PROBA-V CLOUD ROUND ROBIN

Validation Report

Version 1.3

Kerstin Stelzer, Michael Paperin, Grit Kirches

02.05.2017

Table of content

Change Log

Version	**Date**	**Changes**	**Contributions**
1.0	23.02.2017	Version 1.0 provided to ESA	Kerstin Stelzer, Michael Paperin, Grit Kirches
1.1	24.02.2017	Version 1.1 including comments provided by ESA	Rosario Quirino Iannone, Fabrizio Niro, Kerstin Stelzer
1.2	11.04.2017	Version 1.2 updated after Cloud Workshop and fully integration of FUB results	Kerstin Stelzer, discussions points from Workshop participants
1.3	02.05.2017	Including comments provided by ESA	Rosario Quirino Iannone, Kerstin Stelzer

1 Introduction

The scope of this document is to describe the validation results of the Round Robin for the cloud detection Algorithm for Proba-V. The validation data set and the validation methods are described. The validation is performed on the one hand with manually selected pixel collection and on the other hand by comparison of the different cloud flags in randomly selected ProbaV images. In total 8 algorithms are compared, provided by 5 different participants. One algorithm is the operational cloud detection algorithm of the reprocessed ProbaV data.

Figure 1 provides an overview of the overall workflow of the Round Robin exercise with reference scenes, validation and test data set and finally the validation matrices and their compilation to a quality assessment, which is given within this report.

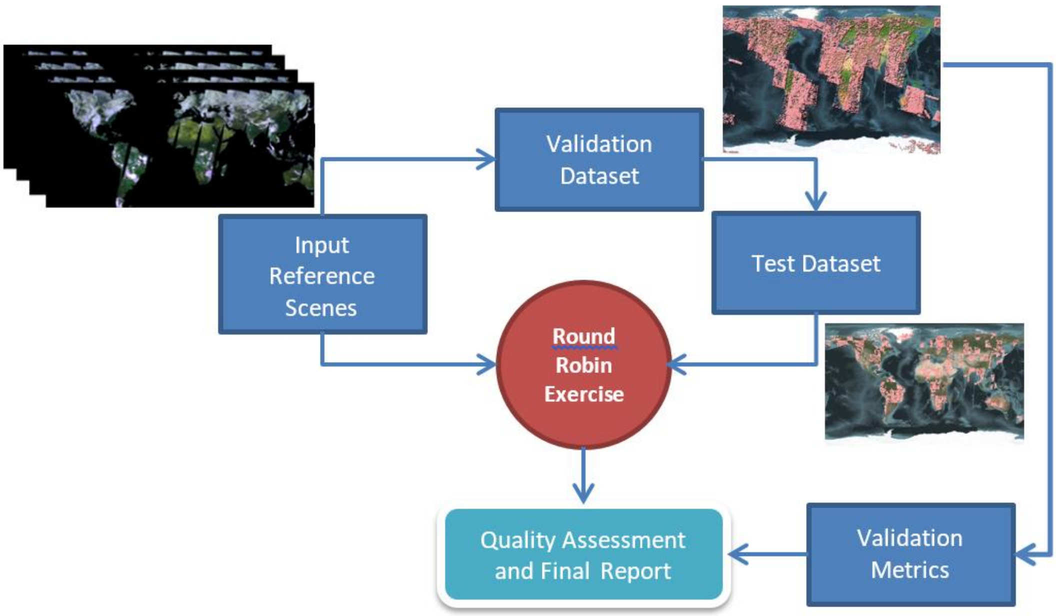


Figure 1: Workflow for the ProbaV Cloud screening Round Robin exercise

2 ProbaV Cloud Detection data sets

For each cloud detection algorithm, an ATBDs is provided by the algorithm developers. The ATBDs are available here: https://earth.esa.int/web/sppa/activities/instrument-characterization-studies/pv-cdrr/project-documents. The overview of algorithms is given in Table 1 specifying the basic technique and the classes that are provided.

Table 1: Overview on algorithm method and classes considered by each algorithm

Algorithm	Method	CLOUD	CLEAR	SEMITRANS	ICE/SNOW	THIN SNOW	CLOUDSHADOW	INVVALID
Algo 1	Discriminant Analysis trained with SEVIRI&MODIS	x	x					x
Algo 2	Discriminant Analysis trained with SEVIRI	x	x					x
Algo 3	Discriminant Analysis trained with MODIS	x	x					x
Algo 4	Multi-spectral and multi-textural thresholding	x		x	x	x		
Algo 5	Neural Network	x	x					x
Algo 6	Self-Organizing Feature Maps	x	x	x				x
Algo 7	Dynamic threshold on the Blue band	x	x		x		x	x
Algo 8	Threshold on Cloud Optical Thickness	x	x	x	x			x

Each data set came with a band containing the cloudmask. The coding for cloud masks were different per algorithms. All algorithms provided an indication for the CLOUD flag but in addition, also semi-transparent clouds (SEMI-CLOUD) and snow covered surfaces (SNOW) were identified by some algorithms. A harmonization of naming and coding was necessary in order to use automatized procedures for the validation. The following text writes a flag always in capital letters, while the reference class or surface description is written in small letters.

3 Validation data set

3.1 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 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.

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

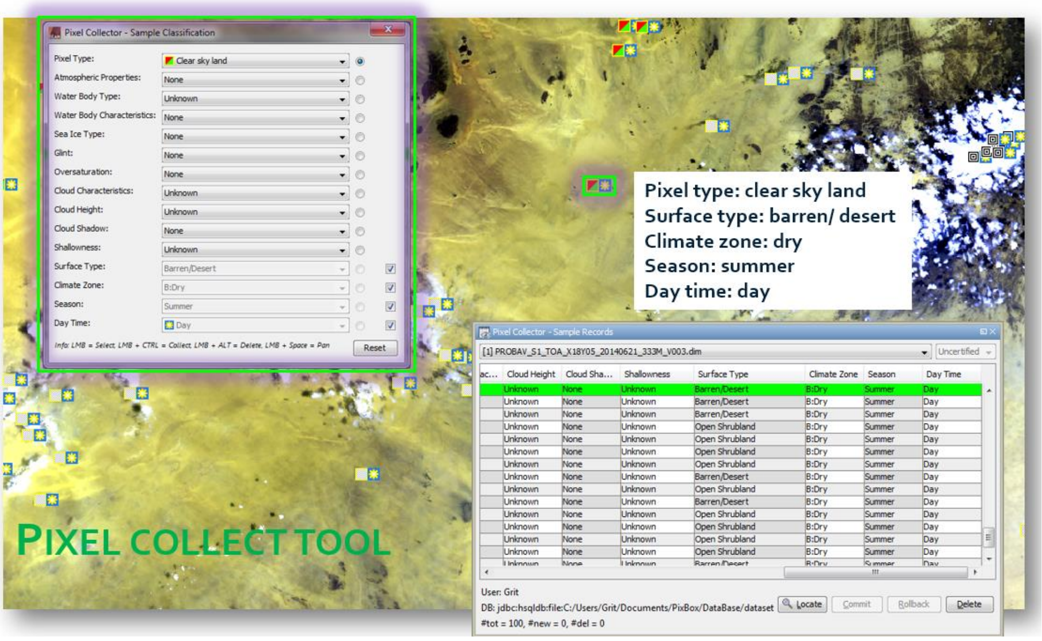


Figure 2: Pixel Collection tool with categories to be assigned to each pixel

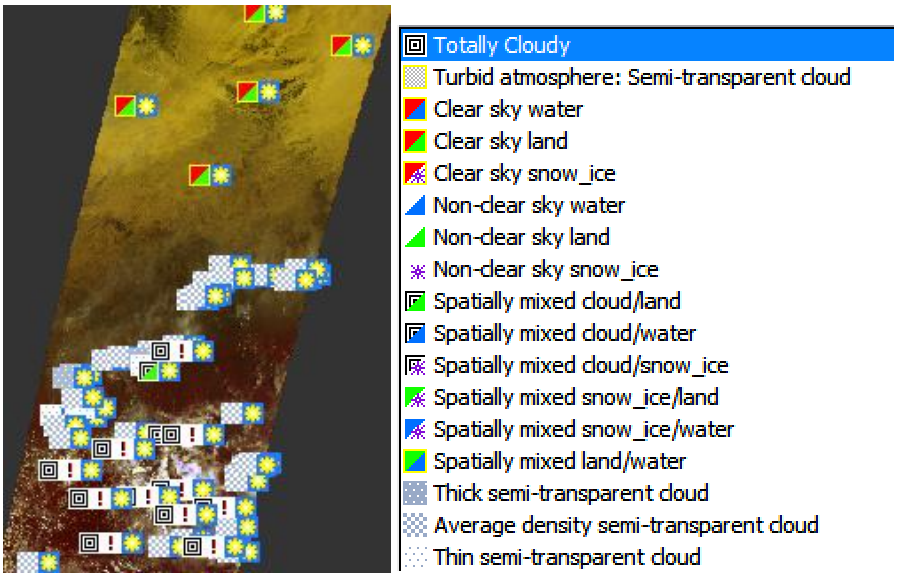


Figure 3: Example image showing the position of collected pixels and labelling.

3.2 Validation data set

The validation data set contains 53000 entries collected from 61 different images. They cover the four days 21.03., 21.06, 21.09. and 21.12.2015. The pixels are collected based on the ProbaV Level 2A products, processing version V001. Figure 4 shows the number of pixels per categories in the validation data set.

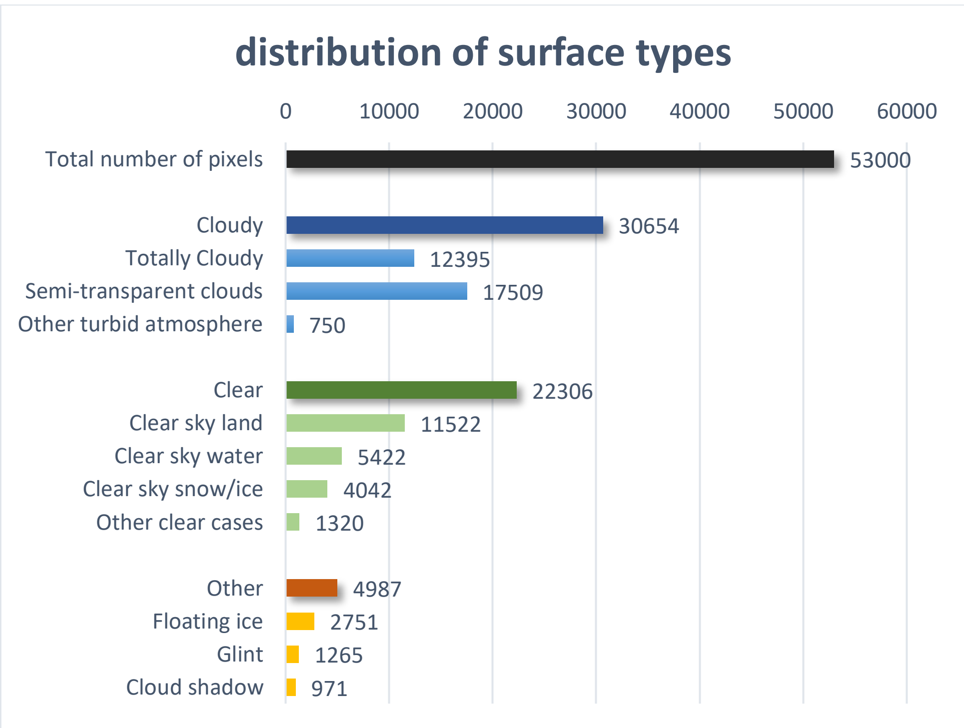


Figure 4: Distribution of surface types within the validation data set (numbers are the counts of pixels per category)

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

Figure 5: Position of globally collected pixels of the validation data set covering clear land, clear water, clear ice, totally cloud and semitransparent clouds

A small subset of the validation data set was provided to the participants of the round robin for information. Thus, the participant were aware of the characteristics of the validation data set. The test data set had 1350 entries.

4 Validation Methods

A test data set was provided to the participants beforehand. The test data set was a subset of the validation data set and informed the participants about the pixel collection procedure and criteria for the categorization of pixels into clear, cloud or semi-transparent clouds. Examples were provided for orientation.

Two methods were applied to the provided data sets.

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

For the visual inspection, all cloud flag bands have been compiled in one product. The RGB (NIR-RED-BLUE) of a certain image (subset) is displayed with an overlay of one cloud masks. They are compiled in a set of images to compare the 8 algorithms. The mask is displayed in black, or if needed in another colour for better visualization. The images are always compiled in a set of 8 images: the RGB without any flag and the 8 algorithms with the cloud flag overlay.

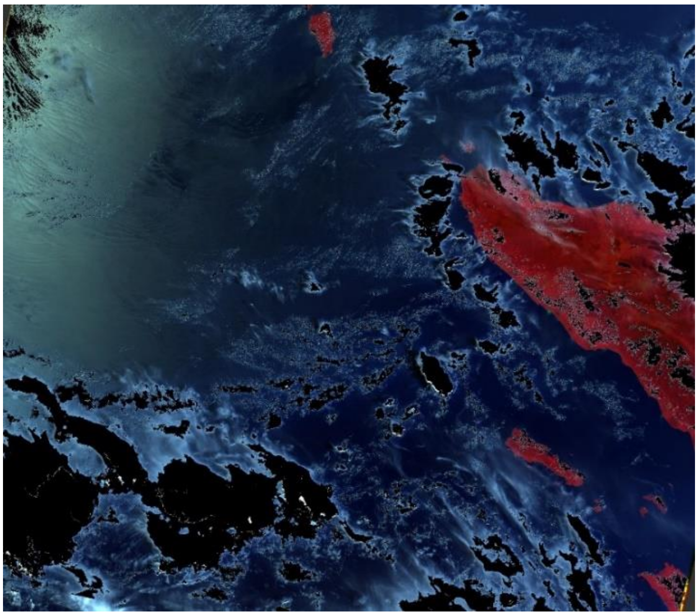


Figure 6: RGB image and CLOUD mask overlay in black

The second method works with confusion matrices and related statistics, namely the User's Accuracy (UA), Producer's Accuracy (PA), Overall Accuracy (OOA) and respective errors (E) (see Figure 7). Further, the values for Scott's Pi, Krippendorf's alpha and Cohen's kappa are derived. All three measures provided the same results and therefore it was decided to provide the values for Krippendorf's alpha in the following assessment. The equations for the used statistical measures are provided in Annex III. High number in User's Accuracy for the CLOUD flag means that the pixel under the CLOUD flag 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), while a low User's Accuracy for the CLOUD flag accuracy indicates that the CLOUD flag has classified also clear surfaces (error of commission).

ProbaV RR Algo_3 - all surfaces

In-Situ Database

CLOUDMASK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	15522	4916	20438	75.9	24.1
CLOUD+SEMI	1320	21465	22785	94.2	5.8
Sum	16842	26381	43223		
P A	92.2	81.4		OAA:	85.57
E	7.8	18.6			

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

5 Results

5.1 Cloud flags in images

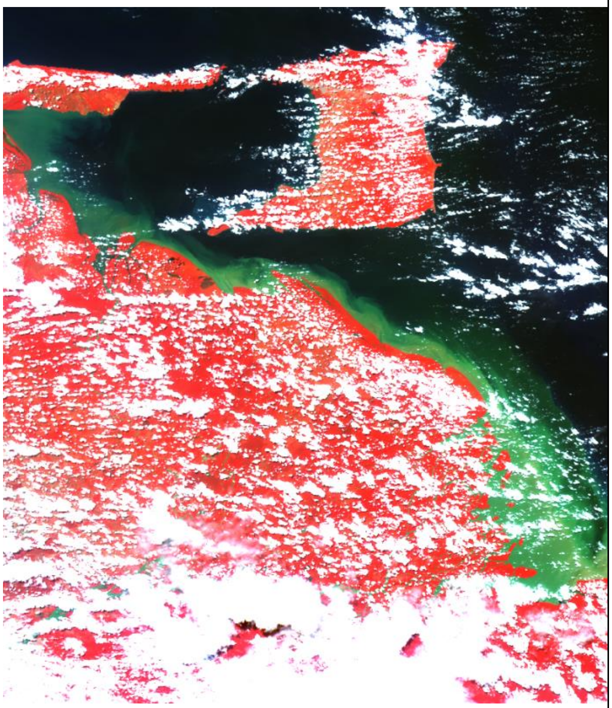
5.1.1 Image show cases

The following pages provide seven show cases for demonstrating the behavior of the algorithms for different surfaces. Attached to this report is a PPT, which provides more examples.

No.	Topic	Product (subsets)
1	Clouds over land and water, turbid coastal water	PROBAV_L2A_20140621_144249_3_333M_V001
2	Thick and semi-transparent clouds over land and behaviour over inland waters	PROBAV_L2A_20140621_144249_3_333M_V001
3	Small cumulus clouds over land	PROBAV_L2A_20140621_144249_3_333M_V001
5	bright surfaces	PROBAV_L2A_20140921_103856_3_333M_V001
6	different cloud types over water, thin clouds over land	PROBAV_L2A_20140321_044547_1_333M_V001
7	Flagging of coastlines and thin clouds	PROBAV_L2A_20140921_154245_3_333M_V001

Each show case is accompanied by a short assessment of each algorithm. A summary per algorithm is given in chapter 5.1.2. For a better overview this is provided as bullet lists.

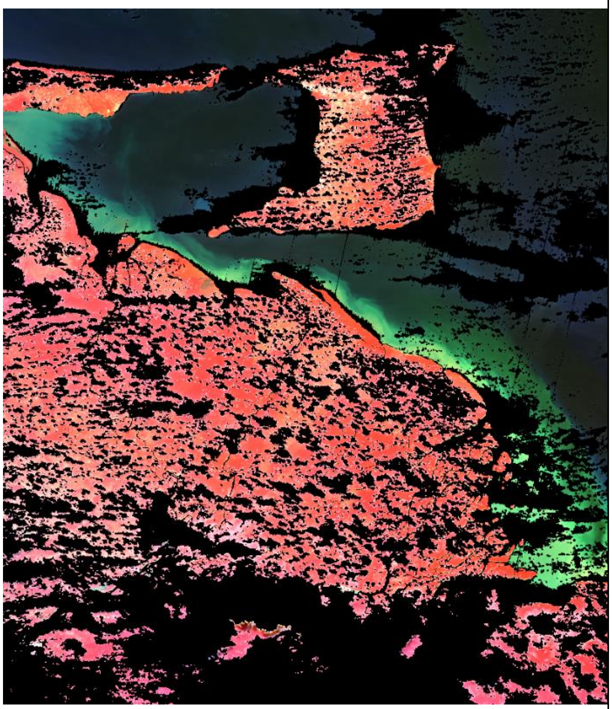
Show case 1: Clouds over land and water, turbid coastal water



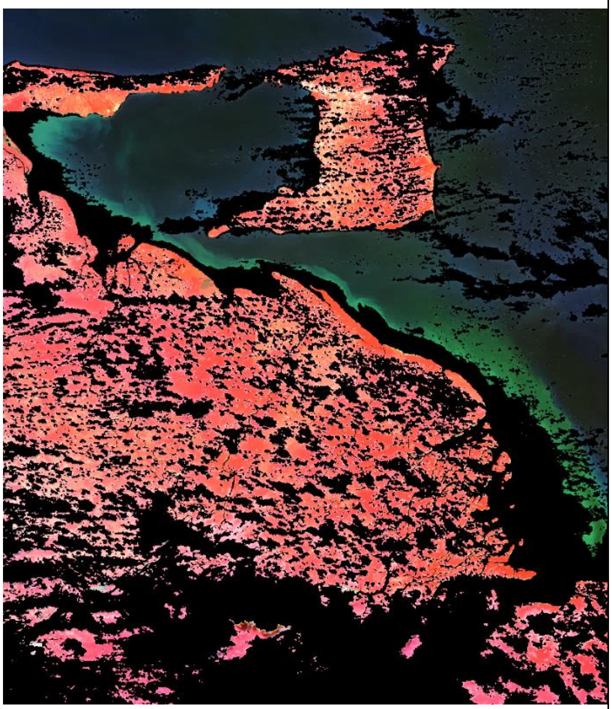
RGB

Brief description of algorithms performance:

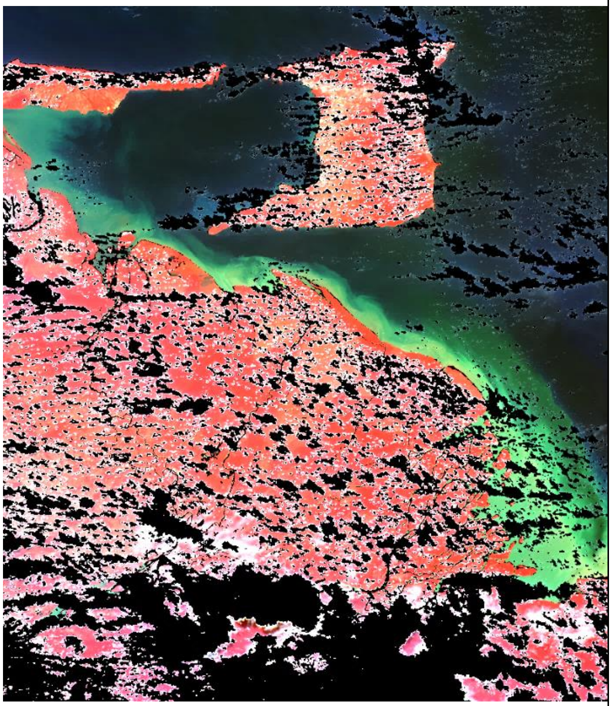
This example shows the performance of the algorithms over clear water, turbid water and over land. Different cloud types are present. Cloud detection over land is less strict in algo 3 and 5 with omission of thin clouds, while algo 7 flags also many clear land pixels. Algo 4 and 7 leave out some opaque clouds, which are detected by their SNOW flag (discussed later). Algo 8 detects the clouds very well over land and water. Over water, algo 6 omits cloud borders, less obvious also in algo 3 and 5. While algo 3, 4 and 5 detect the clouds correctly over turbid waters, algo 1,2, 7 and 8 flag too many pixels over turbid waters.



