

PV-CDRR

ATBD CLOUD DETECTION FOR PROBA V

TECHNICAL NOTE

	Name	Company	Date	Signature
Prepared by :	B. Berthelot	Magellium	08/02/2017	
Checked by :				
Approved by :	B. Berthelot	Magellium	08/02/2017	

Document reference :	CLOUDIC-TN-002-MAG
Issue.Revision :	1.0
Date :	08/02/2017
Client :	ESRIN
Ref., Tender :	AO/1-7043/11/F/MOS

Distribution list

	Name	Company	N. copies
Addresses :	Rosario Quirino Iannone	ESA/ESRIN	1
Internal copy :	Customer file	Magellium	1 (digital)
	Project Manager	Magellium	1 (digital)
	Development team	Magellium	1 (digital)

Document Change Record

Iss.	Rev.	Date	Reason	Comments
1	0	01/01/2017	Creation of the document	

Table of contents

1	Introduction.....	7
1.1	Objectives of the document	7
1.2	Purpose and Scope	7
1.3	Background	7
1.4	Frame of the Intercomparison exercise	7
1.5	Motivations.....	8
1.6	Structure of the Document	8
1.7	Related documents	8
1.7.1	Applicable documents.....	8
1.7.2	Reference documents.....	8
1.7.3	Bibliography.....	9
1.7.4	Acronyms	9
2	Algorithm overview.....	11
2.1	Outline.....	11
2.2	Theoretical description of the Kohonen Self Organising Map.....	12
2.2.1	Theory	12
2.2.2	Training.....	13
2.2.3	Mapping of SOM	14
2.3	Algorithm description.....	14
2.3.1	Reference data collection	14
2.3.2	Algorithm overview	15
2.3.2.1	Module 1 : Cloud detection implementation and assessment for MODIS	15
2.3.2.1.1	Creation of Training data sets	17
2.3.2.1.2	Training of the maps.....	17
2.3.2.1.3	Classes labelling.....	18
2.3.2.1.4	Results.....	19
2.3.2.1.5	Some examples.....	20
2.3.2.2	Module 2: Algorithm improvement: genericity of the method	23
2.3.2.2.1	Motivations.....	23
2.3.2.2.2	Method	23
2.3.2.2.3	Results.....	24
2.3.2.2.4	Example of results.....	25
3	Application to Proba –V	26
3.1	Proba-V characteristics.....	26
3.2	Proba-V spectral responses.....	26
3.3	Inputs.....	27
3.3.1	Datasets	27
3.3.2	Proba-V data content	28
3.3.3	Data pre processing	30
3.4	Method	31
3.5	Example of results.....	32
3.5.1	PROBAV_L2A_20140321_000027_3_333M_V001	33
3.5.2	PROBAV_L2A_20140321_000512_3_333M_V001	34

4	Analysis of the test dataset results	35
4.1	PROBAV_L2A_20140321_094935_1_333M_V001.hdf	36
4.2	PROBAV_L2A_20140621_042255_1_333M_V001.hdf	37
4.3	PROBAV_L2A_20140921_204654_2_333M_V01.hdf	38
4.4	PROBAV_L2A_20141221_155703_2_333M_V001.hdf	39
4.5	Result Assessment	40
5	Practical considerations	42
5.1	Computation time	42
5.2	Numerical Computation Considerations	42
5.3	Constraints, Limitations, Assumptions	42
6	Conclusion	43
7	Acknowledgments	44
8	Appendix 1 Problems encountered	45
8.1	HDF5 Attributes for geometry	45
8.2	Image size with change International date line is crossed	46
8.3	Data reader	47

List of the Tables

Table 1: List of applicable documents.....	8
Table 2: List of reference documents	8
Table 3: MODIS Land bands.....	15
Table 4: PROBA-V mission main characteristics.	26
Table 5: Proba-V HDF5 file description	28
Table 6: Pixel quality indicator values	30
Table 7: Mandatory fields in the submission files.	31
Table 8: source product of the pixel belonging to the test dataset	35
Table 9: Confusion matrix (top, number of points), (bottom, in %).....	40
Table 10: Confusion matrix (top, number of points), (bottom, in %).....	41

List of the Figures

Figure 1: Neural Network	12
Figure 2: MODIS channel 1 to 7 Relative Spectral Response	15
Figure 3: Module 1 of the algorithm (Classification for MODIS)	16
Figure 4: Topological map used to detect clouds on MODIS data	17
Figure 5: Output of the topological map	18
Figure 6: Best matching unit map overview	18
Figure 7: Histogram showing the unit association. Percentage of cloud (tck), thin (thn) cloud and clear pixels (cls).	19
Figure 8 : MODIS topological map. Clear, cloudy, and semi-transparent are in blue, brown, and green color respectively. Uncertain classes are in dark.	19
Figure 9: Rate of agreement for pixels classified by SOM; (a)	20
Figure 10: Result example. (a) Quicklook; (b) BMUS; (c) Modis Cloud mask; (d) SOM cloud mask; (e) Error mask.....	21
Figure 11: Result example. (a) Quicklook; (b) BMUS; (c) Modis Cloud mask; (d) SOM cloud mask; (e) Error mask.....	22
Figure 12: Using MODIS as a reference for transferring the learning	23
Figure 13: MODIS vs Landsat ETM+ Relative Spectrum Response.....	23
Figure 14: Simulation of Landsat ETM spectrum from MODIS spectrum	24
Figure 15: Rate of agreement for pixels classified by Landsat SOM; (a).....	24
Figure 16: Classification of a Landsat image from Kohonen map learnt on Landsat data, and from Kohonen map learnt on Landsat simulated from MODIS data.....	25
Figure 17: Cloud mask comparison	25
Figure 18: Spectral response functions are shown per camera (LEFT, CENTER, and RIGHT) for the BLUE (upper left), RED (upper right), NIR (lower left) and SWIR (lower right) spectral channels. (Source http://proba-v.vgt.vito.be/content/spectral-response-functions)	27
Figure 19: Location of the Proba-V segment center.....	27
Figure 20: Proba-V topological map derived from MODIS topological map.....	32
Figure 21: Image filled with No Data.....	46

1 Introduction

1.1 Objectives of the document

The purpose of this study is to describe the algorithm used for the cloud detection of PROBA V. It is part of an intercomparison exercise managed by IDEAS project under the name "Proba-V Clouds Detection Round Robin".

1.2 Purpose and Scope

The main objective of the document is to describe the algorithm used to detect clouds on PROBA V data.

The purpose is to examine the feasibility of a generic automatic cloud detection method developed for MODIS using neural networks and assessed its accuracy on Proba-V data. Since the method has been proved accurate through testing on pre-existing multi-spectral land observation satellite data, it is then presented to the Proba-V project as a way to eliminate cloud data from their data sets in order to perform subsequent analysis. This method allows the creation of binary masks.

1.3 Background

Proba-V instrument consists of three identical wide-field of view pushbroom passive radiometers, along across track (to cover the entire required swath). PROBA-V observes in four spectral bands: BLUE (centered at 0.463 μm), RED (0.655 μm), NIR (0.837 μm), and SWIR (1.603 μm). These spectral bands are comparable with the SPOT-VEGETATION bands.

The operational PROBA-V cloud detection algorithm is a modified version of the method applied to the SPOT-VGT BLUE and SWIR observations (Lissens et al., 2000). Original algorithm is based on a threshold based method using TOA reflectances and combination of indices. It used three branches dedicated to snow/ice detection, cloud detection and shadow detection.

In this new version (collection 0), modifications were necessary to account for the observation time difference between the blue and SWIR channels. The algorithm has been further improved using reference data (from the Globalbedo project) and a surface classification (CCI Land Cover Project) identifying four classes (land, snow/ice, water, and unknown) used to separate Proba-V pixels in order to apply different rules (Proba-V Product User Manual, 2016).

1.4 Frame of the Intercomparison exercise

It appears that despite the improvements of the algorithm (Collection 1), some issues remain and need to be addressed to correct them. This study is a contribution of the improvements.

A Round Robin exercise on Proba-V cloud detection (PV-CDRR) has been organized by the European Space Agency (ESA) in collaboration with the Belgian Science Policy Office (BELSPO) to identify potential of algorithm dedicated to Proba-V cloud detection.

1.5 Motivations

In the frame of different projects, cloud detection algorithms have been developed in order to provide cloud mask used in the processing chain (L2A, L2B) and flag the data. These methods were mainly based on thresholds use which values are defined according the branch of the decision trees taken. They proved their efficiency.

However, they necessitate long qualification and become increasingly complicated, sophisticated, complex and are not reusable from one sensor to the other integrally because spectral characteristics (and acquisition geometry) are most of the time different.

The idea to develop a generic method appears and has been tested with success between MODIS and LANDSAT 8 data. Therefore, it was challenging to transfer the method to Proba-V.

The algorithm is described in Section 2.

1.6 Structure of the Document

This section provides the frame of the project. Section 2 gives an overview of the cloud masking approach. Its application to Proba-V is described in section 3. Results of the test datasets are provided in section 4. Section 5 provides with some practical considerations.

1.7 Related documents

1.7.1 Applicable documents

Table 1: List of applicable documents

Id.	Ref.
AD1	PROBA-V PRODUCTS USER MANUAL, Erwin Wolters, Wouter Dierckx, Marian-Daniel Iordache, and Else Swinnen, Version 2.0, Date: 30/09/2016
AD2	Proba-V Clouds Detection Round Robin Protocol, Rosario Q. Iannone
AD3	PROBA-V CLOUD ROUND ROBIN, Description of the Test Data Set, v1.2, Kerstin Stelzer, Michael Paperin, Grit Kirches, 13.10.2016
AD4	PV-CDRR_Protocols

1.7.2 Reference documents

Table 2: List of reference documents

Id.	Ref.
RD. 1.	ProbaV_TestDataSet_PixBox_Database_1.1.xls
RD. 2.	

1.7.3 Bibliography

- Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, 1998: Discriminating clear-sky from clouds with MODIS. *J. Geophys. Res.*, 103 (D24), 32 141–32 157.
- Lissens, G., P. Kempeneers, F. Fierens, and J. Van Rensbergen (2000), Development of cloud, snow, and shadow masking algorithms for VEGETATION imagery. In *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International (Vol. 2, pp. 834 - 836)*.
- GlobAlbedo Project, <http://www.globalbedo.org> (ESA)
- CCI Land Cover Project – Algorithm Theoretical Baseline Document Version 2, available online: http://www.esa-landcover-cci.org/?q=webfm_send/75 (ESA)
- Irish, Richard R., John L. Barker, Samuel N. Goward, and Terry, Arvidson. 2006. "Characterization of the Landsat-7 ETM+ Automated Cloud-cover Assessment (ACCA) Algorithm", *Photogrammetric Engineering and Remote Sensing* 72 (10): 1179.
- Louis, J., A. Charantonis and **B. Berthelot**, "Cloud Detection for Sentinel-2", *Proceedings of ESA Living Planet Symposium (2010)*.
- Kohonen, T. (1984). *Self organization and associative memory (2nd ed.)*. Berlin, Heidelberg: Springer-Verlag. 312 pp.
- Chabiron, O, *Cloud detection using SOM. Development of a generic method. Master of mathematics supervised by B. Berthelot*, 2013.

1.7.4 Acronyms

ATBD	Algorithm Theoretical Basis Document
BELSPO	Belgian Federal Science Policy Office
BRDF	Bidirectional Reflectance Distribution Function
CCI	Climate Change Initiative
CE	Commission Error
EO	Earth Observation
ESA	European Space Agency
ESRIN	European Space Research Institute
ETM	Enhanced Thematic Mapper
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FAR	False Alarm Rate
HDF5	Hierarchical Data Format 5
HR	Hit Rate
IDEAS+	Instrument Data quality Evaluation and Analysis Service
MEP	Mission Exploitation Platform
MCM	MODIS Cloud mask
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Form
NIR	Near-InfraRed
NN	Neural Network
NOAA	National Oceanic and Atmospheric Administration
OE	Omission Error

PA	Producer's Accuracy
PDF	Portable Document Format
POD	Probability of detection
PROBA-V	Project for on-board Autonomy-Vegetation
PV-CDRR	Proba-V Cloud Detection Round Robin
QA	Quality Assessment
QC	Quality Control
RTM	Radiative Transfer Model
SPPA	Sensor Performance and Product Algorithm
SWIR	Short Wave Infra-Red
TIR	Thermal Infra-Red
TOA	Top Of Atmosphere
UA	User's Accuracy
VGT	VEGETATION
VIS	VISible
VITO	Vision on Technology (Flemish Institute of Technological Research)

2 Algorithm overview

2.1 Outline

Cloud detection from multispectral sensors is surely the first step of data processing. Apart from energy budget studies where their presence is a key parameter (so that they have to be detected and their physical properties need to be characterized clouds) clouds need to be detected and removed from the processing in order to estimate accurate atmosphere and surface radiances.

The great majority of cloud detection algorithms are based on threshold methods. This is the case of Proba-V, VGT1/2 (Lissens et al, 2000), but also MODIS (Ackerman et al, 2010), LANDSAT (Irish et al, 2006), S2/MSI (Louis et al., 2010) for instance. These methods, by their very nature, have an abrupt transition between two phases, which might lead them to errors over ambiguous cases such as thin clouds over very bright or very obscure surfaces.

This is the main problem. Bright, dense and cold clouds are well detected. Cirrus, semi-transparent clouds are not so easily detected. The transition Cloud/Non cloud is therefore not easy to set. A third class of semitransparent cloud allows to manage the transition between cloudy pixels and clear ones, but is difficult to define. This is the reason why we selected the non-linear classification method of the neuron networks because their end results are smoother when confronted with such cases.

In particular, the method uses Kohonen network.

The method is described in details in the following section but in a few word, a Kohonen network is an automatic non supervised classification method that allows to clusterise pixels into classes that are labeled in a second step by an expert. Once established, the classifier is applied to the series of data to produce the cloud masks.

In our case, the strategy was double. The aim was to develop a well validated topological map created using Kohonen network that could be used generically to be adapted to various sensors. This idea has been implemented and tested by Chabiron et al. in 2013 on MODIS and LANDSAT. The reference topological map has been created using MODIS L1C data and validated against MODIS cloud masks. Then, the topological map has been derived to fit Landsat spectral band, and used to produce Landsat cloud masks. The report contains the validation report.

Therefore, using the merits and inconvenient of the method, it seems interesting to adapt this methodology to Proba-V.

