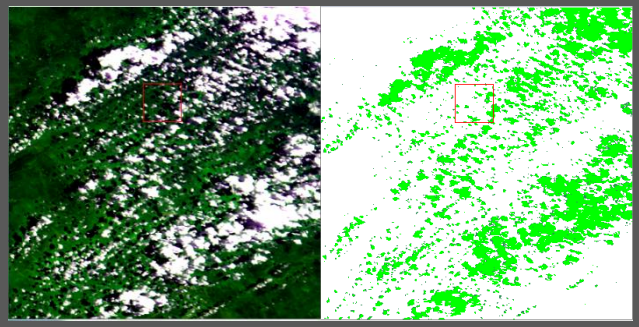
magellium

PUTTING KNOWLEDGE ON THE MAP

PROBA-V CLOUD ROUND ROBIN

Béatrice Berthelot

01/03/2017



Objectives

- The objective was to capitalize on former activities related to cloud detection to propose a method for Proba-V.
- The purpose is to examine the feasibility of an automatic cloud detection method already applied successfully to MODIS and Landsat data using neural networks (Kohonen classification) and assessed its accuracy on Proba-V data.
- The idea was to use the MODIS results as a reference that could be transferred to Proba-V
- ➔Assessment of the genericity of the method

Presentation overview

- Method description
- Results
- Potential for improvement

Background

- 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.

- 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.

- This work showed that a well validated topological map created using Kohonen network could be used generically to be adapted to various sensors.
- So why not Proba-V ?

Kohonen network or the self-organising feature map

- A Kohonen map incorporates in an unsupervised way the topology present in the data.
- Kohonen maps operate in two modes: training and mapping.
  - 1) The first step in the classification is to summarise the information contained in the training data set by producing a set of reference vectors that are representative of the data.
  - 2) The second step consists in associating a referent vector and a label

Implementation

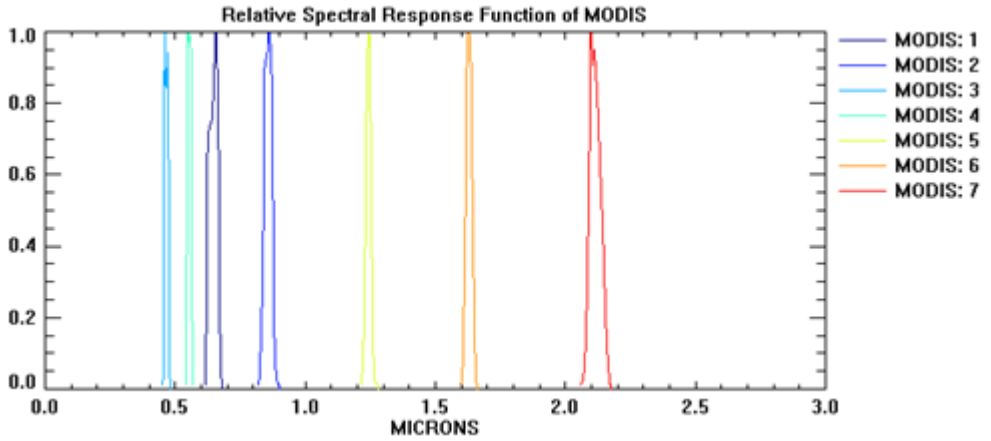
- The algorithm is based on two main modules composed of several sub systems which chain:
- Module 1: MODIS Classification
  - Subsytem 1: Learning
  - Subsytem 2: Classification unsupervised
  - Subsytem 3: Bayesian classification
- Module 2: Adaptation to other sensors
  - Subsystem 1 : Adaptation of the reference topological map to the other sensor

Algorithm development

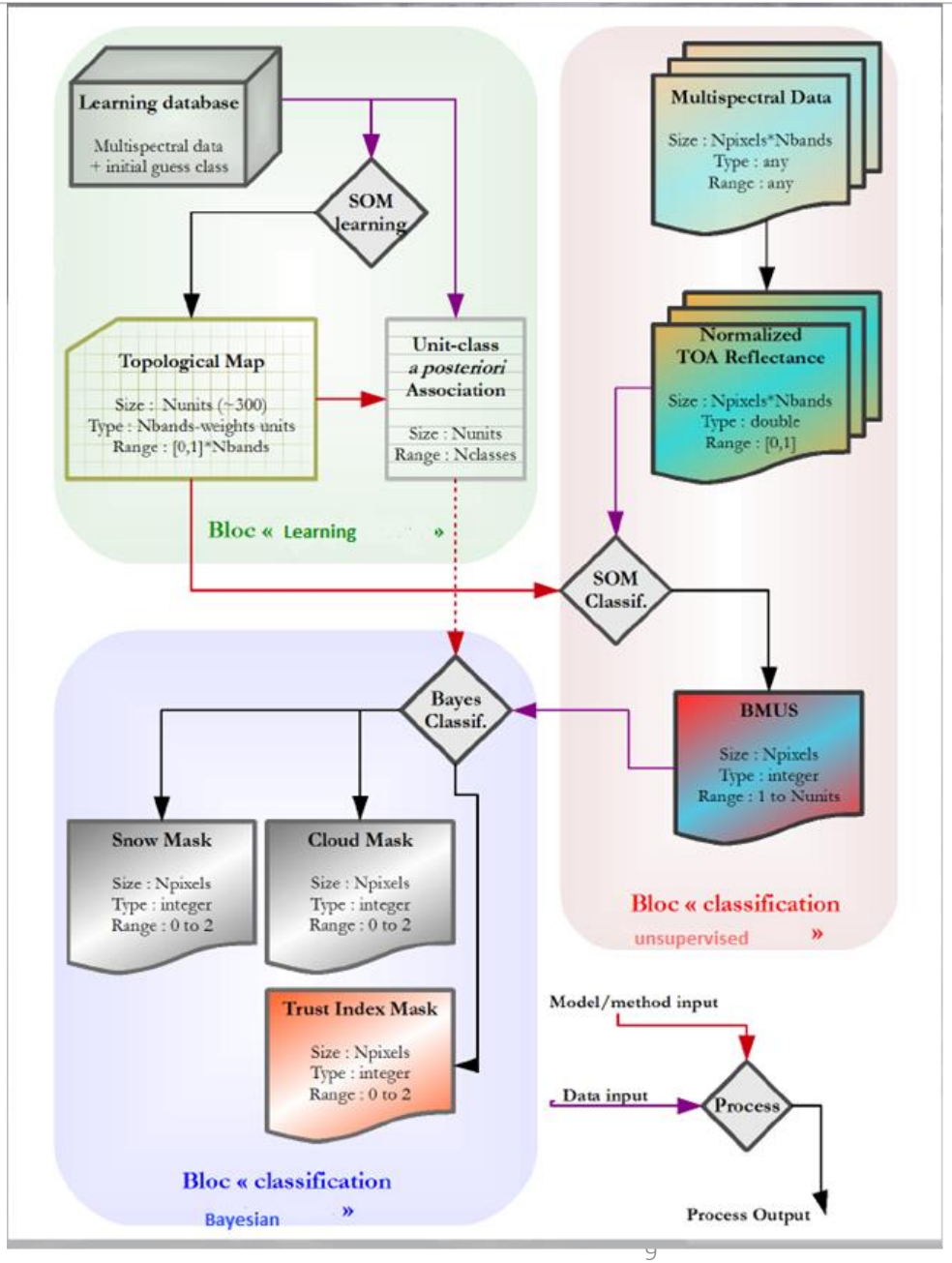
- The development of the method relies on six different phases which are:
- 1) Reference data collection
- 2) Creation of the topological map;
- 3) Classes labelling;
- 4) Validation on reference sensor;
- 5) Derivation of the map for a sensor X ;
- 6) Validation of the results.