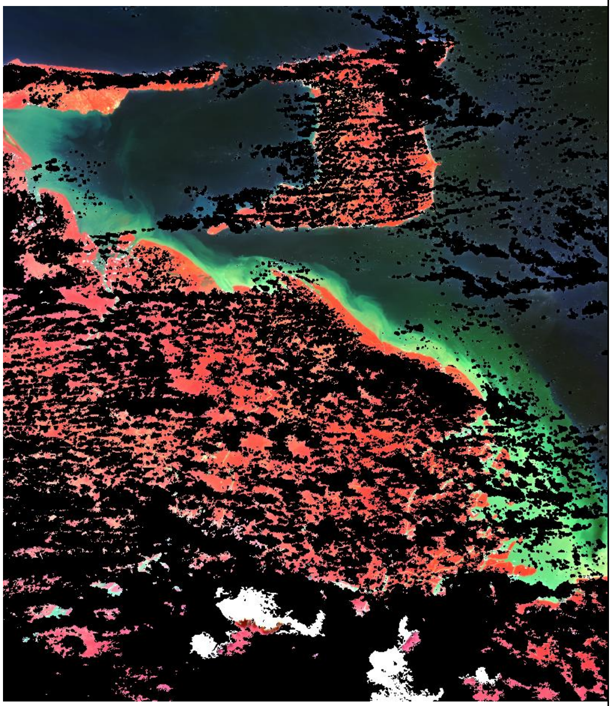
Algo 1 CLOUD



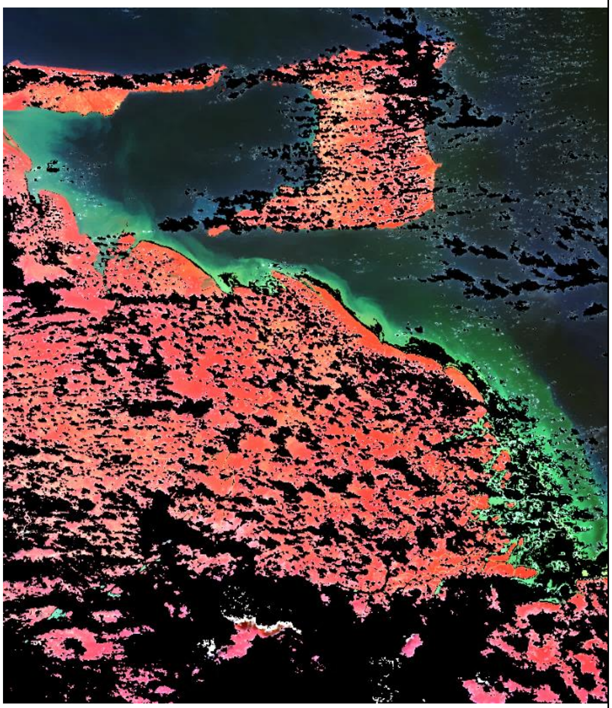
Algo 2 CLOUD



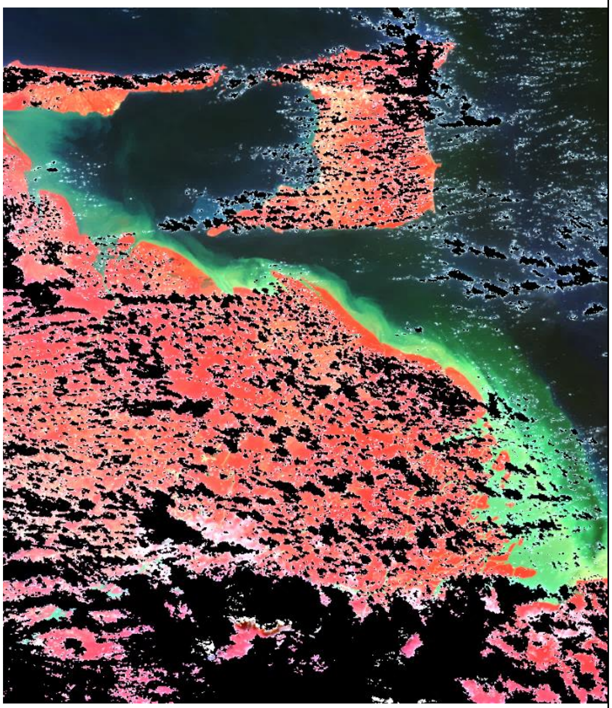
Algo 3 CLOUD



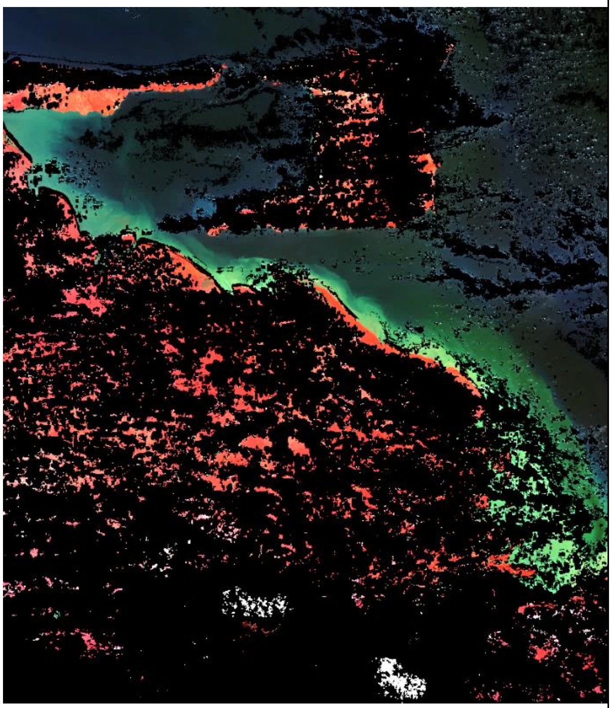
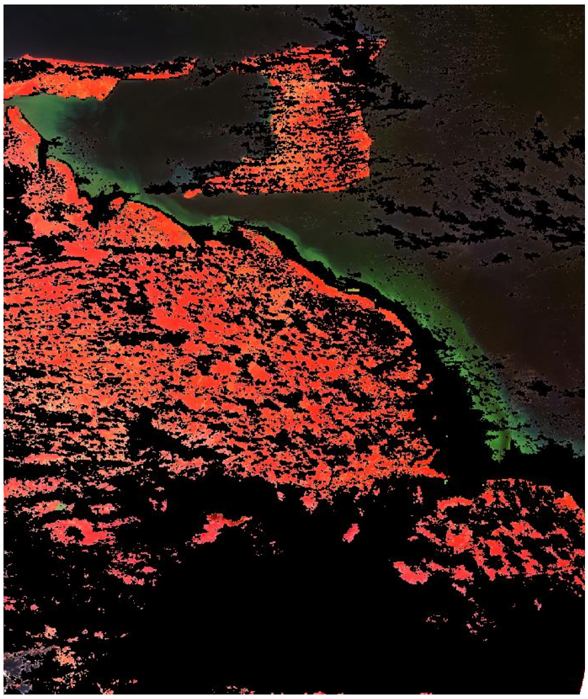
Algo 4 CLOUD + SEMI

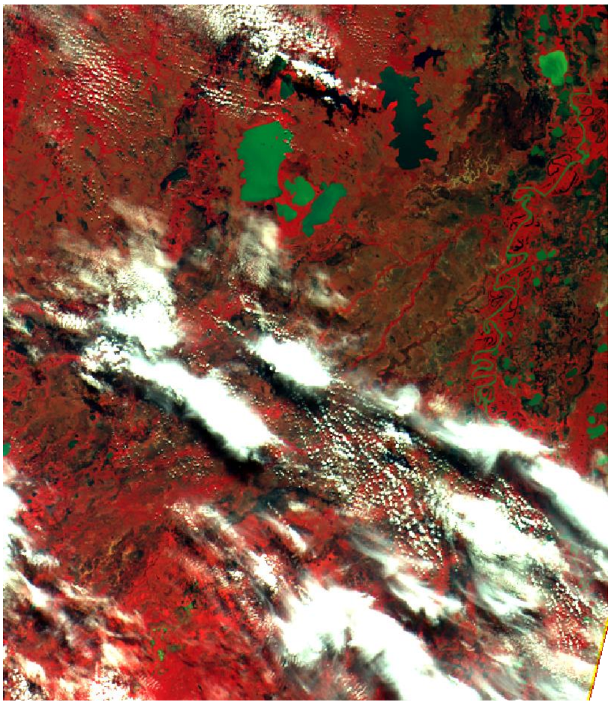
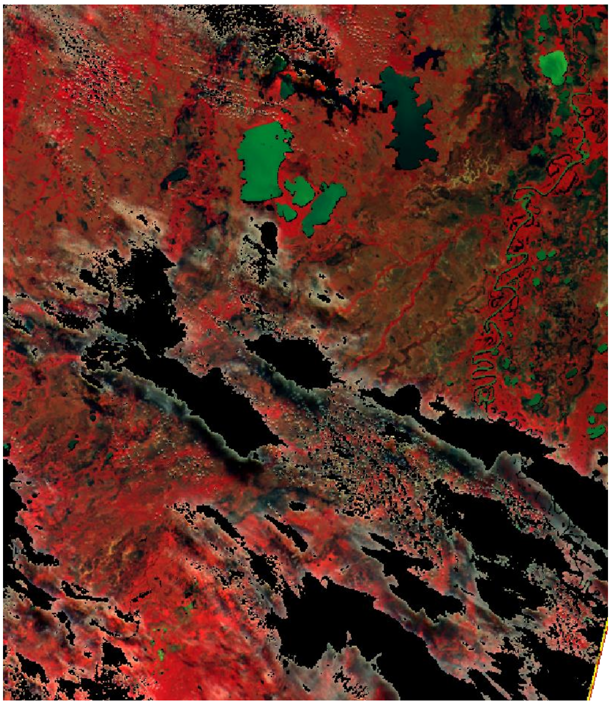
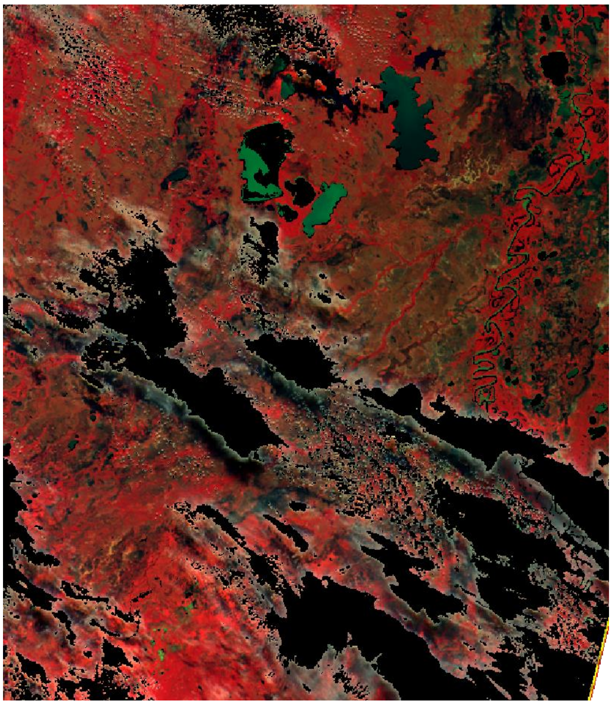


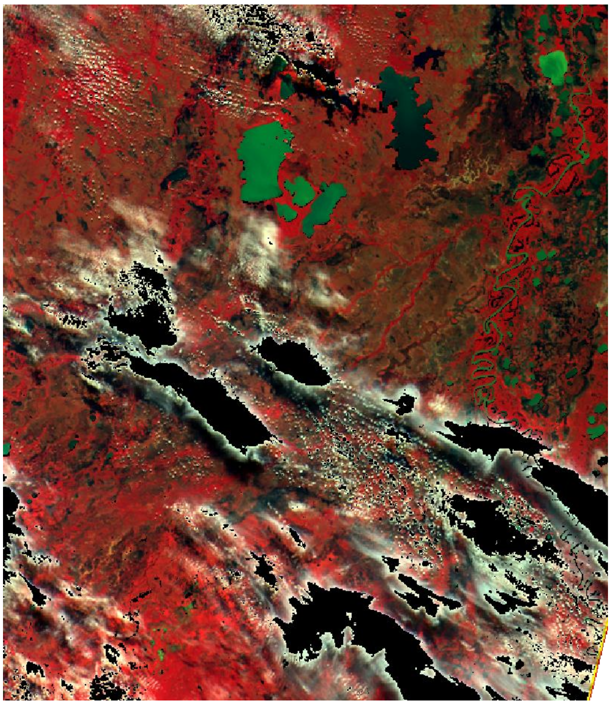
Algo 5 CLOUD



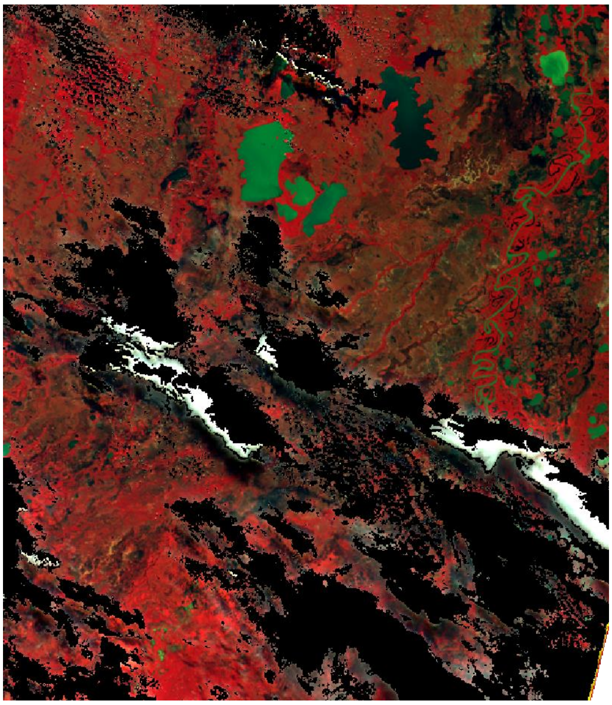
Algo 6 CLOUD + SEMI

	
Algo 7 CLOUD	Algo 8 CLOUD+SEMI

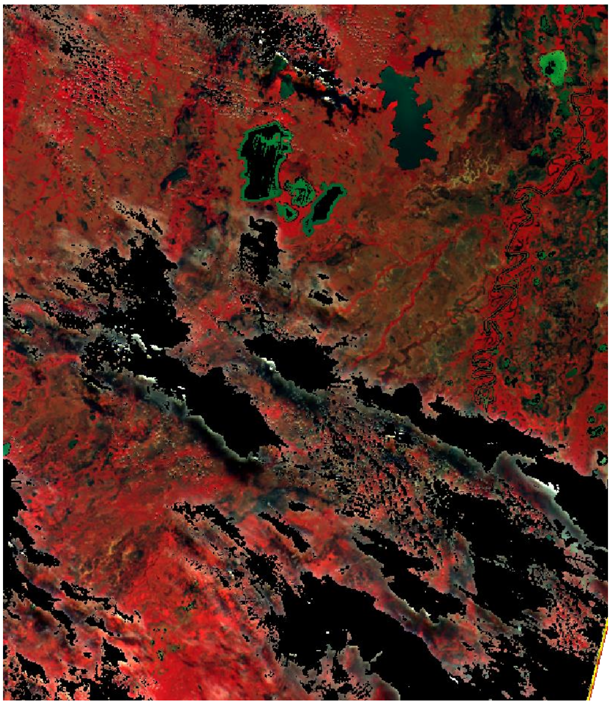
Show case 2: Thick and semi-transparent clouds over land and behaviour over inland waters	
 RGB	Brief description of algorithms performance: Opaque and semi-transparent clouds over land can be assessed with this example. Furthermore, bright inland waters are present which are classified as cloud by some algorithms. The opaque clouds are detected by all algorithms except algo 4 which flags some of the clouds with the SNOW flag. Semi-transparent clouds are best detected by algo 4 and 7, while algo 1, 2, 5 and 8 leave out a border of semi-transparent clouds. Algo 3 and 6 omit the semi-transparent clouds. Algo 2 and 5 flag the bright lake water as cloud, while the dark lake water is not detected as cloud. Algo 1, 3, 4, 6 do not have any problems over bright inland waters.
 Algo 1 CLOUD	 Algo 2 CLOUD



