snow are detected as clouds.

Cloud Mask

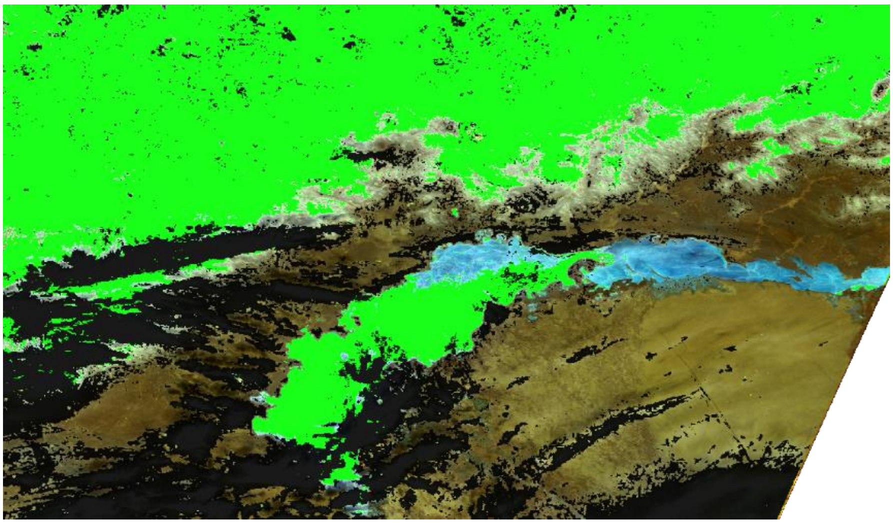
Snow Mask

Clouds and snow over ice-covered lakes - PROBAV_L2A_20140321_062242_2_333M_V101



RGB

The subset shows snow coverage over land in different densities and ice coverage of lake with different ice thicknesses.

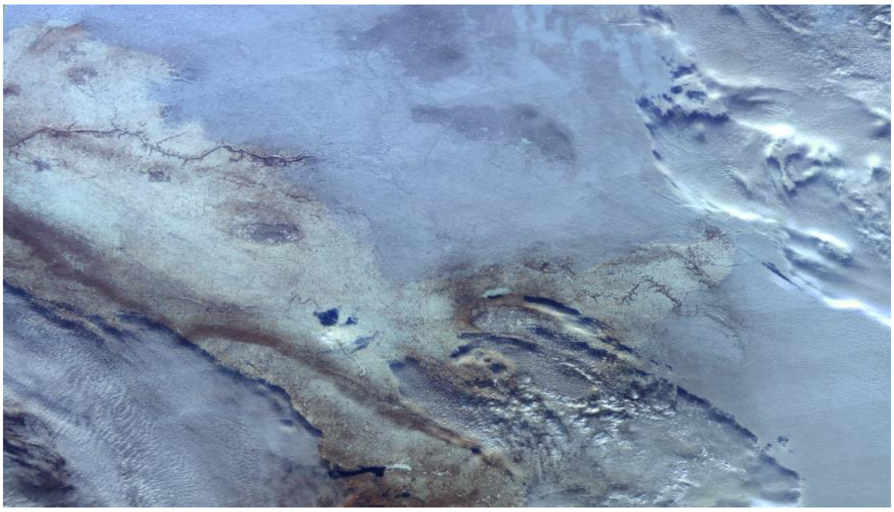
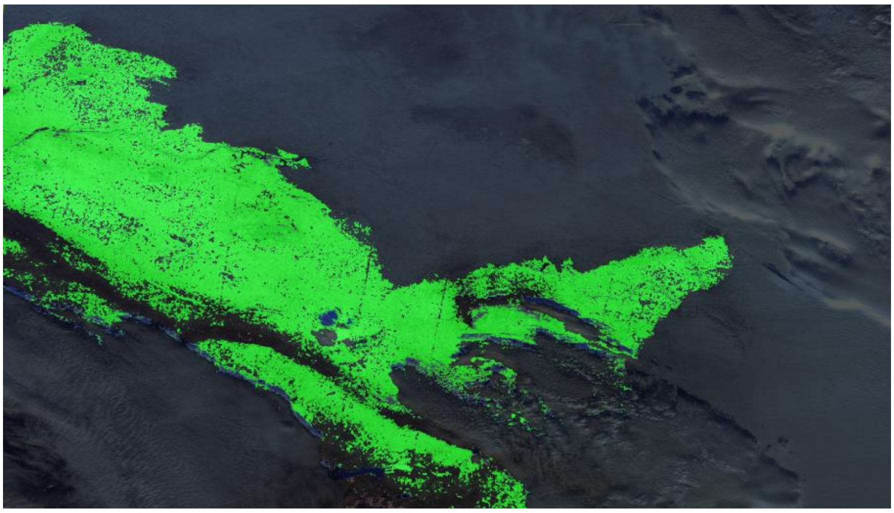
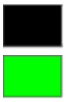


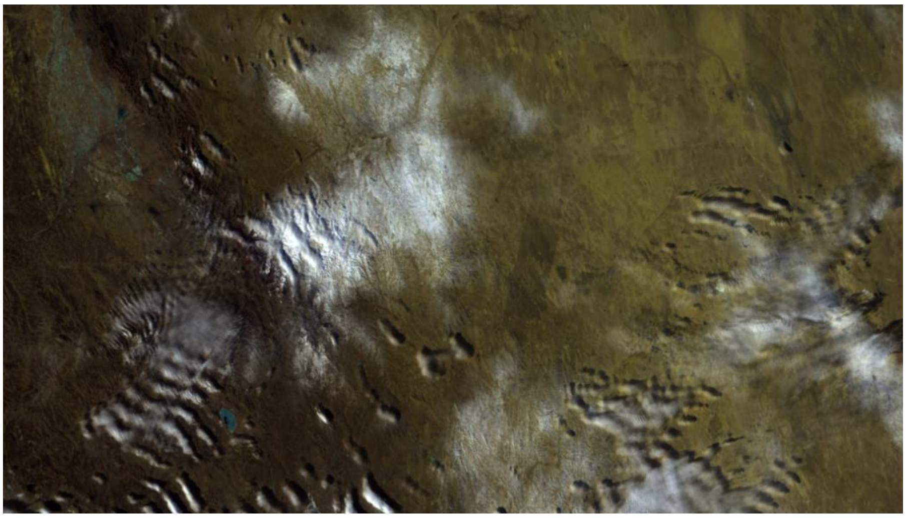
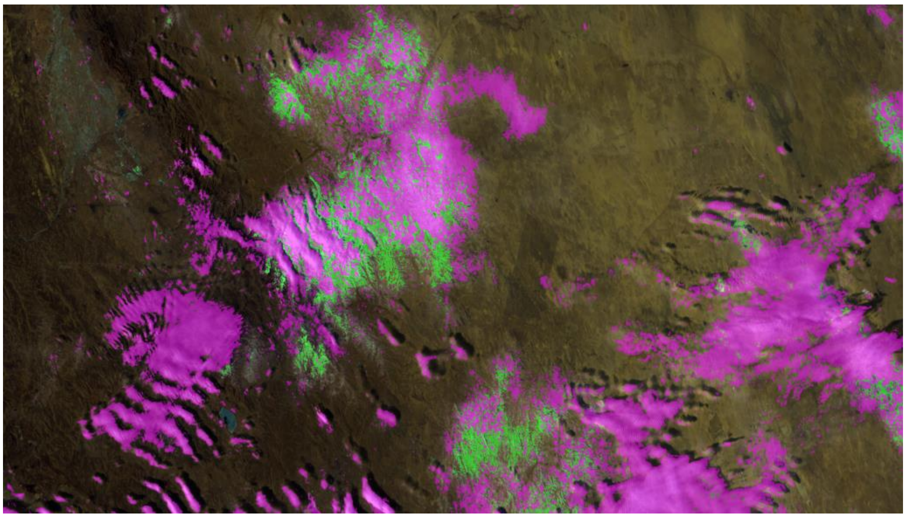
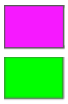
Assessment

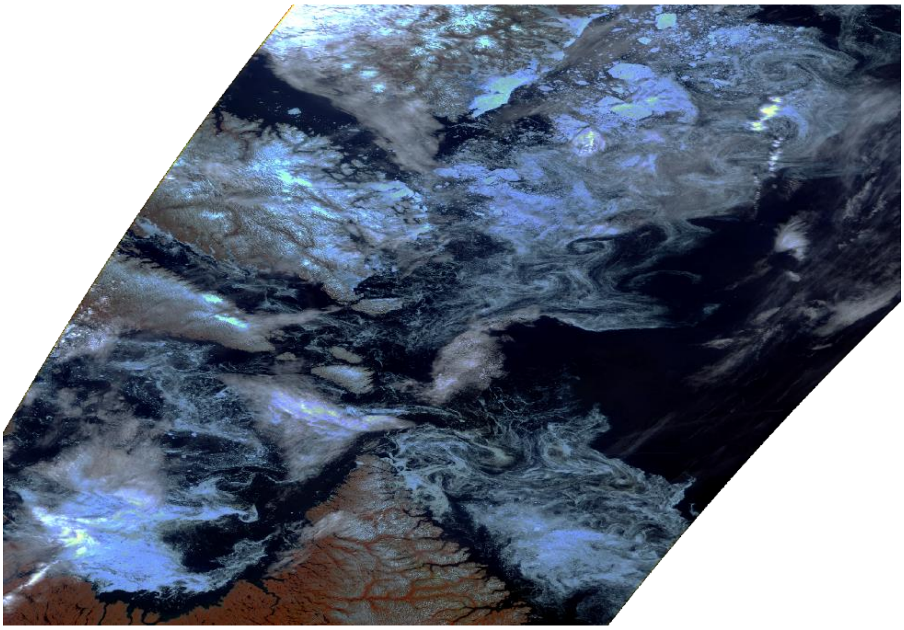
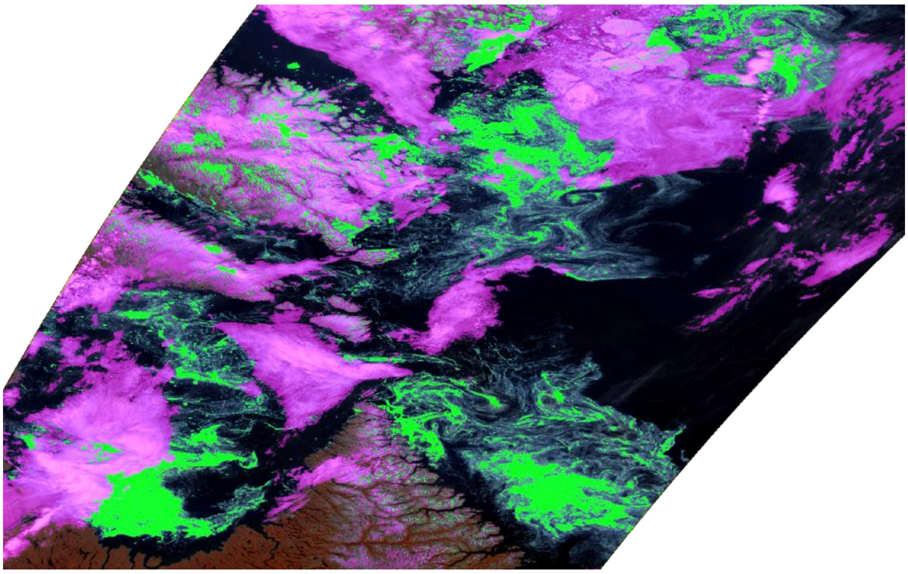
The snow-covered areas are captured by the snow mask, but not the mixed snow/land pixels. The thick ice-covered lake areas are detected. Thin ice coverage of the lake is not detected by the snow mask.