The strategy for clear vs cloud Proba-V identification is the following:

- 1) Adapt the method developed for MODIS to Proba-V;
- 2) Validate the first results against the reference dataset provided for this exercise;
- 3) Apply the method to the full Proba-V dataset.

The details of this process are discussed in the following sections. The theory of the Kohonen classification is detailed in section 2.2. The way it has been implemented for MODIS and validated is described in section 2.3. The implementation for Proba-V is described in section 3.

2.2 Theoretical description of the Kohonen Self Organising Map

2.2.1 Theory

In 1984, Kohonen introduced an elegant mapping technique for examining the structure of high-dimensional data sets: the self-organising feature map. Such a Kohonen map incorporates in an unsupervised way the topology present in the data. It consists of a set of non-interconnected units which are spatially ordered according to some topology; typically a two-dimensional hexagonal or rectangular grid is chosen and is sufficient to elucidate all specific features characterizing the nature of the high-dimensional data.

The neural networks method applied in processing of the Proba-V data sets is a variant of the topological maps method of Kohonen. The topographical maps method is one of the first, non-supervised, automatic neural networks learning models implemented, and it is closely associated through theoretical studies to statistical methods of automatic classification, such as the nonlinear principal component analysis and the density function approximation.

The algorithm projects the entry data space (a multi-dimensional matrix) on a lower dimension field (of typical dimension 1 2 or 3), known as "topological map", and hereby referred to as C. The purpose of this projection is to partition great numbers of observations in groups of similar characteristics. However, the Kohonen algorithm allows the realized groupings to have a neighborhood structure that can be materialized through a topological map.

A Kohonen's map architecture is comprised of two interconnected layers: The input layer and the topological map.

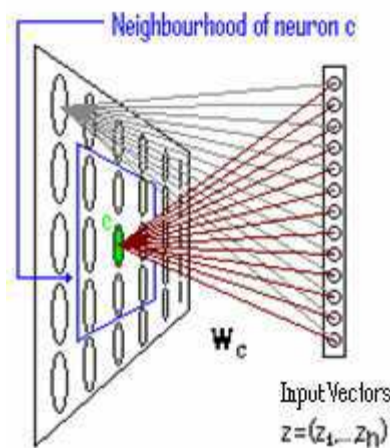


Figure 1: Neural Network

The sole function of the input layer is to present to the network the data (input vectors) to be processed. The neurons take the dimension of the input vectors.

During the research the dimension of each neuron has been 5, 6, 7, 8 or 9 depending on the parameters we were exploring at the time.

The topological map is a grid space of regular structure and of low dimension (2 or 3). Each grid intersection is a neuron. The map geometry is predetermined during its conception. The linking between the neurons of C is done through a non oriented graph structure, which

induces a discrete distance d on C . For each couple of neurons (c,r) of C , $d(c,r)$ is thus defined as being the length of the shortest path between c and r on the C grid.

For each neuron c , this distance function allows for the definition of a neighborhood (as a zone of influence surrounding each neuron). This notion in turn imposes topological constraints upon the final map structure representation.

The kernel function

$$K^{h(t)}(d(c,r)) = \exp\left(\frac{-d(c,r)}{2h(t)}\right)$$

is used to manage the span of each neighborhood. This function is dependent of $h(t)$ called temperature function, which decreases after each iteration of the algorithm, and of the distance d . the variables c and r represent neurons. The influence that one neuron has on another when altered is inversely dependent to their distance and the temperature parameter. It is an iteration counter and therefore is a natural number.

Conforming to the majority of the artificial neural networks algorithms, Kohonen maps operate in two modes: training and mapping.

2.2.2 Training

If we represent by w_c the weight vectors that connect each neuron of the grid space with the values of the input layer, then these vectors are the parameters processed during the training phase of the SOM. The training is done in a predetermined number of iterations.

Each iteration is conducted in two sub-periods. During the first period each point of the input vector is affected to a specific neuron by the affection function:

$$x(z) = \underset{c}{\operatorname{arg\,min}} \|z - w_c\|^2$$

After the best matching map unit c is found, the data point z will be associated to the class which is represented by the neuron c . During the second period (or update period) the neurons modify their respective weight vectors w_c . The self-organizing aspect of the map only allows for an update in the vicinity of the most active areas, meaning the neighborhood of the best matching neuron.

The update of the i -th weight vector w_c (the weight vector W_{ci}) is calculated by:

$${}_{t+1}W_{ci} = \begin{cases} {}_tW_{ci} + \alpha(t) * K^{h(t)}(d(c,g)) * (z_i - {}_tW_{ci}) & \text{for } K^{h(t)}(d(c,g)) > 0.001 \\ {}_tW_{ci} & \text{for } K^{h(t)}(d(c,g)) < 0.001 \end{cases}$$

With $\alpha(t)$ being the training step function, which decreases after each iteration, and the cut-off limit for $K^{h(t)}(d(c,g))$ determined by the user (depending on the required accuracy of the results, the computing power and availability of time).

Near the end of training, the weight vector of each neurone converges towards a value such as he will no longer be able to be active but for a very limited and well determined neighbourhood of observations from the input vector. Each neuron has a part of the input vector's data attributed to it:

$$Z_c = \{z_p \mid \text{with } p \text{ the index value of the attributed data points.}\}$$

Each such grouping of values is represented by the vector of reference of the group, r_{wc} . The vector of reference r_{wc} is none other than the weight vector w_c arriving to the neuron c at the end of the training season.

The purpose of the training of self organising maps is to create **referent vectors** that best represent the underlying probability density of the data set. The network will approximate the neighbourhood relationships of the different sub groupings of the training batch. This imposes to neighbouring neurones on the map to have referent vectors that are close in the Euclidian distance sense, upon the data space.

The initial grid form can be selected manually by the user, and it usually represents the anticipated complexity of the map. The som toolbox for matlab does a principal component analysis of the data space, and then selects the length and width of the grid to be proportional to the main eigenvectors calculated by the principal component analysis.

After the training period, the Kohonen map is a structured presentation of the data set studied, and we can then use it to interpret data.

2.2.3 Mapping of SOM

Mapping is usually an easy evaluation (through minimum distance) of the best matching map unit for each pixel.

2.3 Algorithm description

The development of the method relies on six different phases which are:

- 1) Reference data collection
- 2) Creation of the topological map;
- 3) Classes labelling;
- 4) Validation on reference sensor;
- 5) Derivation of the map for Landsat sensor;
- 6) Validation of the results.

2.3.1 Reference data collection

Reference dataset is built from 65 Aqua MODIS TOA reflectances (MOD021 product, collection 5) and the associated cloud mask (MOD035) for the validation. The instrument is viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands from 0.4 μm to 14.4 μm .

In the study, we used the seven bands dedicated to land studies as it is reported in the table below. Their relative spectral responses is plot in Figure 2

Thermal Infra-Red bands are not used. It was a choice to use only VIS/NIR channels because we wanted to apply the algorithm on sensors which are covering only the range 0.4 to 2.1 μm such as Sentinel 2/MSI, VEGETATION, ..

Table 3: MODIS Land bands

Band number	Bandwidth (nm)
1	620 - 670
2	841 - 876
3	459 - 479
4	545 - 565
5	1230 - 1250
6	1628 - 1652
7	2105 - 2155

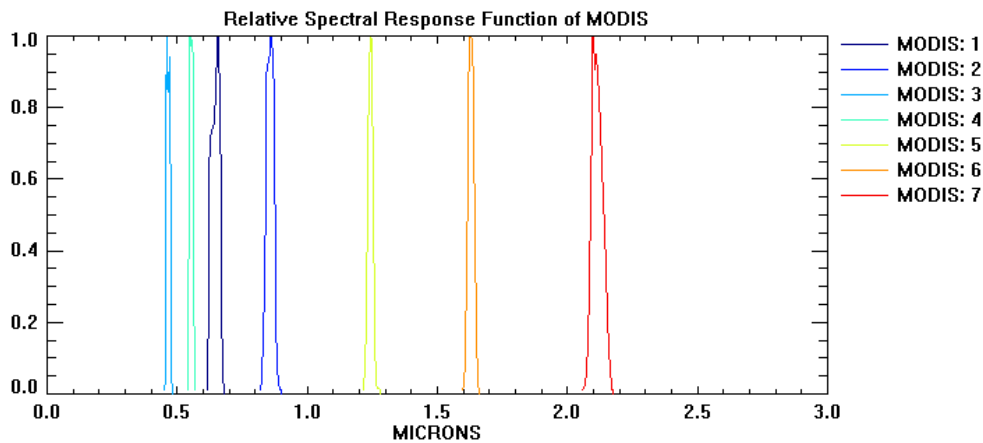


Figure 2: MODIS channel 1 to 7 Relative Spectral Response

2.3.2 Algorithm overview

The algorithm is based on two main modules composed of several sub systems which chain:

- Module 1: MODIS Classification
 - ❖ Subsystem 1: Learning
 - ❖ Subsystem 2: Classification unsupervised
 - ❖ Subsystem 3: Bayesian classification
- Module 2: Adaptation to other sensor
 - ❖ Subsystem 1 : Adaptation of the reference topological map to the other sensor

2.3.2.1 Module 1 : Cloud detection implementation and assessment for MODIS

The algorithm is described by Figure 3 where the three subsystems are detailed. Input MODIS TOA reflectance data is used

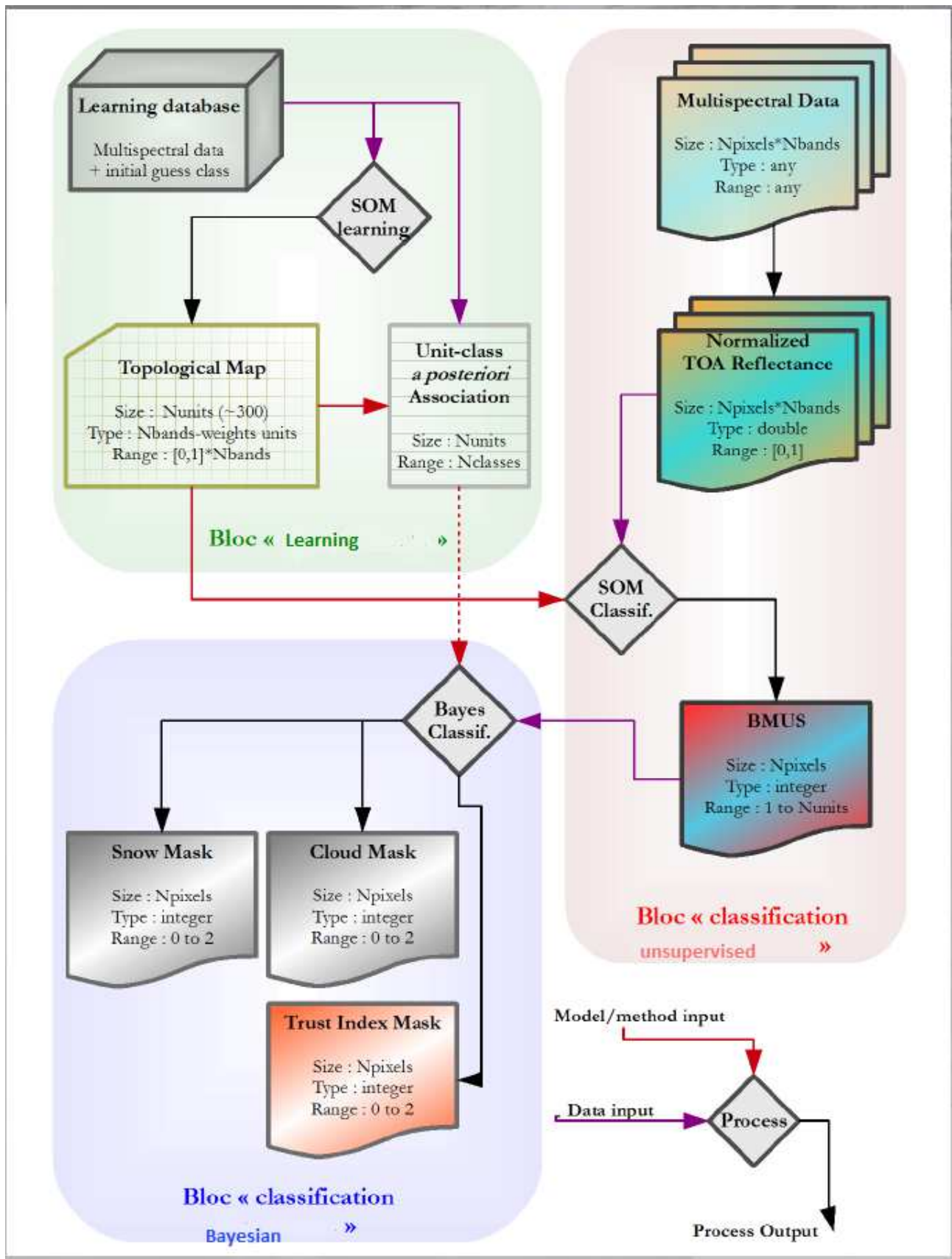


Figure 3: Module 1 of the algorithm (Classification for MODIS)

2.3.2.1.1 Creation of Training data sets

The selection of a data set is a procedure that must be carefully planned in order to improve the output of the program. As mentioned in the previous theoretical explanation, the purpose of the training of self-organizing maps is to create referent vectors that best represent the underlying probability density of the data set. In other words to better detect specific characteristics with a map, we need to give input values that contain a significant number of the desired characteristics.

Therefore, we have built a learning base with MODIS data with precautions. This database consists of 20 different images, from which a total of 1 million pixels are extracted. These pixels were sorted thanks to the MODIS cloud mask to account for bright and semi-transparent clouds, cloud shadows and clear pixel in a given proportion. We therefore have a selection that is not only very reliable but also includes snow (represented by the NDSI index).

2.3.2.1.2 Training of the maps

To perform the training we used the SOM toolbox developed for matlab. The parameters used for the training of the SOMs were rectangular shaped grid and maximum number of neurons 294 in order to keep the latter calculations less time consuming since they depend on the number of neurons.

The map is represented in Figure 4. The smoothest variability of the colors associated to each referent vector represents the low variability of the spectrum for one referent to its neighbors. No abrupt changes are observed. It is also worthy to note how similar values tend to cluster together, and how transition from one class to another is smooth. This is the very advantage of the method.