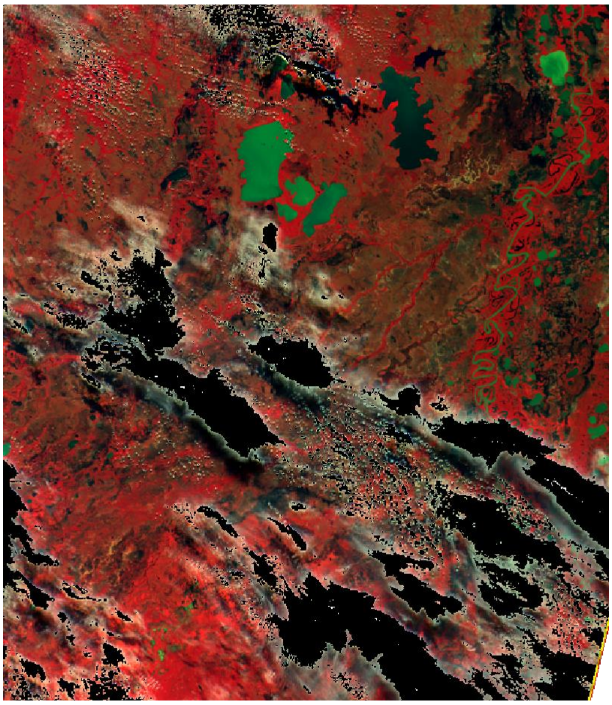
Algo 3 CLOUD



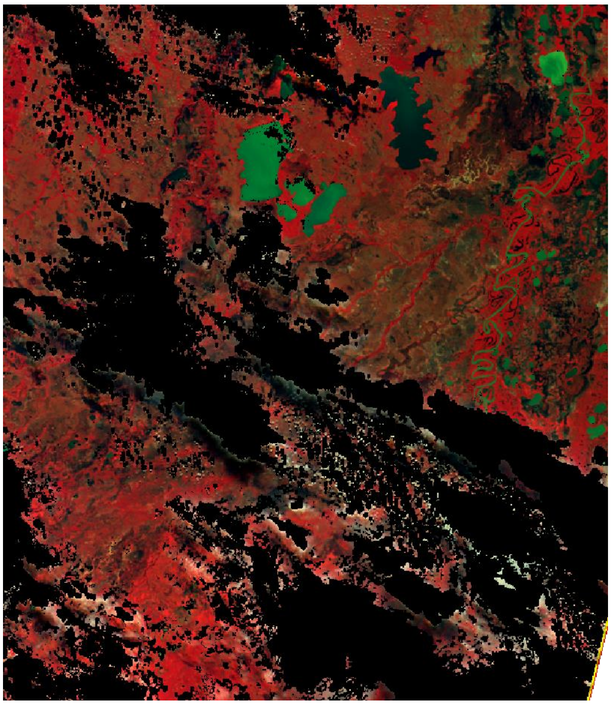
Algo 4 CLOUD + SEMI



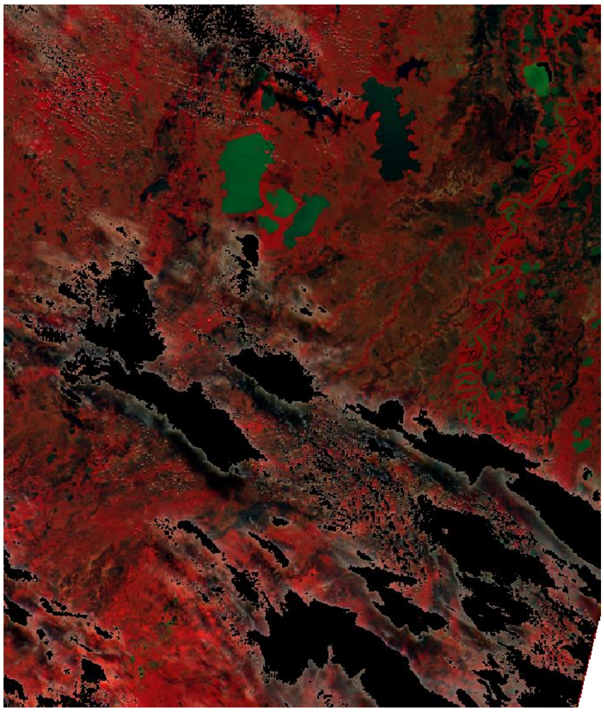
Algo 5 CLOUD



Algo 6 CLOUD + SEMI

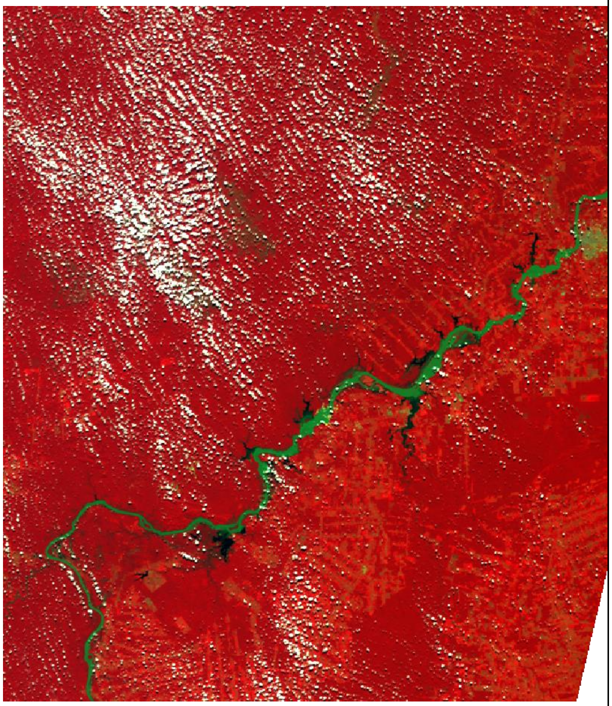


Algo 7 CLOUD



Algo 8 CLOUD+SEMI

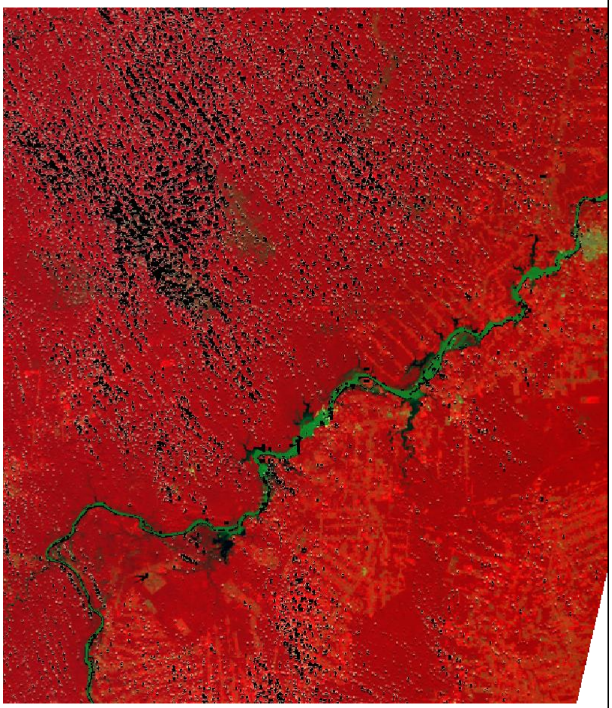
Show Case 3: Small cumulus clouds over land



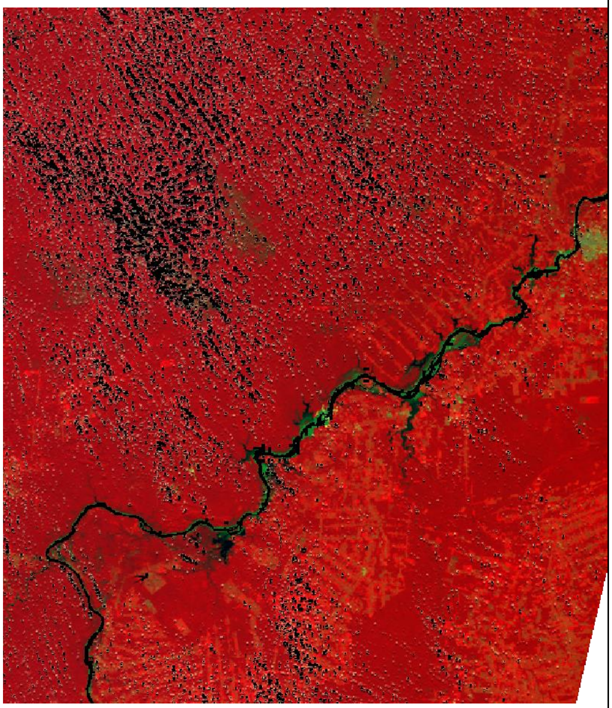
RGB

Brief description of algorithms performance:

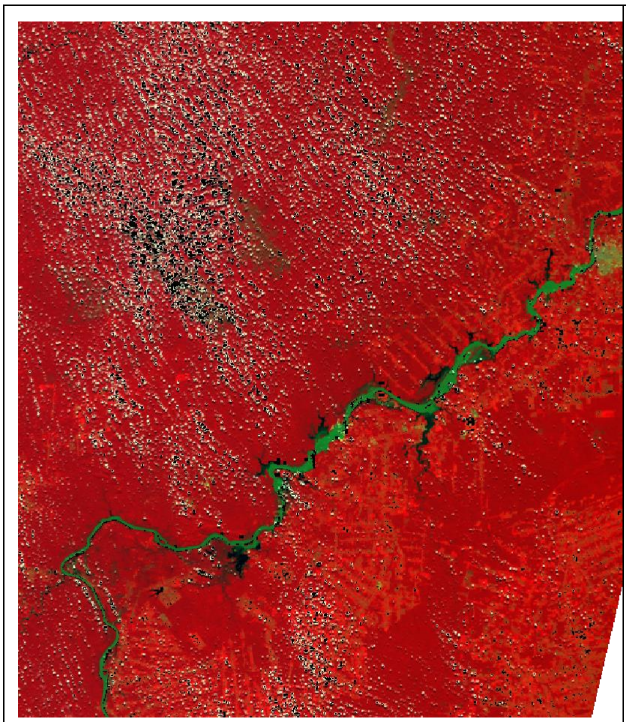
This example shows the performance of the algorithms for small cumulus clouds and over bright inland water (river). The small cumulus clouds are detected differently by the algorithms; while algo 3 detects only few of the small clouds, algo 7 is detecting them all, including a buffer around. This leads to detection also of clear surfaces. The visual impression of algo 4 and algo 8 is best, but algo 4 is not detecting clouds over the river. Those are well detected in algo 1, 6, 7 and 8. Some algorithms detect the river water as cloud (algo 2 and 5).



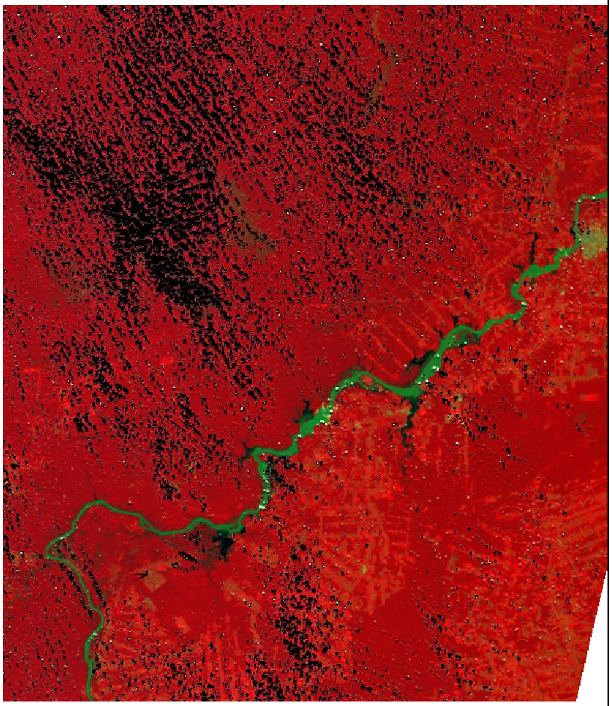
Algo 1 CLOUD



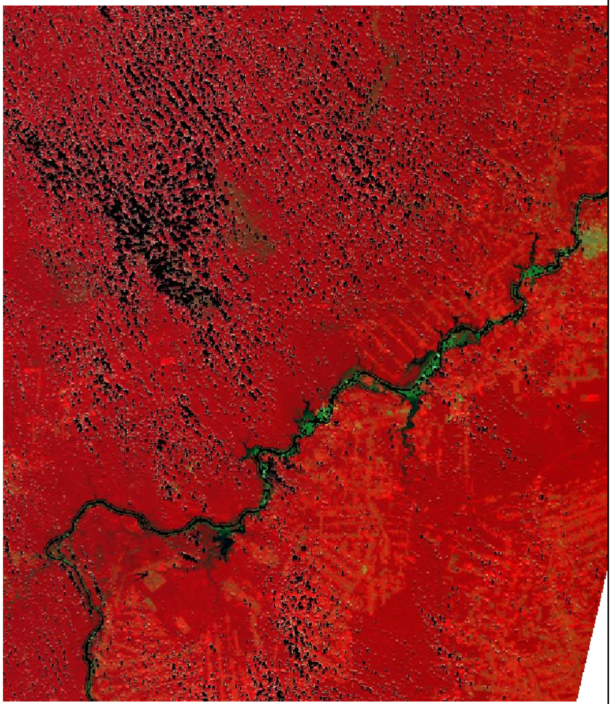
Algo 2 CLOUD



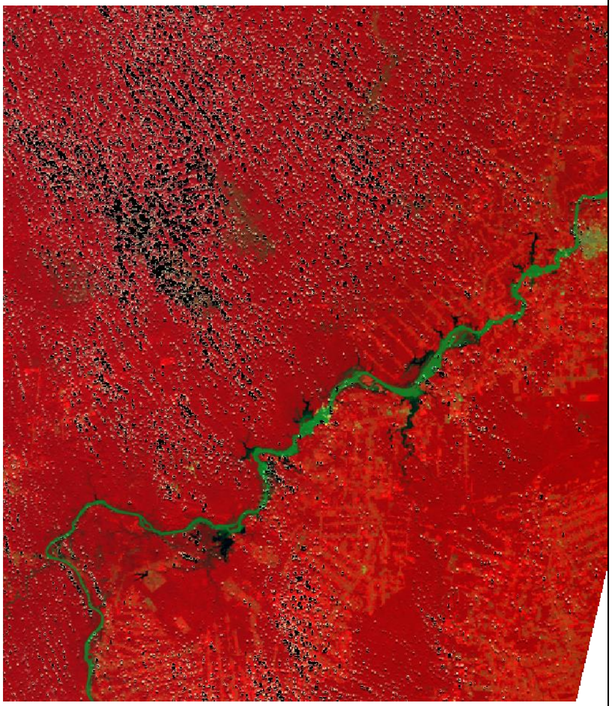
Algo 3 CLOUD



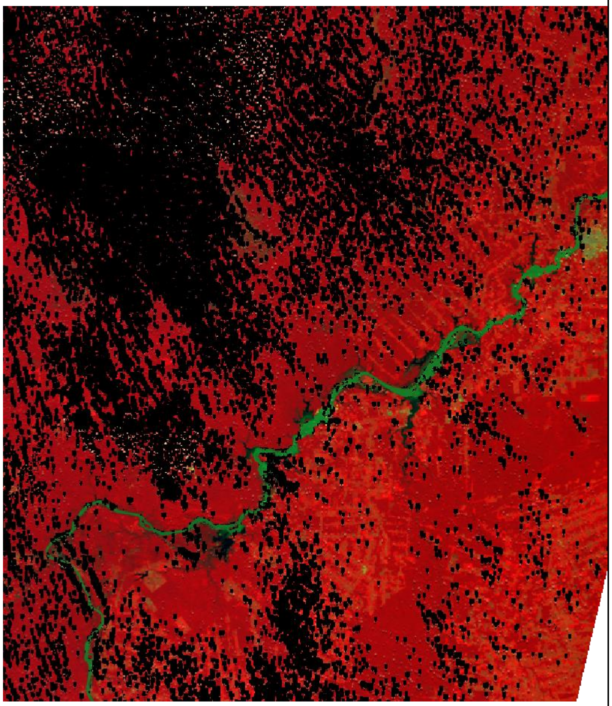
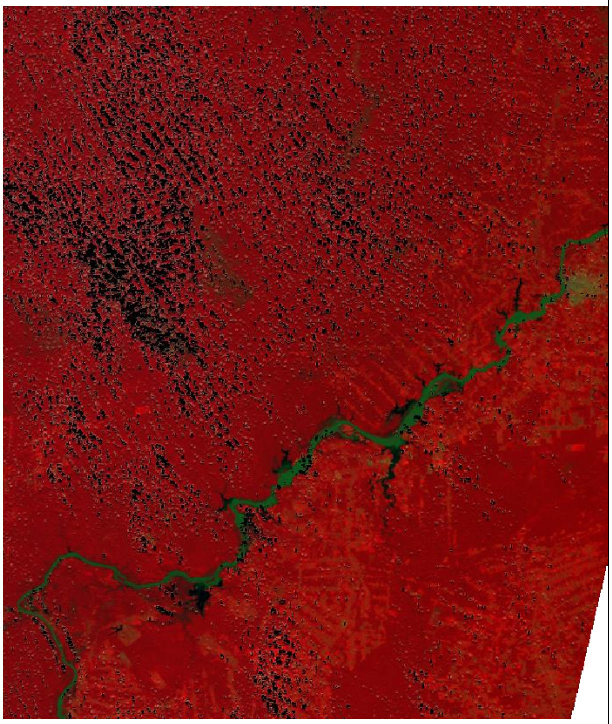
Algo 4 CLOUD + SEMI



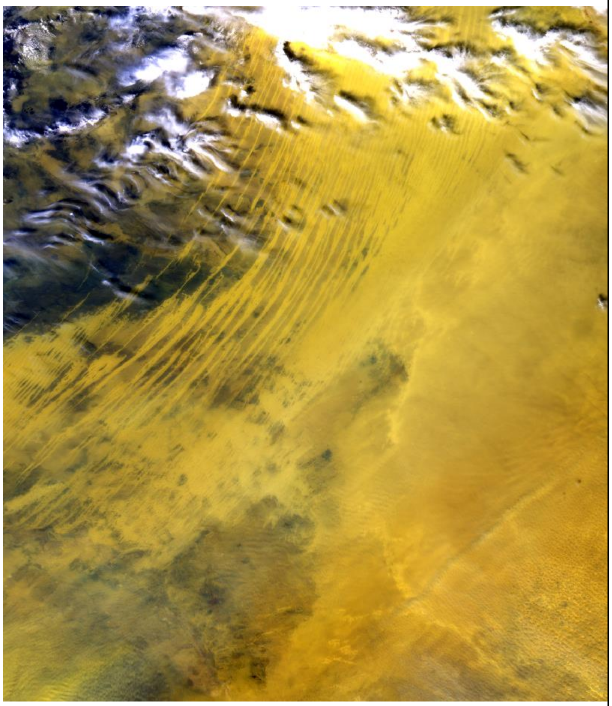
Algo 5 CLOUD



Algo 6 CLOUD + SEMI

	
Algo 7 CLOUD	Algo 8 CLOUD+SEMI

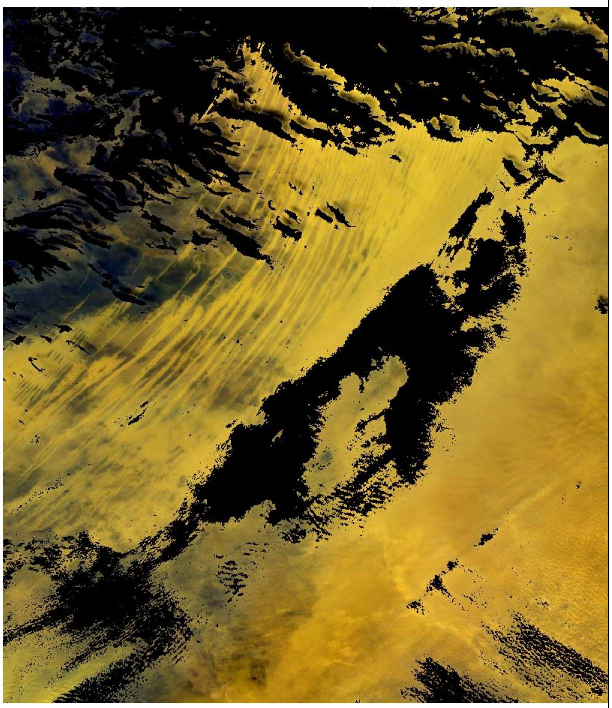
Show Case 5: Bright surfaces



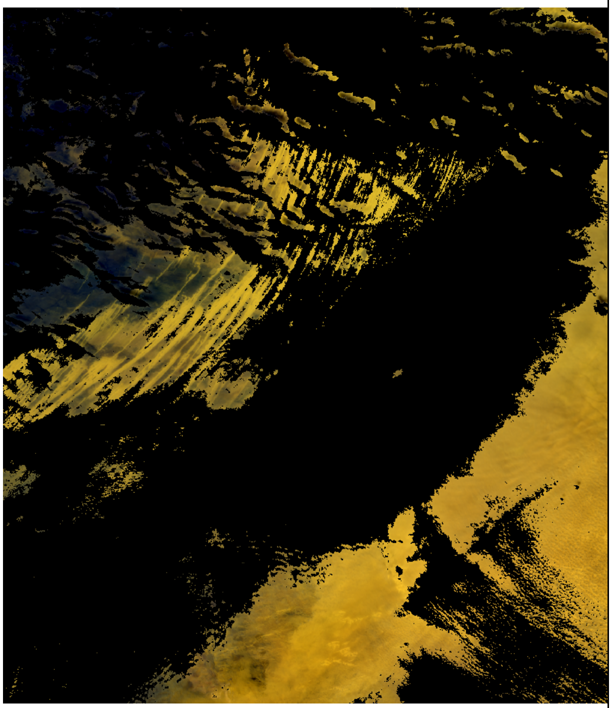
RGB

Brief description of algorithms performance:

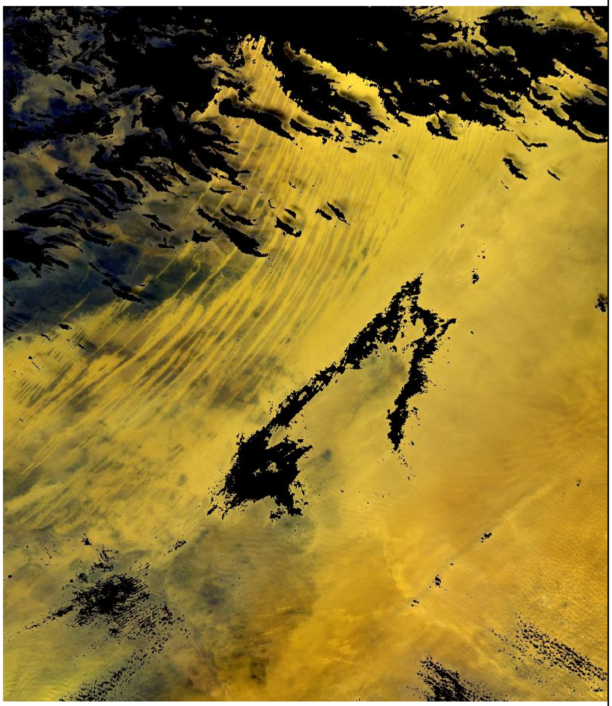
Bright surfaces are challenging for cloud detection algorithms. This show case even has semi-transparent clouds over bright surfaces. The opaque clouds in the north of the image are detected by all algorithms. Algo 8 detects least semi-transparent clouds, followed by algo 4. Both are not erroneously detecting clouds over the bright central part of the image as all other algorithms do. The image shows dust and semi-transparent cumulus clouds, which are very difficult subject for cloud detection.