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.
- In the study, we used the seven bands dedicated to land studies.



Module 1 : Cloud detection implementation and assessment for MODIS

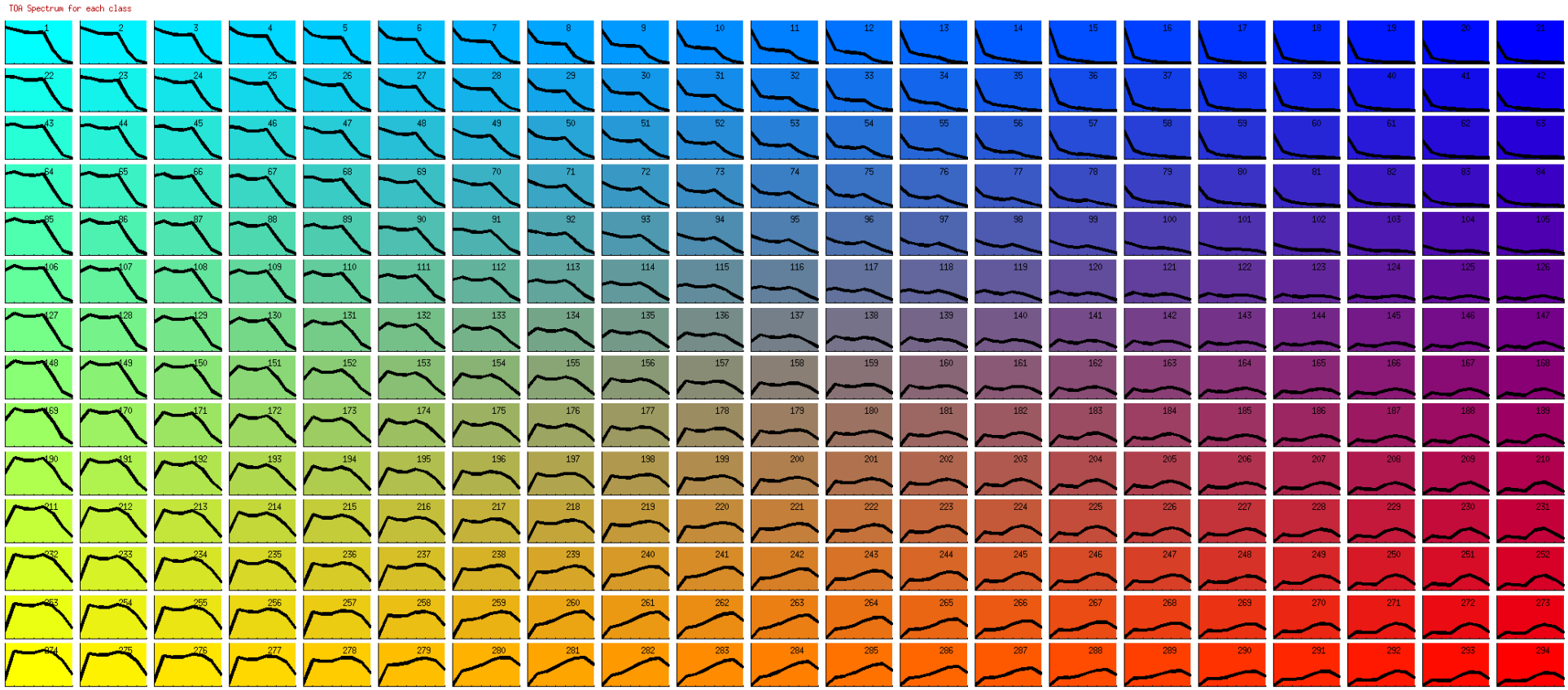


Creation of Training data sets

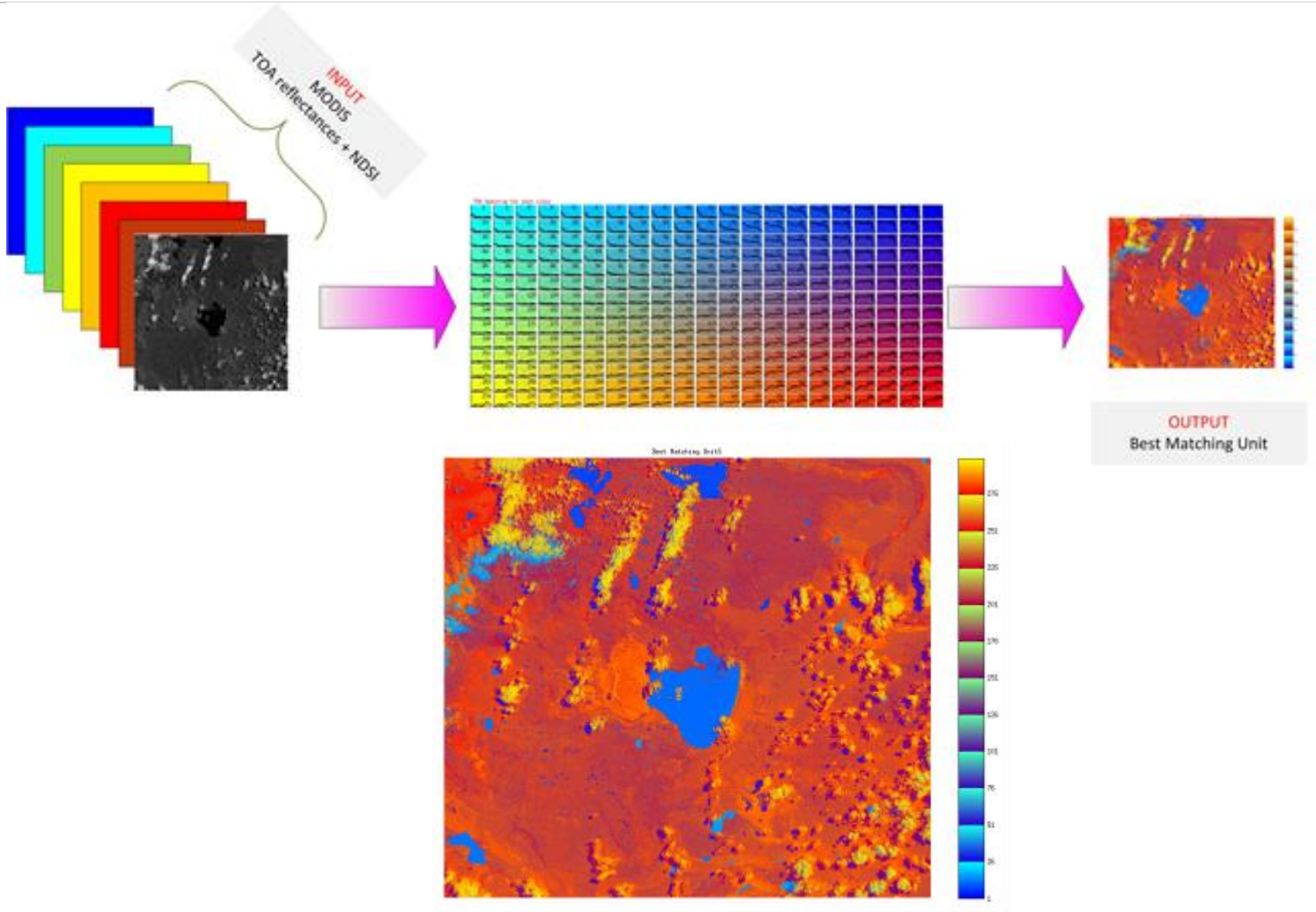
- We have built a learning base with MODIS data with precautions. This database consists of 20 different images, from which a total of 1 million pixels are extracted. These pixels were sorted thanks to the MODIS cloud mask to account for bright and semi-transparent clouds, cloud shadows and clear pixel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

