Submitted for the Proceedings of ESA Living Planet Symposium, Prague, 9-13 May 2016, ESA-SP 740

METH-103 - Cloud and Cloud Shadow masking of high and medium resolution optical sensors – an algorithm inter-comparison example for Landsat 8

Carole Lebreton[(1)], Kerstin Stelzer[(1)], Carsten Brockmann[(1)], Luc Bertels[(2)], Nicholas Pringle[(3)], Michael Paperin[(1)], Olaf Danne[(1)], Els Knaeps[(2)], Kevin Ruddick[(3)]

[(1)] Brockmann Consult GmbH, Max-Planck-Str 2, 21502 Geesthacht, Germany, Email: carole.lebreton@brockmann-consult.de

[(2)] VITO NV, Boeretang 200, 2400 Mol, Belgium, Email: els.knaeps@vito.be

[(3)] Remote Sensing and Ecosystem Modelling (REMSEM) team, Operational Direction Natural Environment (OD Nature), Royal Belgian Institute of Natural Sciences (RBINS), 100 Gulledelle, 1200 Brussels, Belgium, Email: kruddick@naturalsciences.be

ABSTRACT

Image processing for satellite water quality products requires reliable cloud and cloud shadow detection and cloud classification before atmospheric correction. Within the FP7/HIGHROC ("HIGH spatial and temporal Resolution Ocean Colour") Project, it was necessary to improve cloud detection and the cloud classification algorithms for the spatial high resolution sensors, aiming at Sentinel 2 and using Landsat 8 as a precursor. We present a comparison of three different algorithms, AFAR developed by RBINS; ACCAm created by VITO, and IDEPIX developed by Brockmann Consult. We show image comparisons and the results of the comparison using a pixel identification database (PixBox); FMASK results are also presented as reference.

1. ALGORITHMS PRESENTATION

1.1 The FMASK method

The FMask software v3.2 [**1**] [**2**] consists of a two-pass method for detecting clouds followed by an object-oriented method for detecting cloud shadows projected onto water and land surfaces. The first pass consists of a set of spectral tests, all of which much be satisfied for a pixel to be considered as a "potential cloud". The second pass then establishes normalized temperature, spectral variability, and brightness probabilities for pixels over land and water using automatically computed scene-dependent thresholds. A cloud probability mask is derived and combined with the first pass potential cloud mask. The algorithm was implemented at RBINS and provides the reference on which the other algorithms are compared to. Full documentation of the method (applied with default configuration) is provided in the abovementioned papers and the corresponding software is available for public download from https://code.google.com/p/fmask/.

1.2 ACCAm

The ACCAm (Automated Cloud-Cover Assessment modified) cloud detection process is based on the ACCA (Automated Cloud-Cover Assessment) algorithm first published by [**3**] for Landsat 7 and further developed at VITO for Landsat 8. The cloud detection is done using a decision tree of several thresholds on different bands (green, red, NIR, SWIR, and TIR) and band ratios (Figure 1).

The algorithm makes also use of the Landsat 8 'Coastal Aerosol' and 'Cirrus' bands, dedicated for cirrus cloud detection. In ACCA, the detection process consists of 2 parts, where the first one classifies ambiguous pixels, which are then evaluated against a statistically derived threshold on the TIR band in the second part. However, this strategy is far from satisfactory for scenes over water, and has been removed in ACCAm, i.e. ACCAm does not differentiate any more between ambiguous clouds and sure (certain) clouds. The Landsat 8 bands 1 to 9 are first converted to Top Of Atmosphere (TOA) planetary reflectance, whereas bands 10 and 11 are converted to at-satellite brightness temperature.