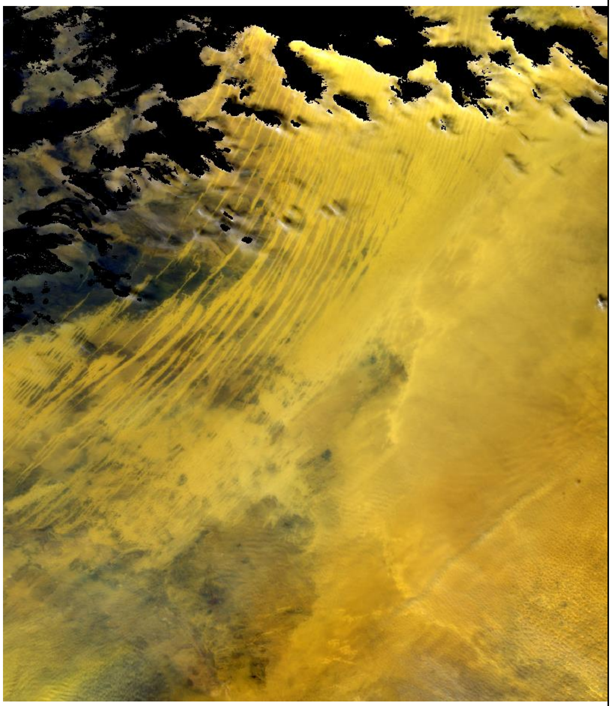
Algo 1 CLOUD



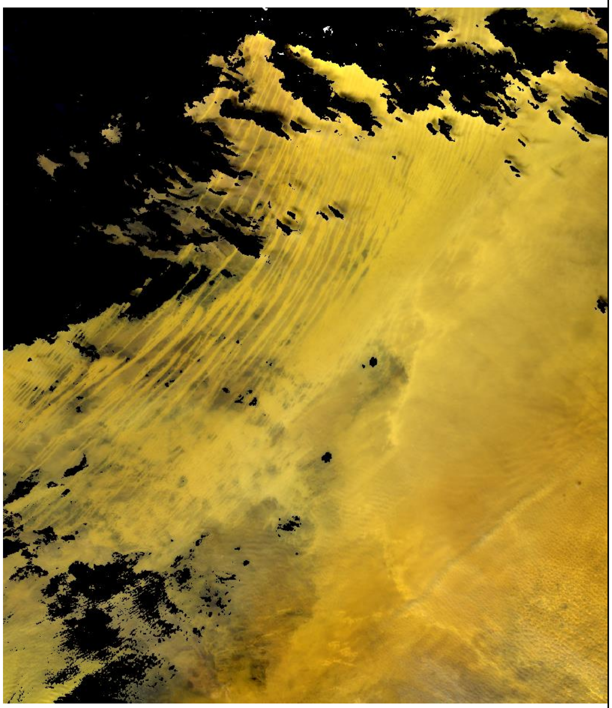
Algo 2 CLOUD



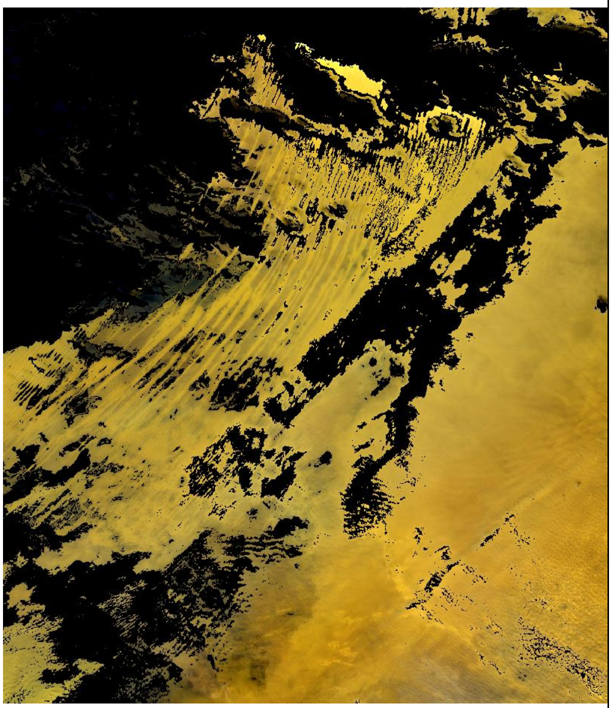
Algo 3 CLOUD



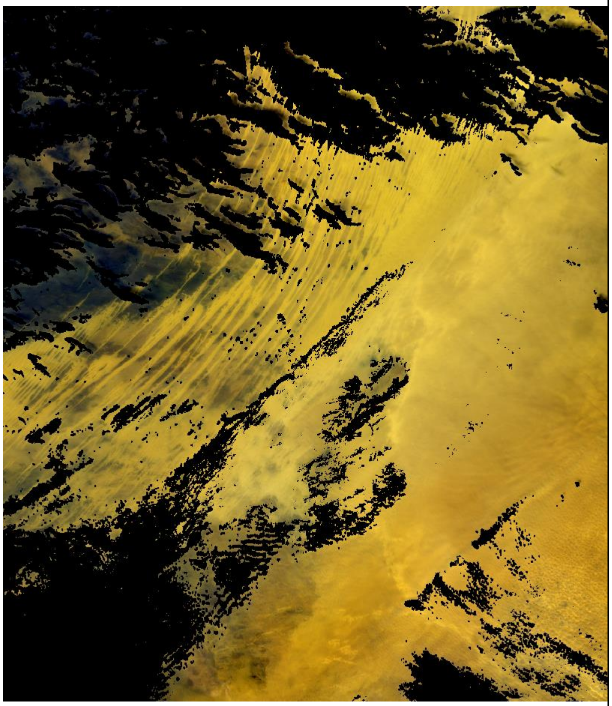
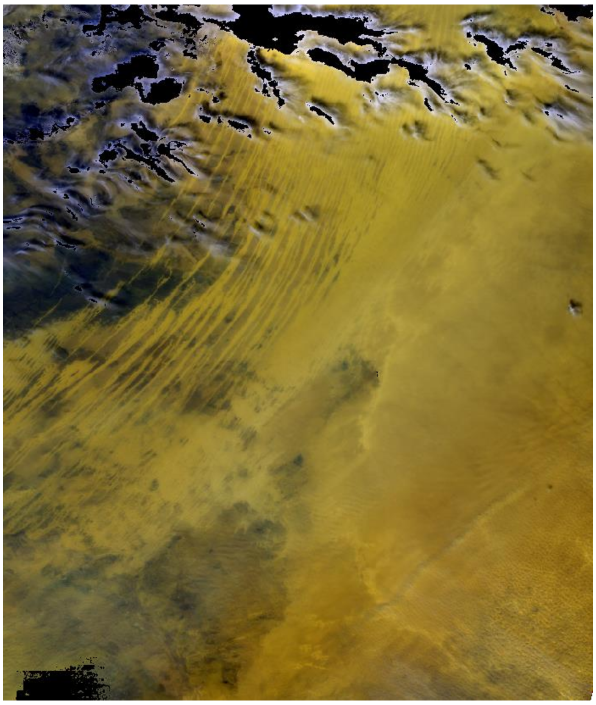
Algo 4 CLOUD + SEMI



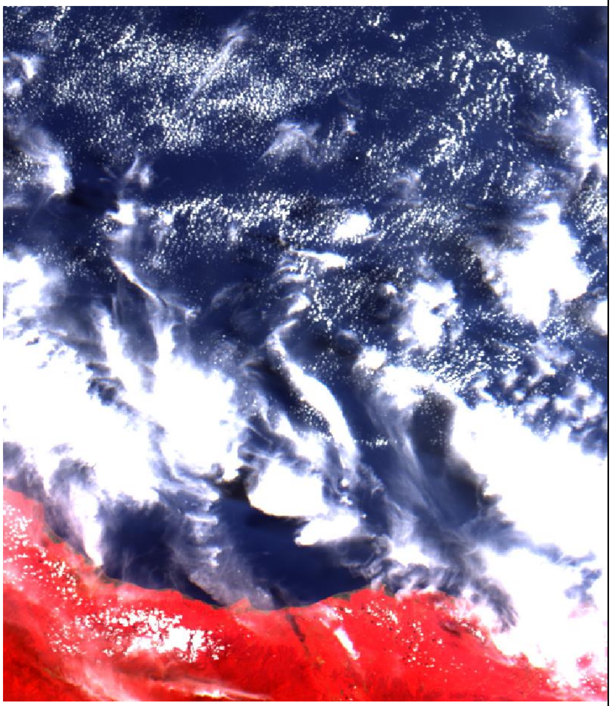
Algo 5 CLOUD



Algo 6 CLOUD + SEMI

	
Algo 7 CLOUD	Algo 8 CLOUD+SEMI

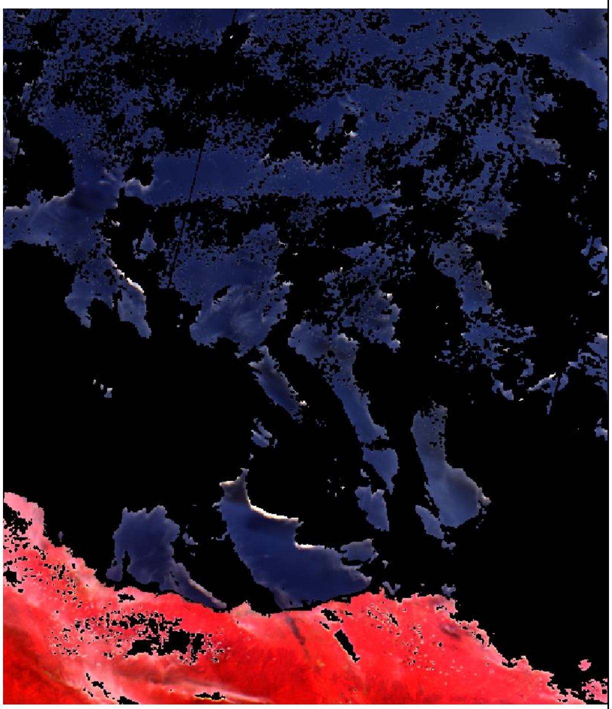
Show Case 6: Different cloud types over water, thin clouds over land



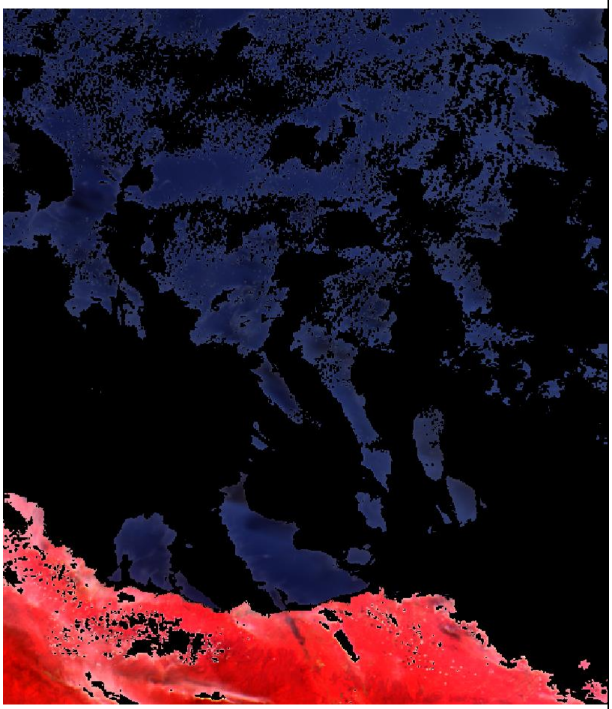
RGB

Brief description of algorithms performance:

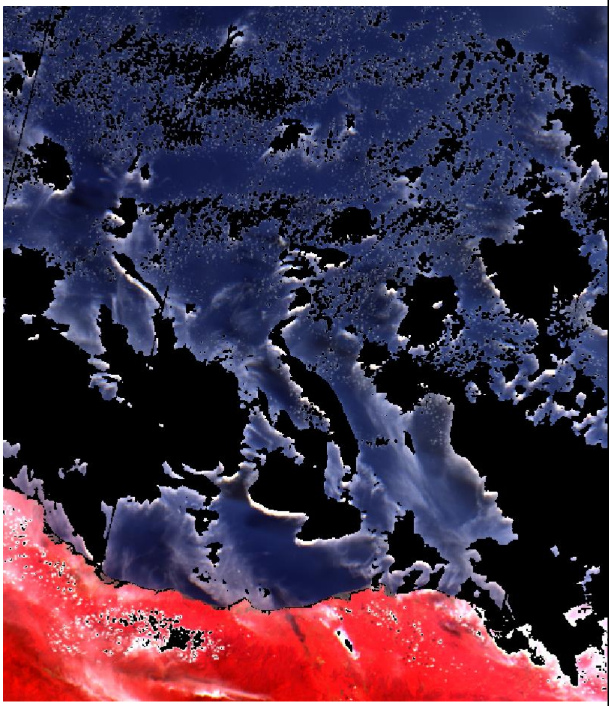
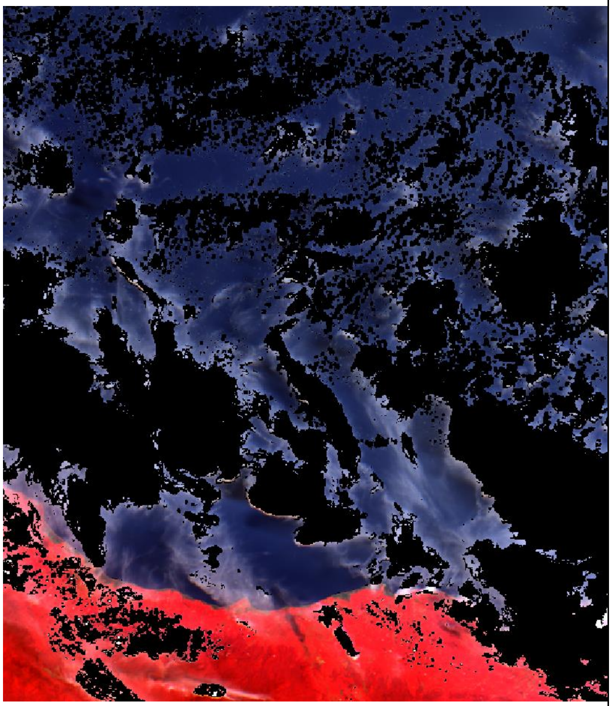
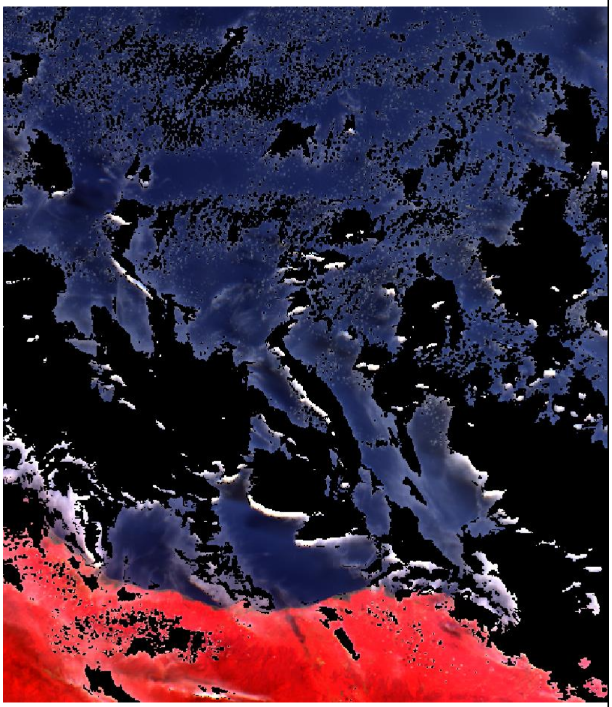
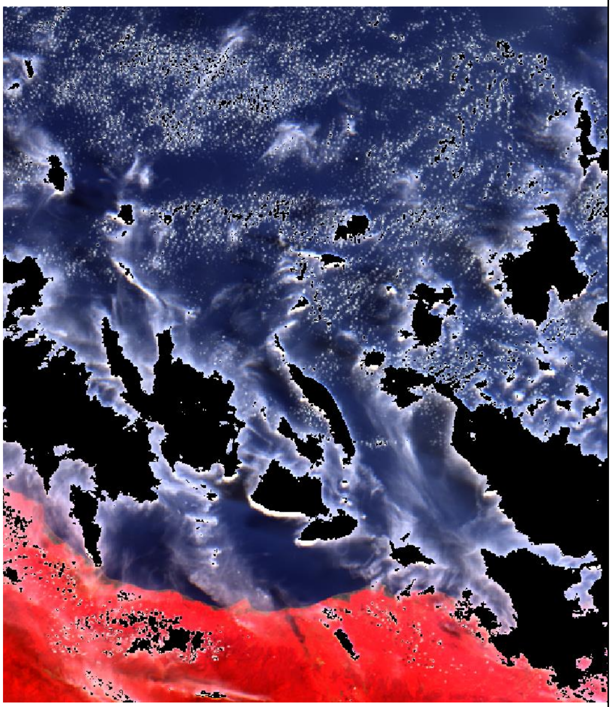
Many different cloud types with different cloud optical thickness are present in this show case, both over land and water. The opaque clouds are detected by all algorithms, but algo 7 has a margin around the clouds which are not detected. Algo 1, 3, 5 and 6 seem to have a shifted cloud flag. Semi-transparent clouds over water are detected by algo 1, 2 and 8, least detected by algo 6. The small cumulus clouds over water are detected least by algo 6, best by algo 4 and algo 8.

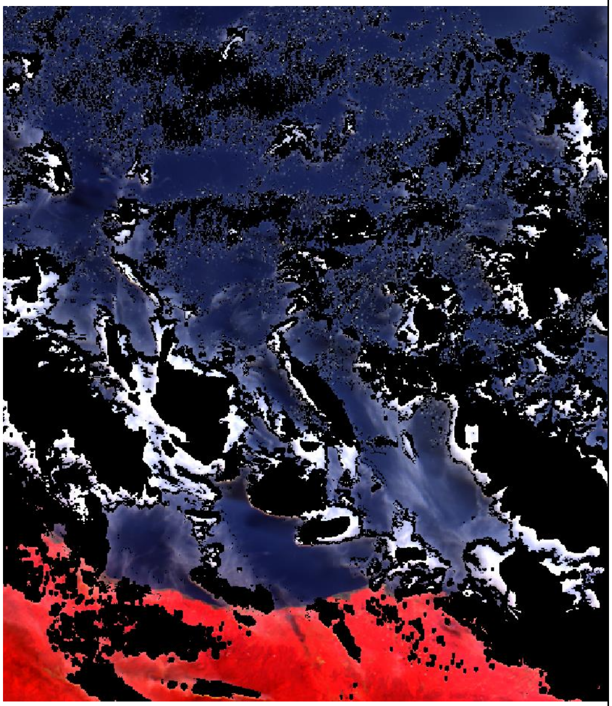
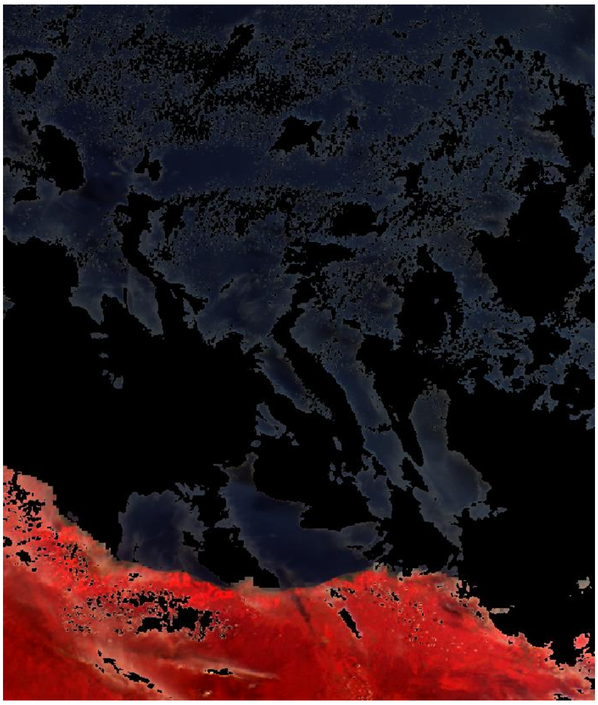


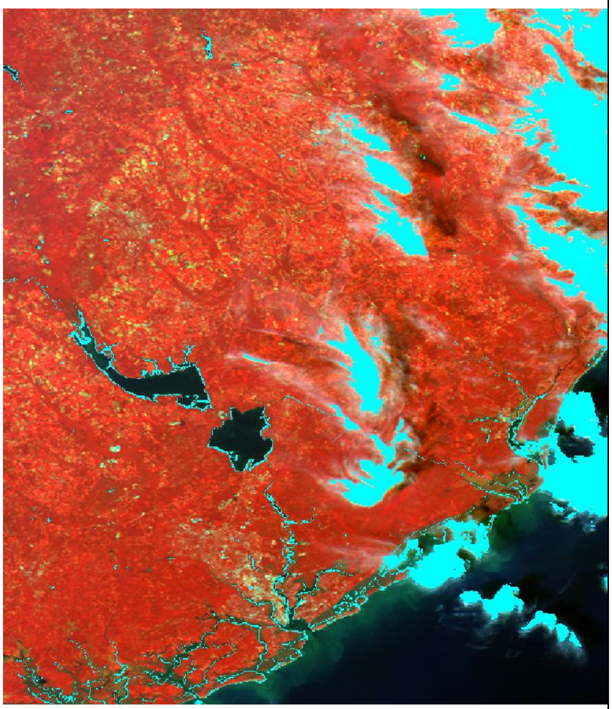
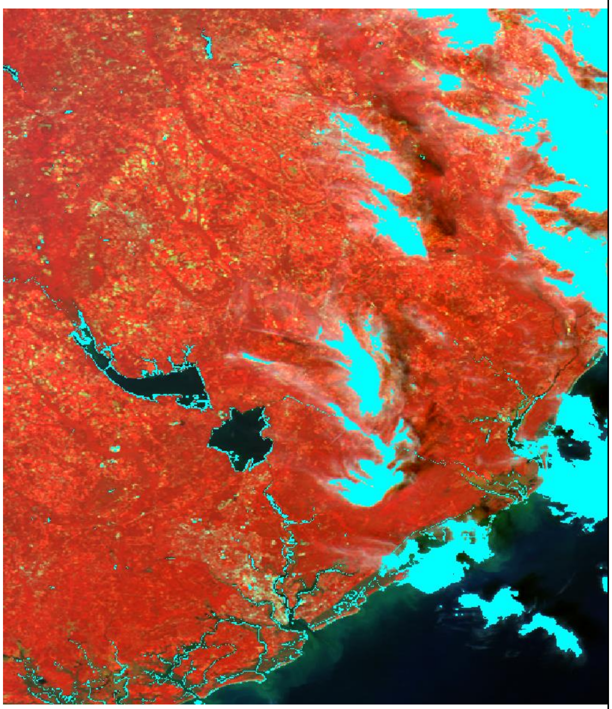
Algo 1 CLOUD