Cloud Mask

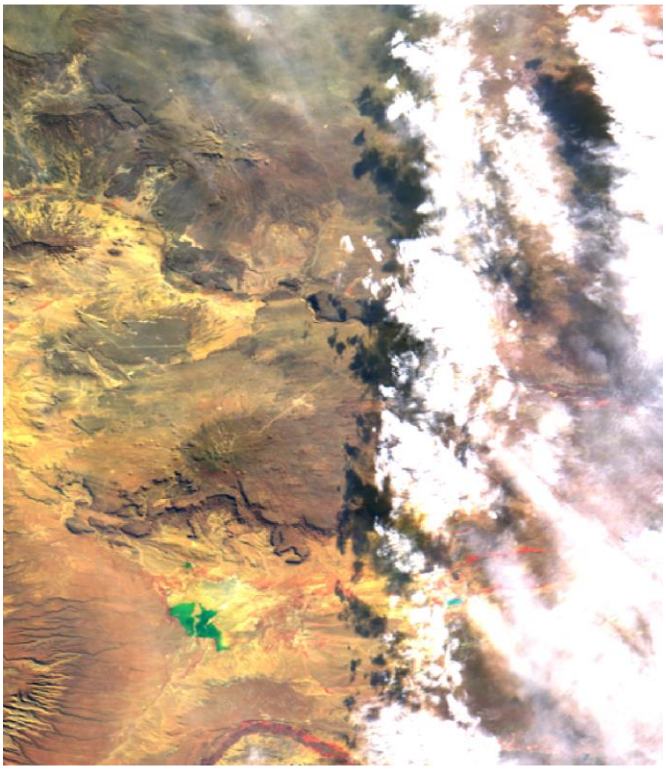
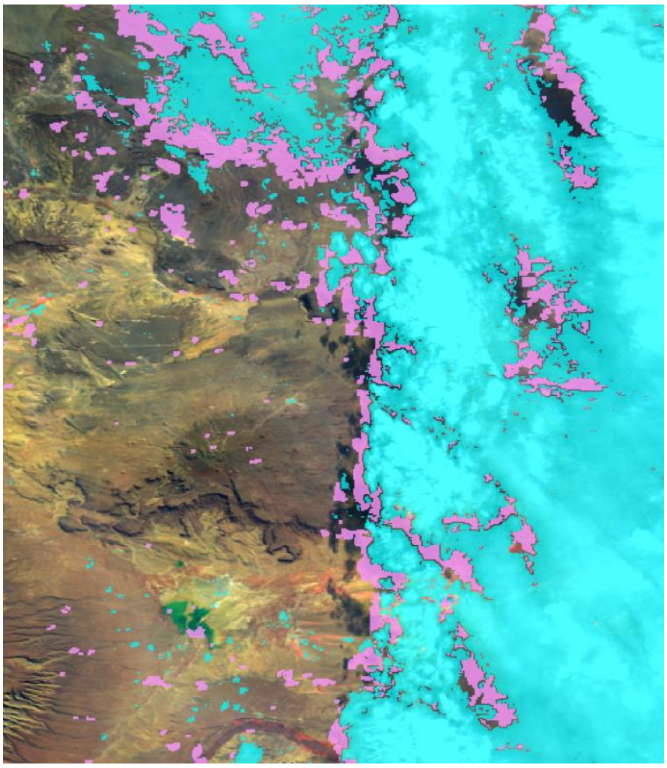
Snow Mask

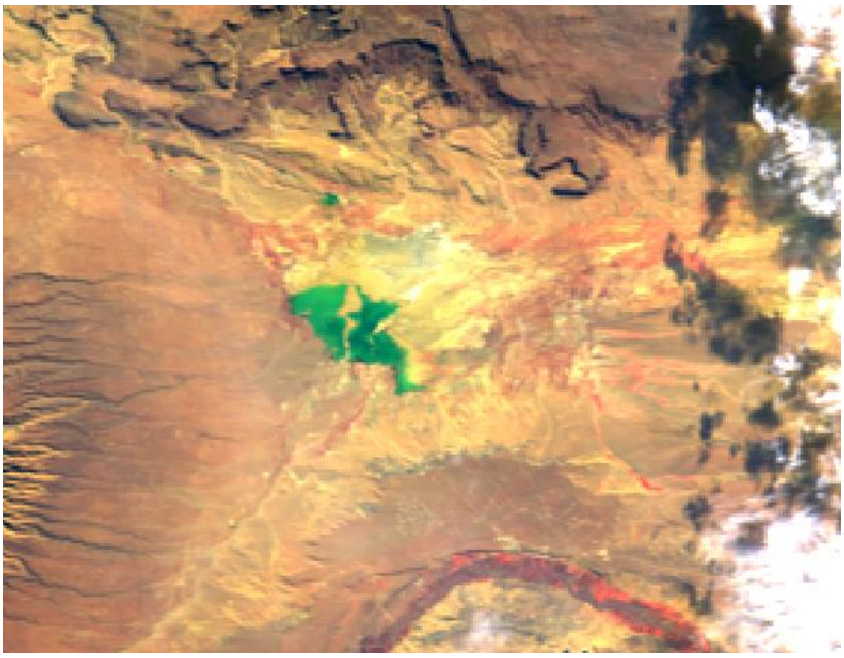
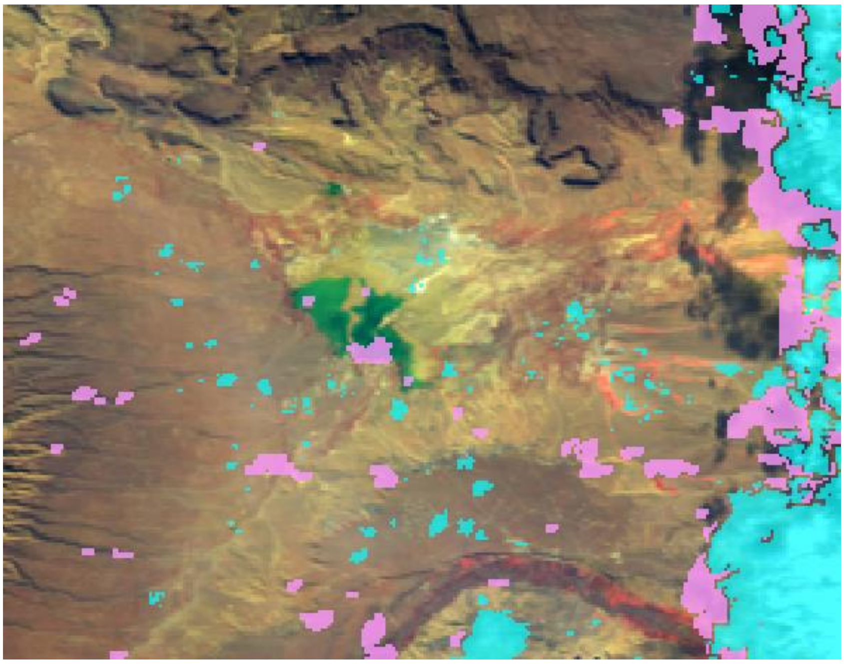
Clouds and snow over ice-covered lakes - PROBAV_L2A_20141221_172608_3_333M_V101	
	**RGB** The subset shows cloud-covered and cloud free snow areas over land. The clouds are mainly semi-transparent.
	**Assessment** The snow-covered areas are captured by the snow mask and the clouds are masked as cloud.  Cloud Mask Snow Mask

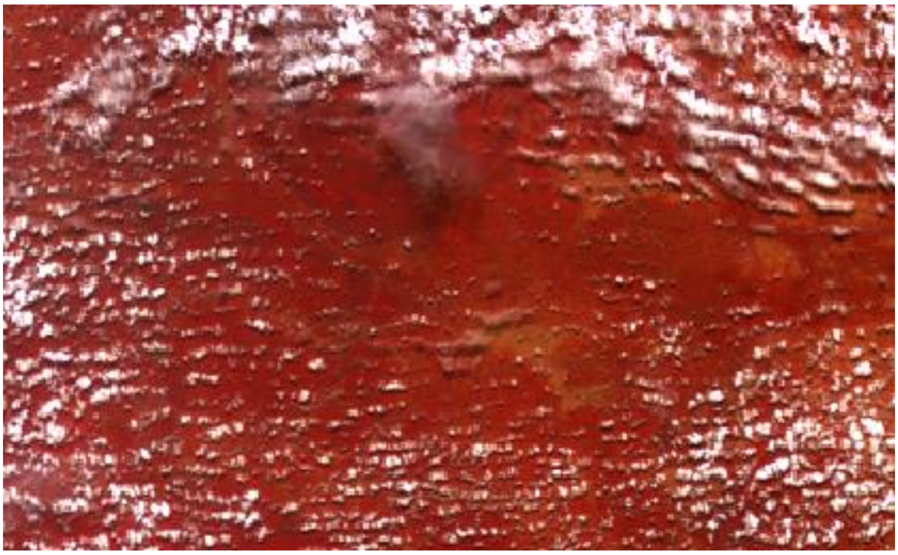
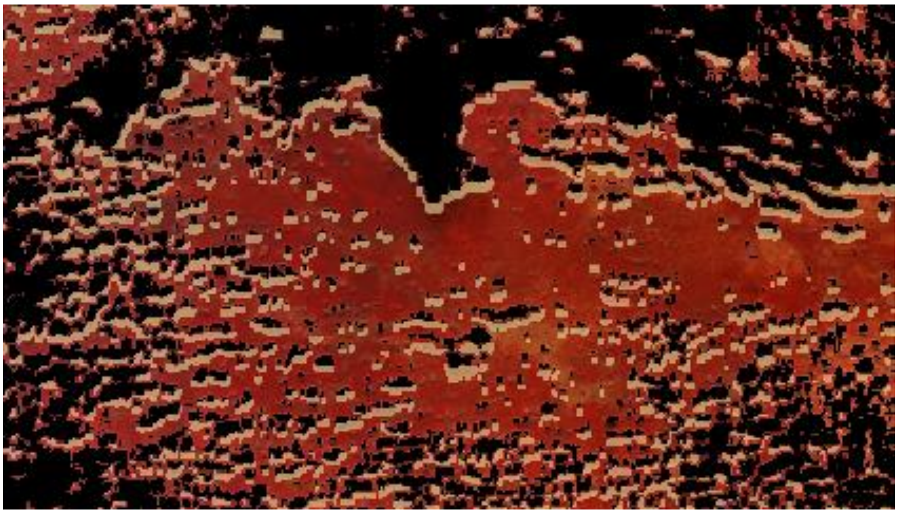
Clouds over thin snow coverage - PROBAV_L2A_20141221_035601_2_333M_V101	
	**RGB** Clouds covering thin cloud-covered surfaces.
	**Assessment** Well separation of clouds and snow. 