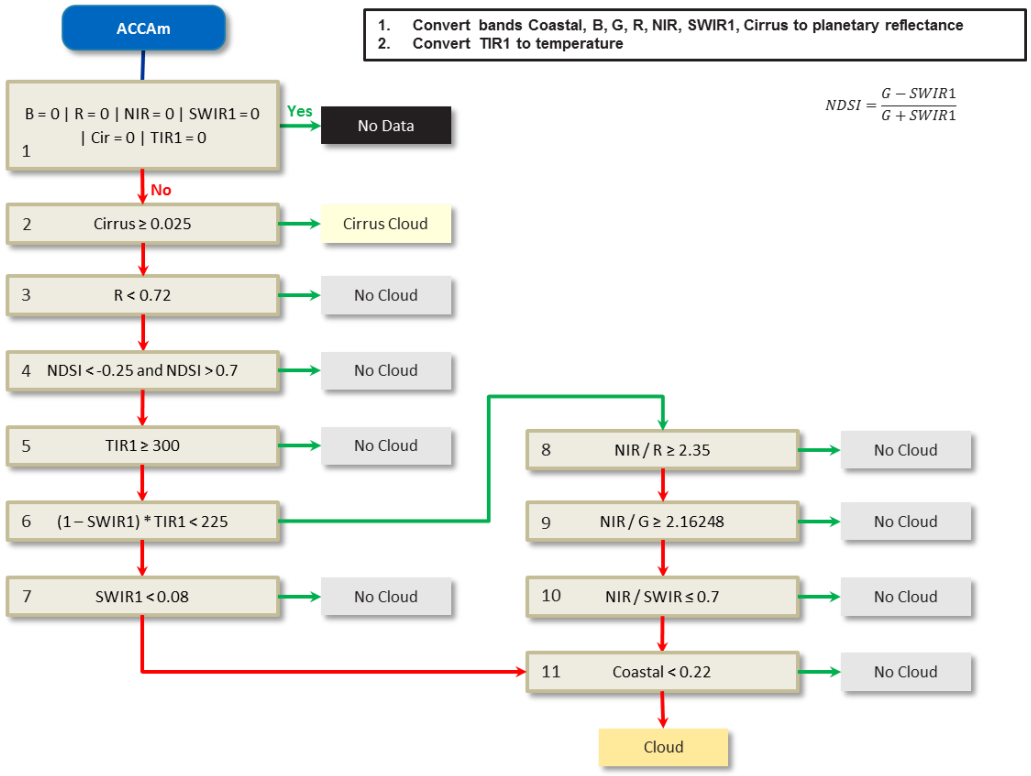

Figure 1. Threshold based decision tree for cloud detection with ACCAm

1.3 The AFAR method

1.3.1 Cloud detection

An overview of the cloud and cloud shadow detection method occurring in AFAR [**4**] is given in Figure 2. AFAR stands for ACOLITE/FMASK Aquatic Refined, since the method is designed for future use with the ACOLITE atmospheric correction software [**5**] and has a strong heritage from the FMASK method "first pass" tests, but also distinct differences, adopted here for optimal performance in aquatic applications.

In the AFAR method, a distinction is made between cirrus and other clouds because of the different typical height and hence different shadow of cirrus clouds. The tests shown in Figure 2 are defined in Table 2. For more on this method, please refer to [**4**].

1.3.2 Cloud shadow projection

AFAR also offers a Cloud shadow detection. AFAR adopts a conservative approach to cloud shadow detection, assuming all detected clouds (except the "possible clouds" which will mainly be sandy or surf-covered beaches) to have a cloud top height of 5km, or 10km for cirrus clouds. AFAR then projects the shadow from the given height in sun viewing direction using the Bresenham's line algorithm [**6**] and masks all pixels as cloud shadow from the cloud itself to the maximum horizontal extent.