Training of the map

- Topological map representation

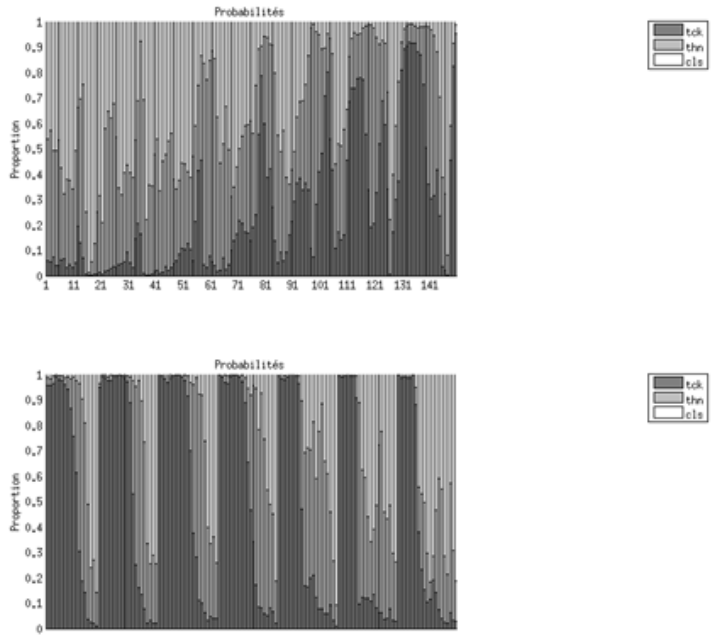


Application of the map



Classes labelling

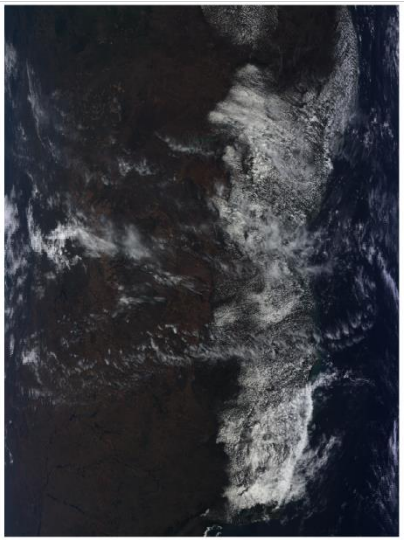
- The objective of the subsystem is to provide a mask containing at least the binary information cloud and clear pixels, and possibly semi-transparent cloud.
- ➔we performed a supervised statistical learning using the existing MODIS cloud masks (bayesian step) to assign a label automatically.
- We associate to each unit of the topological map the value of the mask which has been the most frequently associated.

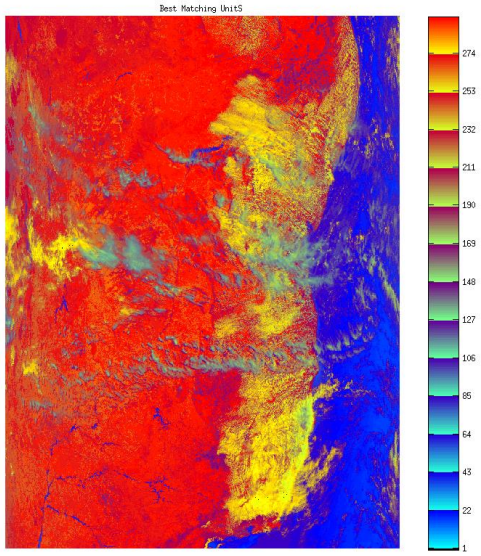


Validation

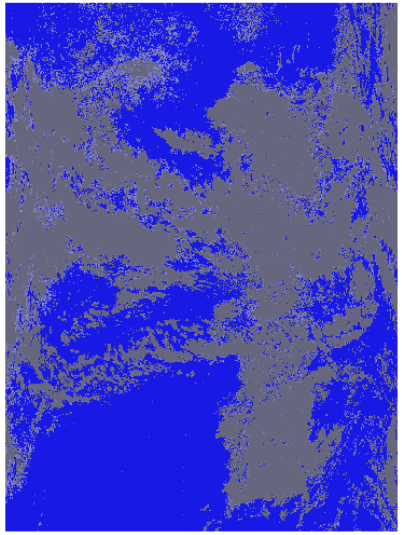
- The test base has been built from 45 selected images.
  - We compare masks of the type:
  - 0: cloud;
  - 1: semitransparent cloud;
  - 2: Clear.
-