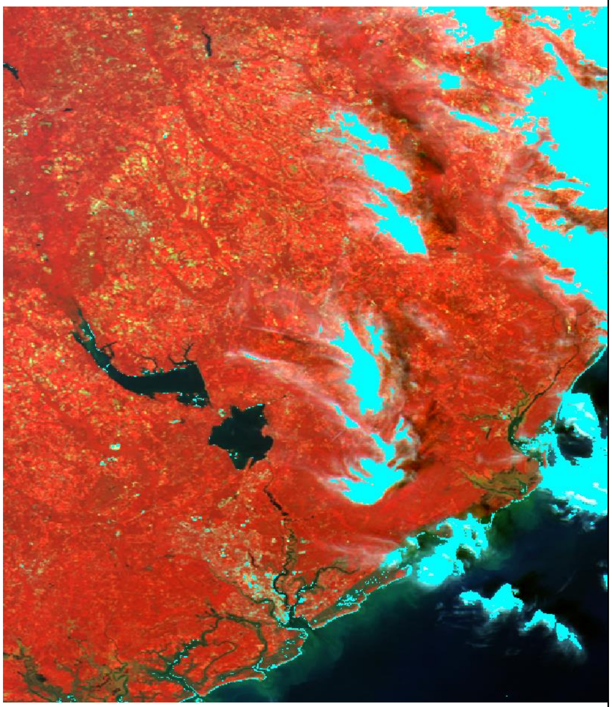
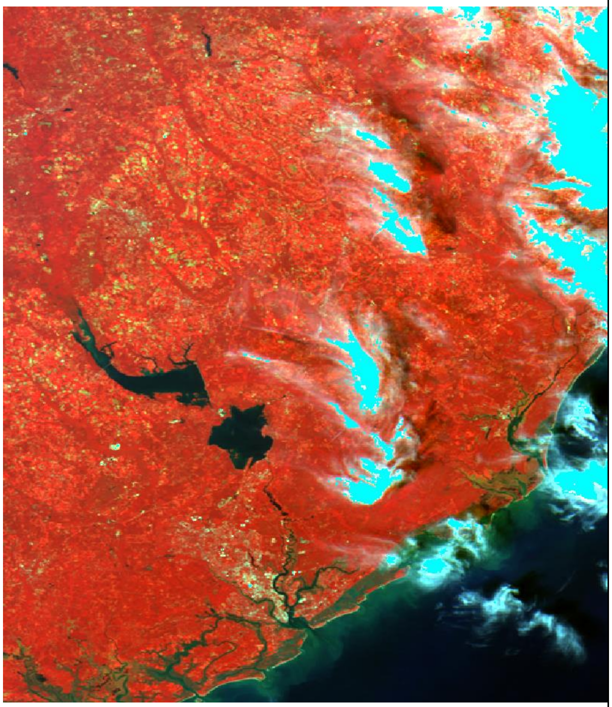


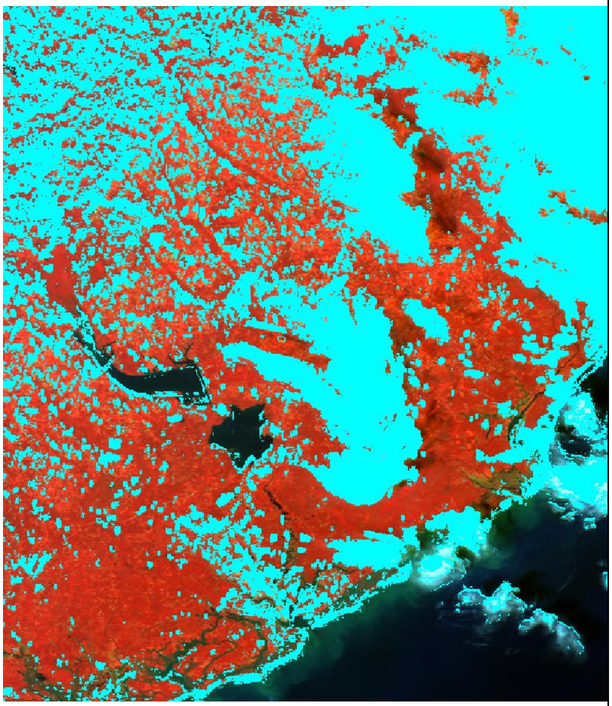
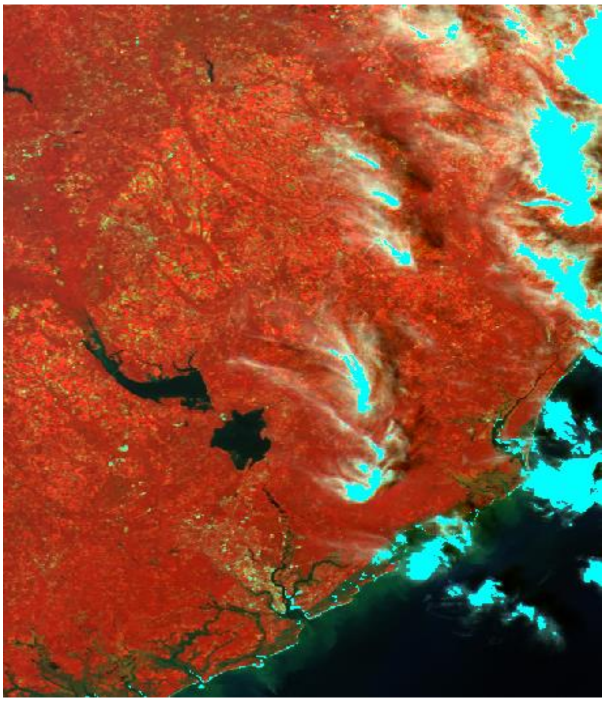
Algo 2 CLOUD

 Algo 3 CLOUD	 Algo 4 CLOUD + SEMI
 Algo 5 CLOUD	 Algo 6 CLOUD + SEMI

	
Algo 7 CLOUD	Algo 8 CLOUD + SEMI

Show Case 7: Flagging of coastlines and thin clouds	
 RGB	Brief description of algorithms performance: This last example demonstrates the behavior of the cloud flags along coastlines and lake borders. The CLOUD flag is shown in cyan for better visualization. Algo 1, 2, 5, 7 and 8 show CLOUD flag over coast-lines and except algo 8 also at inland water borders, while this is most prominent in algo 7. Algo 4 (slightly) and Algo 7 (massive) flag urban areas and algo 7 many of the agricultural areas. Thin clouds are omitted by algo 3 and algo 8 most significant, while the other algorithms only leave out the very thin semi-transparent clouds.
 Algo 1 CLOUD	 Algo 2 CLOUD

 Algo 3 CLOUD	 Algo 4 CLOUD + SEMI
 Algo 5 CLOUD	 Algo 6 CLOUD + SEMI

	
Algo 7 CLOUD	Algo 8 CLOUD+SEMI

5.1.2 Summary of observations

The observations gained from the investigation of images are summarized for each algorithm. The topics opaque clouds, small cumulus clouds, semi-transparent clouds, behaviour over different water types and over snow/ice are considered.

Algo 1

- opaque clouds are well detected
- small cumulus clouds are well detected
- detects small cumulus clouds over bright surfaces, but also large areas without clouds
- semi-transparent clouds over water well detected
- more omission of semi-transparent clouds over land than over water
- turbid water partly classified as cloud
- sun glint completely classified as cloud
- inland water partly classified as cloud
- coastlines often classified as cloud
- snow classified as cloud
- ice partly classified as cloud

Algo 2

- opaque clouds are well detected
- small cumulus clouds are well detected
- semi-transparent clouds over water well detected
- too many clouds over bright surfaces

- more omission of semi-transparent clouds over land than over water
- turbid water often classified as cloud
- sun glint classified as cloud
- inland water often classified as cloud
- coastlines often classified as cloud
- snow detected as cloud ice classified as cloud

Algo 3

- opaque clouds are well detected
- small cumulus clouds are not detected
- a lot of semi-transparent clouds are not detected
- bright surfaces partly classified as cloud
- turbid water very well detected (not classified as cloud)
- sun glint detected as cloud
- inland water ok
- coastlines partly classified as cloud
- snow mainly not classified as cloud
- ice on water not classified as cloud

Algo 4

- opaque clouds well detected, but sometimes classified as SNOW and therefore missing in CLOUD flag
- small cumulus clouds are well detected
- overall very good detection of semi-transparent clouds; more omission of clouds over water than over land
- no wrong cloud over bright surfaces, omits some small clouds and thin clouds out
- turbid water very well classified (not classified as cloud)
- high sun glint partly classified as cloud
- inland water ok coastlines ok
- dedicated SNOW flag working well

Algo 5

- opaque clouds are well detected
- small cumulus clouds are mainly detected
- semi-transparent clouds are well detected
- bright surfaces partly classified as cloud
- turbid water partly detected as cloud
- inland water often detected as cloud
- coastlines partly detected as cloud
- snow mainly not classified as cloud
- ice on water classified as cloud

Algo 6

- opaque clouds are well detected too few clouds detected over turbid waters
- small cumulus clouds are not well detected
- bright surfaces partly detected as cloud
- sun glint almost not classified as clouds
- a lot of semi-transparent clouds are not detected
- inland water very well detected
- coastlines ok
- snow not always classified as cloud
- ice on water not classified as cloud

Algo 7

- a margin is often seen around clouds between opaque and semi-transparent clouds that is not detected by the CLOUD flag
- small cumulus clouds are well detected, though large buffer around clouds are masked
- bright surfaces partly detected as cloud
- turbid water partly detected as cloud
- high sun glint detected as cloud, medium sun glint partly detected as cloud
- coastlines often detected as cloud
- dedicated SNOW flag detects snow, but ice on water partly detected as cloud

Algo 8

- opaque clouds are well detected
- small cumulus clouds are well detected
- bright surfaces not detected as cloud, clouds over bright surfaces are partly omitted
- semi-transparent clouds over water mainly well detected, over land often omitted
- turbid waters are detected as cloud
- inland water is well detected
- coastlines are partly detected as cloud
- high glint detected as cloud

5.2 Confusion matrices

5.2.1 Individual matrices CLEAR and CLOUD

The first confusion matrices show the investigation of the comparison of clear surfaces and cloudy surfaces. Here, the clear cases are cloud-free land and water surfaces, while the cloud pixels are both – opaque clouds and semi-transparent clouds. The algorithms provided different results in terms of classification, i.e. some algorithms specified semi-transparent clouds separately, others not (see Table 1). For those algorithms that provide a separate semi-transparent class, the CLOUD class in the analysis is a combination of classes of opaque cloud and semitransparent cloud.

The interpretation of the confusion matrices is shown in the following example:

ProbaV RR Algo_3 - all surfaces

In-Situ Database

CLOUDMASK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	15522	4916	20438	75.9	24.1
CLOUD+SEMI	1320	21465	22785	94.2	5.8
Sum	16842	26381	43223		
P A	92.2	81.4		OAA:	85.57
E	7.8	18.6			

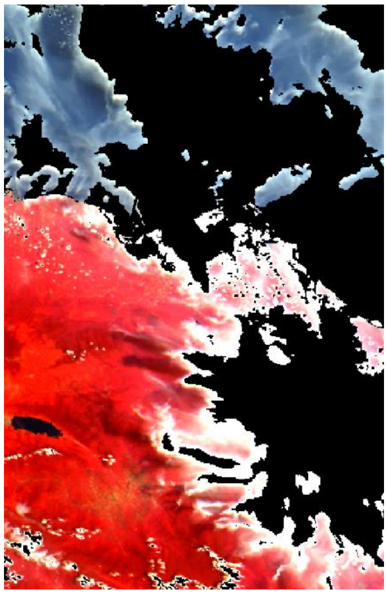


Figure 8: Confusion matrix and image for a cloud conservative algorithm (algo 3)

ProbaV RR Algo_2 - all surfaces

In-Situ Database

CLOUDMASK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	14058	1746	15804	89.0	11.0
CLOUD+SEMI	2784	24635	27419	89.8	10.2
Sum	16842	26381	43223		
P A	83.5	93.4		OAA:	89.52
E	16.5	6.6			

Figure 9: Confusion matrix and image for a clear-sky conservative algorithm (algo 2)

The numbers for Producer's Accuracy and User's Accuracy reflect very well the behavior of both algorithms. While the cloud conservative algorithm (algo 3) shows better Producers' Accuracy for the clear pixels and better Users's Accuracy for the CLOUD class, it is the opposite for the clear-sky conservative algorithm (algo 2). When algo 3 raises the CLOUD flag one can be quite sure that it is really a cloud, while it is not sure that all clouds are classified as CLOUD. Algo 3 gets most of the clouds in the CLOUD flag but also includes a lot of clear pixels.

The following figure is showing the confusion matrices for all provided algorithms. They have slightly different sums in total pixels, which is due to different flags (e.g. SNOW flag for algo 4 and 7) and that some algorithms provide an INVALID flag, which is excluding different number of pixels from the analysis.

ProbaV RR Algo_1 - all surfaces

In-Situ Database

CLOUDMASK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	13468	1866	15334	87.8	12.2
CLOUD+SEMI	2333	23633	25966	91.0	9.0
Sum	15801	25499	41300		
P A	85.2	92.7		OAA:	89.83
E	14.8	7.3			

ProbaV RR Algo_2 - all surfaces

In-Situ Database

CLOUDMASK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	14058	1746	15804	89.0	11.0
CLOUD+SEMI	2784	24635	27419	89.8	10.2
Sum	16842	26381	43223		
P A	83.5	93.4		OAA:	89.52
E	16.5	6.6			

ProbaV RR Algo_3 - all surfaces

In-Situ Database

CLOUDMASK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	15522	4916	20438	75.9	24.1
CLOUD+SEMI	1320	21465	22785	94.2	5.8
Sum	16842	26381	43223		
P A	92.2	81.4		OAA:	85.57
E	7.8	18.6			

ProbaV RR Algo_4 - all surfaces

In-Situ Database

CLOUDMAK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	16161	2854	19015	85.0	15.0
CLOUD+SEMI	688	22847	23535	97.1	2.9
Sum	16849	25701	42550		
P A	95.9	88.9		OAA:	91.68
E	4.1	11.1			

ProbaV RR Algo_5 - all surfaces

In-Situ Database

CLOUDMAK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	15697	2500	18197	86.3	13.7
CLOUD+SEMI	1145	23883	25028	95.4	4.6
Sum	16842	26383	43225		
P A	93.2	90.5		OAA:	91.57
E	6.8	9.5			

ProbaV RR Algo_6 - all surfaces

In-Situ Database

CLOUDMAK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	16069	4051	20120	79.9	20.1
CLOUD+SEMI	771	22332	23103	96.7	3.3
Sum	16840	26383	43223		
P A	95.4	84.6		OAA:	88.84
E	4.6	15.4			

ProbaV RR Algo_7 - all surfaces

In-Situ Database

CLOUDMAK

Class	Clear	Cloud+Semi	Sum	U A	E
CLEAR	13095	1655	14750	88.8	11.2
CLOUD+SEMI	2934	24140	27074	89.2	10.8
Sum	16029	25795	41824		
P A	81.7	93.6		OAA:	89.03
E	18.3	6.4			

ProbaV RR Algo_8 - all surfaces

In-Situ Database

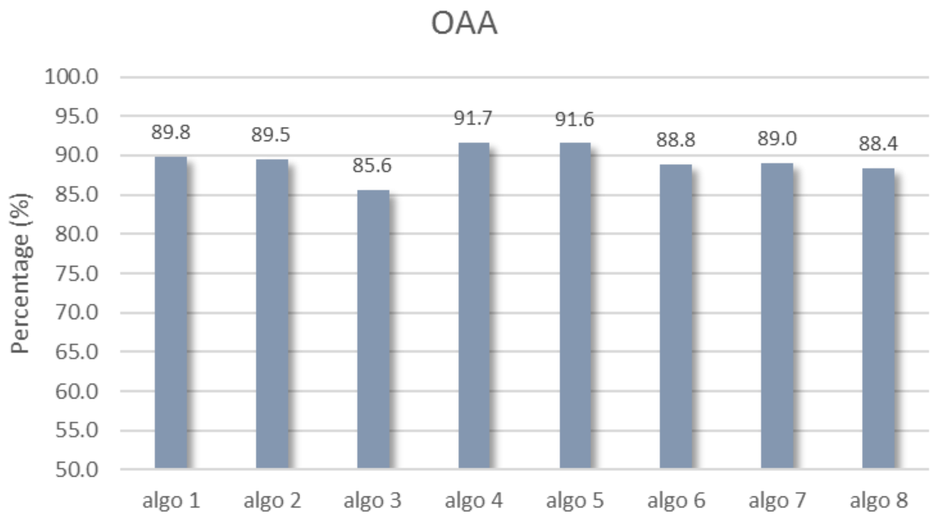
CLOUDMASK

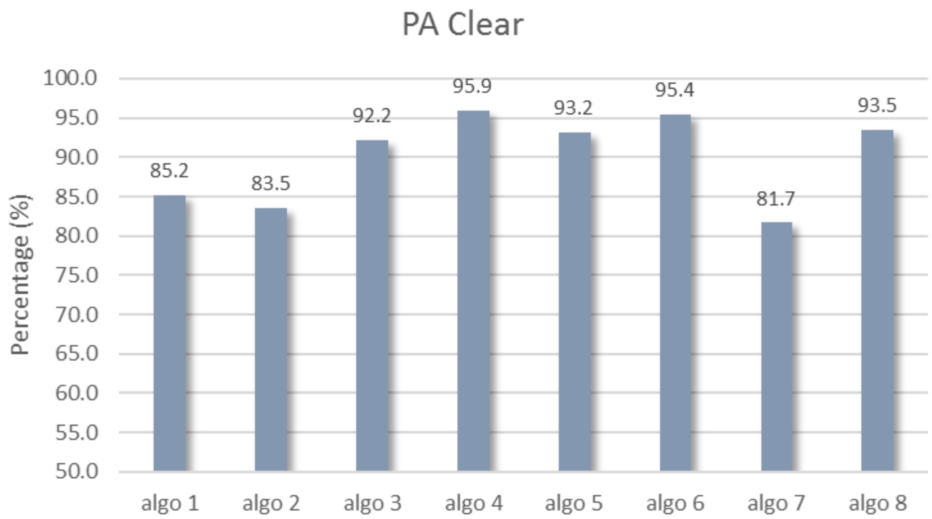
Class	Clear	Cloud	Sum	U A	E
CLEAR	15797	3576	19373	81.5	18.5
CLOUD	1093	19954	21047	94.8	5.2
Sum	16890	23530	40420		
P A	93.5	84.8		OAA:	88.45
E	6.5	15.2			

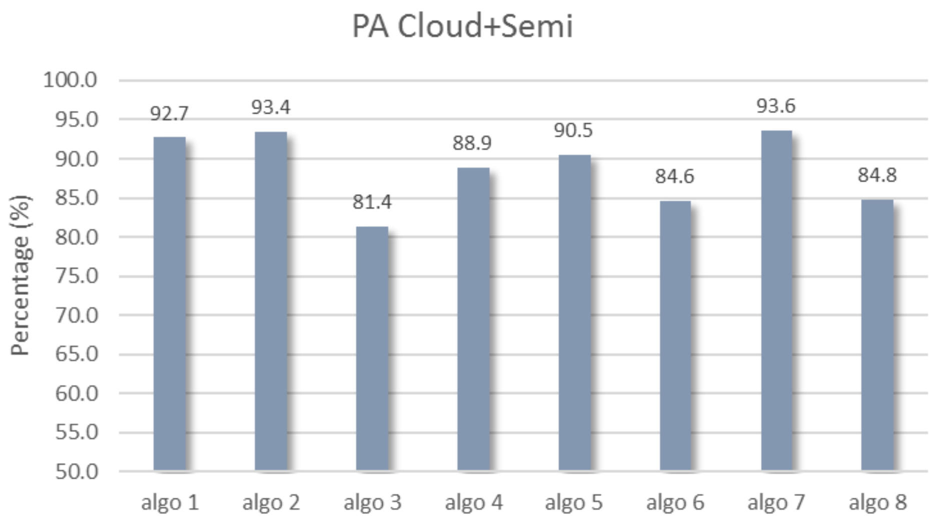
Figure 10: confusion matrices for 8 algorithms; showing the flags for CLEAR and CLOUD compared with clear and cloud manually selected pixels. The cloud pixels include the semi-transparent pixels.