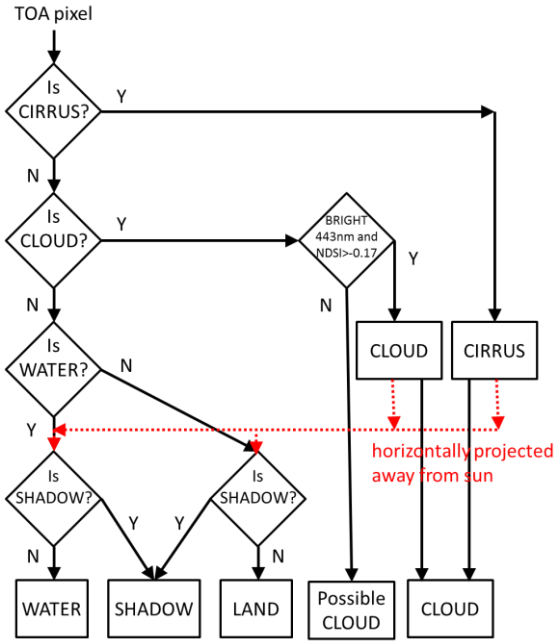

Figure 2. Flow-chart depicting pixel identification process from top of atmosphere (TOA) pixel data to the level 1 pixel classification (rectangular boxes). All tests are performed on a pixel-by-pixel basis except the cloud shadow test, which uses information from neighbouring pixel in the direction of the sun (shown via red arrows).

Table 1. Summary of the spectral tests used in the present study. SWIR=Short Wave Infrared; NDSI=Normalised Difference Snow Index; NDVI=Normalised Difference Vegetation Index.

	Name	Test	Notes
CIRRUS	CIRRUS	$\rho_{1373} > 0.01$	Only detects cirrus clouds
CLOUD (if all tests passed)	Bright SWIR	$\rho_{2201} > 0.0215$	Removes land and cloud
	Low NDSI	$NDSI = \frac{\rho_{550} - \rho_{1650}}{\rho_{550} + \rho_{1650}} < 0.8$	Rejects snow
	Low NDVI	$NDVI = \frac{\rho_{830} - \rho_{660}}{\rho_{830} + \rho_{660}} < 0.8$	Rejects vegetation
	White	$\frac{\lvert\rho_{483} - \rho_{AVE}\rvert + \lvert\rho_{561} - \rho_{AVE}\rvert + \lvert\rho_{655} - \rho_{AVE}\rvert}{\rho_{AVE}} < 0.7$ where $\rho_{AVE} = (\rho_{483} + \rho_{561} + \rho_{655})/3$	May include bare soil, sand, snow/ice
	HazeOptimised	$\rho_{483} - 0.5 * \rho_{655} > 0.08$	May include rocks, snow/ice
	NIR/SWIR	$\frac{\rho_{830}}{\rho_{1650}} > 0.75$	Rejects bright rocks
WATER	WATER	$\rho_{2201} < 0.0215$	
CLOUD (not SAND)	Bright 443nm	$\rho_{443} > 0.2$	Avoid bright white sand
	Not low NDSI	$NDSI = \frac{\rho_{550} - \rho_{1650}}{\rho_{550} + \rho_{1650}} > -0.17$	Avoid bright white sand

1.4 IDEPIX

The cloud screening for Landsat 8 developed at Brockmann Consult (BC) is implemented in IDEPIX. It includes a specially trained neural net for Landsat 8, as well as the following tests: SHIMEZ, CLOST, and OTSU.

The SHIMEZ test ("greyscale" method, adapted from [**7**]) assumes that clouds are grey to white. The assumption is made that the mean (B) of the red, blue, and green bands is greater than a definable threshold, and that the difference between each of two bands is lower than a pre-determined threshold (A).

$$SHIMEZ = abs(blue/red - 1) < A \;\&\&\; abs(blue/green - 1) < A \;\&\&\; abs(red/green - 1) < A \;\&\&\; (red + green + blue)/3 > B \quad (1)$$

A = 0.15 over the day
B = 0.25

CLOST (Cloud Stamping test) generates a virtual band from the product: *coastal_aerosol * panchromatic * blue * cirrus*. A histogram of this product is calculated, which looks quite narrow and sharp. The threshold to separate clouds from non-clouds is within the interval [10^{-5}, 10^{-3}]. The threshold is determined to be at 3% of the maximum value of the histogram. In this case, the threshold value was determined at 10^{-3}.