Clouds and snow over ice-covered lakes - PROBAV_L2A_20140621_160724_2_333M_V101	
	**RGB** The subset shows a mixture of sea-ice, melting ice, partly covered by clouds and snow on land and clouds.
	**Assessment** The thick sea-ice-covered areas are detected by the snow mask and semi-transparent clouds are masked as clouds. Sometimes this is mixed-up. The melting ice areas are not flagged by the snow mask. This mixture of surfaces is very challenging. Cloud Mask Snow Mask

4.1.3 Cloud shadow

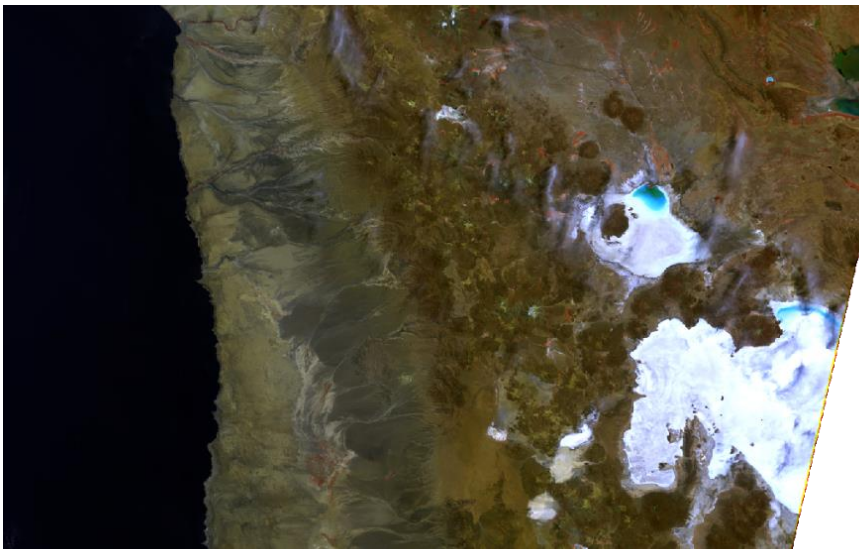
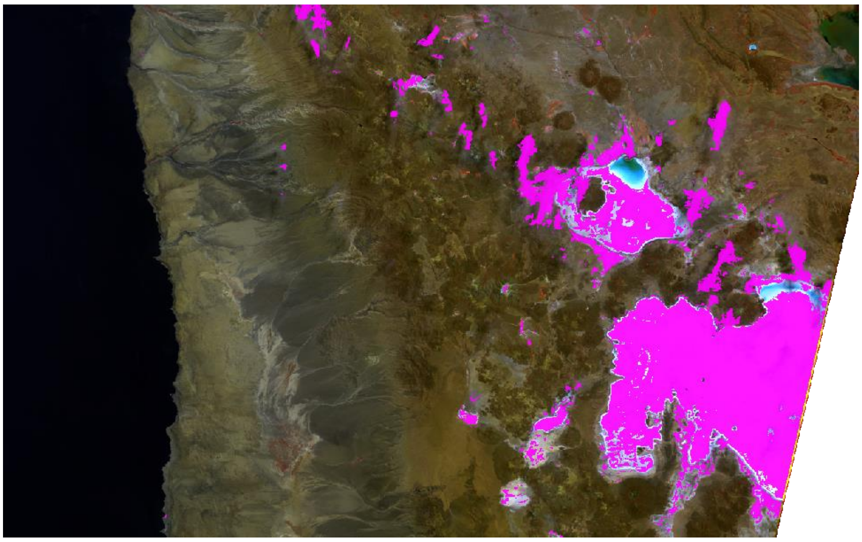
Clouds and snow over land - PROBAV_L2A_20140621_144525_2_333M_V101	
	**RGB** The subset shows mixed clouds (opaque to semi-transparent) over land and mountains covered by snow. Cloud shadows are clearly visible for the opaque clouds and some cloud shadows are covered by semi-transparent clouds.
	The cloud mask is capturing all kind of clouds (plus some clear sky areas). The cloud shadow mask is underestimating the cloud shadow areas. Semi-transparent clouds in the northern part are flagged as cloud and related cloud shadow is not visible in the images but masked by the cloud shadow mask. Small spots in the south west of the subset are masked as cloud and trigger the cloud shadow mask. No clouds and no cloud shadows related to those spots. A zoom is shown in the following figure. Cloud Mask Cloud shadow Mask

Clouds and snow over land - PROBAV_L2A_20140621_144525_2_333M_V101	
	**RGB** Zoom to the cloud flagged spots from example above which are clear sky surfaces.
	**Assessment** Both effects – cloud shadow underestimated for opaque clouds and the flagging of clouds and cloud shadow for clear sky surfaces are clearly visible in the zoomed subset. Cloud Mask Cloud shadow Mask

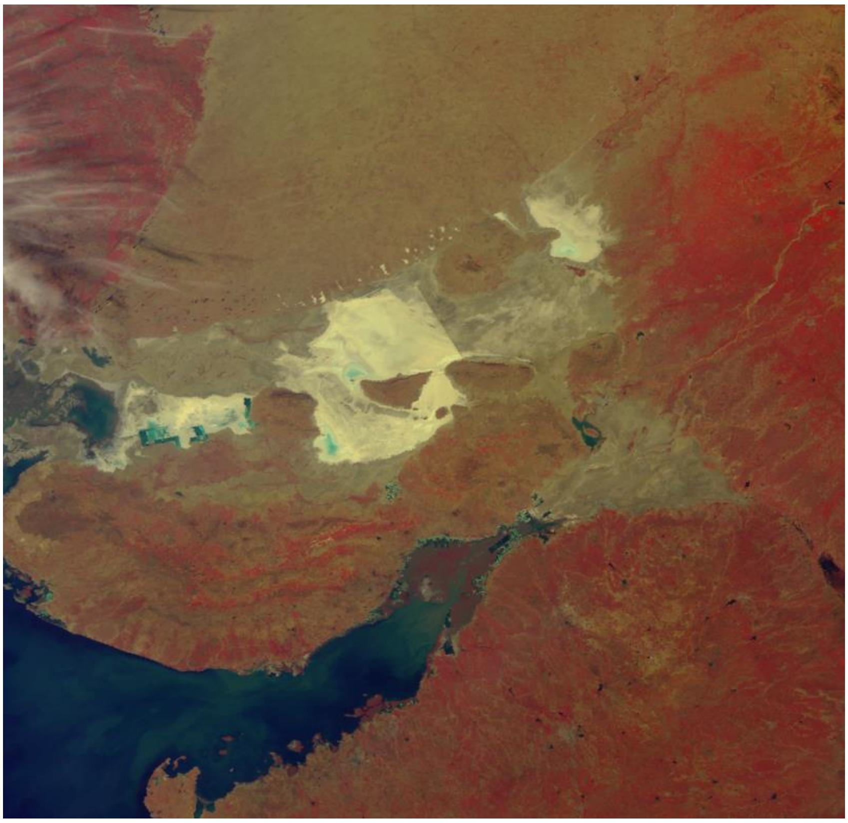
Cloud shadow of cumulus clouds over land PROBAV_L2A_20140321_080337_3_333M_V101	
	**RGB** Cumulus clouds over land and overlaying higher semi-transparent cloud
	**Assessment** The cloud mask is capturing all small cumulus clouds over land and cloud shadow is at right position. Only for the higher clouds, the cloud shadow mask is not fully covered. Cloud Mask Cloud shadow Mask

4.1.4 Special Cases

Some surface types are known to be challenging for cloud detection algorithms. Here, urban areas, salt lakes, sun glint influenced water or bright surfaces such as beaches or deserts are closer assessed. The following images show some examples for these surfaces.

Salt lakes – PROBAV_L2A_20140321_150528_2_333M_V101	
	**RGB** The subset shows salt lakes, which are partly covered by semi-transparent clouds. These lakes are located in Bolivia (salar de Coipasa, salar de Uyuni, laguna Parihuana)
	**Assessment** The semi-transparent clouds are well detected by the cloud mask, but also the cloud-free parts of the salt lakes are detected as cloud. Cloud Mask

Salt lakes - PROBAV_L2A_20141221_053718_3_333M_V101

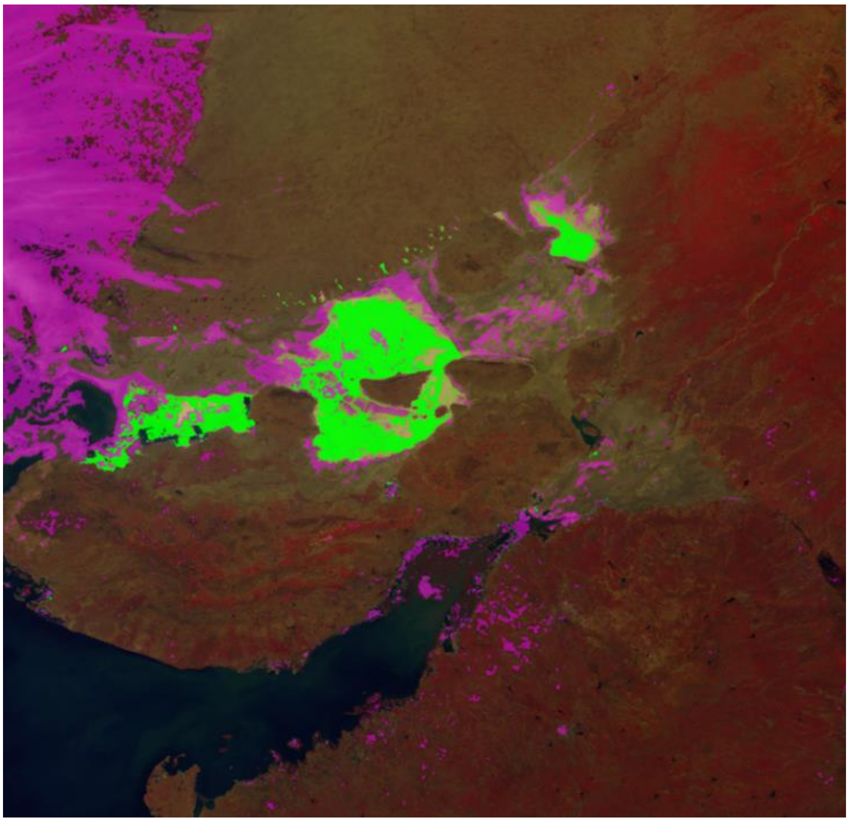


RGB

The subset shows dried out salt lakes, some semi-transparent clouds in the western part.

The bay shows dry fallen inter-tidal areas.