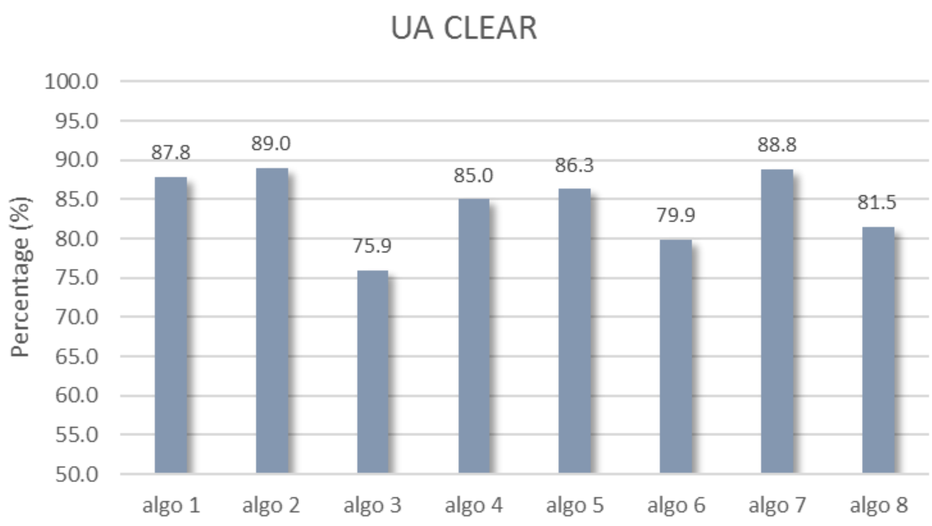
5.2.2 Compilation of accuracy measures

The single confusion matrices characterize each algorithm. For comparison of the outcome, the statistical measures are compiled in the following plots.


OAA
Percentage (%)
100.0
95.0
90.0
85.0
80.0
75.0
70.0
65.0
60.0
55.0
50.0
89.8
89.5
85.6
91.7
91.6
88.8
89.0
88.4
algo 1
algo 2
algo 3
algo 4
algo 5
algo 6
algo 7
algo 8


PA Clear
Percentage (%)
100.0
95.0
90.0
85.0
80.0
75.0
70.0
65.0
60.0
55.0
50.0
85.2
83.5
92.2
95.9
93.2
95.4
81.7
93.5
algo 1
algo 2
algo 3
algo 4
algo 5
algo 6
algo 7
algo 8


PA Cloud+Semi
Percentage (%)
100.0
95.0
90.0
85.0
80.0
75.0
70.0
65.0
60.0
55.0
50.0
92.7
93.4
81.4
88.9
90.5
84.6
93.6
84.8
algo 1
algo 2
algo 3
algo 4
algo 5
algo 6
algo 7
algo 8



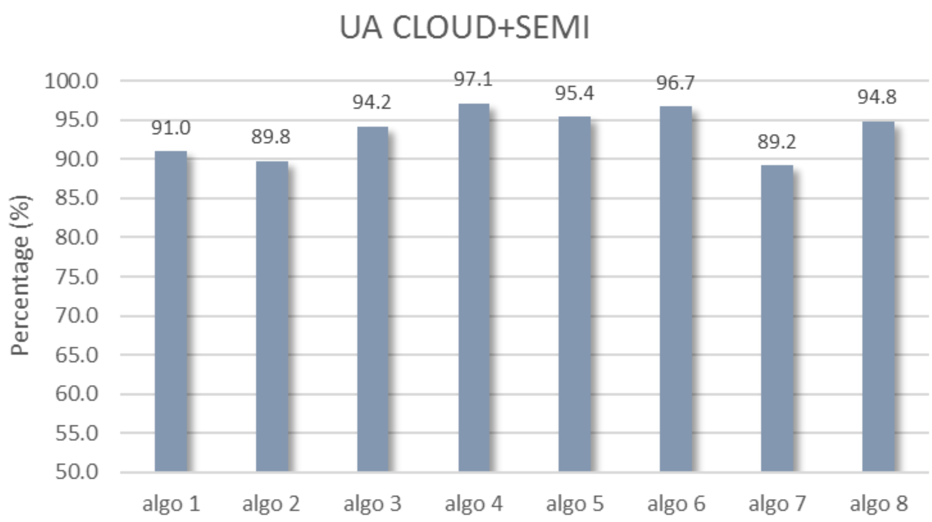


Figure 11: Statistical measures derived from the confusion matrices per algorithm

All three statistical measures for Cohens' Kappa, Krippendorf's Alpha and Scott's Pi were calculated. In all cases, the three measures had the same values. Therefore, only Krippendorf's Alpha is shown here:

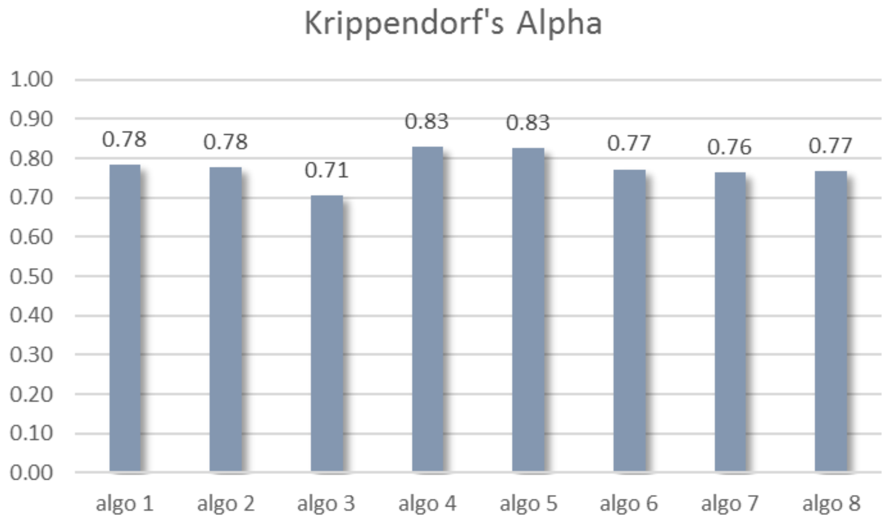


Figure 12: Krippendorf's alpha for all algorithms

Figure 13 and Figure 14 compile the above shown measures for all algorithms. While Figure 13 shows the results for all algorithms per measure, this is inverted in Figure 14: it provides the overview per algorithm. All overall accuracies are larger than 85% and two algorithms have a OAA > 90%. Algo 4 is best in the OAA, PA Clear and UA CLOUD indicating that a classified cloud is a cloud. It is followed by algo 5. Algo 8, 6 and 3 also show high values in the UA CLOUD, but compared to algo 4 and 5, they have low values in PA for clouds indicating that they are omitting clouds.

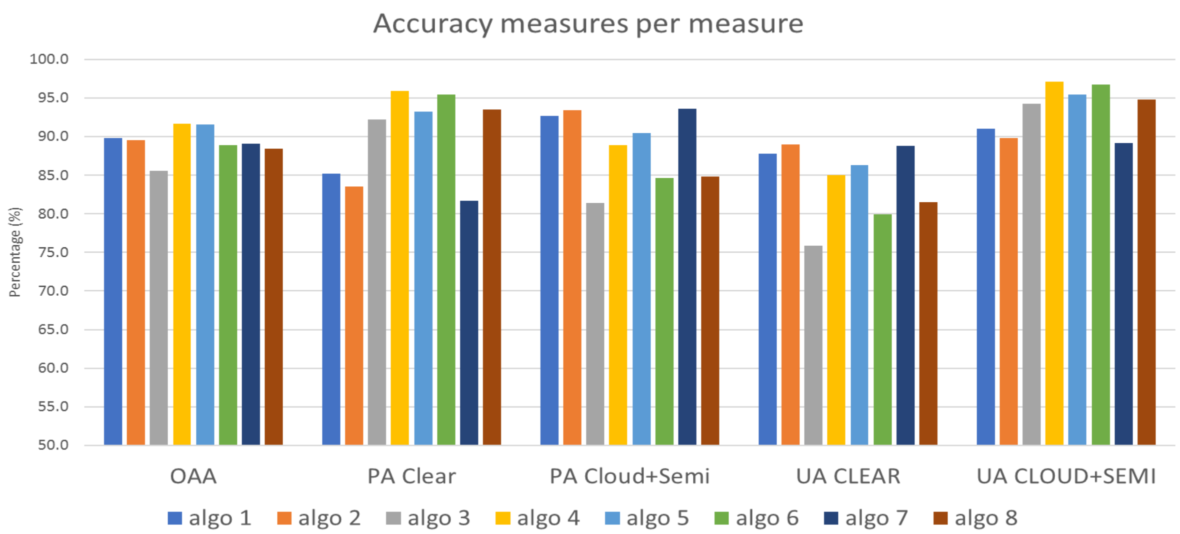


Figure 13: Accuracy measures per measure for clouds including semi-transparent clouds

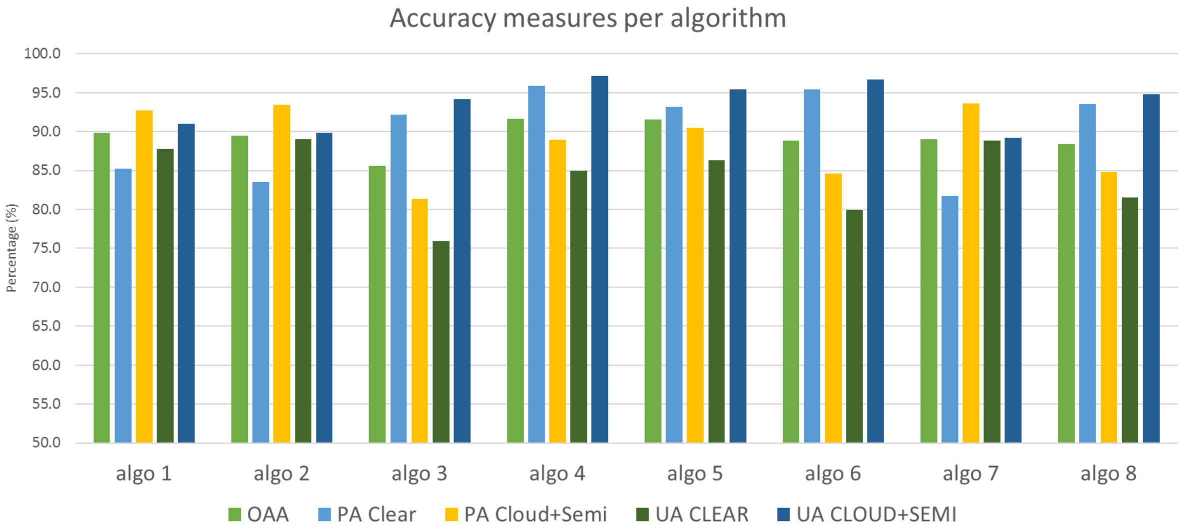


Figure 14: Accuracy measures per algorithm for clouds including semi-transparent clouds

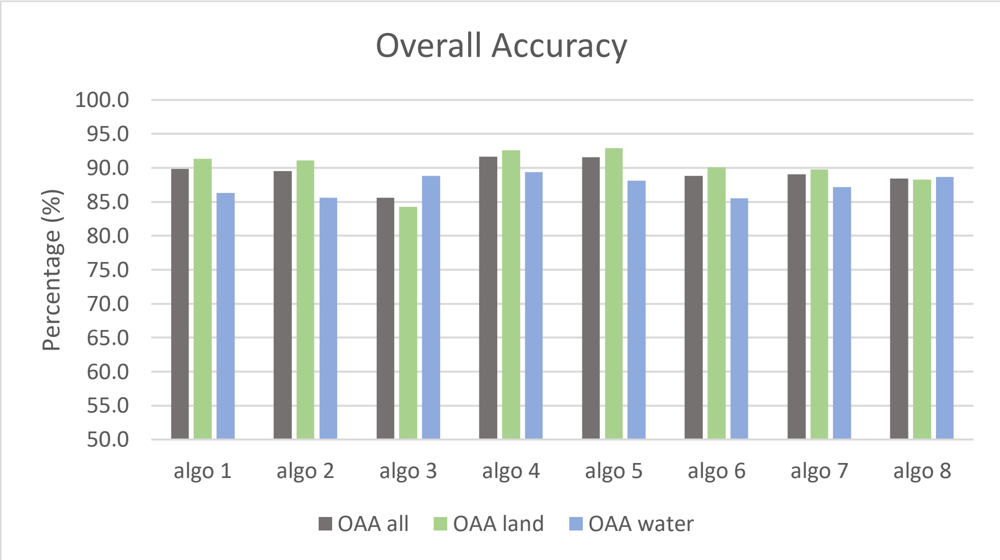
For the cloud conservative algorithms (algo 3, 6 and 8) the differences between the user's accuracy and the producer's accuracy of the same surface is large, because if they classify a cloud it is safe to assume that it is a cloud, but if they classify a clear pixel, it might be also a cloud. Vice versa – they detect a lot of the clear reference pixels but omit the reference cloud pixels. Algo 1, 2 and 7 show highest producer's accuracy for clouds, indicating that the clouds are detected with high probability but that also clear cases are classified as cloud (indicated by the lower producer' accuracy for clear).

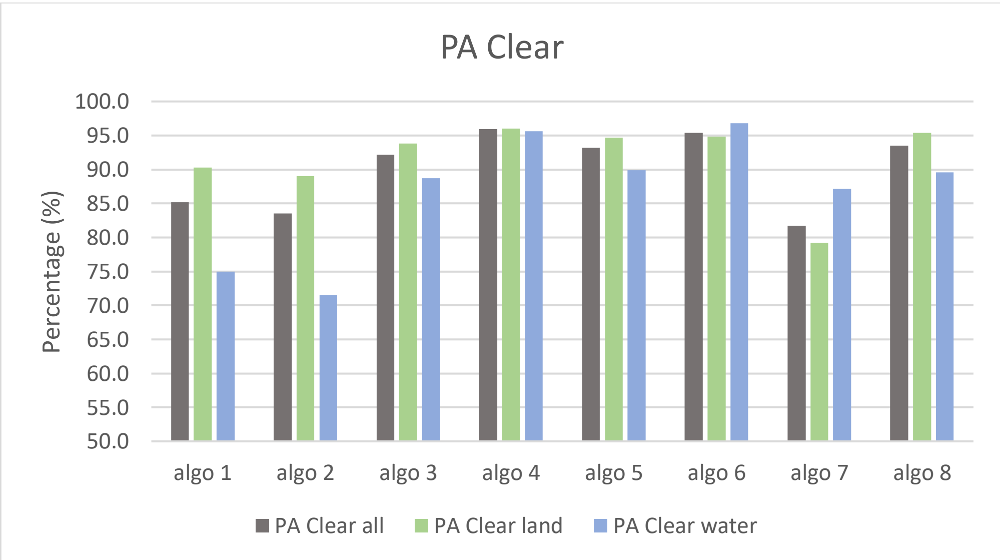
5.2.3 Separation between surfaces

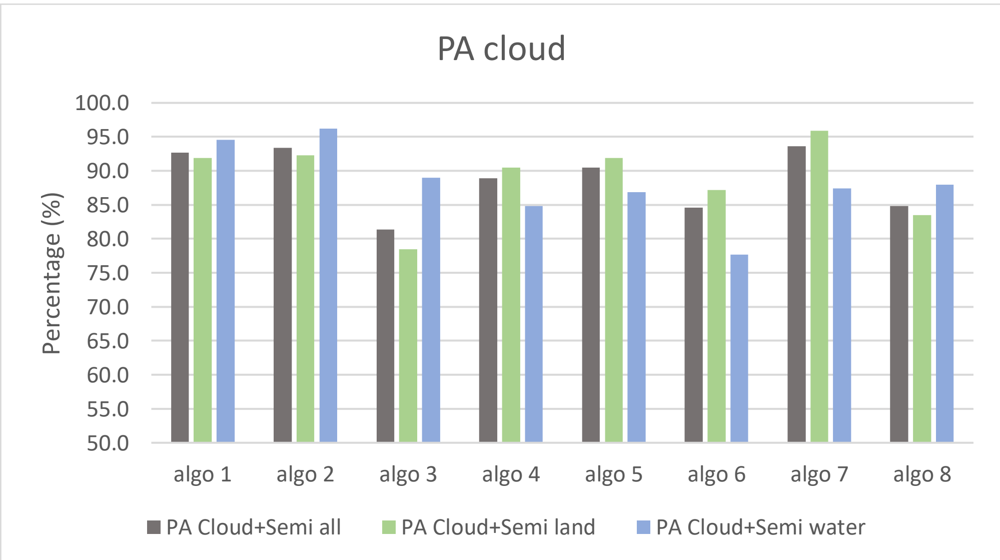
The above analysis considers all surfaces (clouds detection over land and water). The assessment of the images showed that some algorithms differ in their accuracy if the cloud is over land or over water. Therefore, an additional analysis of the statistics separated for land and water has been conducted.

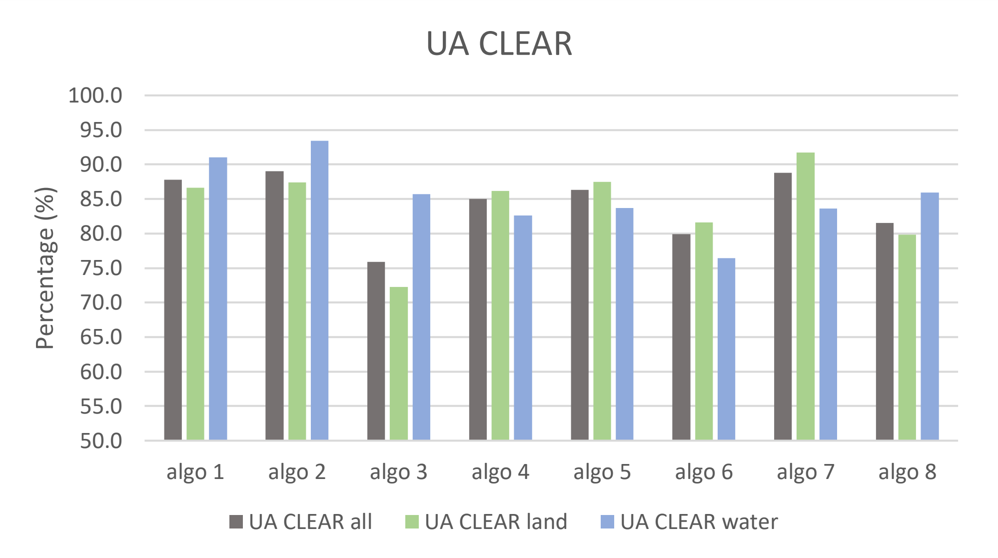
The results are shown in the following figures. For each statistical measure the accuracies are given for all surfaces (grey), for land (green) and for water (blue) for each algorithm.

Figure 15 shows that the overall cloud detection works better over land than over water in all algorithms, except for algo 3. The single statistics show that algo 1-3 detects more clouds over water than over land, while algorithms 4-7 detect less cloud over water than over land, which can be seen by e.g. the lower PA CLOUD over water (algo 4-7) compared to land. Algo 8 shows very similar results for the overall measures (OAA, Krippendorf's Alpha), while the User Accuracy and Producer Accuracies show clear differences between land and water.