Example

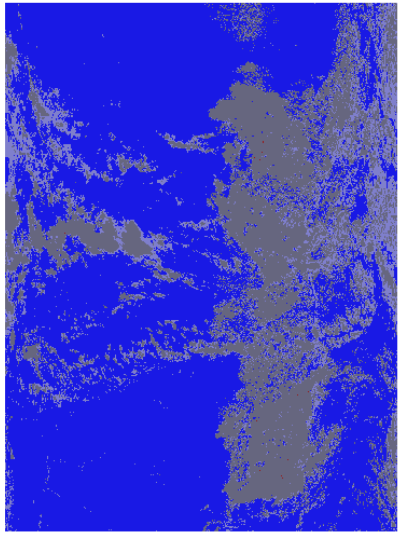


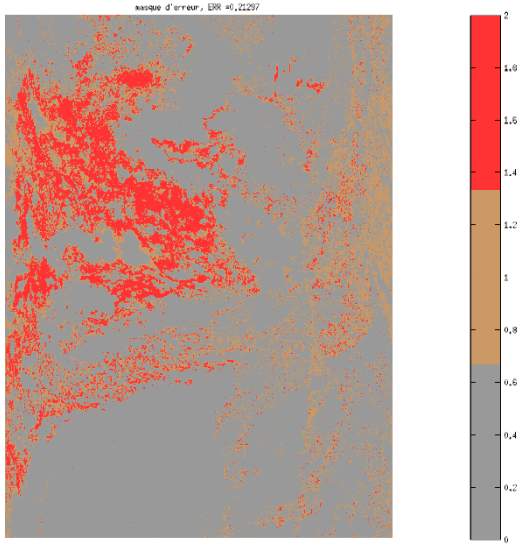


masque MODIS cloud



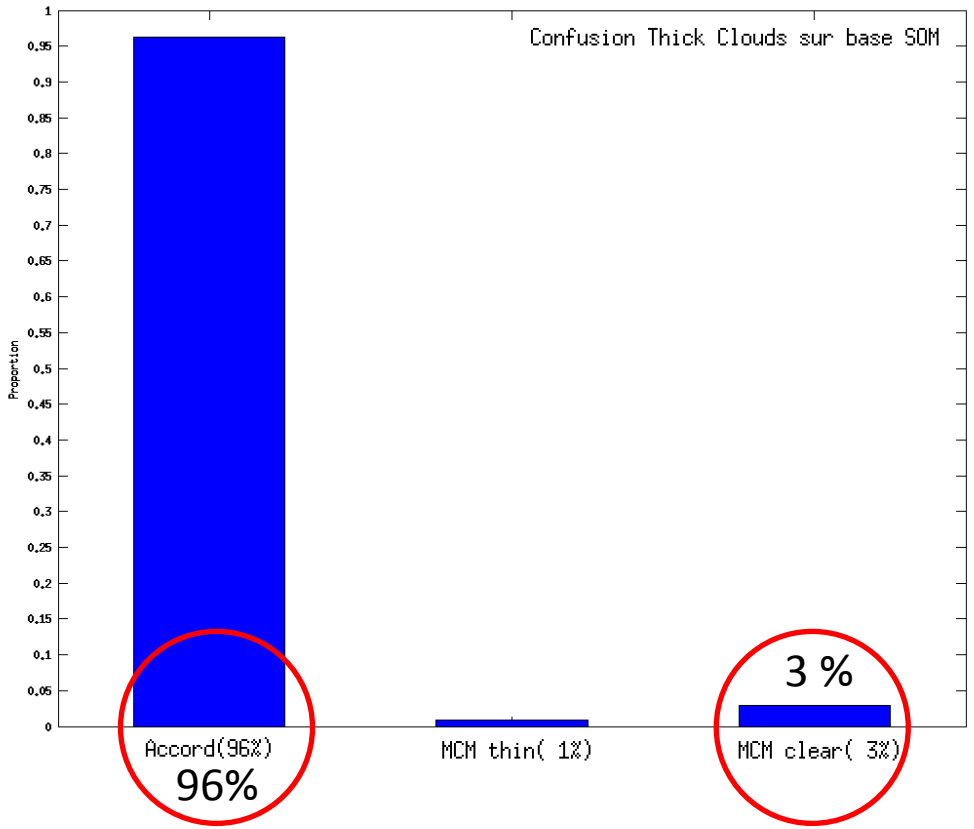
masque BMUS



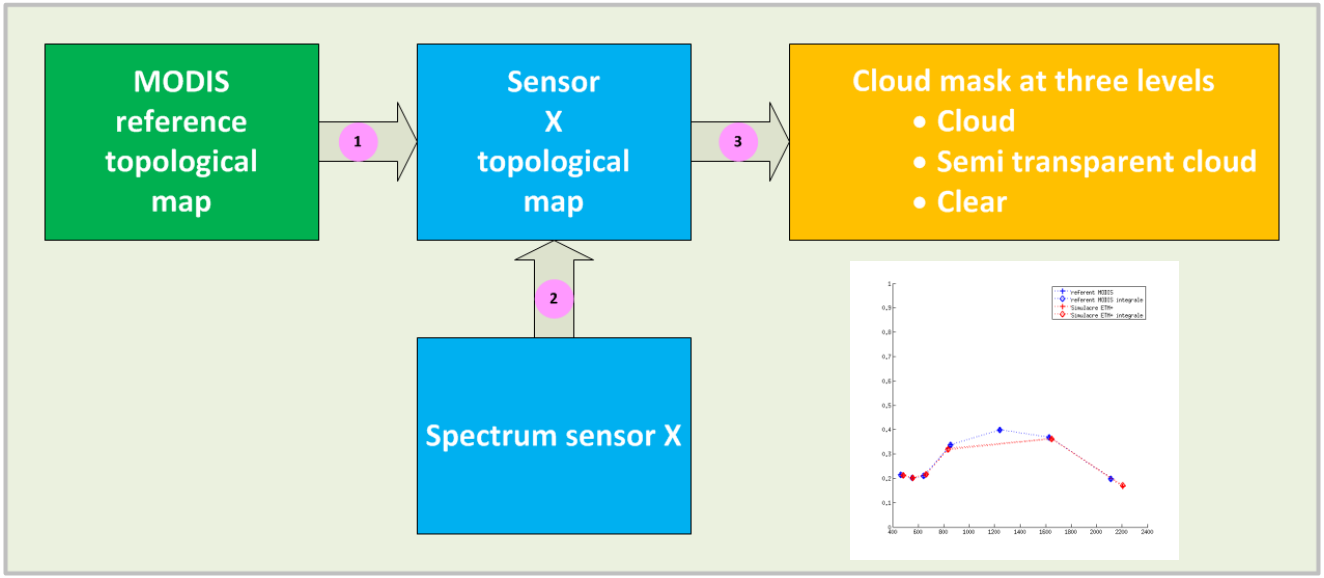


Validation

- Confusion matrix



Module 2: Adaptation of the reference topological map to the other sensors



Advantages

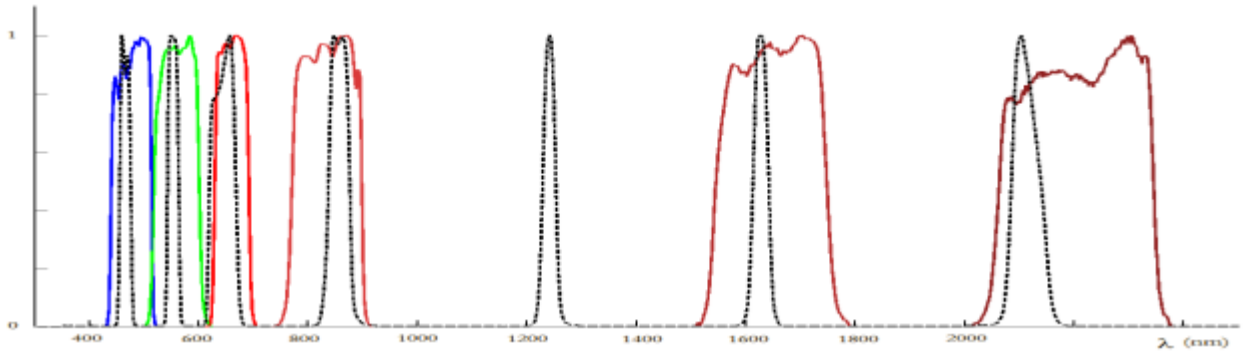
- The topology learning on MODIS data is saved;
- The associations neuron-classes learned on MODIS are saved;
- The sensor X could have less spectral bands than MODIS;
- The cloud spectrum interpolation is still valid because the cloud spectra are smooth.