This lake is located in India Rann von Kachchh



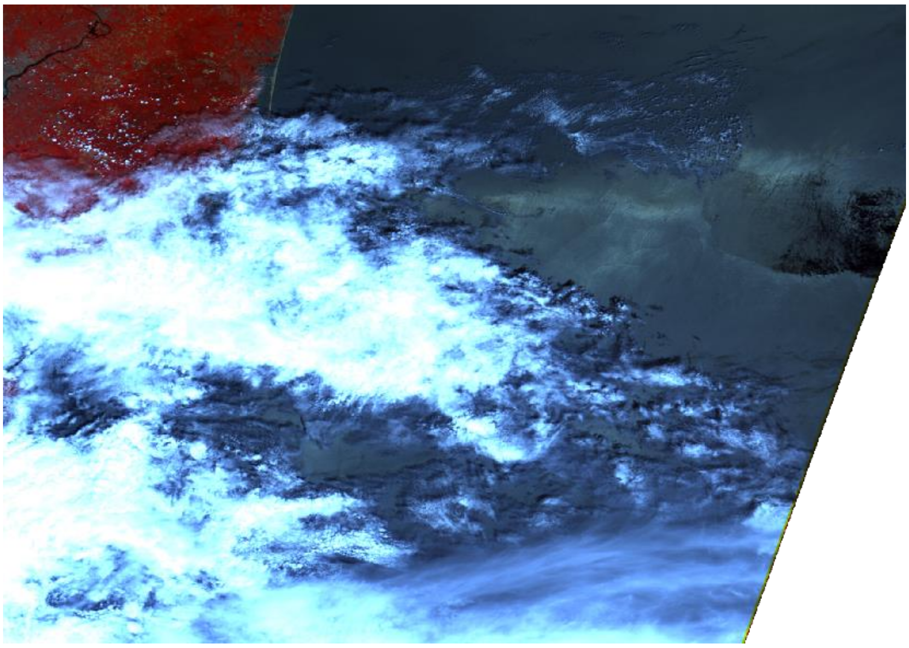
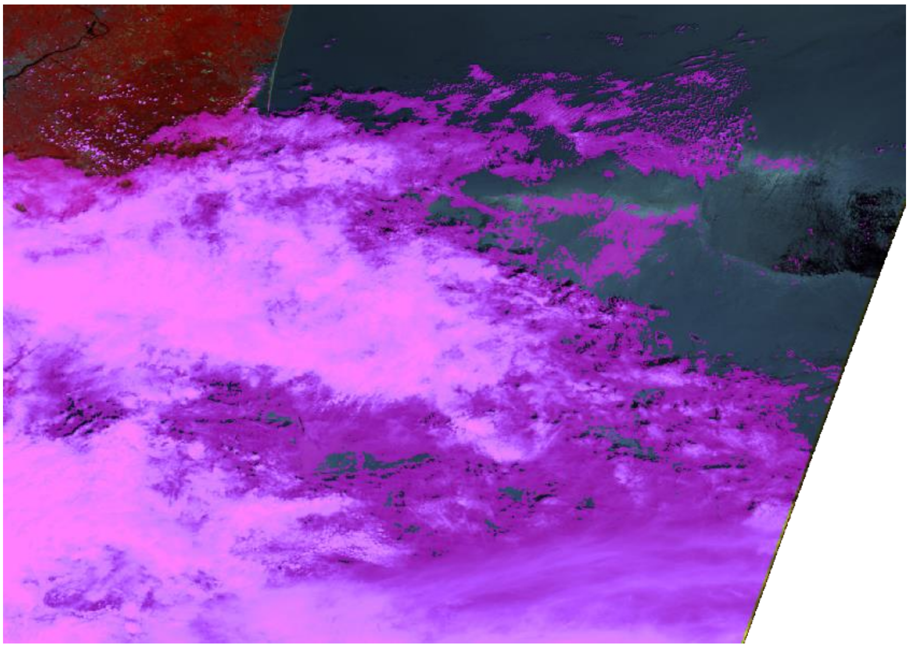
Assessment

The semi-transparent clouds are well detected by the cloud mask, but the cloud mask also detects some parts of the bay that are not cloud-covered, rather dry fallen mud flats.

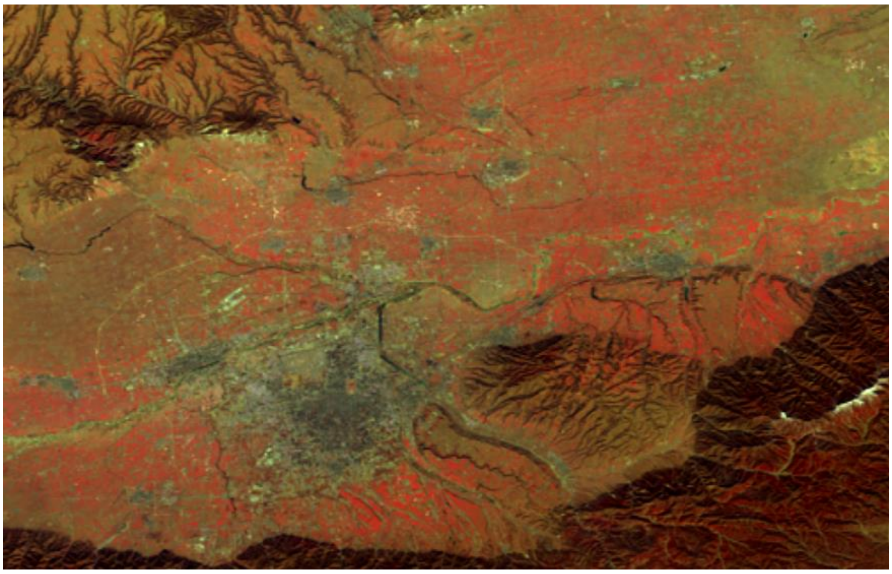
The salt lake is detected as snow.

Cloud Mask

Snow Mask

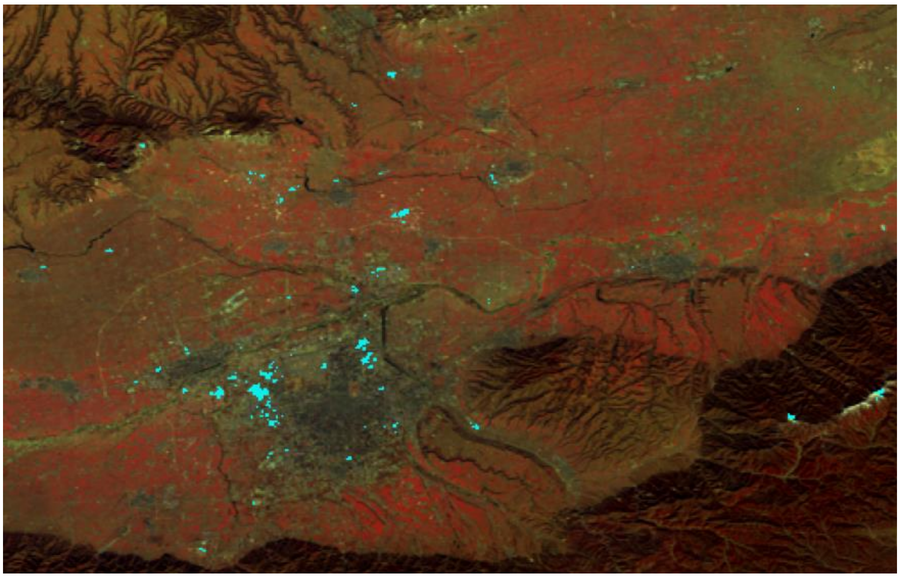
Sun glint affected water - PROBAV_L2A_20140621_160724_2_333M_V101	
	**RGB** Different cloud types over water, effected by sun glint
	**Assessment** The clouds are all detected well by the cloud mask but also some parts of the sun glint are detected as cloud.  Cloud Mask

Urban areas – PROBAV_L2A_20141221_035601_2_333M_V101



RGB

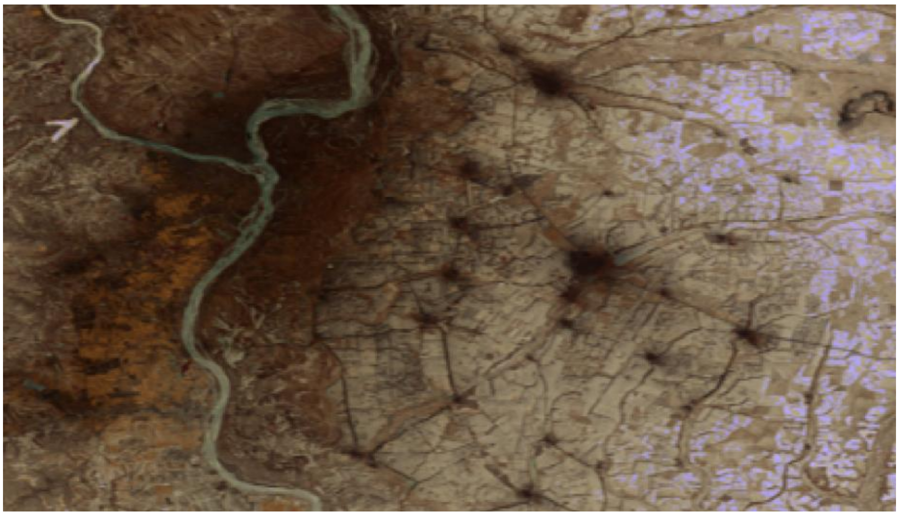
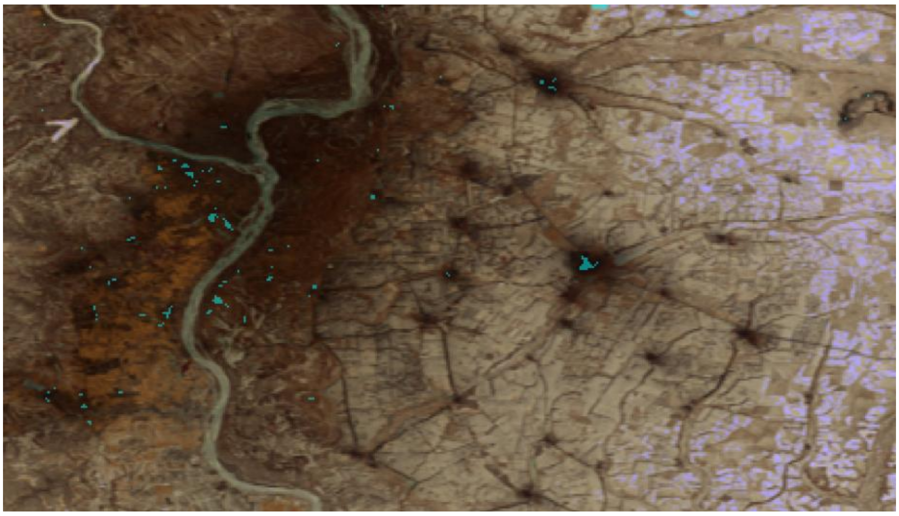
The subset shows urban areas (Xi'an) under clear sky conditions.



Assessment

Parts of the urban areas are flagged as clouds. Those are not necessarily the brightest pixels in the cities.

Cloud Mask

Urban areas in winter – PROBAV_L2A_20140321_030009_2_333M_V101	
	**RGB** The subset shows urban areas in winter; while the surrounding is covered by snow, the urban areas are snow-free due to warmer temperatures. This subset is located at the border between Russia and China in the surrounding of Blagoweshchensk.
	**Assessment** In these special cases (wintertime) it occurs that the urban areas are detected as cloud. Cloud Mask

4.1.5 Cloud detection and PROBA-V FLAGS

The following examples show the dependency of the cloud detection of the (GOOD) flags provided by ProbaV. Figure 8 shows that no cloud is detected if the GOOD SWIR flag is not set in the images. This often occurs at the end of orbits.

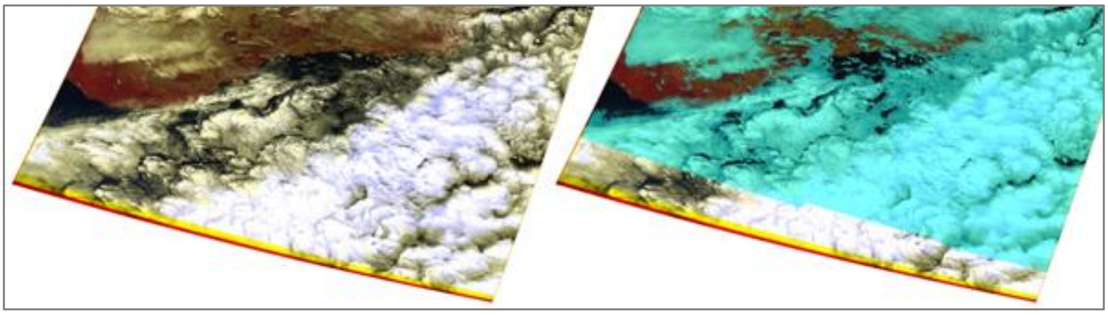


Figure 8: Influence of GOOD SWIR flag

It can occur that bright surfaces (e.g. snow/ice) which are not flagged by the GOOD BLUE flag, are masked as cloud. This is shown in Figure 9. While the GOOD BLUE areas are correctly detected as snow/ice, the large snow-covered area not flagged as GOOD BLUE is detected as cloud.