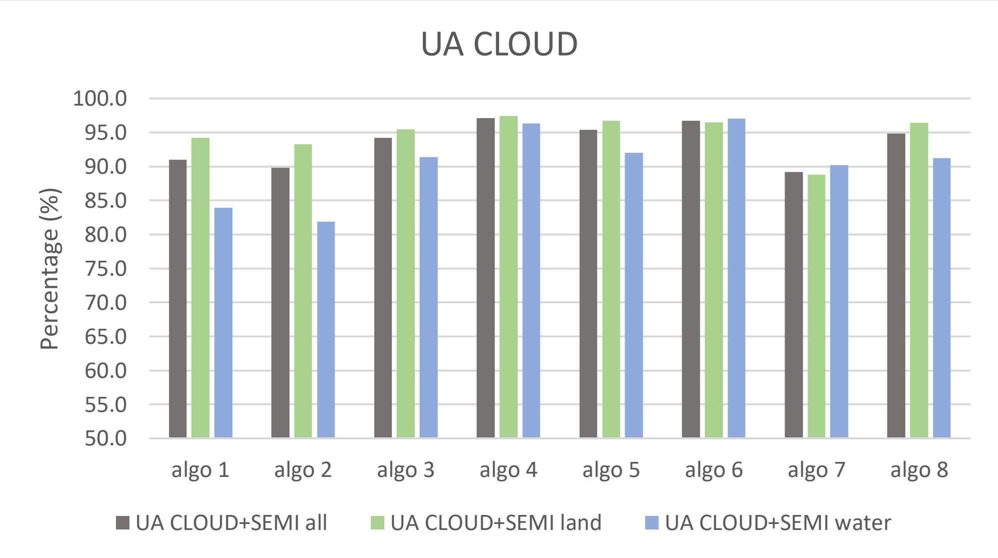

Overall Accuracy
Percentage (%)
100.0
95.0
90.0
85.0
80.0
75.0
70.0
65.0
60.0
55.0
50.0
algo 1
algo 2
algo 3
algo 4
algo 5
algo 6
algo 7
algo 8
OAA all
OAA land
OAA water


PA Clear
Percentage (%)
100.0
95.0
90.0
85.0
80.0
75.0
70.0
65.0
60.0
55.0
50.0
algo 1
algo 2
algo 3
algo 4
algo 5
algo 6
algo 7
algo 8
PA Clear all
PA Clear land
PA Clear water


PA cloud
Percentage (%)
100.0
95.0
90.0
85.0
80.0
75.0
70.0
65.0
60.0
55.0
50.0
algo 1
algo 2
algo 3
algo 4
algo 5
algo 6
algo 7
algo 8
PA Cloud+Semi all
PA Cloud+Semi land
PA Cloud+Semi water



UA CLEAR
Percentage (%)
100.0
95.0
90.0
85.0
80.0
75.0
70.0
65.0
60.0
55.0
50.0
algo 1
algo 2
algo 3
algo 4
algo 5
algo 6
algo 7
algo 8
UA CLEAR all
UA CLEAR land
UA CLEAR water



UA CLOUD
Percentage (%)
100.0
95.0
90.0
85.0
80.0
75.0
70.0
65.0
60.0
55.0
50.0
algo 1
algo 2
algo 3
algo 4
algo 5
algo 6
algo 7
algo 8
UA CLOUD+SEMI all
UA CLOUD+SEMI land
UA CLOUD+SEMI water

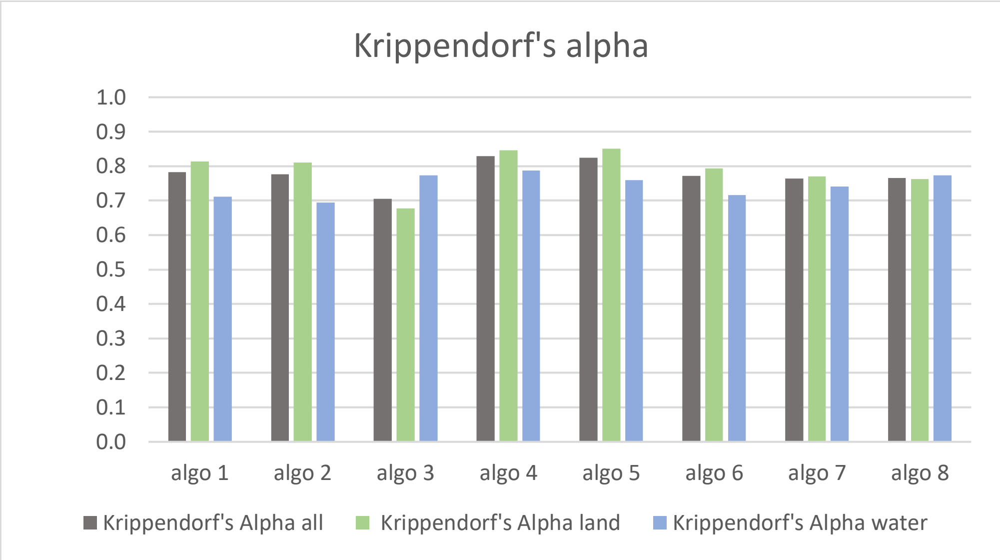


Figure 15: Statistical measures for cloud detection separated by surface (all – land – water).

5.2.4 Classification of semi-transparent clouds

It has been investigated how the algorithms behave with respect to semi-transparent clouds. While three of the provided algorithms provided a dedicated class for semi-transparent clouds, the others algorithms categorized semi-transparent clouds either as cloud or as clear. Figure 16 shows how an algorithm separates between cloud, semitransparent cloud and clear and in Figure 17 it is further investigated how the three density classes of semi-transparent clouds are classified by the algorithms.

In Figure 16 the grey portion of the bars indicates pixels classified as CLOUD, the red ones are classified as CLEAR. The beige portion is shown for algorithms that provide a SEMI-CLOUD class and have therefore a third separation.

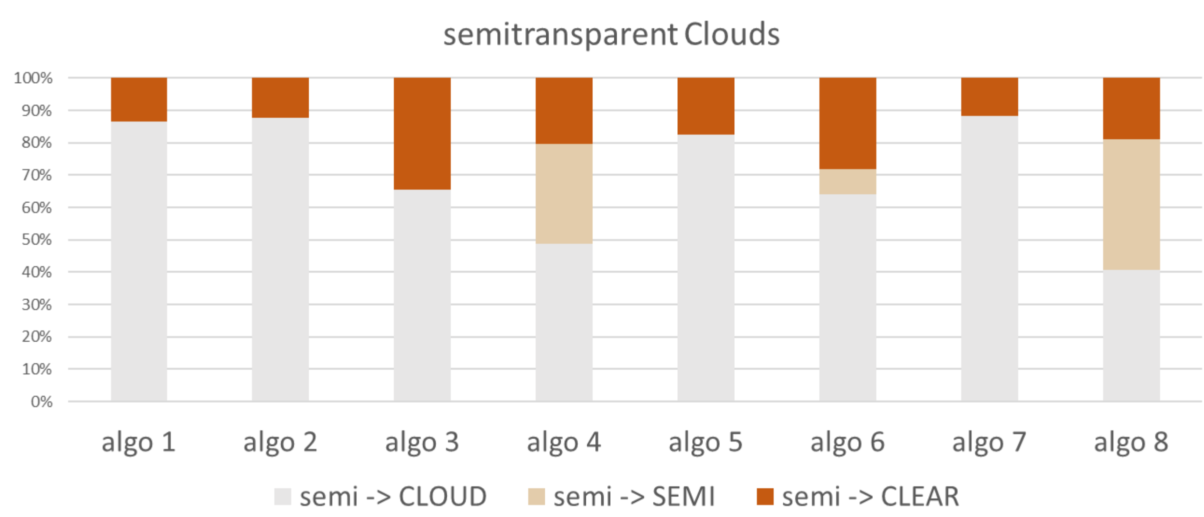


Figure 16: classification of semi-transparent clouds into the classes CLEAR and CLOUD; and SEMI-TRANS if provided by the algorithm

Figure 17 answers the question if a semi-transparent is classified as CLOUD or as CLEAR. Here, it is differentiated if the semi-transparent cloud was thin, middle or thick. The light colours indicate the percentage of pixels classified as CLOUD, the dark colour indicates that percentage of pixels classified as CLEAR. The figure shows for all algorithms that the thick semi-transparent clouds are almost all classified as CLOUD (right bar of each algorithm) while from the thin semi-transparent clouds a large portion is classified as CLEAR. This portion differs between the algorithms; e.g. algo 3 has the largest portion of not classified thin semi-transparent clouds, while algo 7 has the smallest portion classified as CLEAR. This has been also reflected in the images, while algo 7 is the most clear-sky conservative algoirthm in many cases, algo 3 and algo 6 are the least clear-sky conservative ones, followed by algo 8.

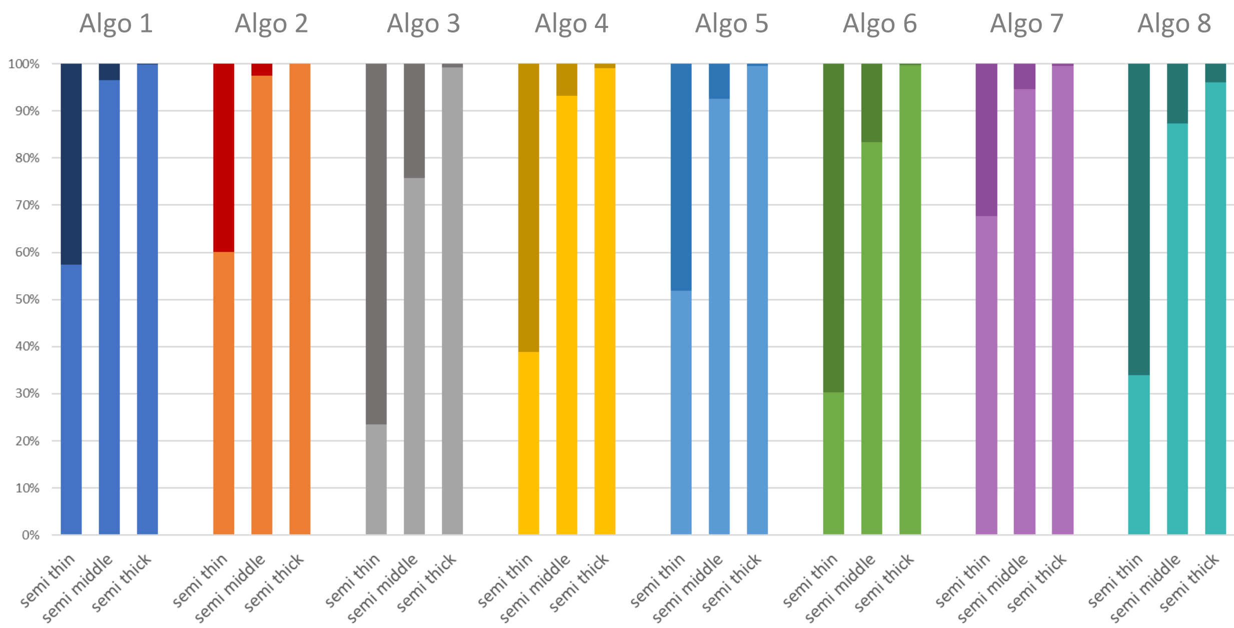


Figure 17: Classification of semi-transparent clouds; bright colours are pixels classified as CLOUD, dark colours are pixels classified as CLEAR.

Algo 4 and algo 6 have dedicated SEMI-CLOUD classes, which were added for this investigation to the cloud class for better comparability.

5.3 Algorithms intercomparison

The following matrices illustrates the similarity between the different algorithms. The numbers are the percentages of cloudy pixels (Table 2) and clear pixels (Table 3) that the algorithms have in common. It shows that algo 1 and 2 are with 99.4% agreement the most similar results for detected clouds, while algo 6 and algo 2 have less agreement (76.2%). Algo 7 has the largest agreement to all algorithms, which reflects the CLEAR conservative character of the algorithm. And vice versa, for the clear pixels in common, the CLOUD conservative algorithms have the highest agreement to all other algorithms (Table 3).

Table 2: Percentage of agreement for pixels identified as CLOUD among the algorithms

	Algo1	Algo2	Algo3	Algo4	Algo5	Algo6	Algo7	Algo8
Algo1	100.0	99.4	80.8	89.2	84.3	77.4	93.0	85.6
Algo2	99.4	100.0	76.4	87.3	82.1	74.2	91.4	84.9
Algo3	80.8	76.4	100.0	94.2	88.8	85.3	95.4	90.4
Algo4	89.2	87.3	94.2	100.0	87.0	80.2	96.6	90.3
Algo5	84.3	82.1	88.8	87.0	100.0	84.5	95.5	90.1
Algo6	77.4	74.2	85.3	80.2	84.5	100.0	98.3	92.6
Algo7	93.0	91.4	95.4	96.6	95.5	98.3	100.0	82.7
Algo8	85.6	84.9	90.4	90.3	90.1	92.6	82.7	100.0

Table 3: Percentage of agreement for pixels identified as CLEAR among the algorithms

	Algo1	Algo2	Algo3	Algo4	Algo5	Algo6	Algo7	Algo8
Algo1	100.0	88.2	98.9	89.5	93.0	98.7	74.3	90.5
Algo2	88.2	100.0	98.7	95.9	96.2	99.5	78.8	96.8
Algo3	98.9	98.7	100.0	74.2	77.6	87.4	59.3	75.7
Algo4	89.5	95.9	74.2	100.0	88.8	94.2	72.6	90.5
Algo5	93.0	96.2	77.6	88.8	100.0	94.2	65.7	82.6
Algo6	98.7	99.5	87.4	94.2	94.2	100.0	60.2	75.6
Algo7	74.3	78.8	59.3	72.6	65.7	60.2	100.0	90.5
Algo8	90.5	96.8	75.7	90.5	82.6	75.6	90.5	100.0

5.4 Specialties of individual algorithms

5.4.1 SEMI-CLOUD flag

Three algorithms have in addition to the cloud flag also the semitransparent cloud flag. This is a plus of an algorithm because it provides the user more options how to flag the clouds. If the user is more interested in clear-sky conservative cloud flag, the SEMI-CLOUD flag can be added to the CLOUD flag and if he is more interested in gaining as much pixels as possible, it can be excluded. In most of the above analyses, the SEMI-CLOUD flag has been added to the CLOUD flag in order to have comparable results with the algorithms that do not provide this additional flag.

The SEMI-CLOUD flag is available for algo 4, algo 6 and algo 8.

5.4.2 SNOW flag

Two algorithms provide a dedicated SNOW flag, providing users the possibility to separate between cloud and snow pixels. It has been observed during the analysis of the images that the SNOW flag was also raised within CLOUDS, which causes problems when only using the CLOUD flag. Algo 4 provides in addition a flag for thin snow layers. It is not further taken into account within this analysis.

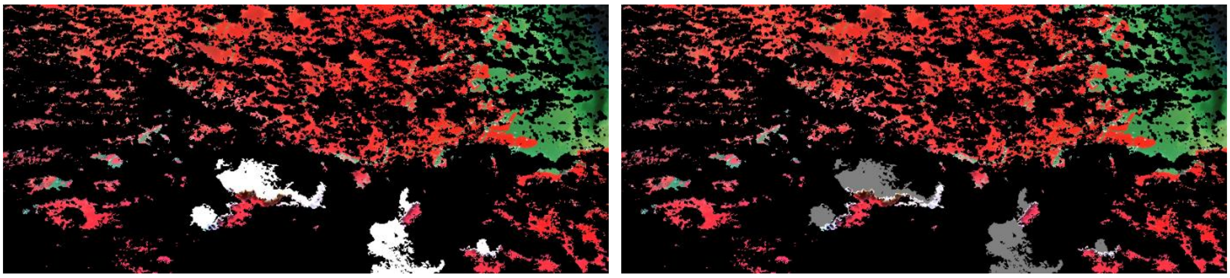


Figure 18: Clouds are not covered by the CLOUD flag but the SNOW flag in algo 4 (grey areas in the right image).

It seems that the SNOW flag is raised in cases where all or part of the ProbaV GOOD flags are OFF.. But this is not the only reason. Figure 19 shows the example with the ProbaV GOOD flags raised (upper right) and the behaviour of the CLOUD flag in algo 4 (lower left) and algo 7 (lower right). This should be further investigated by the algorithm developers. It was discussed that the SNOW flag is raised for ice clouds due to spectral similarity to ice

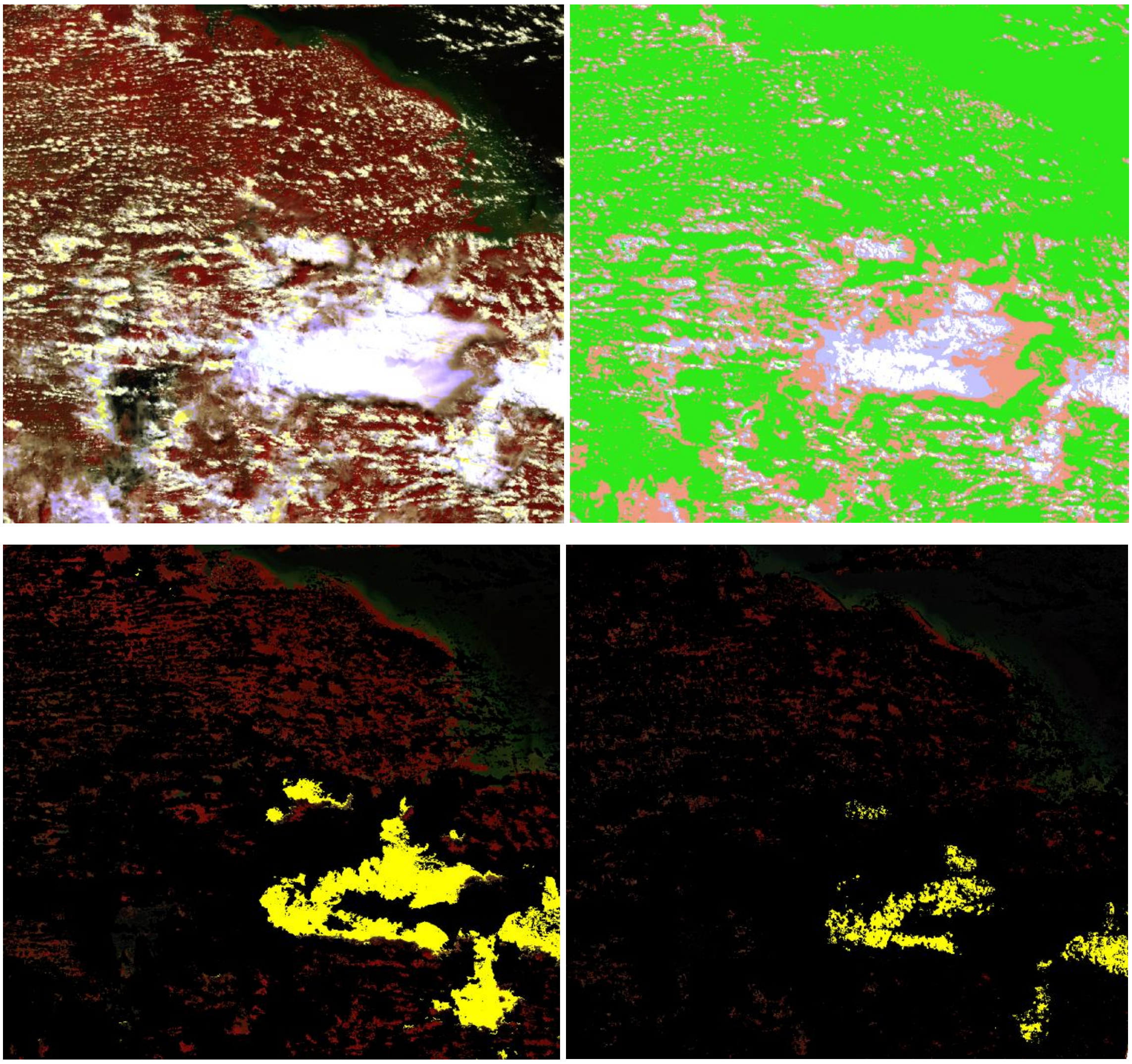


Figure 19: Opaque clouds detected as SNOW when GOOD flag is not raised: upper left: RGB, upper right: GOOD flags (green: GOOD BLUE, orange: GOOD RED, violet: GOOD NIR); lower left: algo 4, lower right: algo 7 (CLOUD flag in black, SNOW flag in yellow)

5.4.3 Artefacts in CLOUDMASK

Algo 5 shows in some images a horizontal or vertical line in the CLOUD flag.

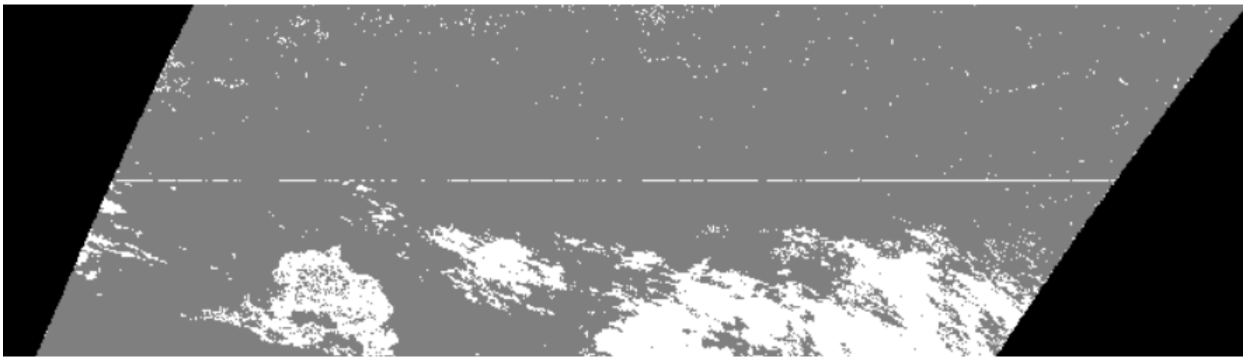


Figure 20: Algorithm 5 PROBAV_L2A_20140321_044547_1_333M_V001_subset3

6 Summary

This report provides an overview on the performance of the eight cloud detection algorithms. The results of the statistical and visual inspection of the different CLOUDMASKs show the performance of the different algorithms for cloud detection under different aspects. While the images are well suited for visual impression and individual interpretation, the statistics provide a measurable result.