Inconvenients

- 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.

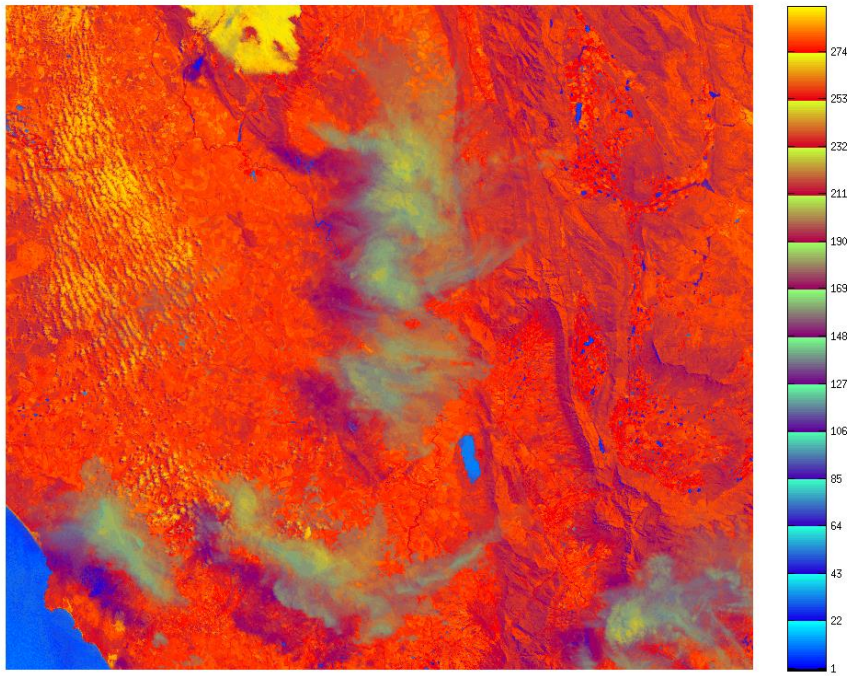
Assessment and validation using Landsat dataset

- We have computed a topological map for Landsat and one for MODIS
- We have cloud masks for landsat to validation the method
- RSR comparison