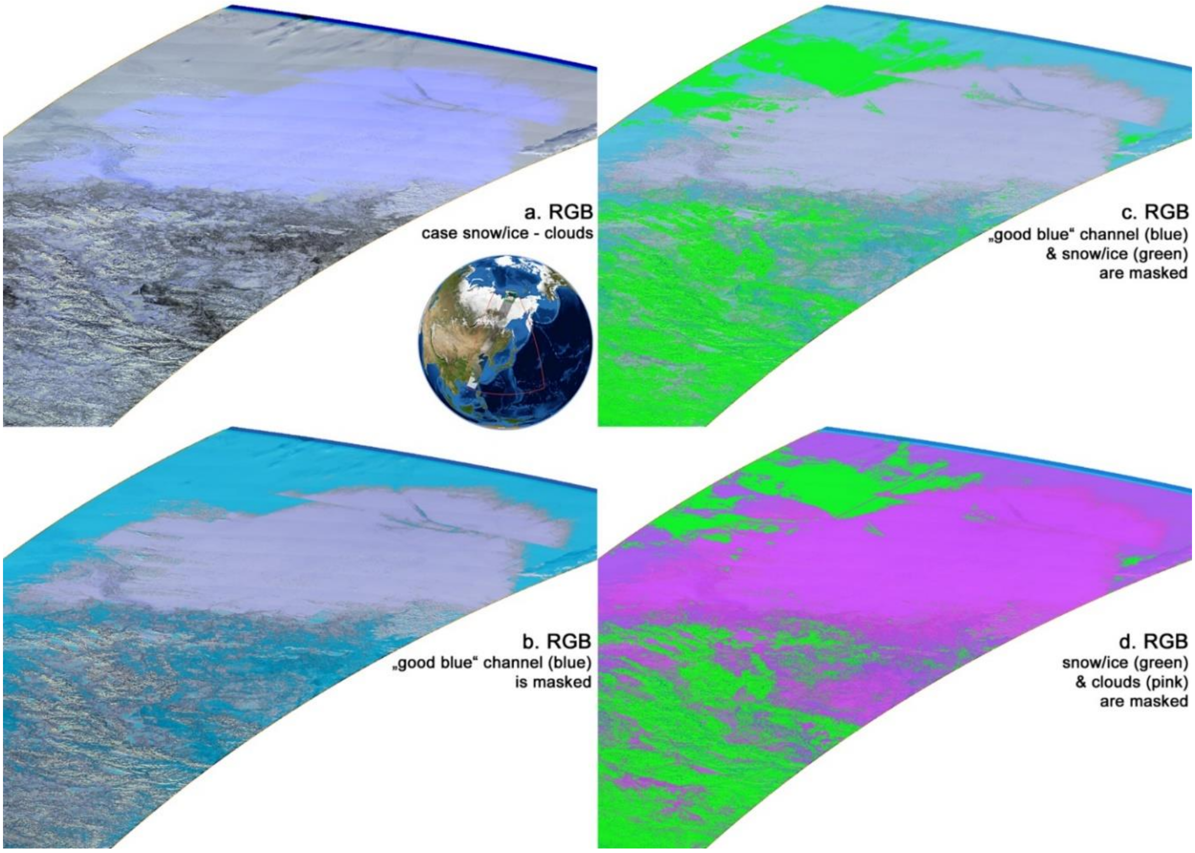


Figure 9: Influence of GOOD BLUE flag on the detection of clouds over snow areas. (a) RGB, (b) GOOD BLUE not set over large ice-covered area, (c) snow/ice flag is raised over snow-covered areas, (d) the area not flagged as GOOD BLUE is detected as cloud, though it is also cloud free snow/ice. PROBA-V product: PROBAV_L2A_20140321_030009_2_333M_V101.

4.1.6 Summary Assessment

The visual inspection certifies the cloud and snow detection an overall very good performance. In order to provide an overview of the performance for the cloud detection over different surfaces, the assessment has been translated into three ranking categories for the different cloud and surface types (cloud mask) as well as to different snow surface types (snow mask). The ranking contains the categories

- very good (++) – misclassification occurs only seldom
- good (+) - sometimes misclassification occurs
- medium (O) - systematic misclassification occurs

Table 1: Visual Cloud Mask Assessment for different cloud types

Cloud types	Underlying surface types	Ranking
Opaque clouds		++
Small cumulus clouds	over land	++
	over water	++
	over sun glint	O
Semi-transparent clouds	Thick	++
	Middle	++
	thin	+

Table 2: Visual Cloud Mask Assessment for clouds over different surfaces types

Clouds detection over different surfaces	Ranking
Coastal and open waters	++
general land surfaces	++
Inland waters	++
Dessert	+
Snow / ice	+
Sea ice	+
Urban areas	O
Salt lakes	O

Table 3: Visual Snow Mask Assessment for different snow/ice surface types

Snow inland/water	Snow/ice type	Ranking
Inland snow / ice (cloud free)	Closed snow/ice coverage	++
	Spatially mixed snow/ice and land	O
	Ice on inland waters (thick)	++
	Ice on inland waters (thin)	+
Sea Ice (cloud free)	Closed ice coverage	++
	Spatially mixed ice/water	O

The assessment of the CLOUD SHADOW mask can be summarized as medium quality, as it often underestimates the cloud height and therefore the mask is not covering all parts of the shadow area.

4.2 Validation with PixBox reference dataset

The Pixbox data set is used for statistical analysis of the cloud mask results. A confusion matrix has been generated for clear and cloudy surfaces from the cloud classification and the Pixbox reference data set. The confusion matrix in Figure 10 shows that 92.3% of the cloudy pixels in the images are correctly flagged as clouds (PA Cloud), while 91.9% of the clear pixels are correctly not flagged as cloud (PA Clear). From the other side, if the cloud detection algorithm is providing a CLOUD mask, 93.3% have been identified as cloud in the validation data set (UA Cloud), and 90.7% of the not flagged pixels are clear surfaces in the validation data set (UA Clear). The omission error of a cloud pixel is 7.7%. The overall accuracy reaches 92.12%. Krippendorfs alpha results in 0.84.

SM_FLAGS.CLOUD

Class	Clear (4,5,6,13, 15)	Cloud (2,16,17,18)	Sum	U A	E
No (0)	16156	1659	17815	90.7	9.3
Yes (1)	1424	19977	21401	93.3	6.7
Sum	17580	21636	39216		
P A	91.9	92.3		OAA:	92.14
E	8.1	7.7			

Figure 10: Confusion matrix for clear and cloudy pixels. The numbers of different categories are explained in Table 4; PA = Producer's Accuracy, UA = Producers Accuracy

The confusion matrix shows different clear sky surfaces contributing to the Clear category and different cloud types contributing to the cloud category. The explanation of different clear sky surfaces and cloud types is given in Table 4.

Table 4: Legend of cloud and clear sky categories

Cloudy cases		**Clear sky cases**	
2	totally cloudy	4	Clear sky water
16	thick semi-transparent cloud	5	Clear sky land
17	medium semi-transparent cloud	6	clear sky snow_ice
18	thin semi-transparent cloud	13	spatially mixed snow_ice/land
		15	spatially mixed land/water

More details for the single cloud types and clear surface types is provided in Table 5 and illustrated in pairs of bar chats in Figure 11, Figure 12 and Figure 15, Figure 16. Here, the first bar chart shows the absolute cases for the different categories (Figure 11, Figure 15) and indicates how many pixels were flagged as CLOUD (white part of the bars). The related second figure shows the relative distribution of CLOUD masked pixels and CLEAR pixels in the different categories (Figure 12, Figure 16). While Figure 11 and Figure 12 show the results for cloudy or cloud influenced categories, Figure 15 and Figure 16 show the results for clear sky categories.

Table 5: Distribution of CLOUD and CLEAR masked pixels in different surface categories

Unique category	**Group**	**CLEAR**	**CLOUD**	**Sum**	**% (clouds)**
Totally Cloudy	Totally cl	168	9916	10084	98.3
Thick semi-transparent cloud	Semi cl	20	1975	1995	99.0
Average density semi-transparent cloud	Semi cl	223	6163	6386	96.5
Thin semi-transparent cloud	Semi cl	1248	1923	3171	60.6
Spatially mixed cloud/land	Mixed cl	297	1609	1906	84.4
Spatially mixed cloud/water	Mixed cl	299	467	766	61.0
Clear sky land	Clear	9133	239	9372	2.6
Clear sky water	Clear	4216	173	4389	3.9
Clear sky snow_ice	Clear	2293	945	3238	29.2
Spatially m14ixed snow_ice/land	Clear	402	65	467	13.9
Spatially mixed snow_ice/water	Clear	475	34	509	6.7
Spatially mixed land/water	Clear	112	2	114	1.8

Totally cloudy pixels and thick semi-transparent pixels are masked as cloud in more than 98% of all cases, averaged density semi-transparent clouds still in 96.5% of all cases. For thin semi-transparent clouds, the CLOUD mask was raised in 60% of the cases. This result reflects very well the ambiguity of this category. Thin semi-transparent clouds can be almost invisible and should not be masked as clouds. For the spatially mixed cases, which are mainly cloud borders, the detection seems to perform better over land areas (84%) compared to water surfaces (61%).

It has been further assessed, if the detection of the thin-semi-transparent clouds is depending on the latitude or the sun geometry. . The diagrams shown in Figure 13 and Figure 14 do not indicate a dependence between the detection of thin-semi-transparent clouds and the latitude or the sun geometry.

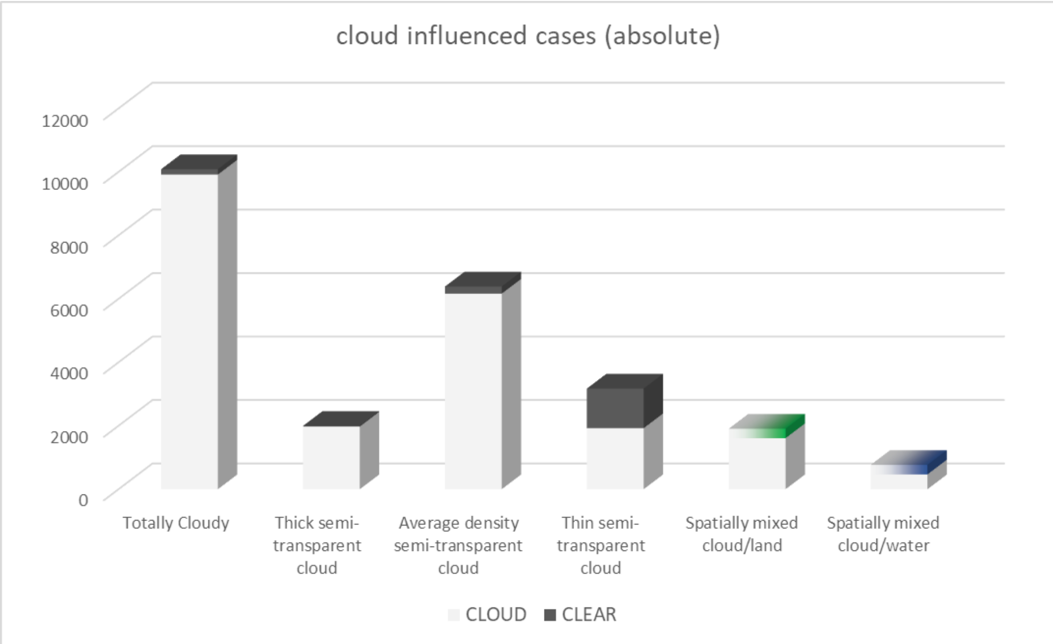


Figure 11: Absolute numbers of CLEAR and CLOUD masked pixels for different cloud types