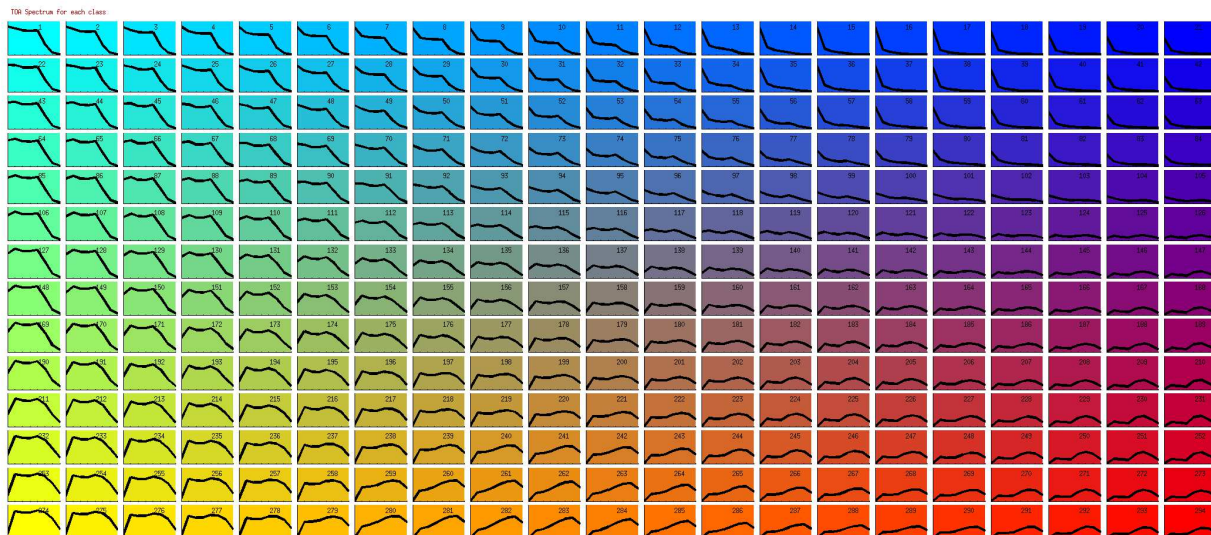


Figure 4: Topological map used to detect clouds on MODIS data

This map structure allows to locate cloud in the left bottom of the map, snow in the top left of the map, sea in the top right of the map, and clear land pixels in the bottom right.

An example of the application of the map is shown in Figure 5. The TOA reflectances and NDSI index are presented as input to the network. The classification is performed and pixels are distributed according to the map on 294 classes. The details of the best matching unit map are seen in Figure 6.

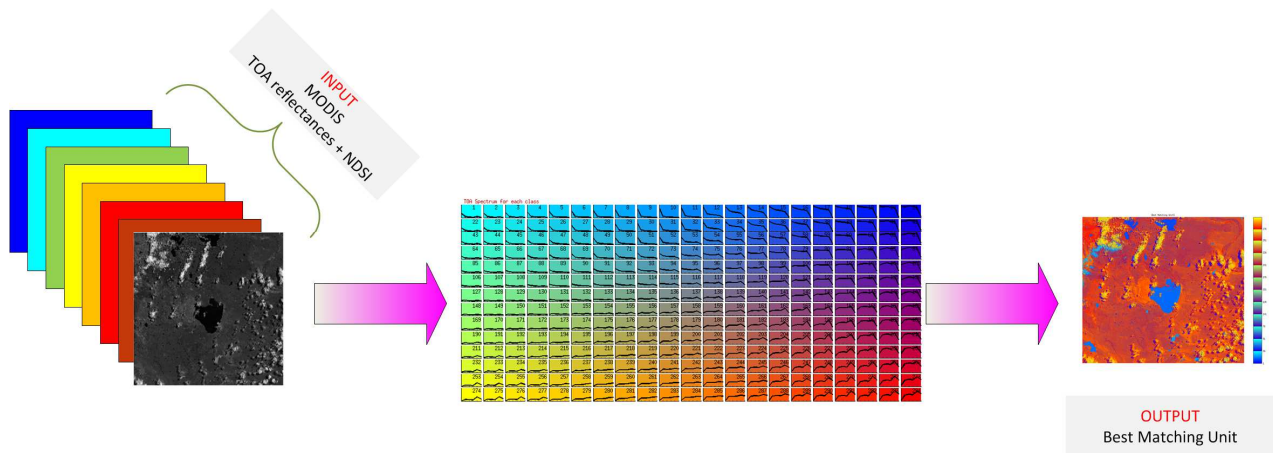


Figure 5: Output of the topological map

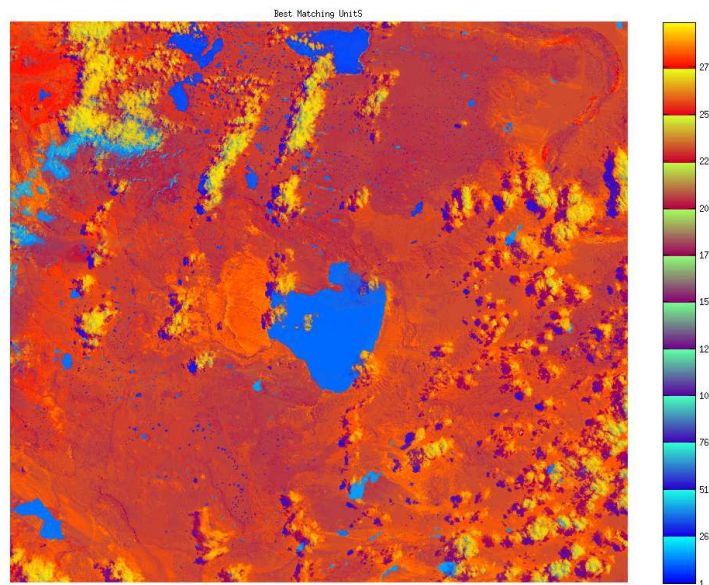


Figure 6: Best matching unit map overview

2.3.2.1.3 Classes labelling

The objective of the subsystem is to provide a mask containing at least the binary information cloud and clear pixels, and possibly semi-transparent cloud. The topological map contains all the information on these 3 output types but the labelling of the 294 classes remains a difficult step to do, even if the structure of the map allows to identify the great part where the clouds are located. The separation between the classes cloud and partially cloudy, and partially cloudy and clear is delicate and could be uncertain even if the labels are defined by an expert.

To solve the problem and increase the accuracy of the results, we performed a supervised statistical learning using the existing MODIS cloud masks. MODIS cloud mask is not only a binary mask. It provides additional information such as the confidence flag which indicates if a thin cirrus cloud is detected for instance, and the confidence we can have in the cloud mask. We use this information as a priori information. For each unit of the map, we count the number of pixels classified in this unit for all three types. Then, we associate to each unit of the topological map the value of the mask which has been the most frequently associated.

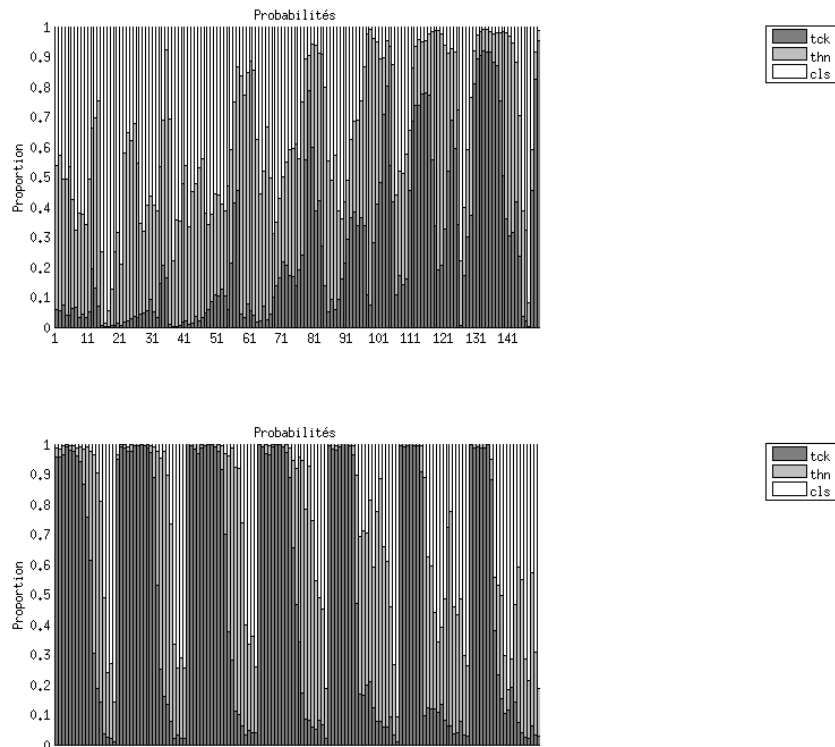


Figure 7: Histogram showing the unit association. Percentage of cloud (tck), thin (thn) cloud and clear pixels (cls).

The topological map created from MODIS and labelled is represented in the figure below.

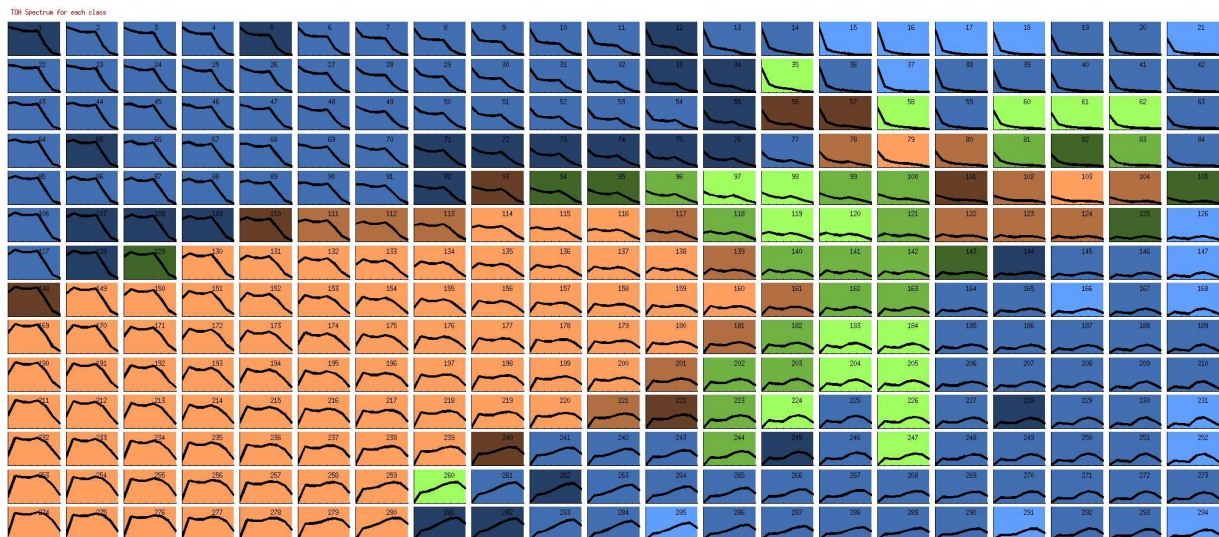


Figure 8 : MODIS topological map. Clear, cloudy, and semi-transparent are in blue, brown, and green color respectively. Uncertain classes are in dark.

2.3.2.1.4 Results

The test base has been built from 45 selected images as varied as possible, in terms of the types of soil encountered and cloud cover. The MODIS cloud masks are very powerful, so the more our masks look like, the more they can be considered accurate.

We therefore compare masks of the type:

- 0: cloud;
- 1: semitransparent cloud;
- 2: Clear.

Since we have the MODIS Cloud mask (MCM) for the validation, the assessment of the quality of the SOM application is performed statistically by computing confusion matrix between clear detection agreement, and cloud detection agreement for both methods. These figures allowed to visualize the performance of the algorithms.

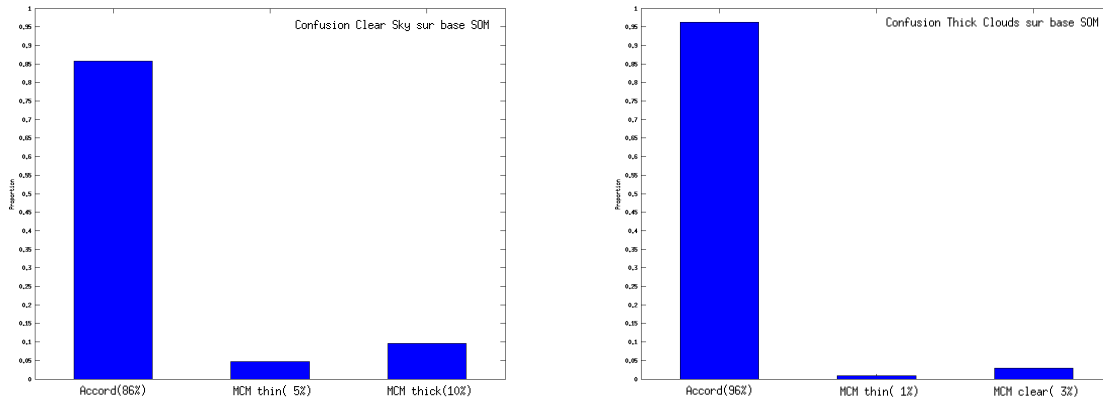


Figure 9: Rate of agreement for pixels classified by SOM; (a) On clear sky ; (b) On clouds.

On the basis of the pixels classified by SOMCloud (Figure 9) in the extreme classes, we note that although the agreement is good (86% in the case of clear sky and 96% in the case of clouds), the differences are observed on the extremes.

Notably, 10% of the pixels classified in the category "sky clear" by SOMCloud are classified in "clouds" by MCM. We investigated the causes of this type of disagreement in the images and came to the following conclusion: SOMCloud masks detect larger "air holes" in the cloud structure that MCM tends to fill. When one looks at the images and especially the BMUs, one tends to give reason to SOMCloud. MCM's tendency to "fill the gaps" is due to the fact that its algorithm includes, in addition to spectral tests, spatial tests. The value of the mask of a pixel thus depends on the neighboring pixels, which explains why the holes in the clouds are sometimes filled. It is quite possible to implement a simple postprocessing to smooth the SOMCloud masks, depending on the level of caution required.

2.3.2.1.5 Some examples

Two examples are therefore provided to illustrate the methodology implemented for MODIS data. The quicklook images, the results of the classification (BMUS), the cloud mask (MODIS and SOM) and a difference map are shown. In the first image, we can see that a large part of the image is detected as clouds by MODIS whereas it is classified as clear by the SOM. In this case, MODIS seems to overdetect cloud over bright areas. In the second image, results are closest except at the cloud edges where the SOM detects more semitransparent clouds. The area around the lake in the south part is covered by a cirrus which is not detected by MODIS.