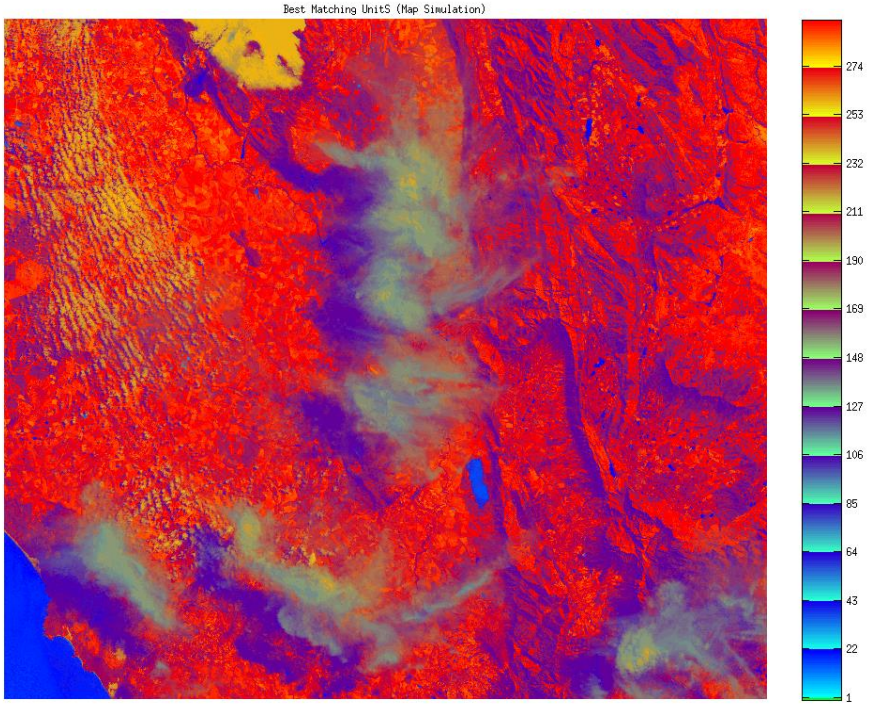


Results comparison

- We computed cloud mask from Landsat data and from Landsat simulated from MODIS



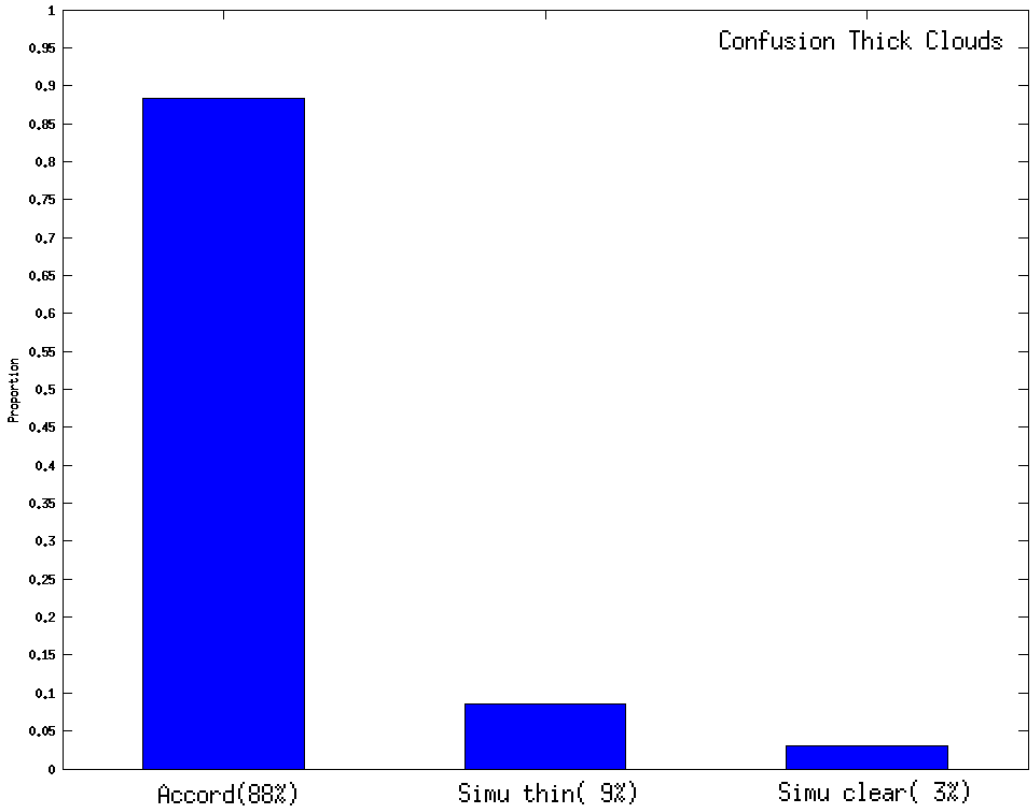
Classification using Landsat SOM



Classification using MODIS SOM

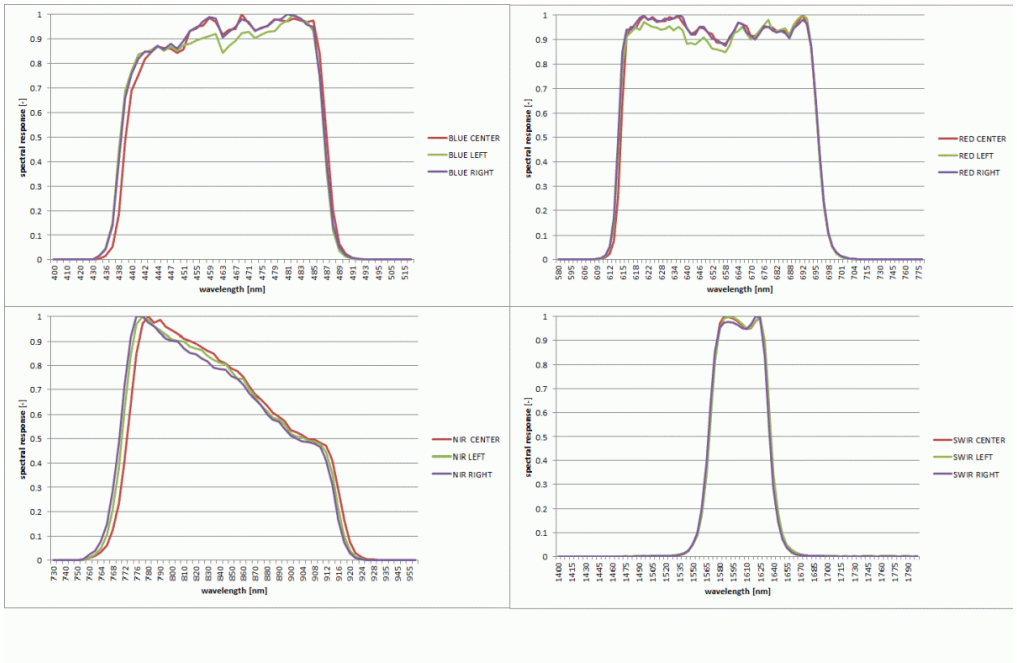
Validation and quality assesment

- Confusion matrix (10 Landsat images)



Application to Proba-V

- MODIS referent vectors of the SOM are spectrally resampled for Proba-V channels using the RSR (middle camera)