The OTSU's thresholding method assumes that the image contains two classes of pixels following a bi-modal histogram - foreground pixels and background pixels. It then calculates the optimum threshold separating the two classes so that their combined spread is minimal. This test is quite useful in that it provides a dynamic thresholding based on the current image being processed, and can thus (in theory) adapt to the conditions of the image. This has for example been tested in [**8**]. The software itself and its description can be found at the following links:

http://www.labbookpages.co.uk/software/imgProc/otsuThreshold.html;
http://en.wikipedia.org/wiki/Otsu%27s_method;
http://rsb.info.nih.gov/ij/plugins/otsu-thresholding.html; http://habrahabr.ru/post/112079/

The final cloud mask within IDEPIX is then a combination of all 4 masks as described in Eq.(1).

$$Cloud = (NN_{result} == Cloud\ sure)\ OR\ (CLOST == True)\ OR\ (SHIMEZ == True)\ OR\ (OTSU == True) \quad (2)$$

The neural net is also retrieving semi-transparent clouds, which are then put into a separate class.

The information from the Cirrus band can also be added to the Cloud mask defined in Eq.(2). The threshold there is not easily defined; this is left to the user to decide.

2. METHODS – PIXBOX VALIDATION

PixBox is a combination of a software designed for manual pixel selection and labelling and the respective data bases generated for further analysis. PixBox has been developed at BC in the framework of the BEAM-VISAT software environment and is currently being ported to the Sentinel Application Platform (SNAP). PixBox was developed as a means to generate databases for validating cloud classification for different sensors. Later on, the databases were also used for the training of neural nets (NN) that offer an automated pixel classification, oriented but not only towards cloud detection. The databases for NN training and validation are different data sets. Meanwhile pixel collections are available for AVHRR, MERIS FR and RR, SPOT VEG, MODIS Aqua and Terra, Landsat-8, ProbaV and Sentinel-2. Within the collection process, each pixel is labelled by a number of attributes, such as clear water, clear land, totally cloud, semi-transparent cloud, ice/snow, haze, dust, floating vegetation.

The validation database for Landsat 8 represents a collection of 21 analysed scenes, for a total of 37060 pixels classified. Pixels were collected using the visible channels from Landsat 8 (bands 2, 3, and 4). Figure 3 shows the locations of the Landsat acquisitions. One scene is located outside the shown subset in Singapore.

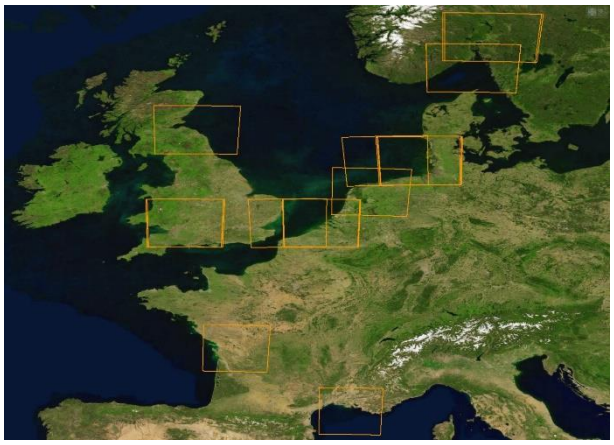

Figure 3. Location of the Landsat 8 scenes used in this comparison exercise.

The pixel type distribution within the pixel collection is presented in Table 2 and Table 3 shows the representation of shadow areas within the database.

Table 2. Pixel Type distribution in the manually collected In-Situ Database

Pixel Type	Number of pixels collected	Percentage classified
Totally cloudy	6691	~18%
Semi-transparent cloud	6748	~18%
Clear sky water	11976	~32%
Clear sky land	10547	~28%
Clear sky snow/ice	726	~2%
Spatially mixed snow_ice/water	372	~1%

For the validation the same images that have been used for the pixel collection are processed with the different algorithms and a confusion matrix is generated by matching the manually labelled pixels with the cloud classification derived by the different algorithms. For this comparison, a subsample of 18 scenes out of the 21 contained in the database have been processed for all methods, for a total of 33960 pixels classified for each method.