2.3.2.1.5.1 29/09/2011

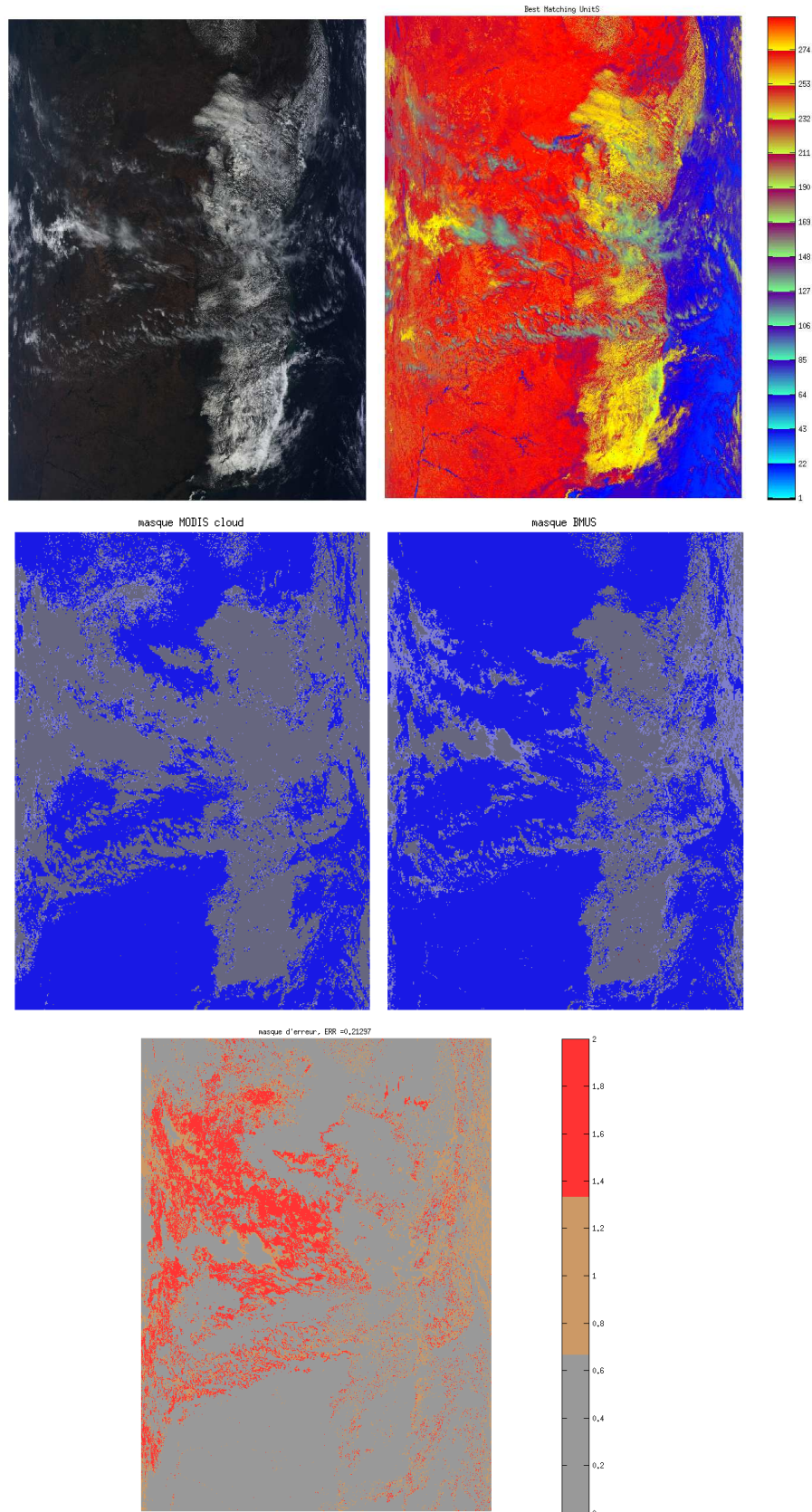


Figure 10: Result example. (a) Quicklook; (b) BMUS; (c) Modis Cloud mask; (d) SOM cloud mask; (e) Error mask

2.3.2.1.5.2 09/05/2009

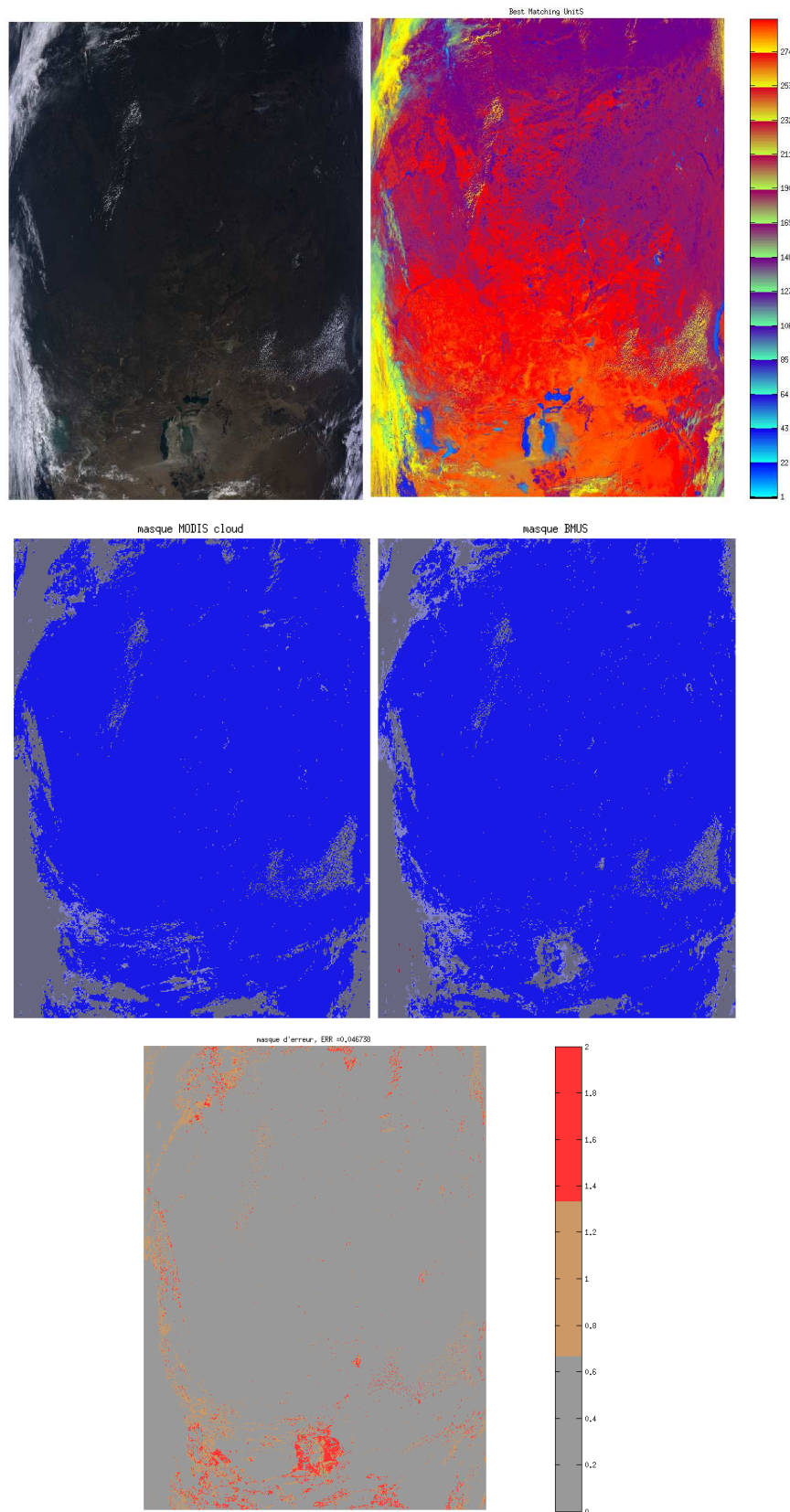


Figure 11: Result example. (a) Quicklook; (b) BMUS; (c) Modis Cloud mask; (d) SOM cloud mask; (e) Error mask

2.3.2.2 Module 2: Algorithm improvement: genericity of the method

2.3.2.2.1 Motivations

The methodology described in the previous section has been also applied to Landsat cloud detection with success, showing the methodology transferability. Therefore, it seemed interesting to simulate topological map without any training for several reasons, the first one being to provide a solution to detect clouds for sensors for which no learning is possible: for preparing methodology in the early phases of development, before sensor launch, or simply to save time and perform tests.

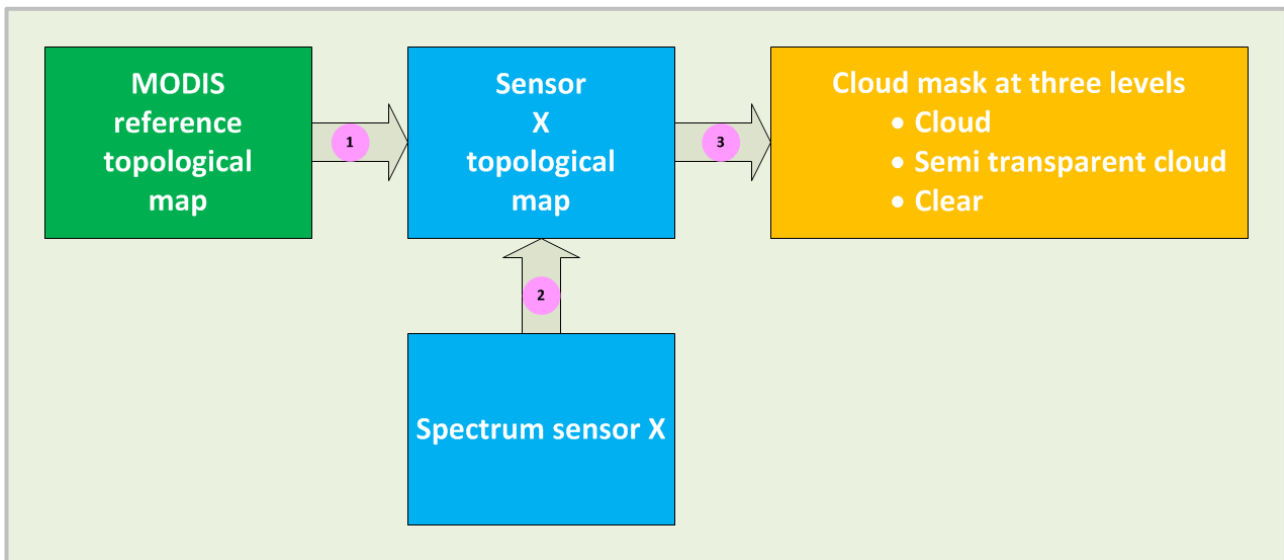


Figure 12: Using MODIS as a reference for transferring the learning

2.3.2.2.2 Method

The methodology described above has been previously implemented for Landsat cloud detection, and validated with a dataset provided by USGS. Therefore, we have simulated landsat topological map from MODIS one, and assessed the accuracy of the transfer using landsat validation dataset.

Using a kriging method, the 294 MODIS reference spectra of the topological map have been interpolated at a high spatial resolution, then convoluted with LANDSAT relative spectrum response to build the new Landsat topological map. The band correspondence is shown on Figure 13 and an example is shown on Figure 14

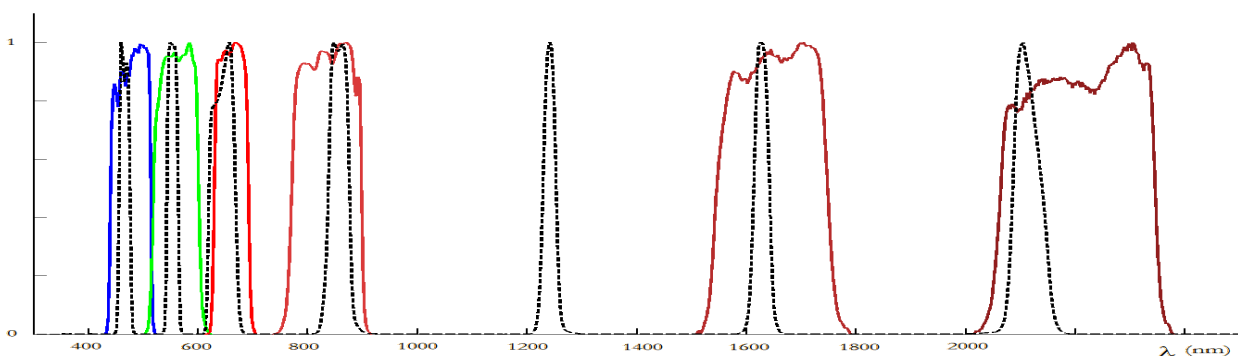


Figure 13: MODIS vs Landsat ETM+ Relative Spectrum Response

Advantages

- The topology learning on MODIS data is saved;
- The associations neuron-classes learned on MODIS are saved;
- The sensor X could have less spectral bands than MODIS;
- The cloud spectrum interpolation is still valid because the cloud spectra are smooth.

Limits of the method

- The interpolation needs to be controlled because land surface spectra are not as smooth as clouds;
- The extrapolation to additional spectral bands located out of the range of MODIS is inaccurate.

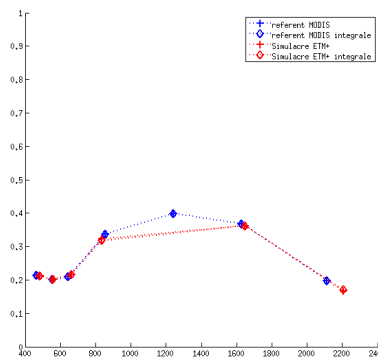


Figure 14: Simulation of Landsat ETM spectrum from MODIS spectrum

2.3.2.2.3 Results

The test base has been built from 10 Landsat images. Cloud masks have been generated i) from the Landsat topological map itself, and ii) from the topological map simulated from MODIS. The cloud masks computed from the simulated Landsat data are in agreement at 95 % with the Landsat reference cloud mask (VCCA).