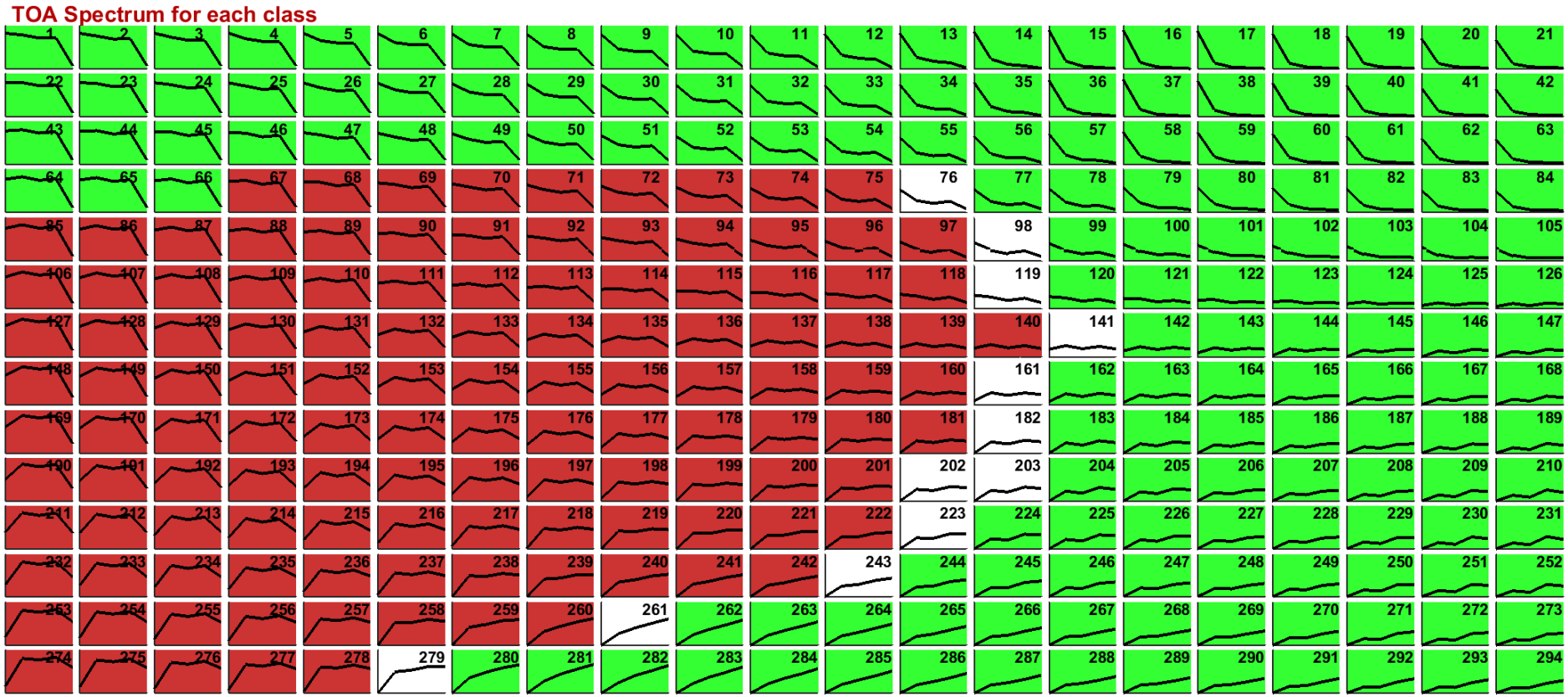


- Spectrum based on 7 TOA reflectances ➔ 4 for Proba-V

Proba-V topological map

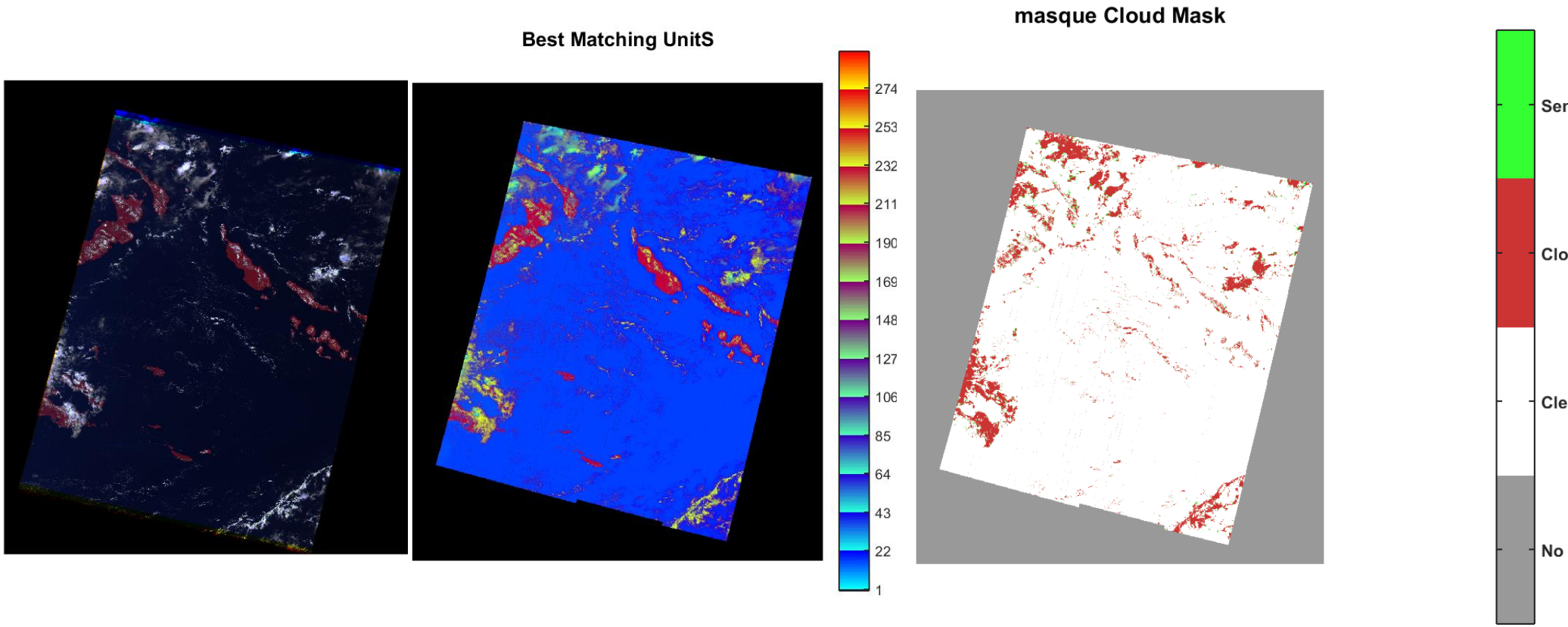
- Labelling

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



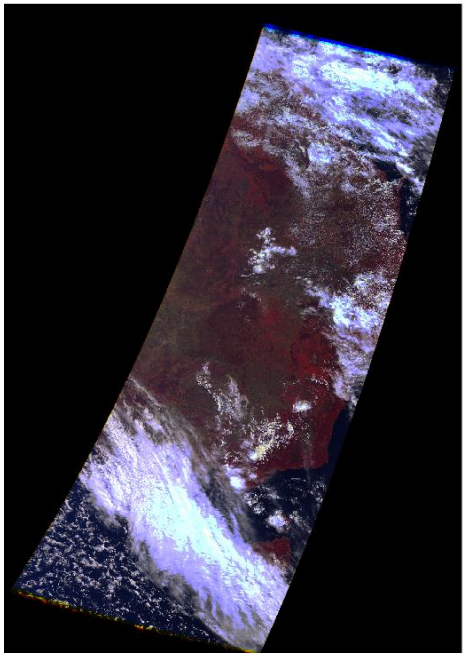
Application

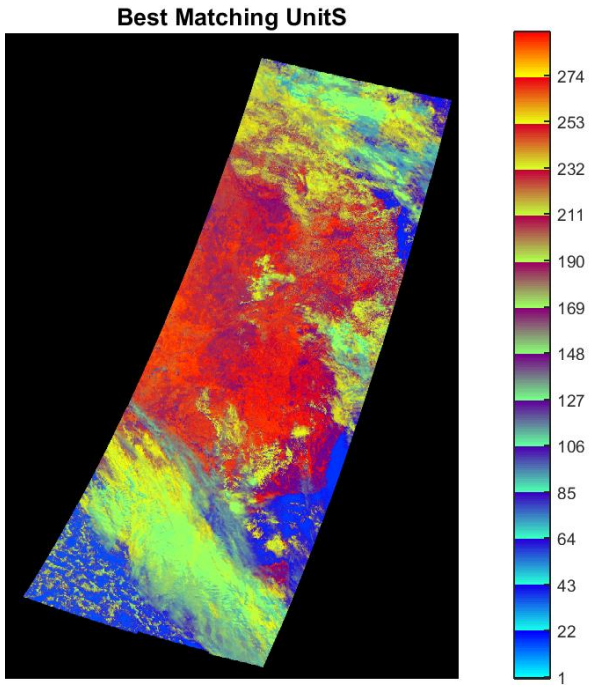
- **PROBAV_L2A_20140321_000027_3_333M_V001**