Table 3. Percentage and distribution of classified shadow areas

Shadow Area	Number of pixels collected (total 7063)	Percentage classified (total 19%)	
		From total	Within class
Over clouds	845	2	12
Over land	2703 (+192 under semi-transparent clouds)	8	41
Over water	2990 (+322 under semi-transparent clouds)	9	48

Comparison of classified pixels between the PixBox database (hereafter referred to as In-Situ database) and each method is achieved with the help of confusion matrices. The fields of the confusion matrix are filled using logical expressions that compare the different categories. Results are presented showing the number of pixels allocated to (a) specific class(es) with regards to the same one(s) in the in-situ database. These numbers allow the operator to know precisely how many pixels are rightly and wrongly (with regards to the in-situ database) classified. Associated to these numbers in the confusion matrix are the producer accuracy, user accuracy, error and overall accuracy numbers. These numbers (in percent) allow to estimate how well a classification is performing with regards to the person doing the (in-situ) classification (producer accuracy) or the person applying the classification (user accuracy).

User accuracy (UA) represents the reliability of the classification, i.e. when a user takes a map and clicks on a specific class, how reliable the classification is in ordering the specific pixel in the given class. Producer Accuracy (PA) on the other hand estimates how well a pixel can be correctly classified. The overall accuracy is given by the percentage of correctly classified pixels on the basis of the total number of included pixels.

3. COMPARISON RESULTS

3.1 Statistics - Confusion matrices

3.1.1 Cloud detection

Figure 4 presents the User Accuracy and Producer Accuracy for each method either over water classified pixels or for cloud pixels (both clouds sure and ambiguous). Overall, all methods perform quite well at detecting a cloud (UA of 100, 74.9, 82.9, 71.4% for ACCAm, AFAR, FMask, and IDEPIX respectively). In this respect, ACCAm is the best method at identifying a cloud when there is a cloud and no clear pixel is wrongly classified as cloud. The PA for clouds is very good and better than FMask for AFAR and IDEPIX (93.9% and 88.4% respectively), but much less so for ACCAm with 49.2% only. This means that ACCAm is not detecting all clouds but classifies some as clear water pixels.

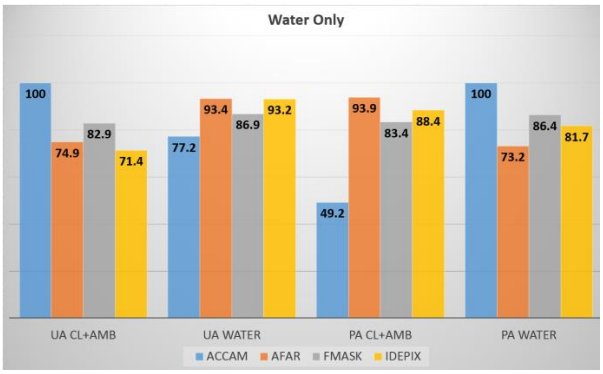

Figure 4 Aggregated confusion matrices results for UA and PA cloud detection over water from all four methods (ACCAm in blue, AFAR in orange, FMask in grey, and IDEPIX in yellow.

3.1.2 Cloud shadow detection

Figure 5 shows the statistics of the confusion matrix from AFAR and FMask cloud shadow detection performances against the in-situ database. AFAR is very successful at recognising a cloud shadow when there is one (PA of 92.3%), while FMask misses more than half of the clouds shadow (PA cloud shadow of 46.6%). AFAR is however classifying too many non-cloud shadow pixels as such (UA for cloud shadow of 56% for AFAR and 83% for FMask), because of the conservatively estimated cloud top height.

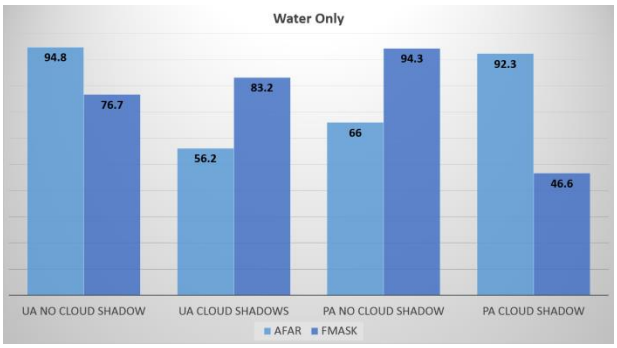

Figure 5. Cloud Shadow UA and PA confusion matrix results from AFAR (light blue) and FMask (dark blue)