The distribution of the decisions taken by the method is represented in the next figures:

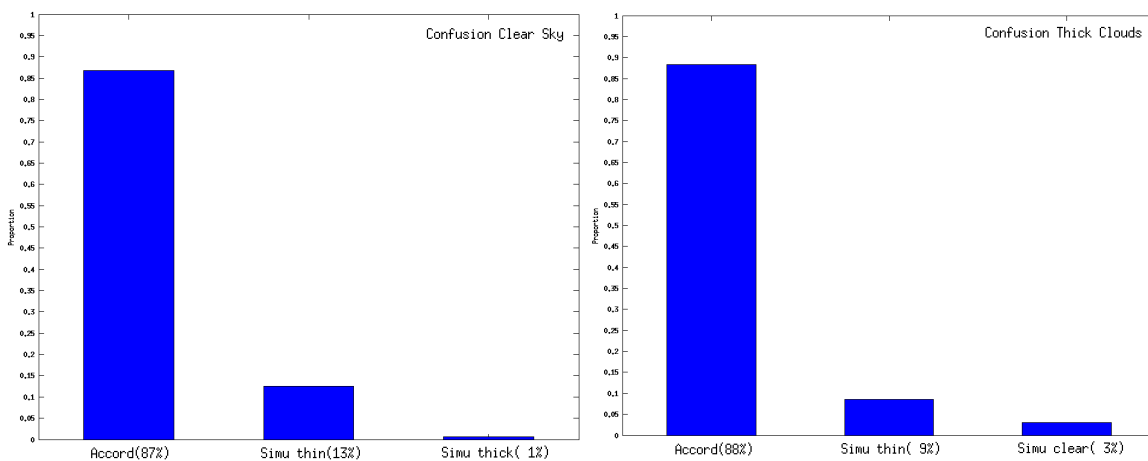


Figure 15: Rate of agreement for pixels classified by Landsat SOM; (a) On clear sky; (b) On clouds.

The level of agreement between the two maps is good (almost 90 %) for the cloud detection, the misclassified pixels are low (1% and 3%). The distribution is different regarding the adjacent classes (semitransparent clouds).

2.3.2.2.4 Example of results

The two following classification maps are the results of the direct learning from Landsat data (on the left) and simulated Landsat data (on the right). As it can be observed, the map are very similar and cloud are well detected when the Landsat based Modis simulated topological map is used.

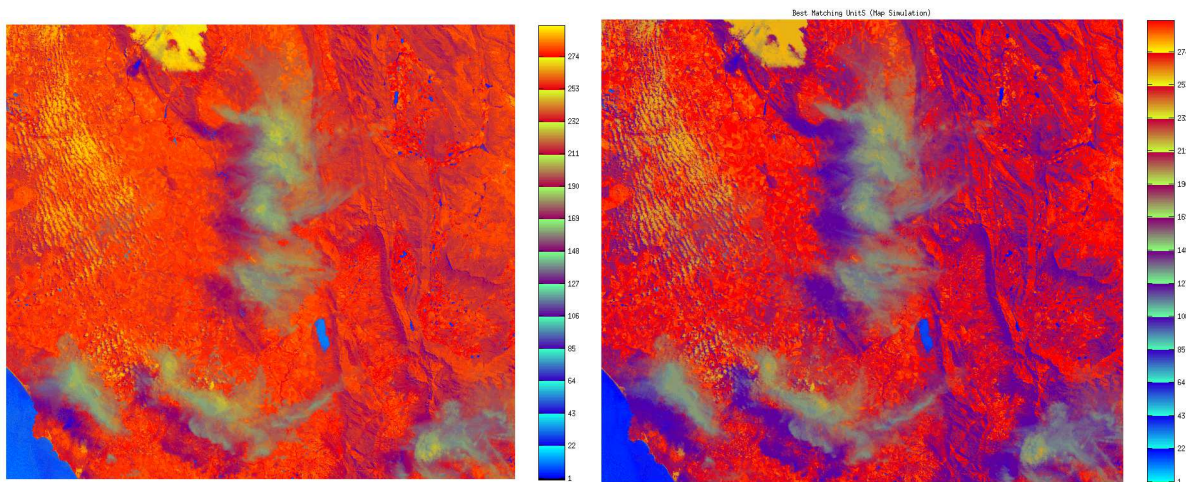


Figure 16: Classification of a Landsat image from Kohonen map learnt on Landsat data, and from Kohonen map learnt on Landsat simulated from MODIS data.

The cloud masks are reported below. The cloud mask reference (VCCA) provided by USGS, the Landsat map, and Landsat cloud mask estimated from MODIS. Some differences are observed for the semitransparent clouds, but the results are really closed.

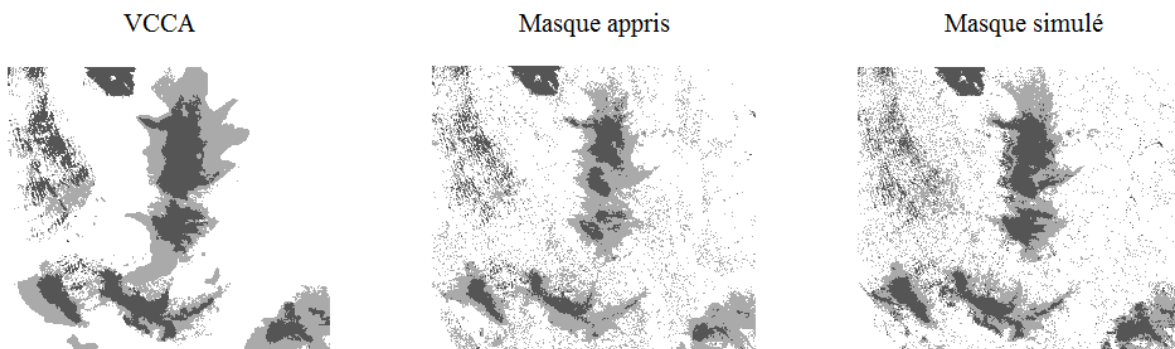


Figure 17: Cloud mask comparison

3 Application to Proba -V

The methodology developed and tested for MODIS and Landsat is applied to Proba-V data. The sensor characteristics are described in sections 3.1 and 3.2, the dataset in section 3.3.

3.1 Proba-V characteristics

Proba-V satellite is a supporting mission for the European COPERNICUS program originally conceived as a "gap-filler mission" between SPOT-VGT and Sentinel-3. PROBA-V has the capability to collect data at a medium (hecta-metric) spatial resolution of 100 m at 2-3 days revisit interval in addition to daily imagery at 300 m. The 100 m spatial resolution will possibly enhance land monitoring programs compared to the 250 m resolution offered for example by MODIS. Its mission characteristics are reported in Table 4.

Table 4: PROBA-V mission main characteristics.

Launch Date	02:06 GMT on 7 May 2013	
Altitude	820 km	
Orbit type	sun-synchronous	
Local overpass time	10:45 h	
Field Of View	102°	
Swath width	2295 km	
Number of cameras	3	
Temporal resolution	Daily near-global coverage (90%) and full global coverage is achieved every 2 days if data from all three cameras are combined (e.g., at 300 m Ground Sampling Distance)	
PROBA-V spectral bands	Centered at (nm)	Width span (nm)
BLUE	463	46
RED	655	79
NIR	845	144
SWIR	1600	73

3.2 Proba-V spectral responses

The spectral responses functions are plot for the three cameras hereafter. We use the centered one as reference for our computation.

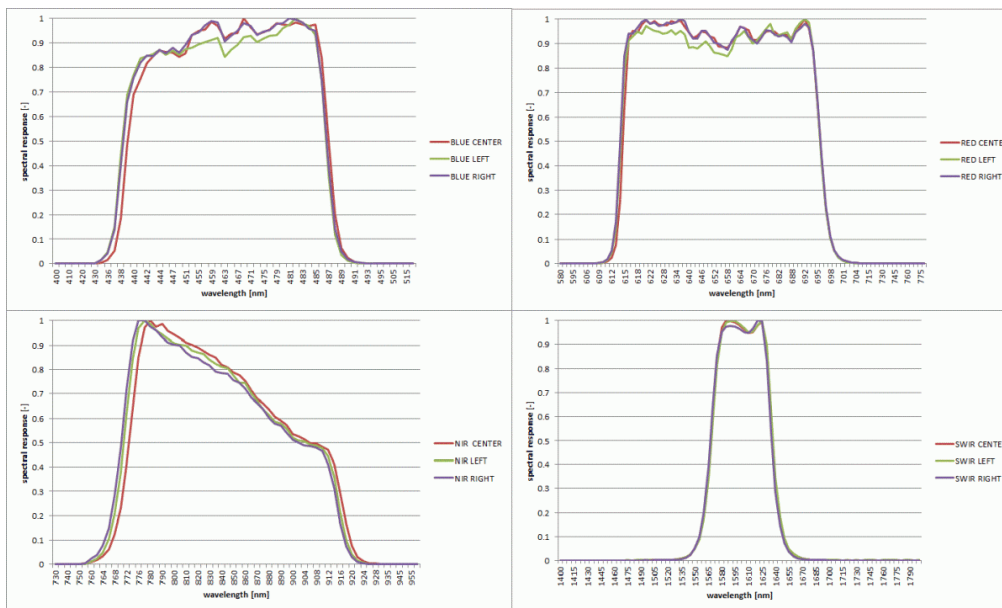


Figure 18: Spectral response functions are shown per camera (LEFT, CENTER, and RIGHT) for the BLUE (upper left), RED (upper right), NIR (lower left) and SWIR (lower right) spectral channels. (Source <http://proba-v.vgt.vito.be/content/spectral-response-functions>)

3.3 Inputs

3.3.1 Datasets

The dataset contains 331 images segments acquired on four dates (21/03/2014, 21/06/2014, 21/09/2014, 21/12/2014) covering land and sea regions. The locations of the center of the segments are represented on Figure 19.

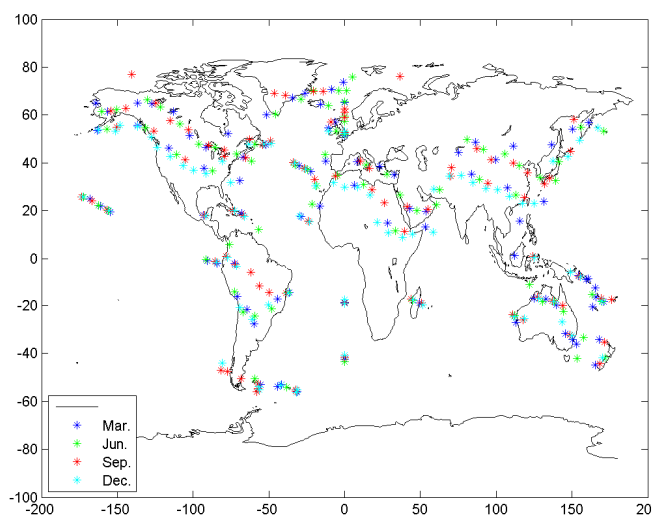


Figure 19: Location of the Proba-V segment center

3.3.2 Proba-V data content

The file format is HDF5. It contains metadata and a group 'LEVEL2A', which contains one group called '/LEVEL2A/'. The subgroup and structure are reported hereafter.

Table 5: Proba-V HDF5 file description

Group '/LEVEL2A'	
Group '/LEVEL2A/GEOMETRY'	
Dataset 'CONTOUR'	
Dataset 'SAA'	'DESCRIPTION': 'Solar Azimuth Angles grid' 'SCALE': 0.666667 'OFFSET': 0.000000 'UNITS': 'Degrees' 'NO_DATA': 255.000000
Dataset 'SZA'	'DESCRIPTION': 'Solar Zenith Angles grid' 'SCALE': 2.000000 'OFFSET': 0.000000 'UNITS': 'Degrees' 'NO_DATA': 255.000000 'MAPPING':
Group '/LEVEL2A/GEOMETRY/SWIR'	
Dataset 'VAA'	'DESCRIPTION': 'Viewing Azimuth Angles grid of SWIR - detector' 'SCALE': 0.666667 'OFFSET': 0.000000 'UNITS': 'Degrees' 'NO_DATA': 255.000000 'MAPPING':
Dataset 'VZA'	'DESCRIPTION': 'Viewing Zenith Angles data SWIR - detector' 'SCALE': 2.000000 'OFFSET': 0.000000 'UNITS': 'Degrees' 'NO_DATA': 255.000000 'MAPPING':
Group '/LEVEL2A/GEOMETRY/VNIR'	
Dataset 'VAA'	'DESCRIPTION': 'Viewing Azimuth Angles grid of VNIR - detector' 'SCALE': 0.666667 'OFFSET': 0.000000 'UNITS': 'Degrees' 'NO_DATA': 255.000000 'MAPPING':
Dataset 'VZA'	'DESCRIPTION': 'Viewing Zenith Angles data VNIR - detector' 'SCALE': 2.000000 'OFFSET': 0.000000

	'UNITS': 'Degrees' 'NO_DATA': 255.000000 'MAPPING':
Group '/LEVEL2A/QUALITY' Attributes: 'PERCENTAGE_CLOUD': 15.351056 'PERCENTAGE_SNOW': 0.103714 'PERCENTAGE_LAND': 10.679968 'PERCENTAGE_MISSING_DATA': 0.000257	
Dataset 'SM'	'SCALE': 1.000000 'DESCRIPTION': 'LeastSignificantBit 0-2 = { 000 = clear, 010 = undefined, 011 = cloud, 100 = ice/snow, 001 = shadow }; LeastSignificantBit 3 = { 0 = sea, 1 = land }; LeastSignificantBit 4 = { 0 = bad SWIR, 1 = good SWIR }; LeastSignificantBit 5 = { 0 = bad NIR, 1 = good NIR }; LeastSignificantBit 6 = { 0 = bad RED, 1 = good RED }; LeastSignificantBit 7 = { 0 = bad BLUE, 1 = good BLUE }; LeastSignificantBit 8 = { 0 = No SWIR coverage, 1 = SWIR coverage }; LeastSignificantBit 9 = { 0 = No NIR coverage, 1 = NIR coverage }; LeastSignificantBit 10 = { 0 = No RED coverage, 1 = RED coverage }; LeastSignificantBit 11 = { 0 = No BLUE coverage, 1 = BLUE coverage }' 'OFFSET': 0.000000 'UNITS': '-' 'NO_DATA': 2 'MAPPING'
Group '/LEVEL2A/RADIOMETRY'	
Group '/LEVEL2A/RADIOMETRY/BLUE'	
Dataset 'TOA'	'DESCRIPTION': 'Top Of Atmosphere reflectance radiometry data (BLUE - channel)' 'SCALE': 2000.000000 'OFFSET': 0.000000 'UNITS': '-' 'NO_DATA': -1.000000 'MAPPING':
Group '/LEVEL2A/RADIOMETRY/NIR'	
Dataset 'TOA'	'DESCRIPTION': 'Top Of Atmosphere reflectance radiometry data (NIR - channel)' 'SCALE': 2000.000000 'OFFSET': 0.000000 'UNITS': '-' 'NO_DATA': -1.000000 'MAPPING':

Group '/LEVEL2A/RADIOMETRY/RED'	
Dataset TOA	'DESCRIPTION': 'Top Of Atmosphere reflectance radiometry data (RED - channel)' 'SCALE': 2000.000000 'OFFSET': 0.000000 'UNITS': '-' 'NO_DATA': -1.000000 'MAPPING':
Group '/LEVEL2A/RADIOMETRY/SWIR'	
Dataset TOA	'DESCRIPTION': 'Top Of Atmosphere reflectance radiometry data (SWIR - channel)' 'SCALE': 2000.000000 'OFFSET': 0.000000 'UNITS': '-' 'NO_DATA': -1.000000 'MAPPING':

The 'MAPPING' attributes are described as follows:

'MAPPING': 'Geographic Lat/Lon', '0.5', '0.5', '148.50595238095238', '-1.00', '0.00297619047619', '0.00297619047619', 'WGS84', 'Degrees'

It contains the size of the pixels, information on the projection

The information on cloud is contained in the status map channel, on these bits.