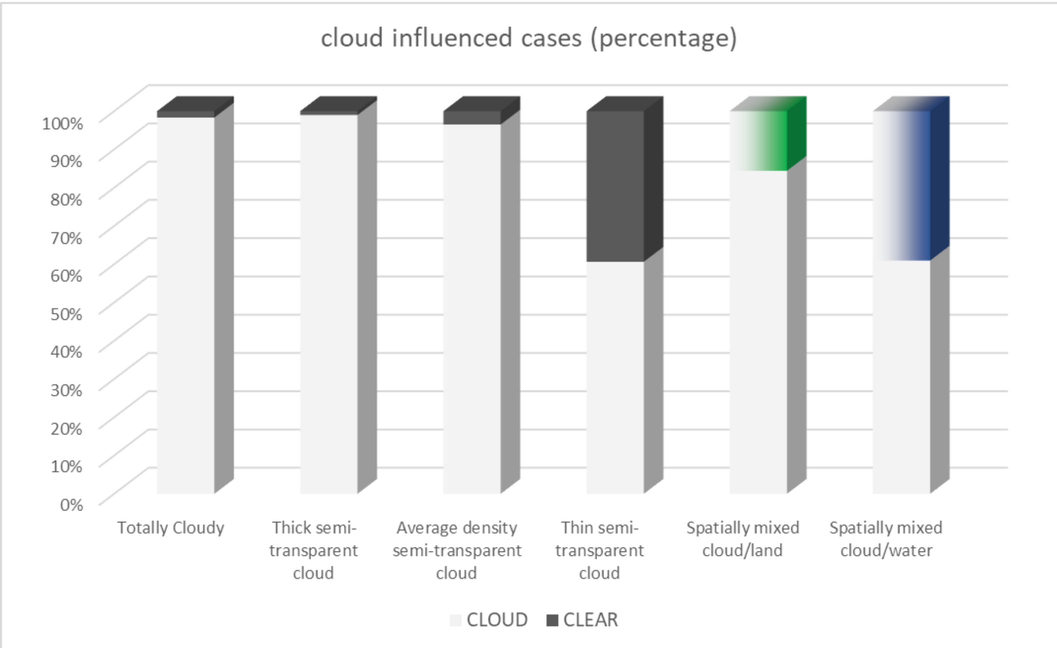


Figure 12: Relationship between CLEAR and CLOUD masked pixels for different cloud types

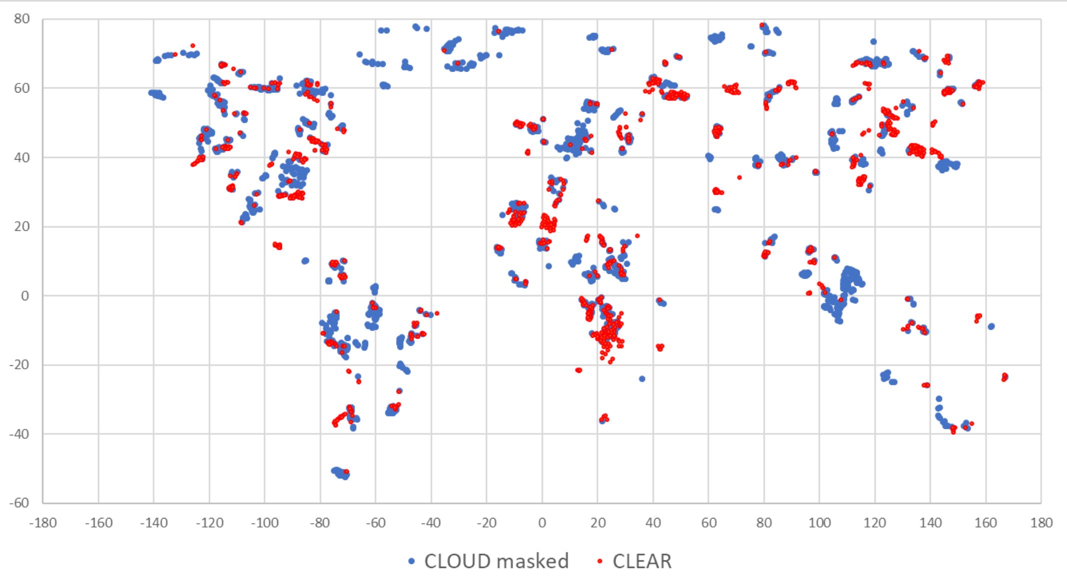


Figure 13: Global spatial distribution of thin semi-transparent clouds blue dots: cloud masked thin semi-transparent clouds, red dots: not masked thin semi-transparent clouds

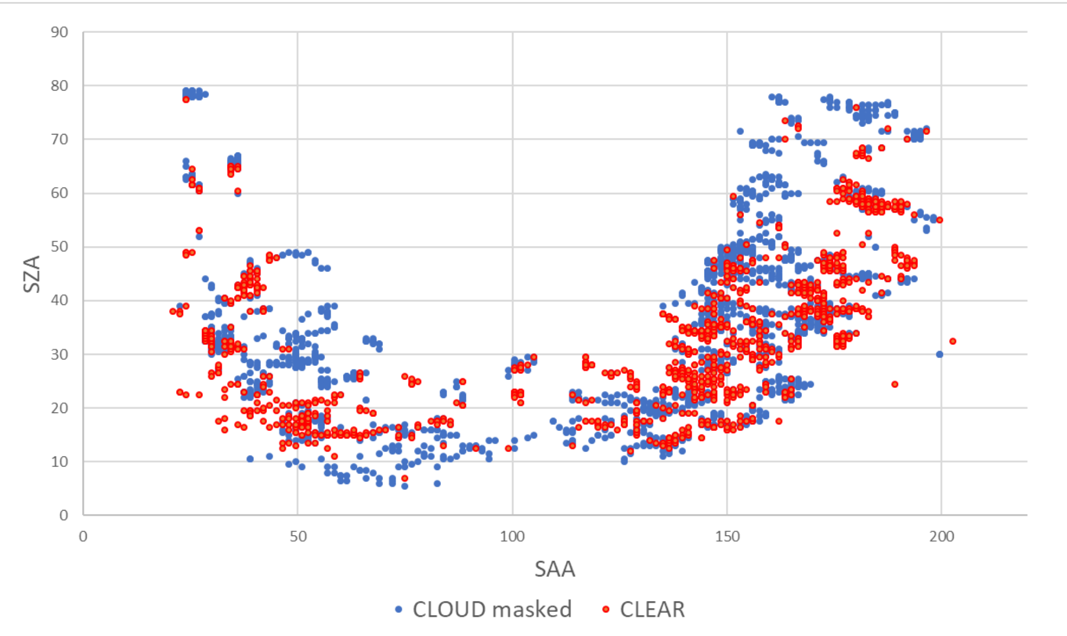


Figure 14: Distribution of thin semi-transparent clouds depending on solar azimuth and solar zenith angle. blue dots: cloud masked thin semi-transparent clouds, red dots: not masked thin semi-transparent clouds

For the clear sky cases, land and water surfaces seem to perform equally good. Only 2.6 / 3.9% of the cases are masked as cloud. This is less distinct for clear sky over snow/ice surfaces, where 29% of the clear snow pixels are masked as cloud. This result is expected because of the difficulties to differentiate between snow and cloud pixels. Thus, also within the group of spatially mixed surfaces, the mixed land/ice pixels are the most difficult cases for the cloud detection (14% masked as cloud). Another critical clear sky case are shorelines, especially bright beaches. Here, only 1.8% of the cases are wrongly detected as cloud. But the total number of cases (114) is very small.

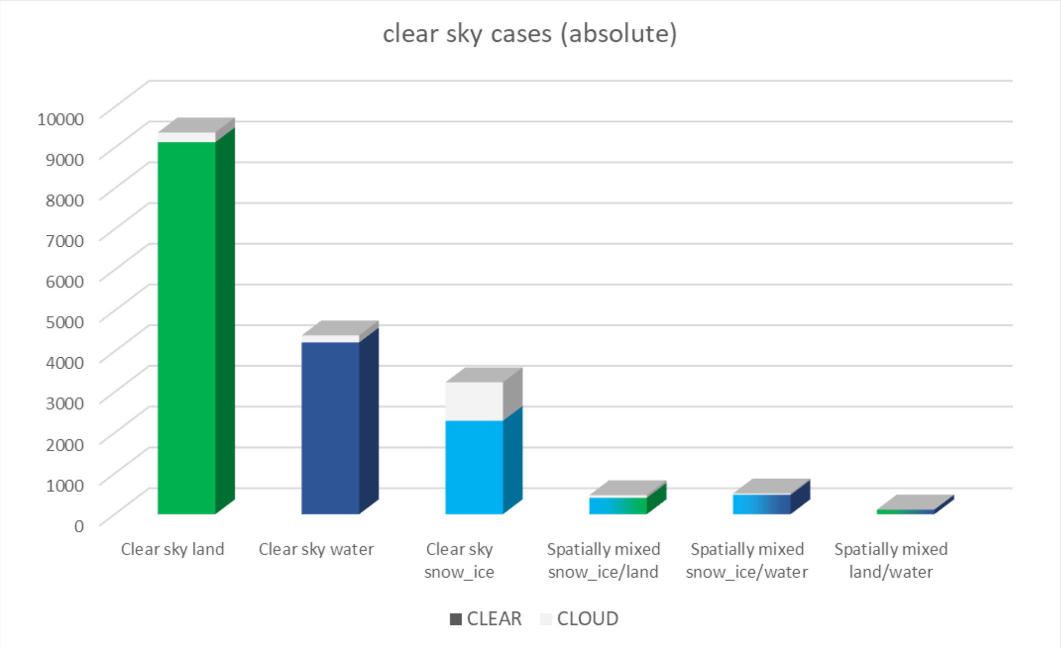


Figure 15: Absolute numbers of CLEAR and CLOUD masked pixels per clear sky categories

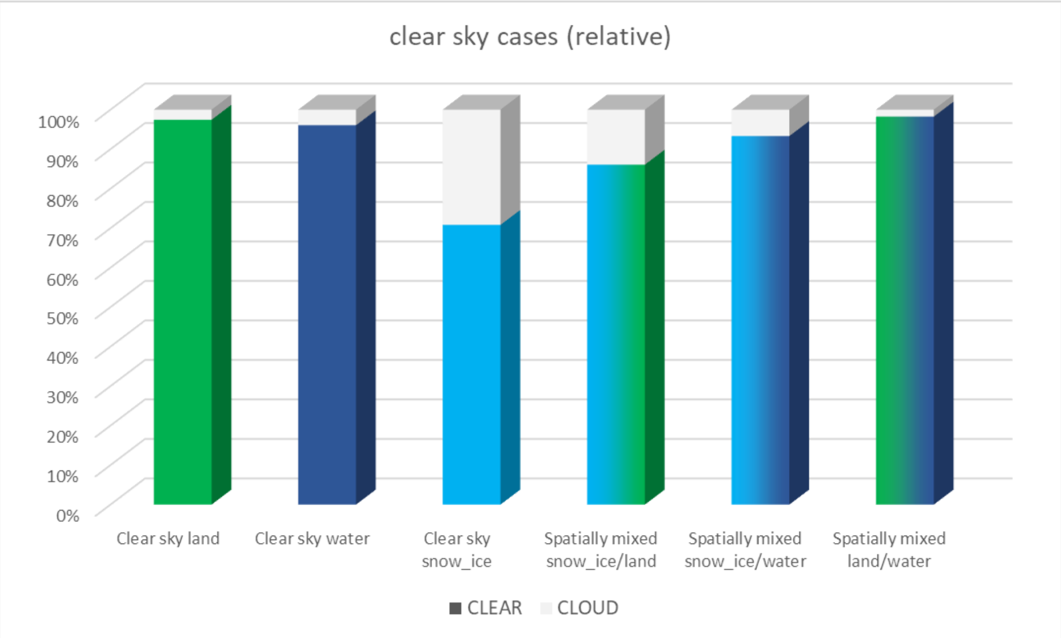


Figure 16: Relationship between CLEAR and CLOUD masked pixels per clear sky categories

The performance of the cloud detection has also been analysed for different water types as well as for salt lakes and sun glint, as shown in Figure 17 (absolute numbers) and Figure 18 (relative). No pronounced differences between the different water types occur. For salt lakes and sun glint the number of commission errors increases, especially salt lakes are flagged as cloud in 71% of the cases. Note that the absolute number of reference pixels for salt lakes is only 70.