3.2 Image comparisons

A visual inspection helps to relate with the statistics derived from the confusion matrices. We present here two examples of the 18 scenes that have been analysed in this study (Figure 6 and Figure 7). The example of the Singapore image is a challenging scene as it is very hazy and the transition of haze and semi-transparent cloud is difficult to detect. Different colours are dedicated to the different cloud categories (sure clouds, semi-transparent and ambiguous clouds). The images confirm that the ACCAm is the least conservative cloud detection and some clouds, especially at the cloud border are not classified as cloud. The cloud sure flag from IDEPIX is smaller covering less clouds, which are detected partly as semi-transparent clouds. The AFAR results show a slightly different behaviour for the sure cloud class than the other algorithm, but regarding sure cloud and ambiguous cloud together, the same areas are flagged.

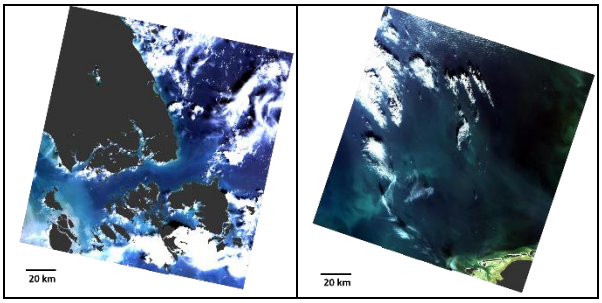

Figure 6. True colour (RGB) images of two Landsat 8 scenes. Left, Singapore taken on Nov. 5th 2014, right North Sea taken Sept. 7th 2014

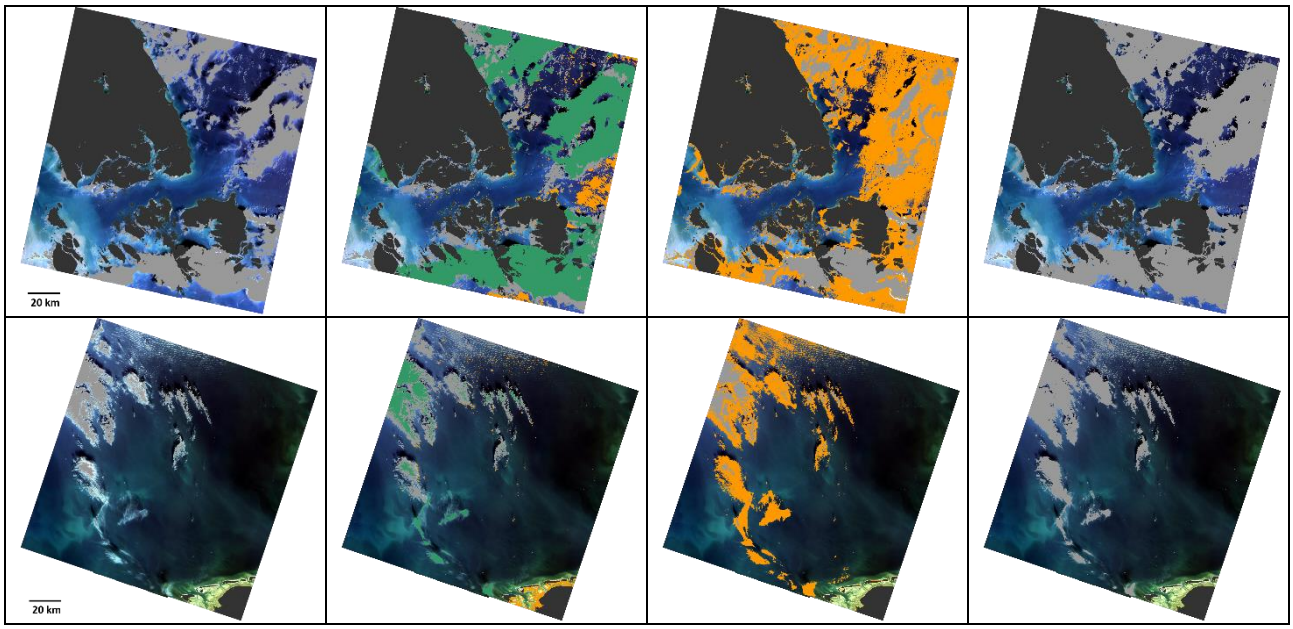

Figure 7. Top Landsat 8 classified for each method (from left to right: ACCAm, AFAR, IDEPIX, FMask). Cloud sure in light grey, ambiguous cloud in green, semi-transparent clouds in orange, land in dark grey. Top Singapore scene fromNov. 5th 2014,bottom North Sea scene from Sept. 7th 2014.