Table 6: Pixel quality indicator values

Bit (LSB to MSB)	Description	Value	Key
0-2	Cloud/Ice Snow/Shadow Flag	000	Clear
		001	Shadow
		010	Undefined
		011	Cloud
		100	Ice

3.3.3 Data pre processing

No preprocessing is applied. The radiometric channels are all coded between 0 and 2000, with no offset. Reflectances are computed by divided the DN by the scale.

As the matlab software is not able to read the Proba-V hdf5 (nor idl), the files are read using a dump and transformed into single channels, in ENVI standard format. There is 1 file per dataset.

Here are for instance the associated channels to the hdf5 acquisition **PROBAV_L2A_20140321_000027_3_333M_V001.HDF5**

The nomenclature is based on the name of the dataset and the group to where it belongs.

PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-SAA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-SAA.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-SWIR-VAA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-SWIR-VAA.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-SWIR-VZA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-SWIR-VZA.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-SZA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-SZA.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-VNIR-VAA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-VNIR-VAA.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-VNIR-VZA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-GEOMETRY-VNIR-VZA.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-QUALITY-SM.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-QUALITY-SM.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-RADIOMETRY-BLUE-TOA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-RADIOMETRY-BLUE-TOA.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-RADIOMETRY-NIR-TOA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-RADIOMETRY-NIR-TOA.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-RADIOMETRY-RED-TOA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-RADIOMETRY-RED-TOA.img
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-RADIOMETRY-SWIR-TOA.hdr
PROBAV_L2A_20140321_000027_3_333M_V001-LEVEL2A-RADIOMETRY-SWIR-TOA.img

3.4 Method

The methodology described in section 2.3 is derived to Proba-V. Only the second module is adapted since the module 1 is the one used for the genericity of the method. The MODIS reference topological map is transformed into a Proba-V topological map using the spectral resampling of the referent vectors and the four relative spectral responses of Proba-V channels.

Once the new map computed, its application to the data is easy and fast. Figure 20 shows the new topological map based on the three classes requested to be included in the cloud mask product (Table 7). Based on deep analysis, we found that some referent vectors were not well classified. Their labels have been changed.

Table 7: Mandatory fields in the submission files.

Cloud Mask Flag	description	Color in the topological map
0	clear	
1	cloud	
2	semi-transparent	

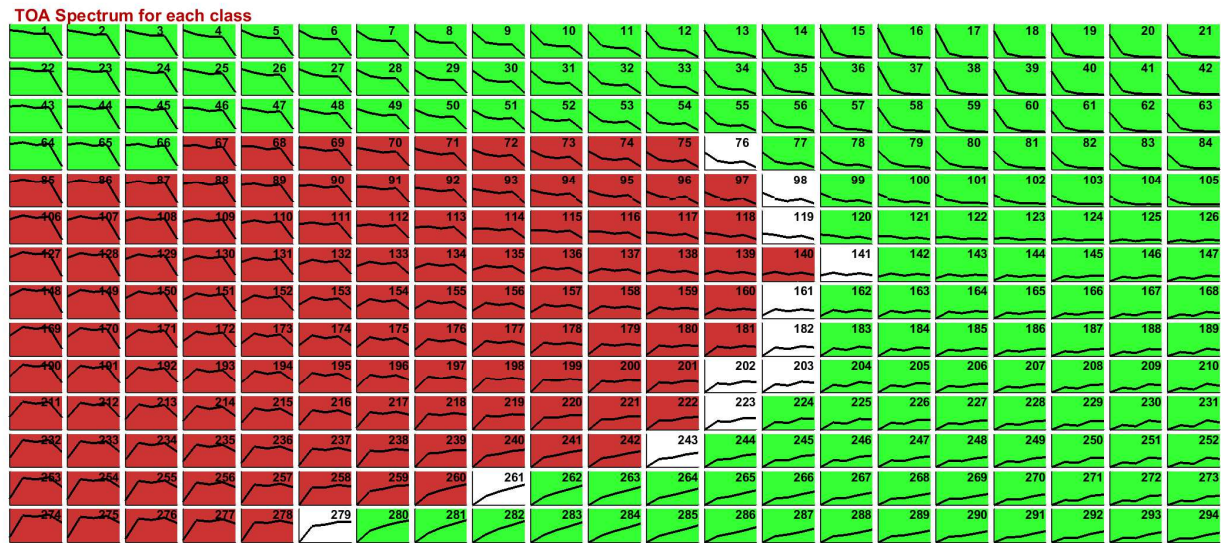


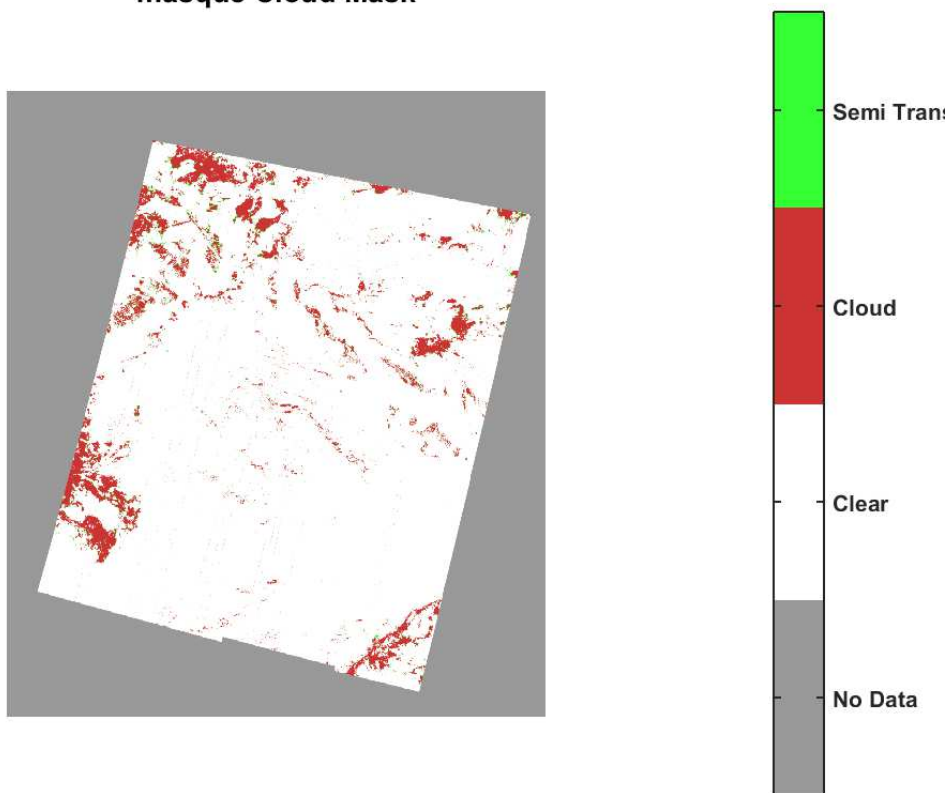
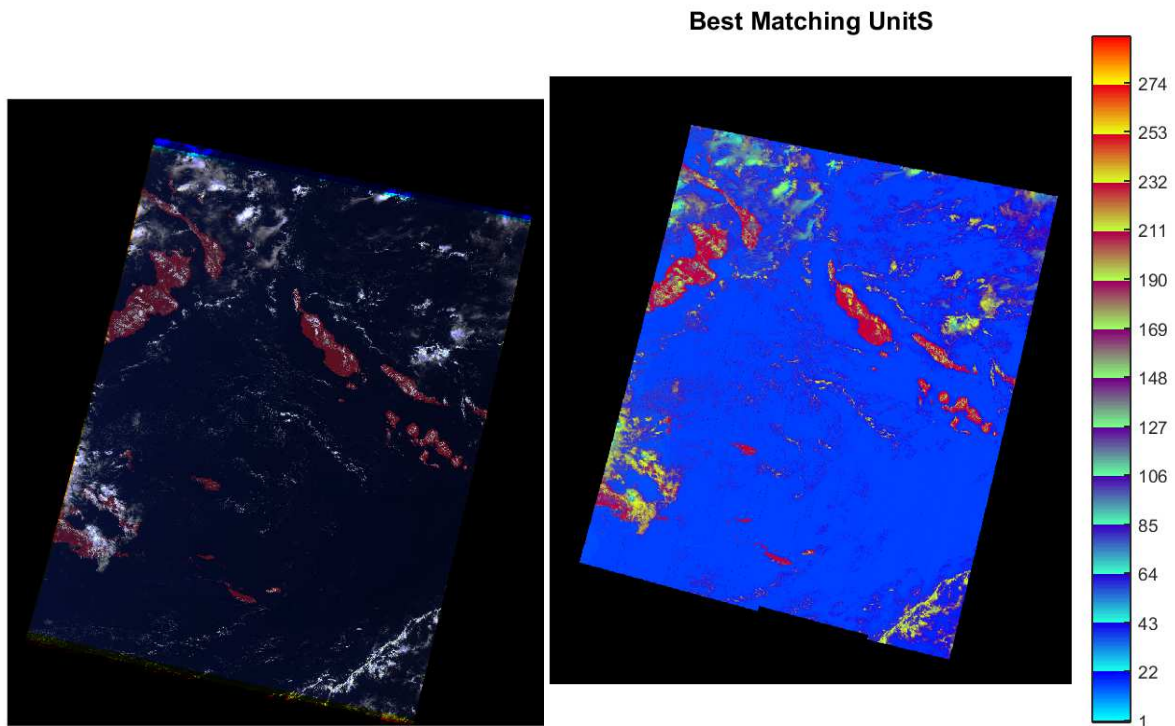
Figure 20: Proba-V topological map derived from MODIS topological map.

3.5 Example of results

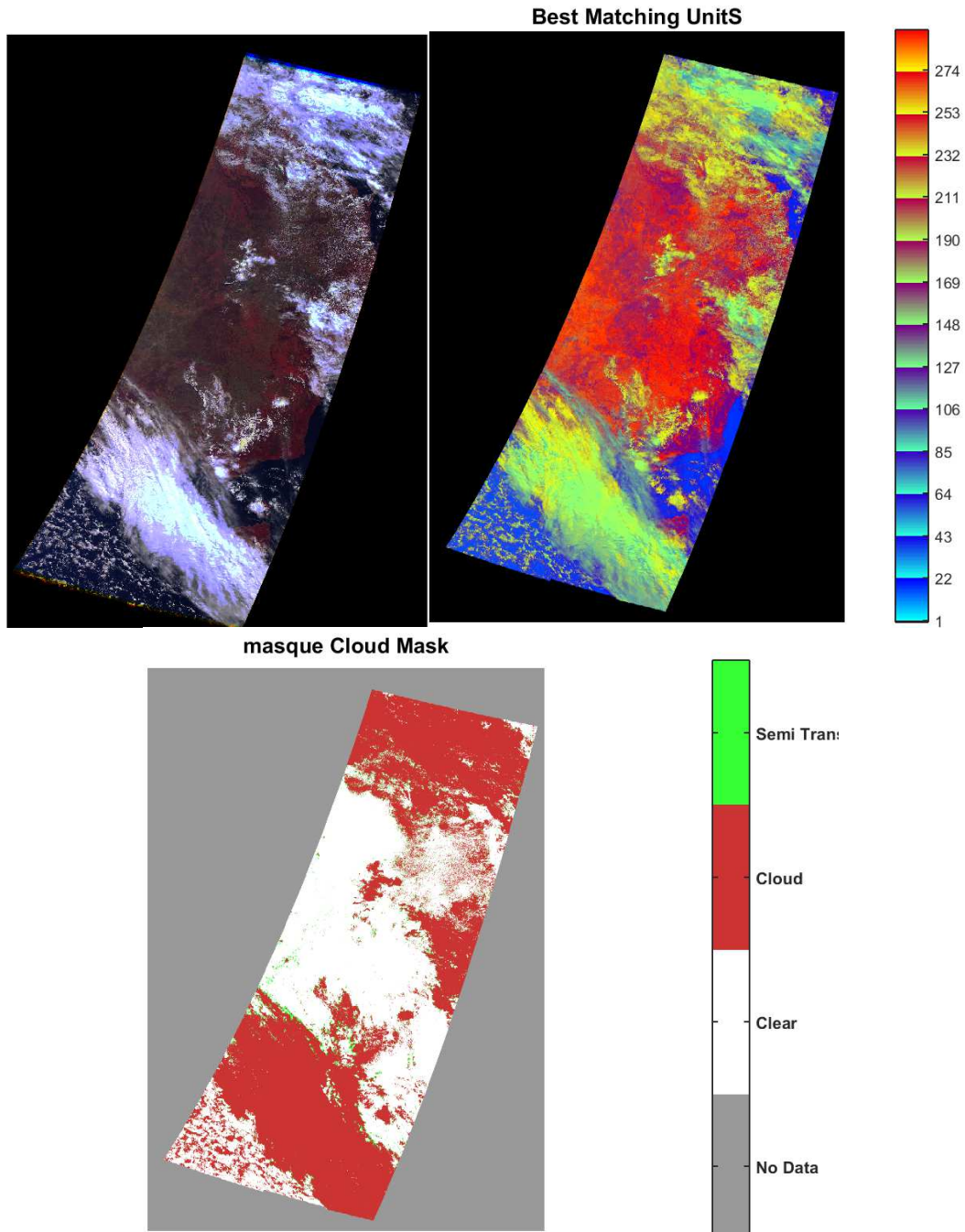
The two following examples display the quicklooks, the classification maps, and the associated cloud masks of two Proba-V images. These two images contain both sea and land, with clouds over both regions.