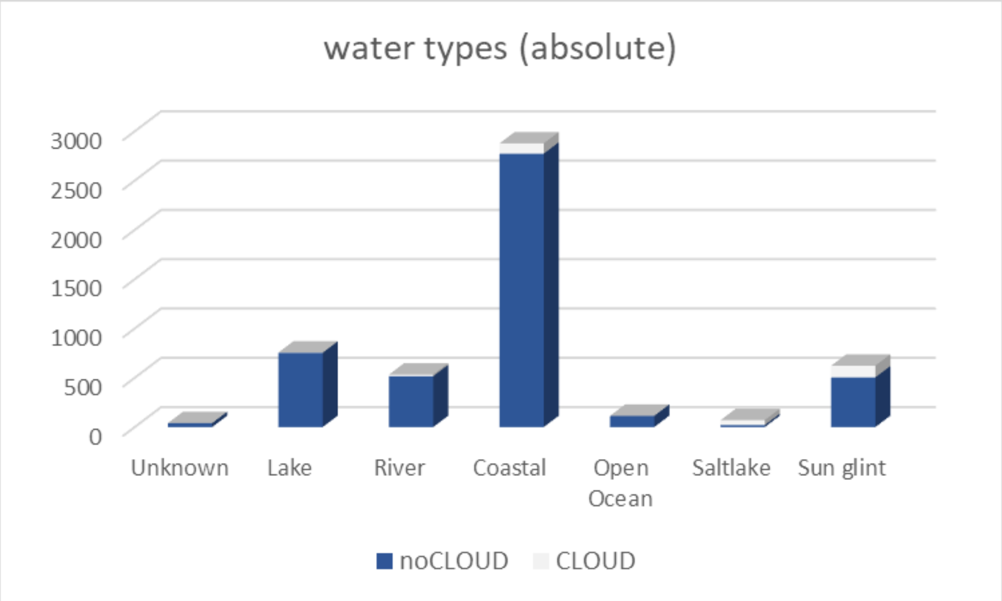


Figure 17: Absolute numbers of CLEAR and CLOUD masked pixels over different clear sky water types

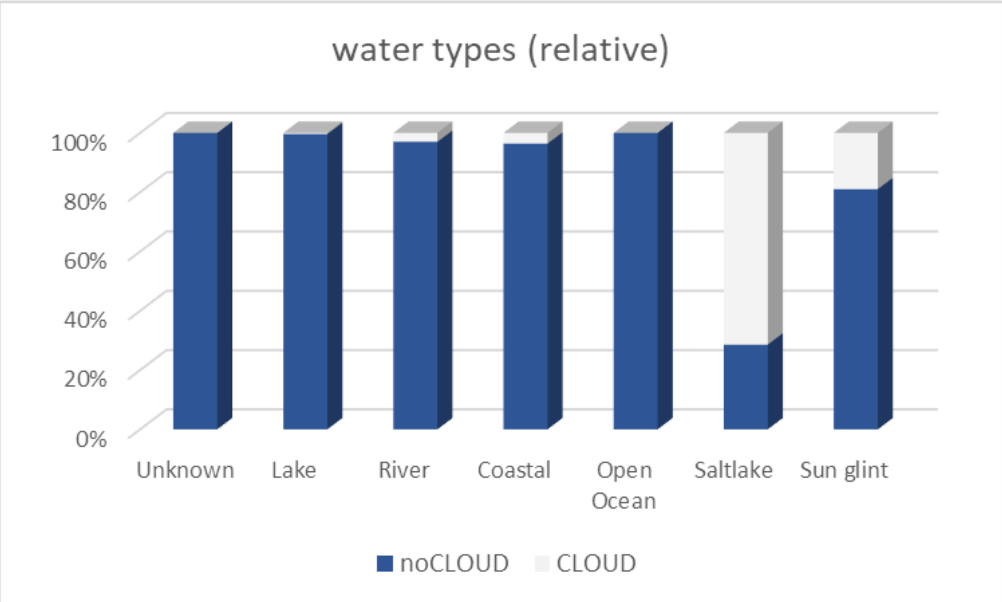


Figure 18: Relationship between CLEAR and CLOUD masked pixels over different clear sky water types

Regarding the cloud shadow detection, in total 783 cloud shadow pixels were collected in the reference data base. From these pixels, 66% have been masked with the CLOUD SHADOW mask (producer's accuracy). On the other hand, from the 736 pixels of the pixels masked as cloud shadow, 70% were also labelled as cloud shadow in the reference data base (user's accuracy). The total number of reference pixels for cloud shadow is rather small.

4.3 L3 10days averages

Will be completed in next version of the report.

5 Summary

The validation of the cloud detection algorithm for PROBA-V 333m products provided by University of Valencia and implemented by VITO is showing the very good results of the algorithm. The two applied methods – the visual inspection as well as the statistical analysis with the PixBox reference data set show how the algorithm performs for different cloud types and different surface types. The overall very good performance has some drawbacks over salt lakes and urban areas (overestimation). Ambiguous cases such as very thin semi-transparent clouds are detected in 50% of the cases approximately. How good this result is, depends on the L2 processing of the data, as some algorithms may cope for such clouds or very turbid atmosphere, while others cannot. However, cloud detection is a trade-off between detecting 100% of all kinds of clouds (True positives) on the one hand and overestimating clouds over clear sky areas (False positives) on the other hand. As shown in the statistical analysis, the validated algorithm resulted in very low numbers of false positives.

The separation between clouds and snow/ice is performing well, which is visible especially in the image interpretation. Sparse snow coverage or melting ice on water is often not detected, but also for these cases a trade-off is necessary between detecting everything and overestimating.

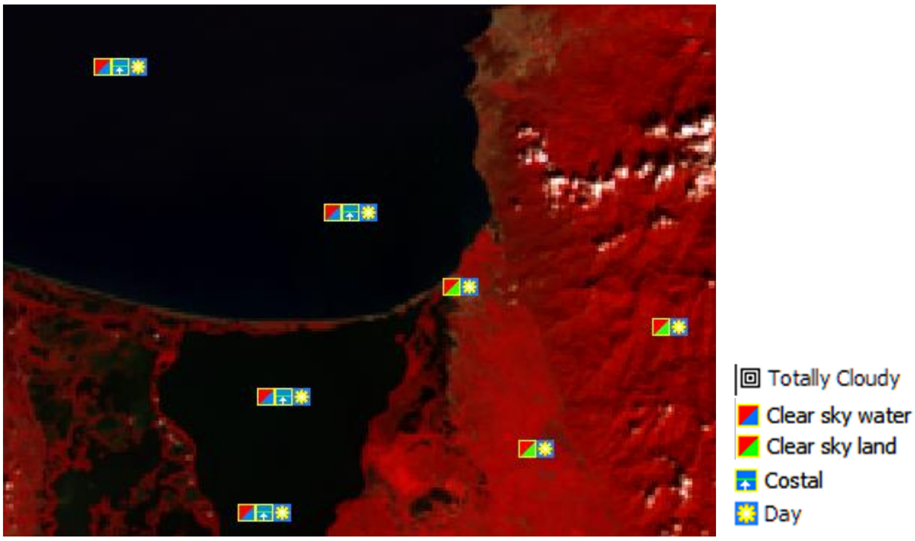
The performance of the cloud shadow mask was not in focus of this investigation but can be assessed as medium quality, often underestimating the cloud shadow area.

6 ANNEX – Example of pixel collection within PixBox reference data set

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:

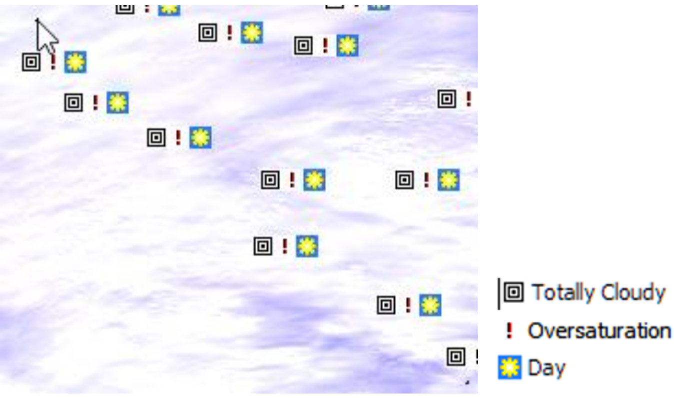


Figure 19: Examples of clear surface (land and water and ice) and totally cloudy pixels (below).

Semi-Transparent cases over land:

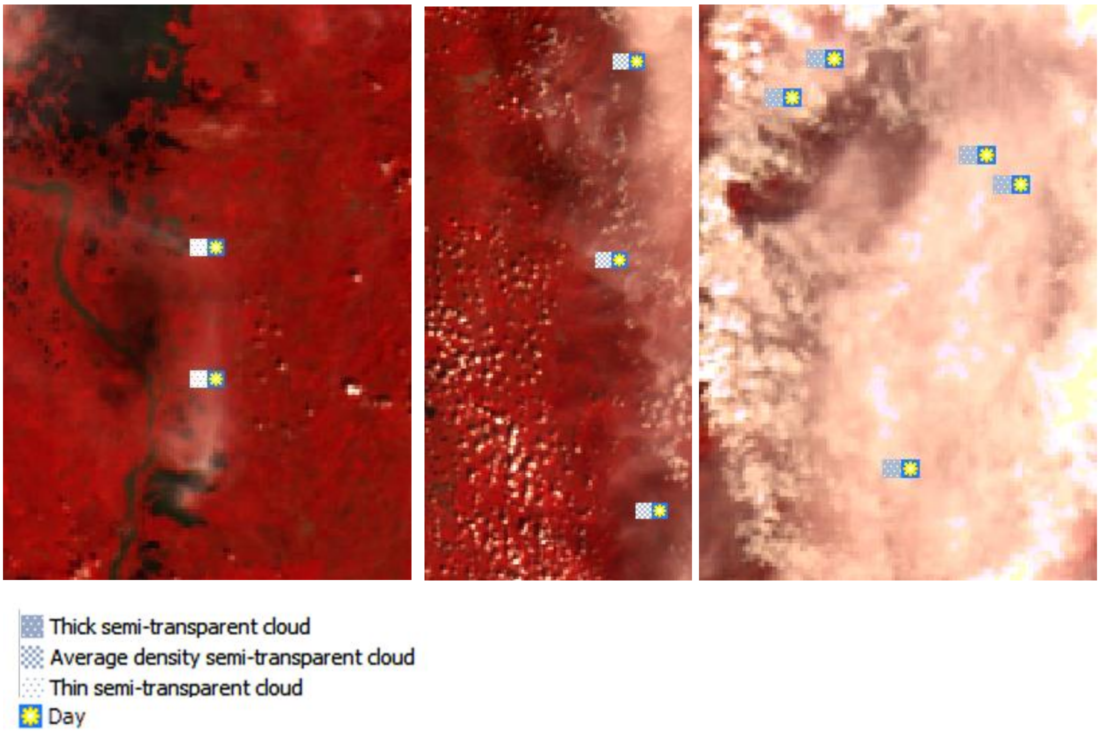


Figure 20: Examples for semi-transparent clouds over land: left: thin; middle: medium; right: dense