4. CONCLUSION

Landsat 8 pre-processing algorithms have been developed at RBINS, VITO, and BC. Most efforts have concentrated on finding a good cloud mask, as well as a land-water mask. All processing partners have developed a cloud masking algorithm, based partly on the similar ideas and methods but with some distinct differences. All three were shortly presented in this paper.

In these studies, although statistics allow for a general overview of the goodness of fit or classification capability of a given algorithm, it is also essential to look at the images and the more specific coverage and detection of each algorithm. For example, ACCAm is the least conservative algorithm with very good performing in detecting clear pixels (with a 100% Producer accuracy for clear water detection, and a 100% User Accuracy for cloud (sure and ambiguous) detection), looking at RGB pictures allows us to understand that the algorithm does not detect all cloudy pixels, e.g. the cloud borders (the reason why PA for Cloud is only 49.2%).

This comparison is good to show that all algorithms developed and further used in the project perform well over water for different cloud conditions. ACCAm is not masking any Clear Water pixels, while AFAR and IDEPIX offer the valuable possibility of detecting ambiguous clouds (such as semi-transparent clouds) or cirrus clouds, which are all encompassed in the cloud mask of ACCAm. AFAR offers static classification, i.e. a pixel is either classified as water or cloud or ambiguous or cirrus or land, whereas IDEPIX allows for multiple pixel classifications, e.g. a water pixel that is semi-transparent. Moreover, for IDEPIX, the masking of semi-transparent clouds is left to the user, who can decide to process the flagged pixels further or not. Within HIGHROC, we will implement IDEPIX together with the Cloud shadow detection process from AFAR attached to it as standard processing for Landsat 8; the other methods will be available as back-up, fall-back or other option.

Although not shown here but all algorithms have also been successfully tested over land, which is important for inland waters and for nearshore coastal waters where clouds over land may have a shadow on water.

Work is underway in all partners' institutions to port the algorithms to Sentinel 2. BC is gathering pixels in order to train the neural net for S2, which will then be included in IDEPIX for SNAP (currently being built). AFAR from RBINS will be made available via ACOLITE also

for S2, and thresholds for S2 bands are presently being evaluated for ACCAm at VITO.

5. BIBLIOGRAPHY

1 Zhu Z, Woodcock CE. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment. 2012 March 51;118:83-94.

2 Zhu Z, Wang S, Woodcock CE. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sensing of Environnment. 2015;159:269-277.

3 Irish RR. Landsat 7 automatic cloud cover assessment. NASA; 2000.

4 Pringle N, Vanhellemont Q, Ruddick K. Cloud and cloud shadow identification for MERIS and Sentinel-3/OLCI. In: proceedings of the Sentinel-3 for Science Workshop held in Venice-Lido, Italy, 2-5 June 2015, ESA Special Publication SP-734; 2015; Venice.

5 Vanhellemont Q, Ruddick K. Advantages of high quality SWIR bands for ocean colour processing: Examples from Landsat-8. Remote Sensing of Environment. 2015;161:89-106.

6 Bresenham JE. Algorithm for computer control of a digital plotter. IBM Systems Journal. 1965;4:25-30.

7 Gomez-Chova L, Camps-Valls G, Galpe-Maravilla J, Guanter L, Moreno J. Cloud-screening algorithm for ENVISAT/MERIS multispectral images. IEEE Transactions on Geoscience and Remote Sensing. 2007 December;45(12):4105-4118.

8 Choi H, Bindschadler R. Coud detection in Landsat imagery of ice sheets using shadow matching technique and automatic normalised difference snow index threshold value decision. Remote Sensing of Environment. 2004 May 30;91(2):237-242.

9 Zhu Z, Woodcock CE. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. Remote Sensing of Environnment. 2014;152:217-234.