As it can be observed, the clouds are well detected, with continuity from land to sea. Semitransparent clouds are contiguous in these images with the clouds. No aberrant over or miss detection are observed.

3.5.1 PROBAV_L2A_20140321_000027_3_333M_V001



3.5.2 PROBAV_L2A_20140321_000512_3_333M_V001



4 Analysis of the test dataset results

In order to test the algorithm, a test dataset composed of four images has been provided with the results. The test data set consists of 1350 pixels with their labels (AD.3). The pixels are manually collected, classified and labelled by an expert user. The expert decided 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.

The distribution of the categories has been provided (Section 3 of AD3).

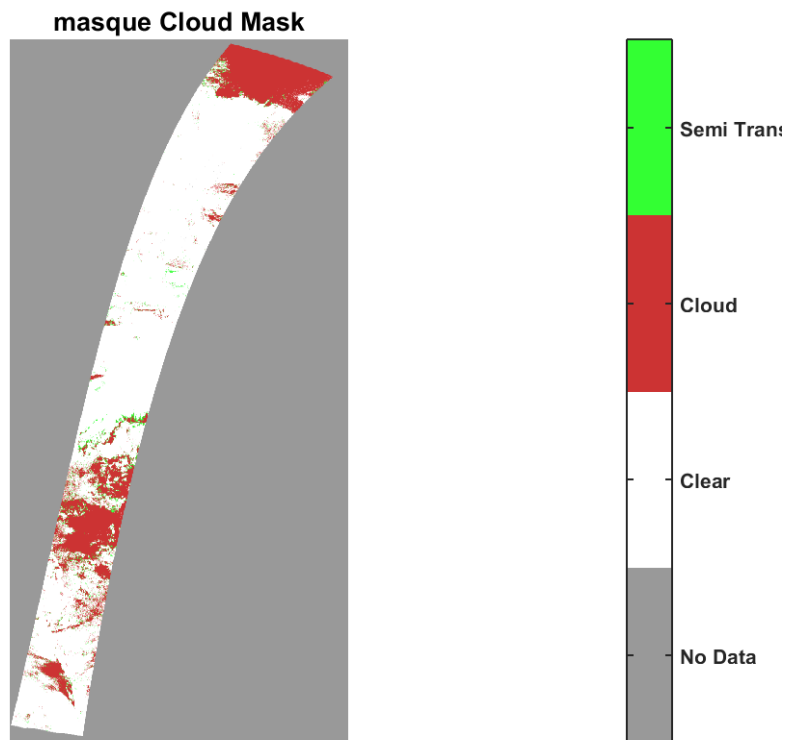
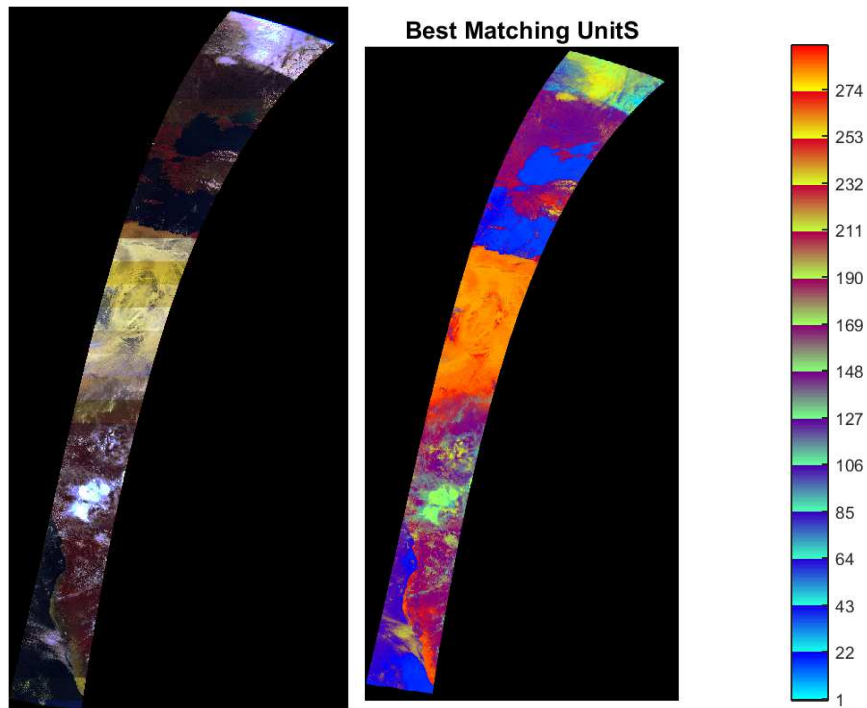
They have been extracted from the four following images.

Table 8: source product of the pixel belonging to the test dataset

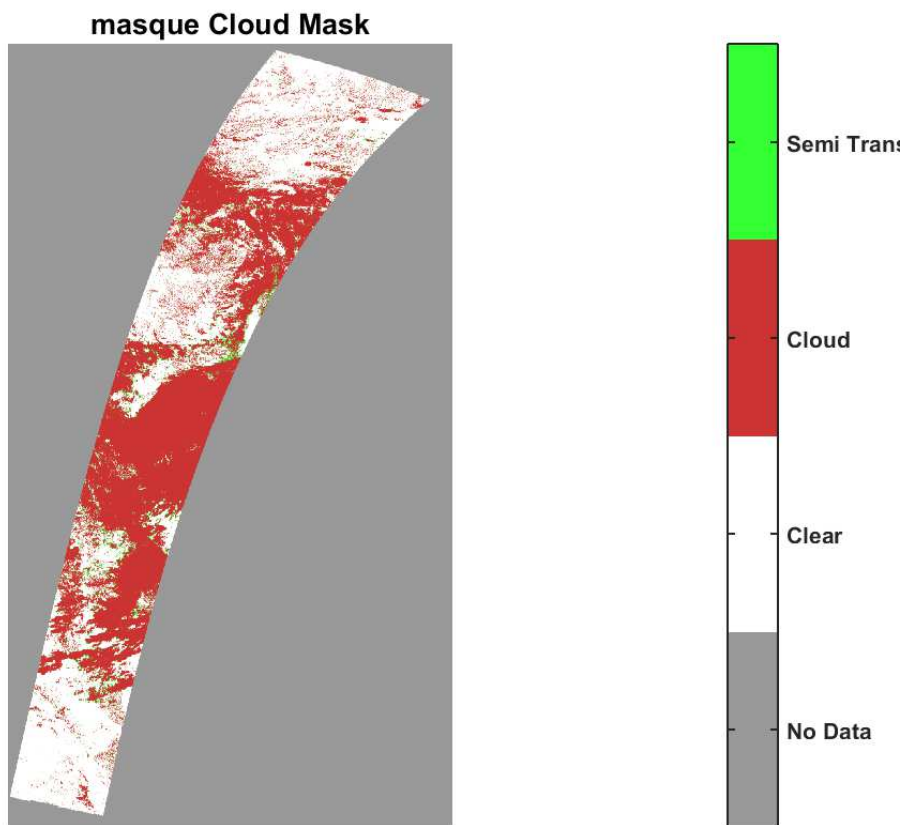
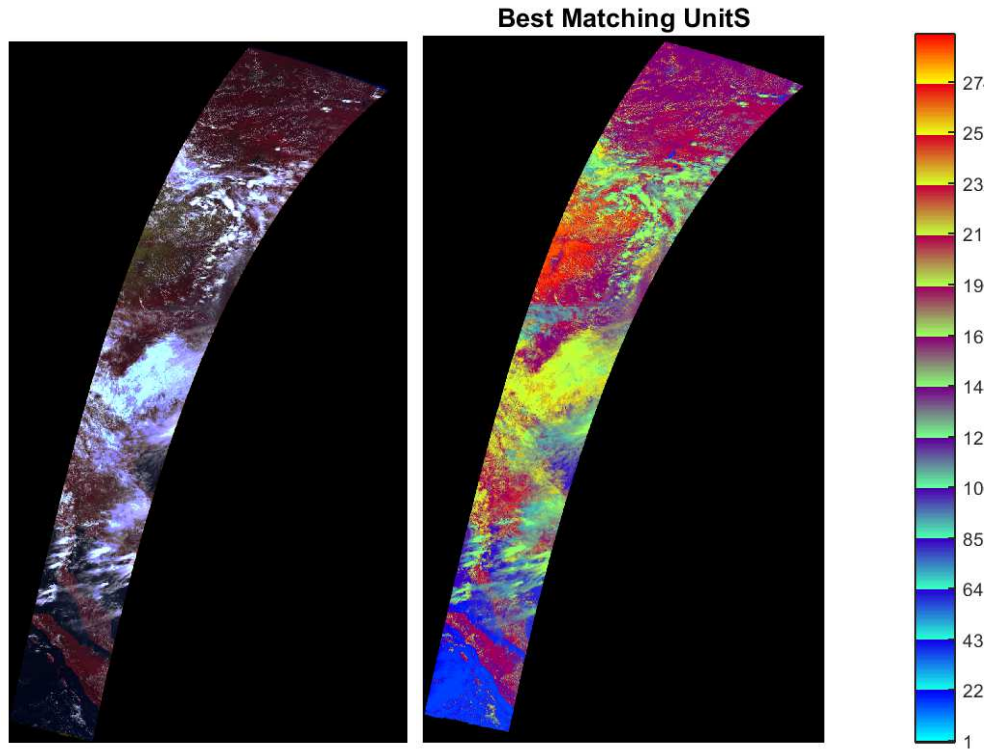
PRODUCT_ID	NAME
1005212284	PROBAV_L2A_20140321_094935_1_333M_V001.hdf
1007605388	PROBAV_L2A_20140621_042255_1_333M_V001.hdf
1007605536	PROBAV_L2A_20140921_204654_2_333M_V001.hdf
1007605798	PROBAV_L2A_20141221_155703_2_333M_V001.hdf

The cloud masks of the four images, the classification maps and the quicklooks are presented hereafter and discussed in the next section.