Semi-Transparent cases over water:

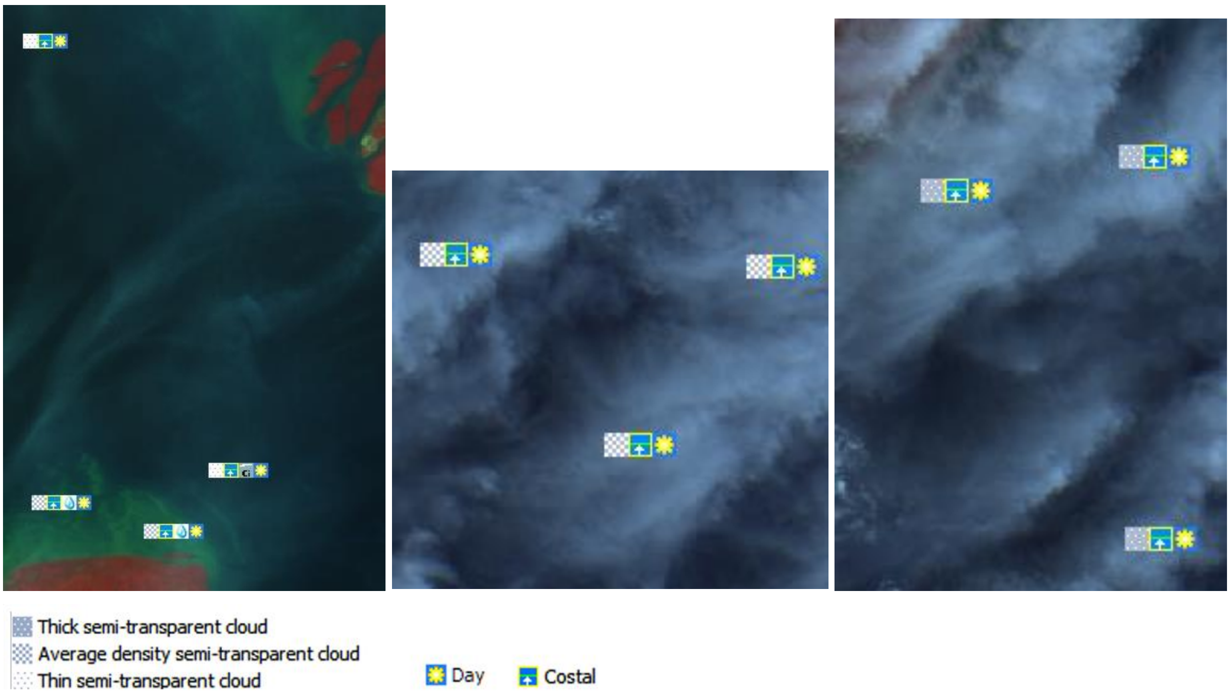


Figure 21: Examples for semi-transparent clouds over water: left: thin; middle: medium; right: dense

Semi-transparent clouds over ice:

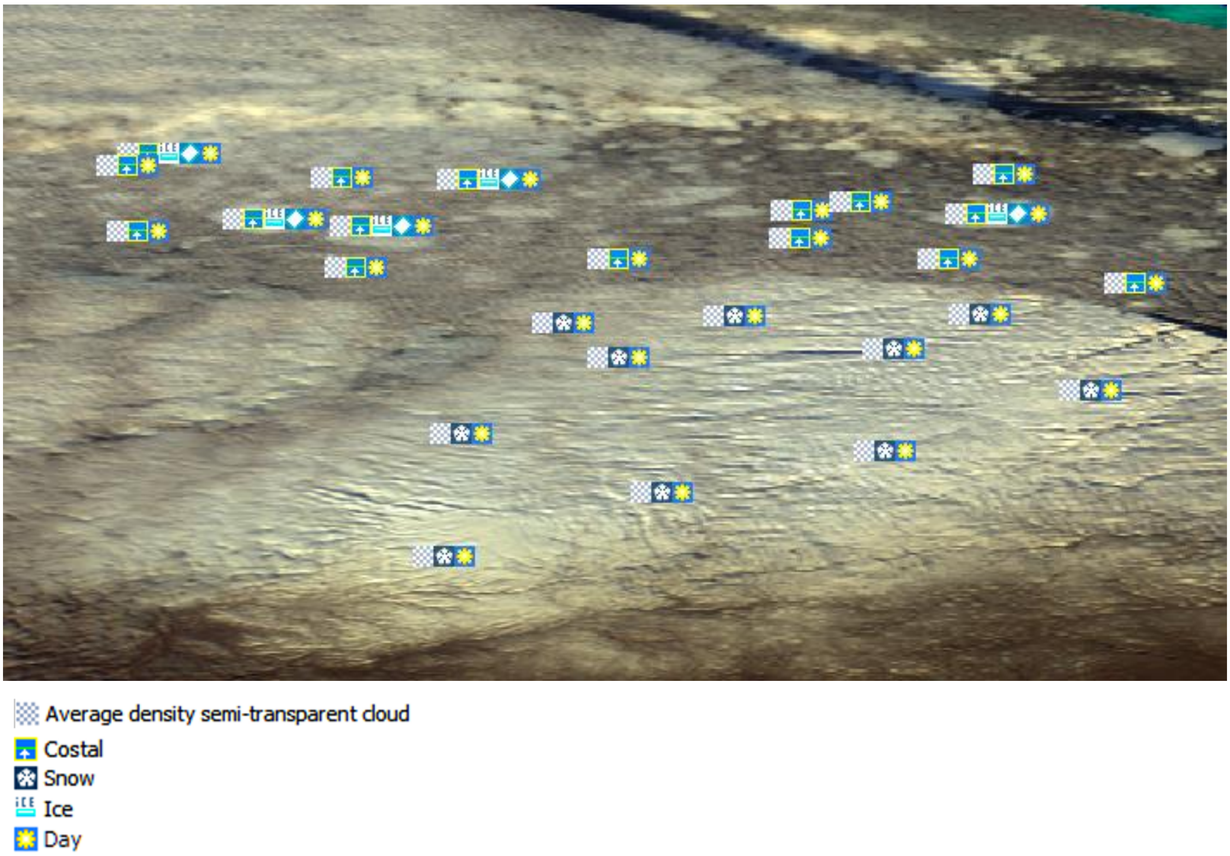


Figure 22: Examples for semi-transparent clouds over ice (medium)

Spatially mixed cases:

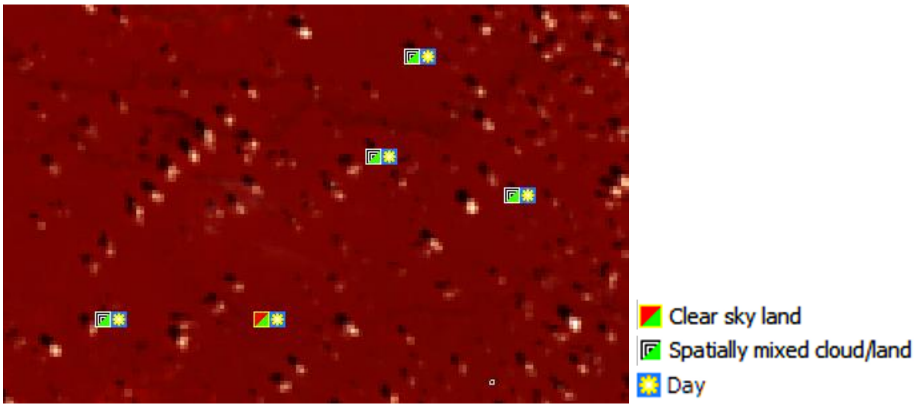


Figure 23: Examples of spatially mixed pixels cloud/land

And some more examples over land of different cases:

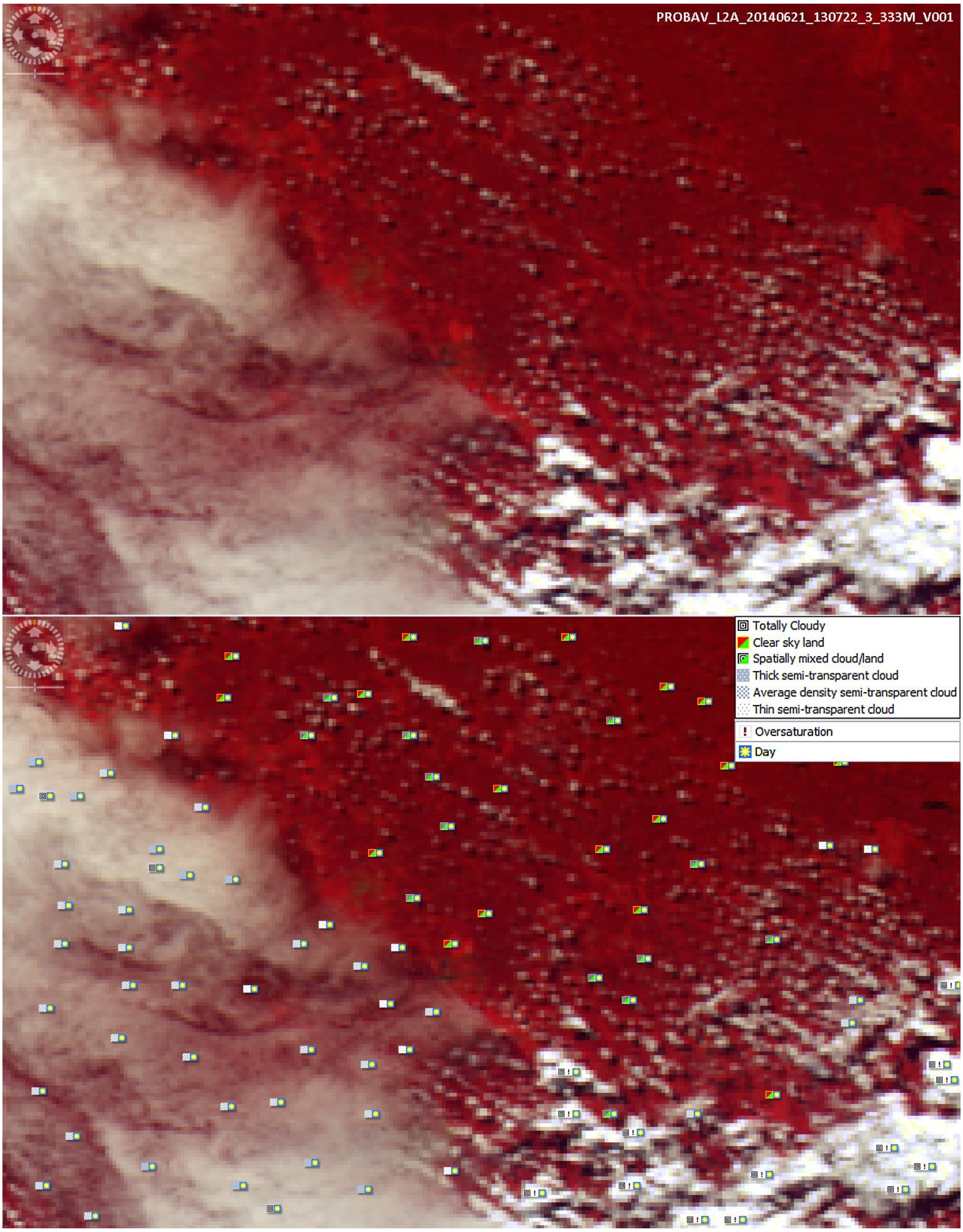


Figure 24: Different cloud categories over land