Overall all algorithms are of very good quality. Each algorithm has strong and weak points. Algo 4 and algo 5 show very high values in the statistics. While algo 5 is slightly more cloud conservative, algo 4 is leaving out thick and bright clouds from the cloud mask. Algo 1 and algo 2 are very similar, algo 2 is more clear-sky conservative as algo 1 and flagging clear surfaces as cloud while having a good detection of semi-transparent clouds. Algo 3 and 6 are the ones detecting least clouds, but have very little commission errors, e.g. at coastlines or in inland waters or bright surfaces. Algo 7 is most clear-sky conservative and therefore detects small clouds, cloud borders and semi-transparent clouds but is flagging many clear pixels. Further, it shows artefacts at cloud borders over water. Algo 8 tends to be a cloud conservative algorithm over land, and a clear conservative algorithms over water.

Table 4: Summary assessment of the 8 algorithm of he Round Robin

	Algo 1	Algo 2	Algo 3	Algo 4	Algo 5	Algo 6	Algo 7	Algo 8
opaque clouds	1	1	1	2	1	1	4	1
small cumulus clouds	1	1	3	1	2	3	1	2
semi-trans water	1	1	3	2	1	3	1	1
semi-trans land	2	2	2	1	1	3	1	2
bright surfaces	2	3	2	1	2	2	2	1
turbid water	2	3	1	1	2	1	2	3
sun glint	3	3	3	2	1	1	2	2
inland water	2	3	1	1	3	1	3	1
coastline	3	3	2	1	2	1	3	1

1	well detected
2	Slightly or partly misclassification
3	Major misclassification
4	Serious problem

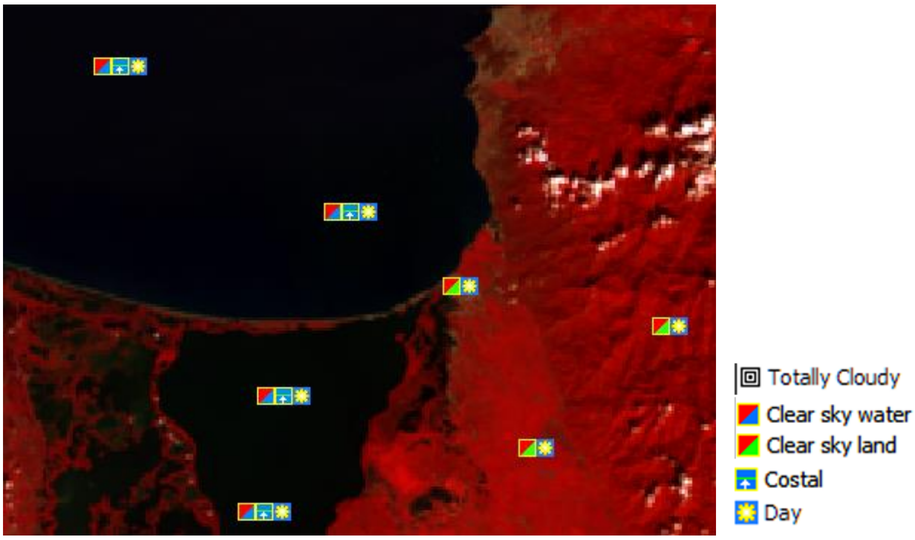
7 ANNEX

7.1 Annex I: Examples for the pixels within the validation 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:

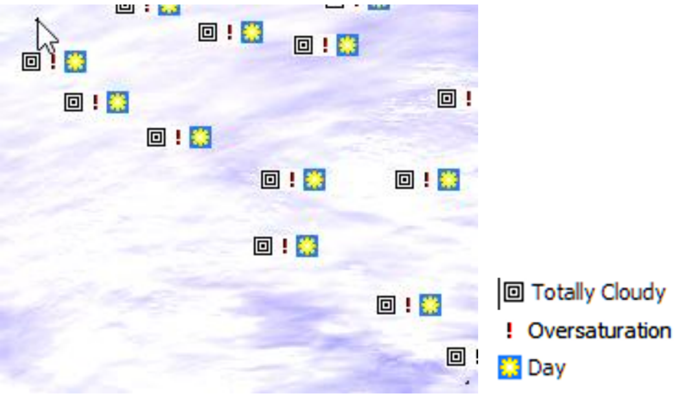


Figure 21: examples of clear surface (land and water and ice) and totally cloudy pixels (below).

Semi-Transparent cases over land:

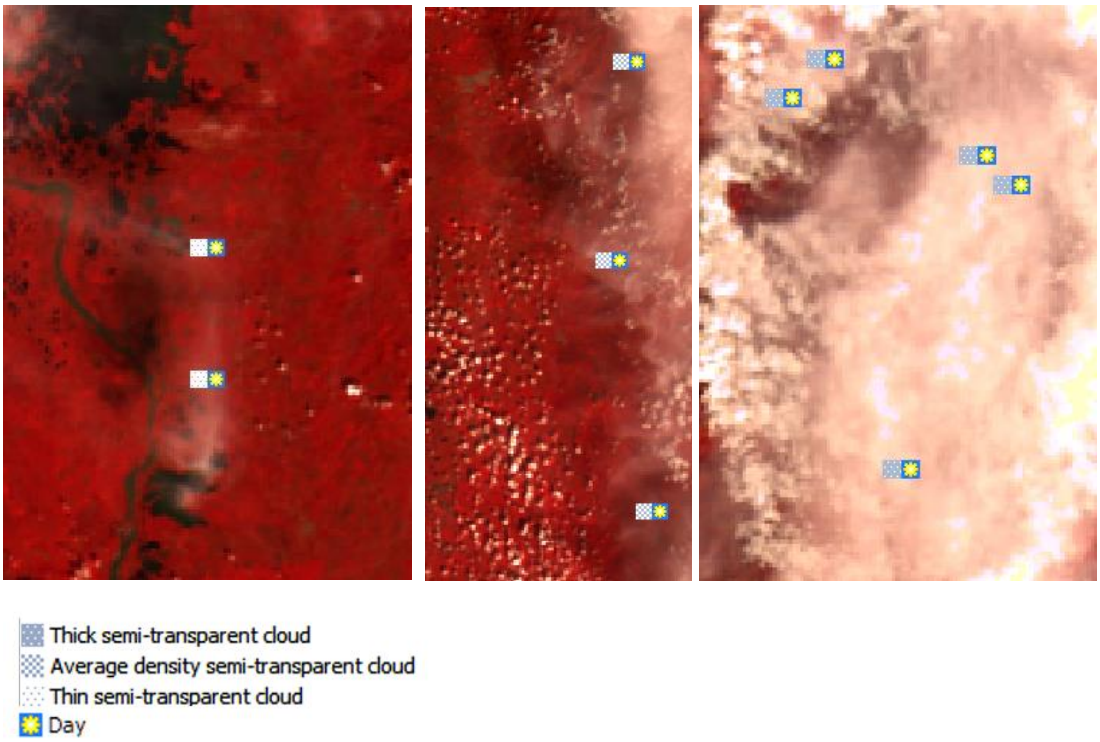


Figure 22: examples for semi-transparent clouds over land: left: thin; middle: medium; right: dense

Semi-Transparent cases over water:

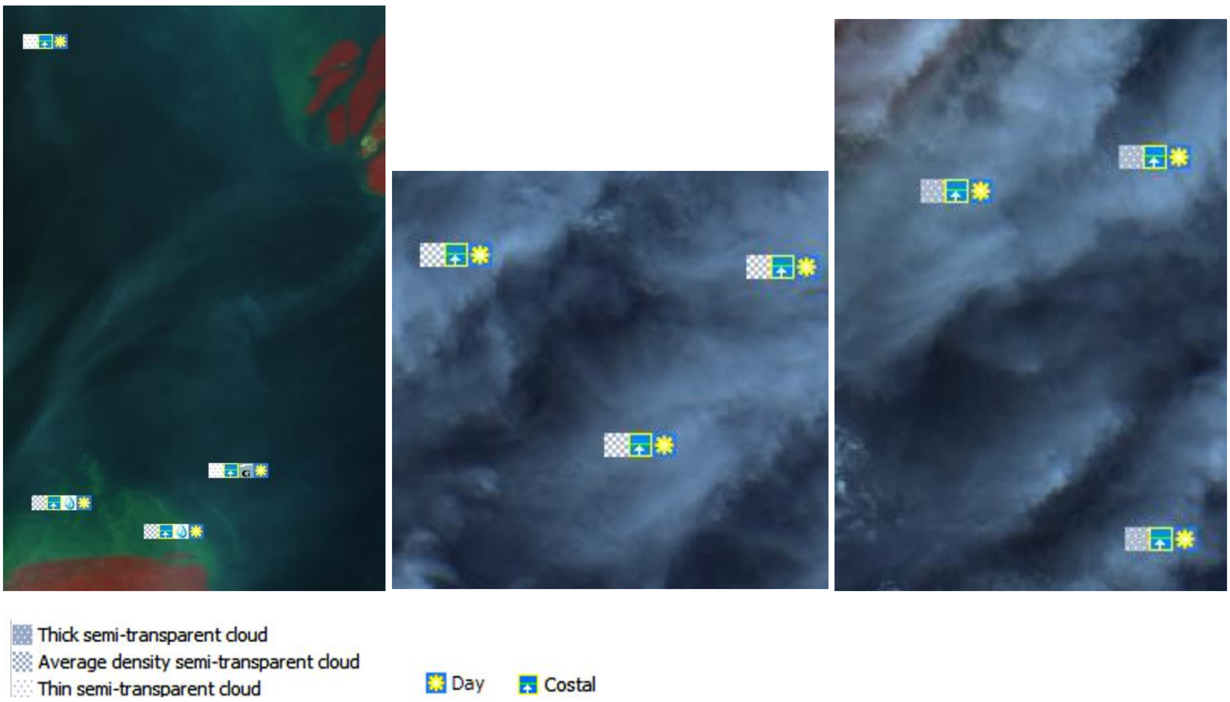


Figure 23: examples for semi-transparent clouds over water: left: thin; middle: medium; right: dense

Semi-transparent clouds over ice:

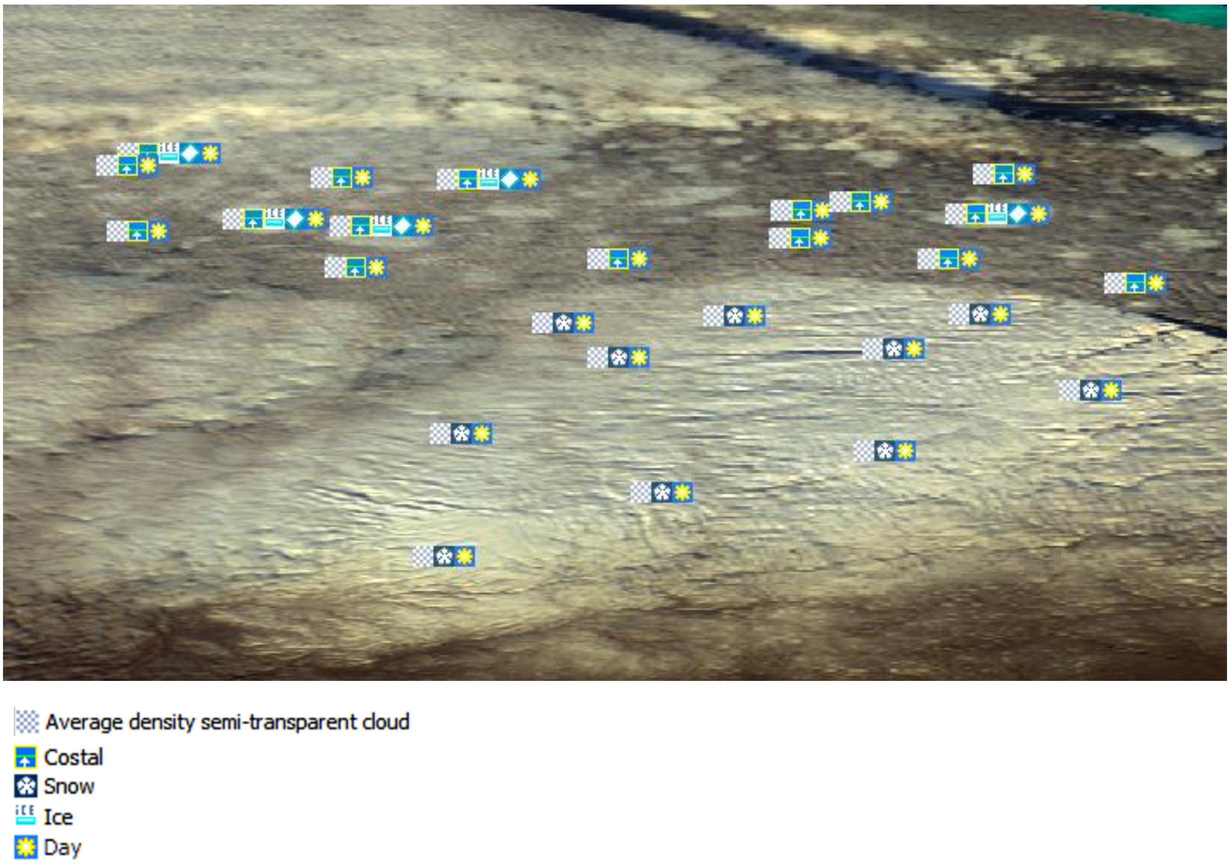


Figure 24: examples for semi-transparent clouds over ice (medium)

Spatially mixed cases:

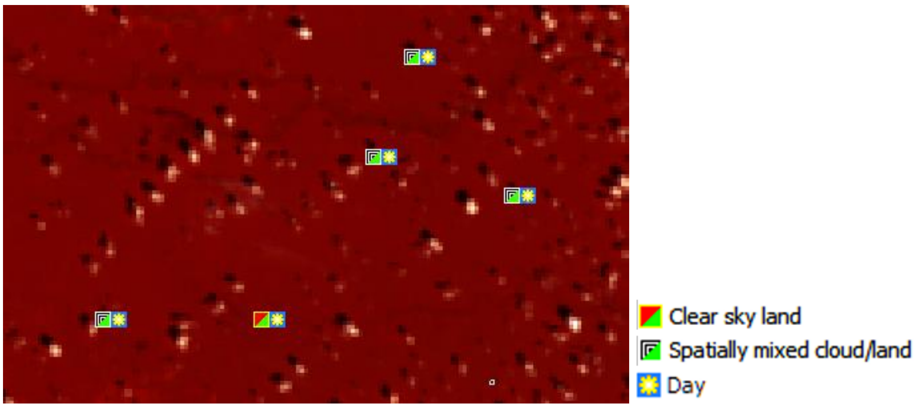


Figure 25: examples of spatially mixed pixels cloud/land

And some more examples over land of different cases:

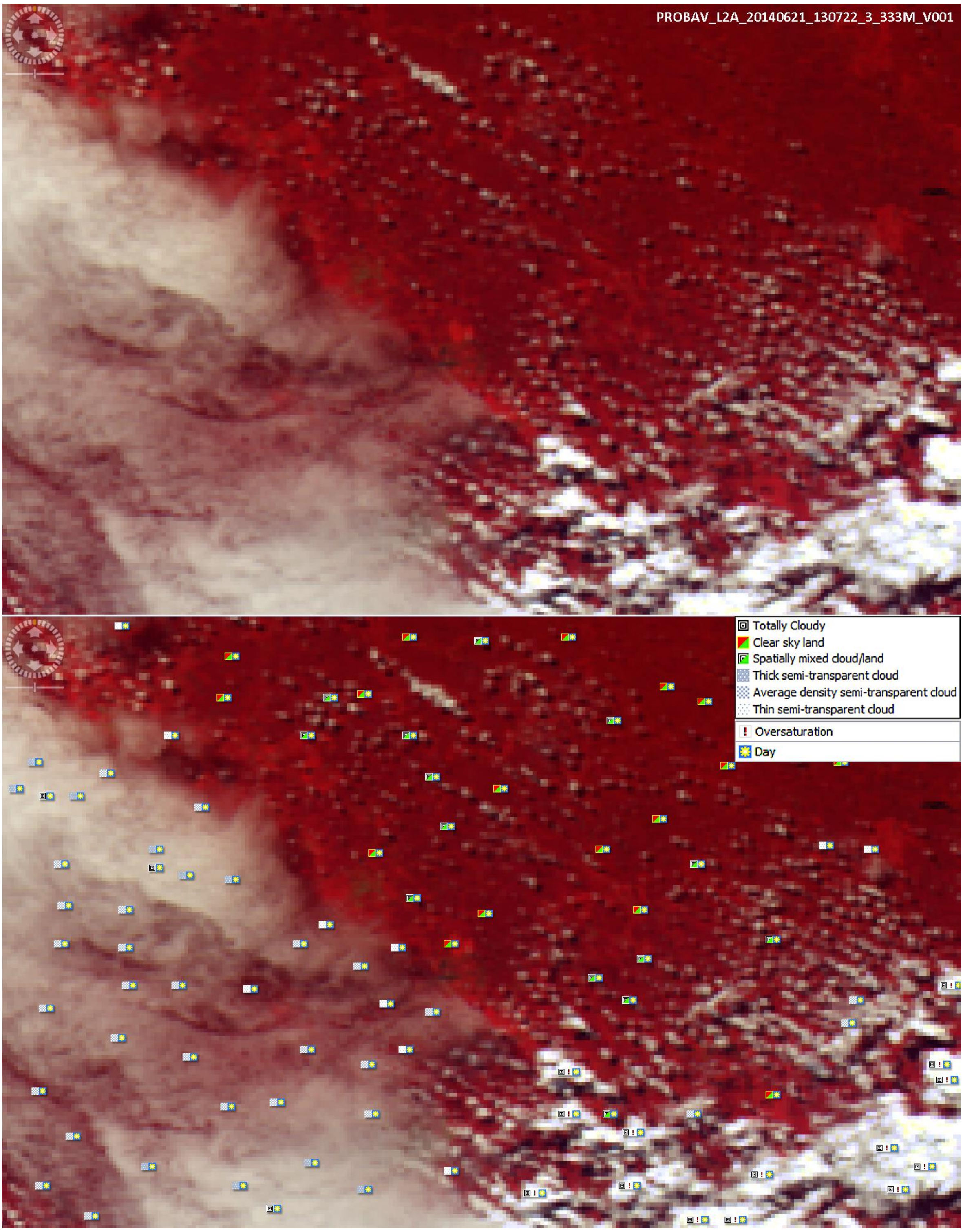


Figure 26: different cloud categories over land

7.2 Annex II: full list of criteria for the pixel characterization within the validation data set

Table 5: 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 differenciated 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)

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_CHARAC-TERISTICS_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

GLINT_ID	NAME
0	No Glint
1	Glint

OVERSATURATION_ID	NAME
0	None
1	Saturation in at least one band

CLOUD_CHARACTERIS-TICS_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

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

7.3 Annex III: statistical measures – formulas

In-Situ Database

CLOUDMASK

Class	Clear	Cloud	Sum	U A	E
CLEAR	13095	45	13140	99.7	0.3
CLOUD	2934	11921	14855	80.2	19.8
Sum	16029	11966	27995		
P A	81.7	99.6		OAA:	89.36
E	18.3	0.4			

A number of accuracy measures are available for assessing the quality of a classification (cloud/no cloud).

$$OAA = \frac{N\ (all\ correctly\ classified\ cases)}{N(total)} * 100$$

$$UA = \frac{N(correctly\ classified\ class)}{N(total\ number\ in\ class)} * 100$$

$$PA\ = \frac{N(ground\ truth\ category\ classified\ correctly)}{(total\ number\ in\ ground\ truth\ cagegory)} * 100$$

Krippendorff's Alpha coefficient for a Coincidence matrix:

$$\mathbb{K} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} with \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = N$$

$$\mathbb{C} = \mathbb{K} + \mathbb{K}^T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} with \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = N$$

$$\kappa = 1 - \frac{D_0}{D_e}$$

where

$$D_0 = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \cdot_{metric} \delta_{ij}^2$$

$$D_e = \frac{1}{N(N-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{i\diamond} \cdot h_{\diamond j} \cdot_{metric} \delta_{ij}^2$$

$$with\ h_{i\diamond} = \sum_{j=1}^{n} a_{ij}\ and\ h_{\diamond j} = \sum_{i=1}^{n} a_{ij}$$

while D_0 is the measured not-agreement and D_e the expected not-agreement.

If

$$_{metric}\,\delta_{ij}^2 = {}_{nominal}\,\delta_{ij}^2 = \begin{cases} 0\ if\ and\ only\ if\ (iff) j = i \\ 1\ if\ and\ only\ if\ (iff) j \neq i \end{cases}$$

then

$$D_0 = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\ \ with\ i \neq j$$

$$D_e = \frac{1}{N(N-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{i\diamond} \cdot h_{\diamond j}\ \ with\ i \neq j$$

$$with\ h_{i\diamond} = \sum_{j=1}^{n} a_{ij}\ and\ h_{\diamond j} = \sum_{i=1}^{n} a_{ij}$$

References:

Krippendorff, 2007 Klaus Krippendorff, Computing Krippendorff's Alpha Reliability, Annenberg School for Communication - Departmental Papers (ASC) - University of Pennsylvania, June 2007 http://repository.upenn.edu/asc papers/43

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