4.1 PROBAV_L2A_20140321_094935_1_333M_V001.hdf

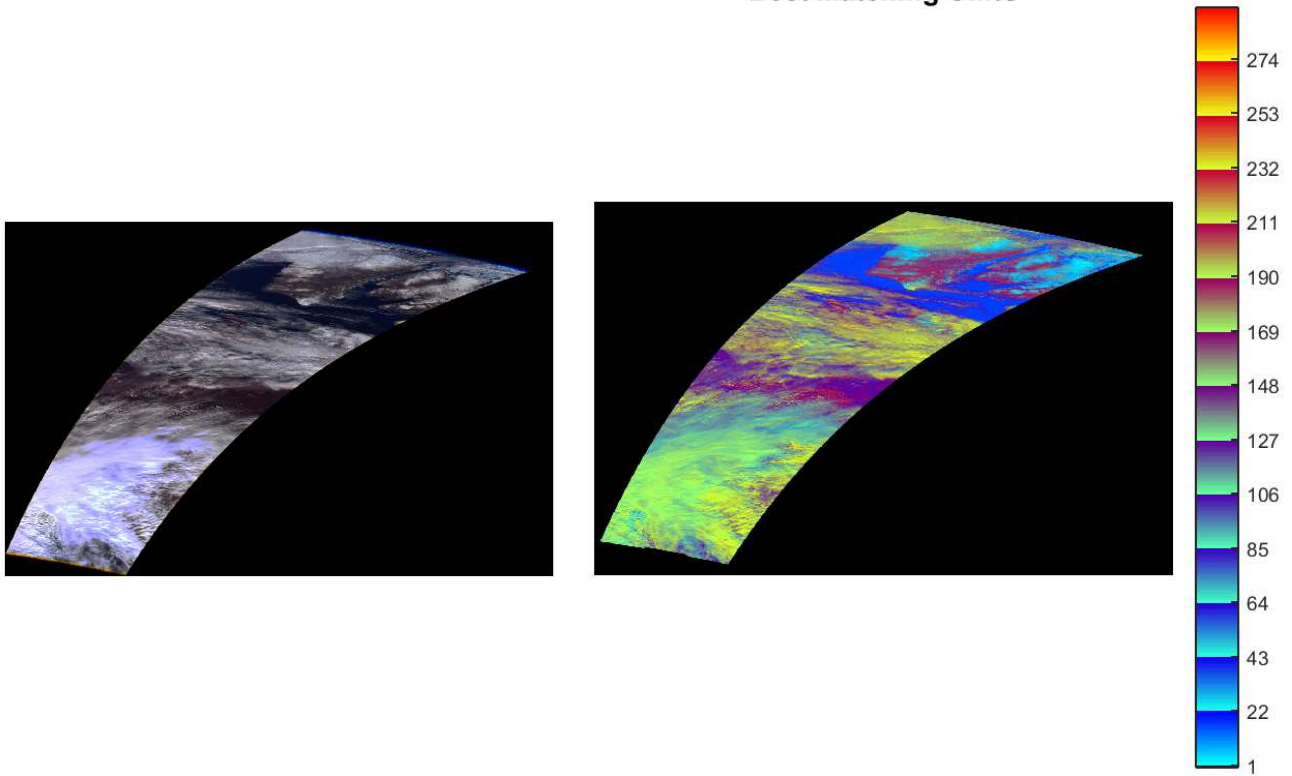


4.2 PROBAV_L2A_20140621_042255_1_333M_V001.hdf

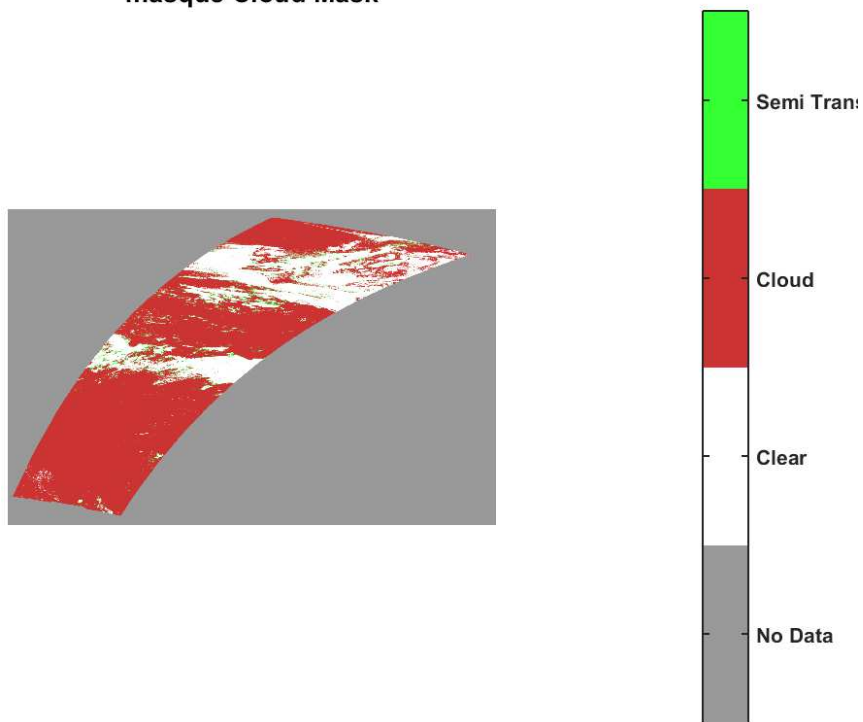


4.3 PROBAV_L2A_20140921_204654_2_333M_V01.hdf

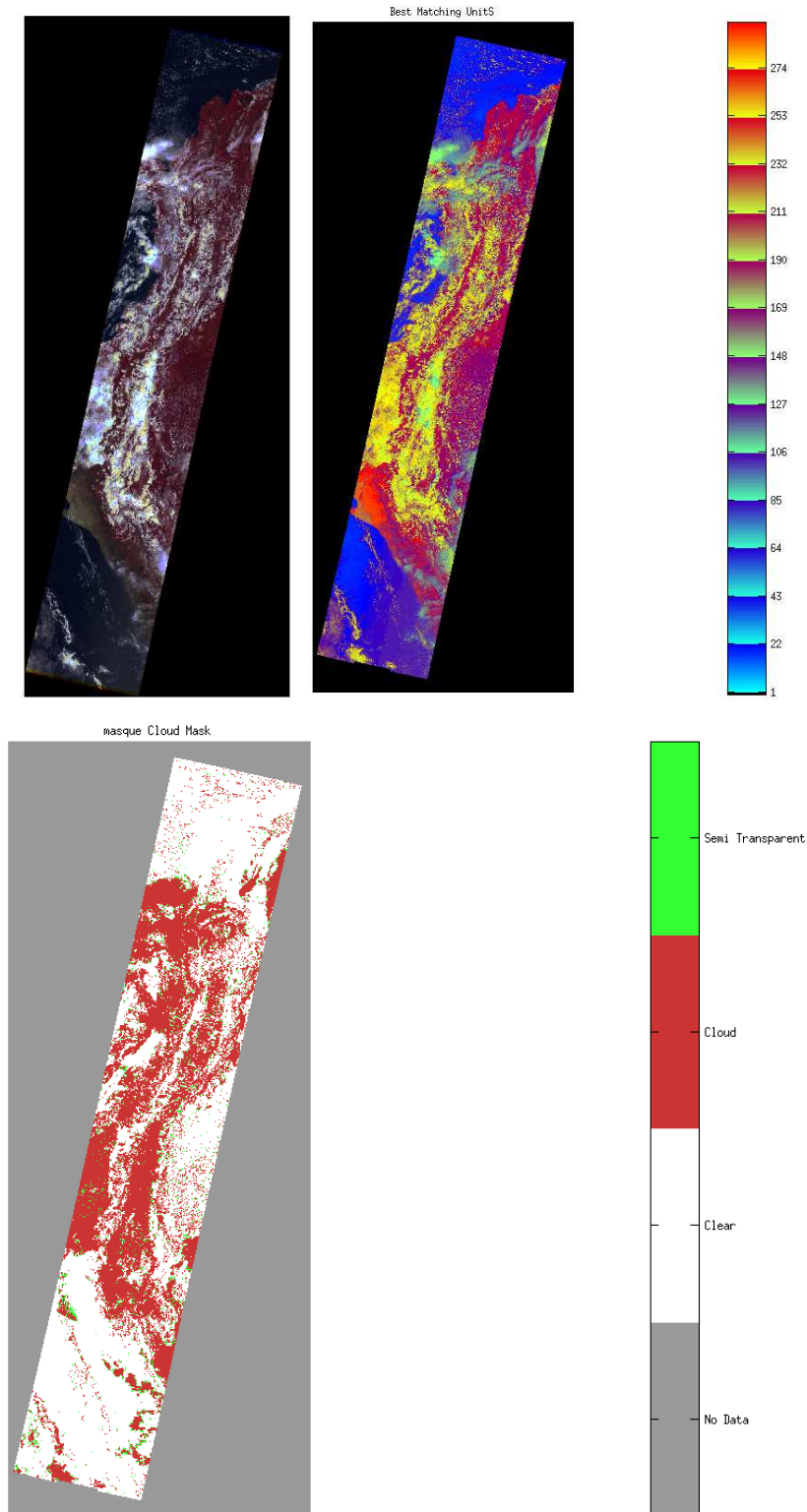
Best Matching Units



masque Cloud Mask



4.4 PROBAV_L2A_20141221_155703_2_333M_V001.hdf



4.5 Result Assessment

The confusion matrix of the results is provided hereafter. It allows to compare the detection of clouds from the method to the test data set provided in the exercise.

First case: if the binary mask is based on clouds, clear and nodata fields, the confusion matrix provides the following results (Table 9).

86.15% of the results are in agreement, which is a very good result.

We have a difference of 10 %, meaning that the SOM did not detect clouds where the Test dataset detect some. On the contrary, 3.63% of the pixels are detected as clouds by the SOM whereas they are detected as cloud free in the test dataset.

The analysis of the classes should provide answers whereas it is a wrong detection (cloud vs clear, or cloud vs semitransparent clouds).

We can also note that No data pixels have been assigned to clear pixels in the test dataset. They were excluded from our processing.

Table 9: Confusion matrix (top, number of points), (bottom, in %)

			Test Dataset		
			-1	0	1+2
			No data	Clear	Cloud + Semitransparent
SOM	No data	-1	0	5	0
	Clear	0	0	458	133
	Cloud Semitransparent	1 +2	0	49	705

			Test Dataset		
			-1	0	1+2
			No data	Clear	Cloud + Semitransparent
SOM	No data	-1	0.00%	0.37%	0.00%
	Clear	0	0.00%	33.93%	9.85%
	Cloud Semitransparent	1 +2	0.00%	3.63%	52.22%

When the cloud mask is explained by dense clouds and semitransparent clouds, the results are the following:

65.93 % of the masks are identical. The main differences come from the detection of semitransparent clouds from the SOM which are detected as clouds (19.41%) in the test dataset. The labelling of the classification maps has been set to be **conservative** (it is better to overdetect than underdetect clouds for the downstream applications except cloud characterization).

The rate of mis detection is low, less than 3 %.

If semitransparent clouds are assimilated as clouds, the

Table 10: Confusion matrix (top, number of points), (bottom, in %)

			Test Dataset			
			-1	0	1	2
			No data	Clear	Cloud	Semitransparent
SOM	No data	-1	0	5	0	0
	Clear	0	0	458	24	109
	Cloud	1	0	39	415	262
	Semitransparent	2	0	10	11	17

			Test Dataset			
			No data	Clear	Cloud	Semitransparent
			-1	0	1	2
SOM	No data	-1	0	0.37%	0.00%	0.00%
	Clear	0	0	33.93%	1.78%	8.07%
	Cloud	1	0	2.89%	30.74%	19.41%
	Semitransparent	2	0	0.74%	0.81%	1.26%

Could it be corrected ?

When working with topological maps, the frontiers cloud/semitransparent clouds, and semitransparent cloud/clear pixels are easy to set. The labels could be pushed from one or more classes in the topological map. This is the role of the expert which introduces its knowledge.

5 Practical considerations

5.1 Computation time

As all neural networks, the computation time of the Kohonen classification is very good.

5.2 Numerical Computation Considerations

We do not anticipate problems with numerical stability and/or round-off errors.

5.3 Constraints, Limitations, Assumptions

No constraints have been identified. We assumed that the coregistration of the channels as well as radiometric calibration is well performed.

6 Conclusion

This project aimed to compute Proba-V cloud mask without any information than the one provided by the instrument itself. The image classification using the Kohonen relies only on the use of the TOA reflectances. It is based on a two-step method. The first one concerns the determination of the map, the second its application to Proba-V. The results are coherent, and visual comparisons do not let appear strong misdetection although only four spectral bands are available against the 7 of MODIS.

We had in the frame of the MODIS/Landsat cloud detection implementation experienced more tests (to assess the spatial homogeneity for instance using the standard deviation in input, to specialise the map for shadow detection (cloud and orography)). We had also specialised the map for snow detection.

Some post processing could also be added (such as mask dilatation).

Some information has not been taken into account mainly by lack of time, but ways for improvements have been identified in future work.

7 Acknowledgments

The author wishes to thank ESA for the organization of this very interesting exercise.

8 Appendix 1 Problems encountered

Some problems with the data are reported here.

8.1 HDF5 Attributes for geometry

For instance, two attributes for the group '/GEOMETRY' are wrongly computed both for the latitude/longitude, and X/Y position.

The computation should be (respectively the same correction has to applied for the longitude, X, Y coordinates)

$$'TOP_LEFT_LATITUDE' - ('TOP_LEFT_LATITUDE' - 'BOTTOM_LEFT_LATITUDE')/2$$

Whereas it is

$$('TOP_LEFT_LATITUDE' - 'BOTTOM_LEFT_LATITUDE')/2$$

This applies to the computation of the center in latitude/longitude, but also in distance. This error avoids to locate by its center the proba-V acquisition.

Group '/LEVEL2A/GEOMETRY'

Attributes:

```
'TOP_LEFT_LONGITUDE': 148.505951
'TOP_LEFT_LATITUDE': -1.000000
'TOP_RIGHT_LONGITUDE': 159.845245
'TOP_RIGHT_LATITUDE': -1.000000
'BOTTOM_LEFT_LONGITUDE': 148.505951
'BOTTOM_LEFT_LATITUDE': -14.232142
'BOTTOM_RIGHT_LONGITUDE': 159.845245
'BOTTOM_RIGHT_LATITUDE': -14.232142
'CENTER_LONGITUDE': 5.669643
'CENTER_LATITUDE': 6.616071
'TOP_LEFT_X': 49898.000000
'TOP_LEFT_Y': -336.000000
'TOP_RIGHT_X': 53708.000000
'TOP_RIGHT_Y': -336.000000
'BOTTOM_LEFT_X': 49898.000000
'BOTTOM_LEFT_Y': -4782.000000
'BOTTOM_RIGHT_X': 53708.000000
'BOTTOM_RIGHT_Y': -4782.000000
'CENTER_X': 1905.000000
'CENTER_Y': 2223.000000
```

8.2 Image size with change International date line is crossed

When the images are acquired over International date line (-180 /180), the images are mainly filled by No Data values, produced big images whereas only a small part contains physical values.

This is the case for instance for the image "PROBAV_L2A_20141221_225542_1_333M_V001.HDF5", which has 120960 columns and 4139 lines. The management of these acquisitions is much heavy than the other one since it required more computing memory. To overcome this problem, the images have been processed being split in a lot of lines.

Here are the latitude/longitude of the image corner, and the cloud mask.

'TOP_LEFT_LONGITUDE': -180.000000
'TOP_LEFT_LATITUDE': -34.633930
'TOP_RIGHT_LONGITUDE': 179.997025
'TOP_RIGHT_LATITUDE': -34.633930

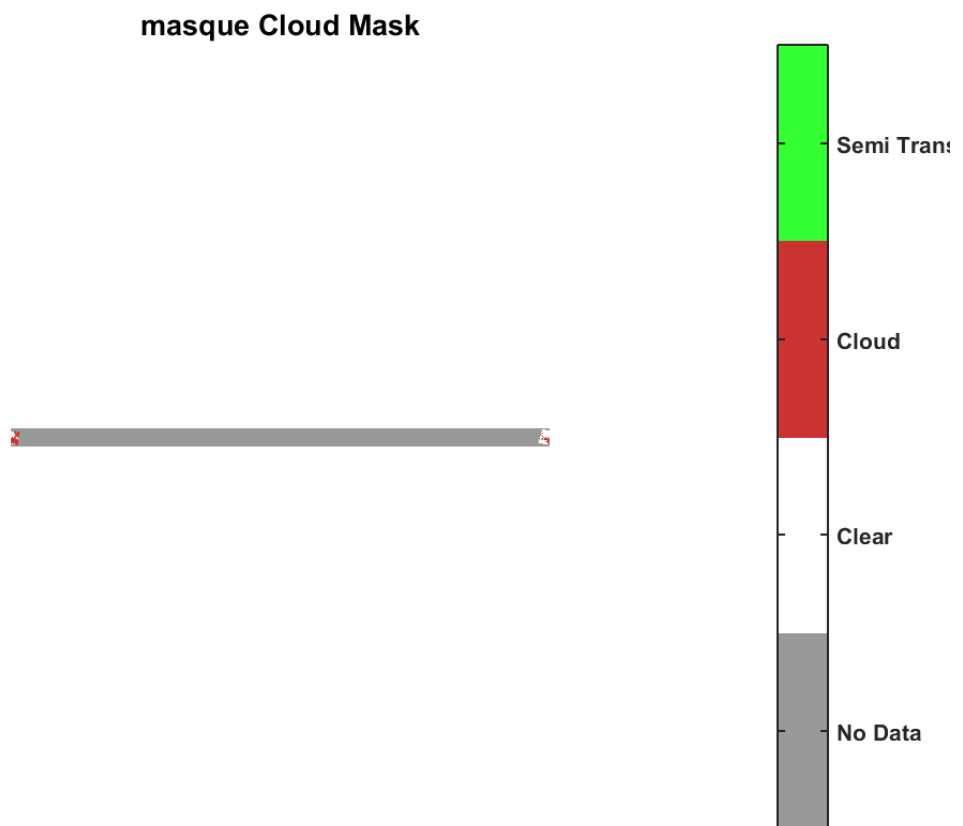


Figure 21: Image filled with No Data

8.3 Data reader

Matlab and Idl were not able to read the datasets of the hdf5 Proba-V images. The reason is probably the zip compression software that does not work with both applications.