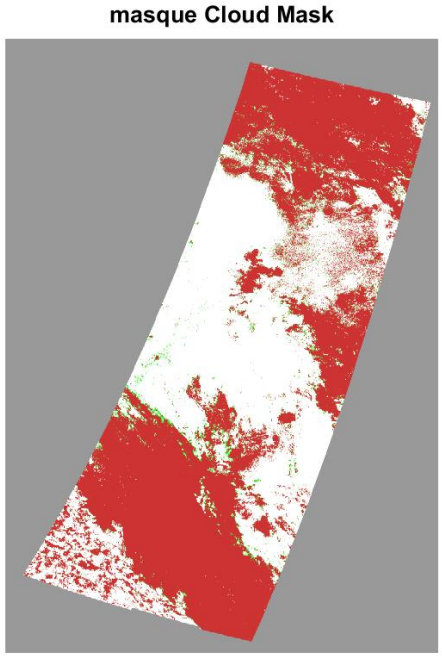


Assessment using the TDS

- **PROBAV_L2A_20140321_000512_3_333M_V001**

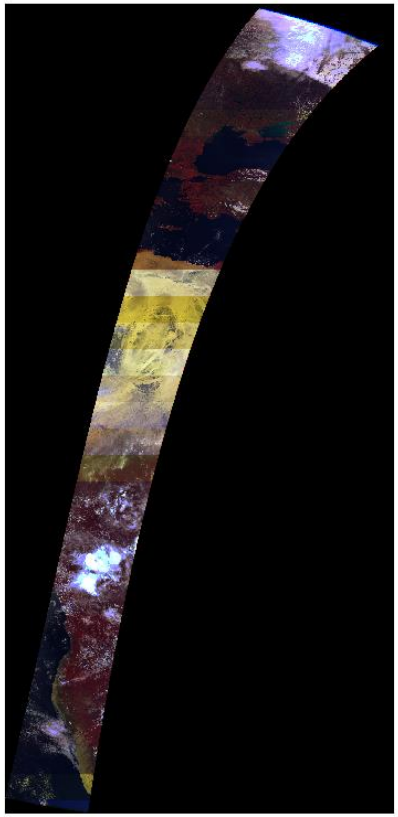


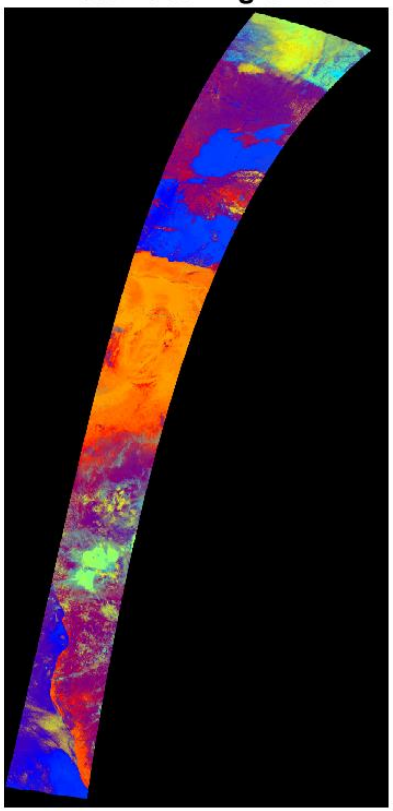


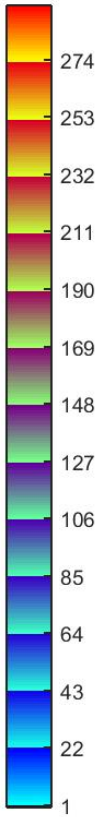


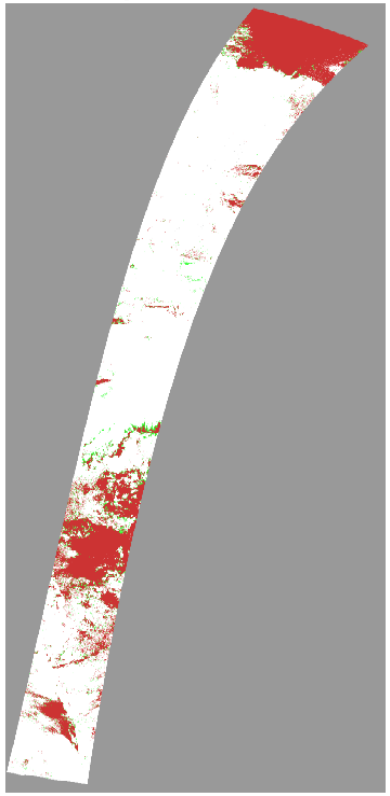
Assessment using the TDS

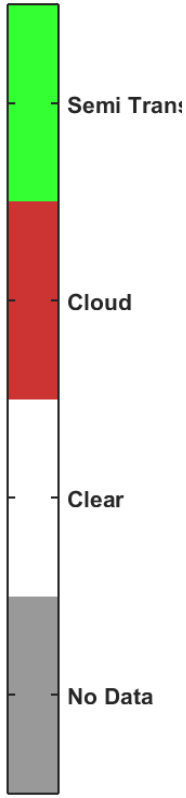
- PROBAV_L2A_20140321_094935_1_333M_V001.hdf





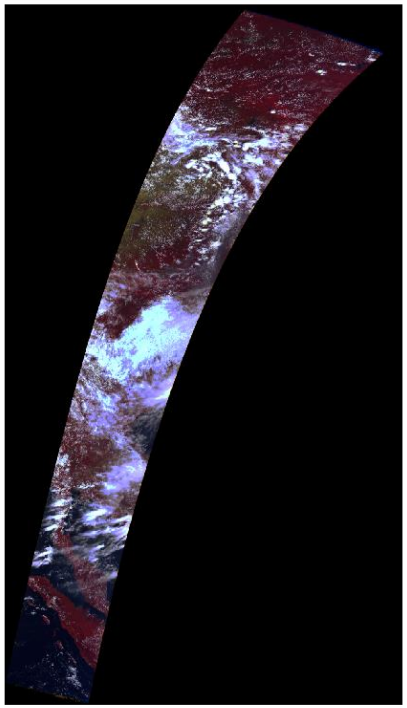


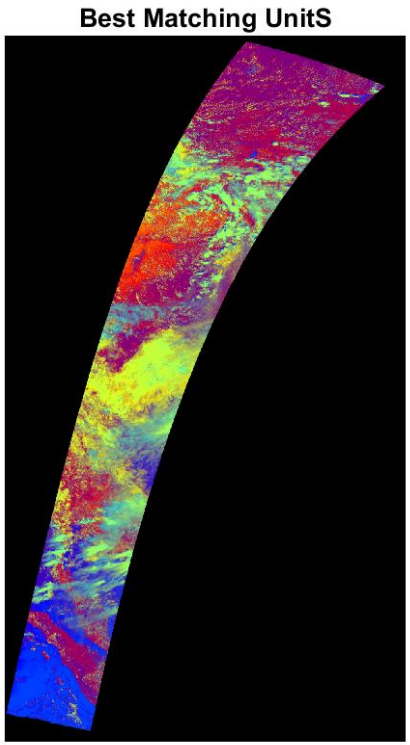


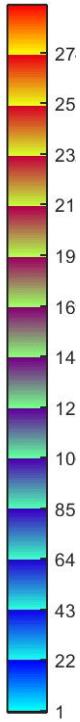


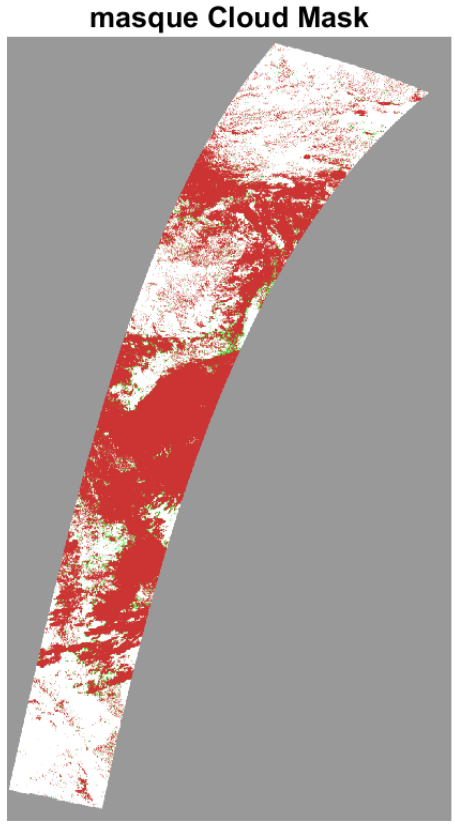
Assessment using the TDS

- **PROBAV_L2A_20140621_042255_1_333M_V001.hdf**



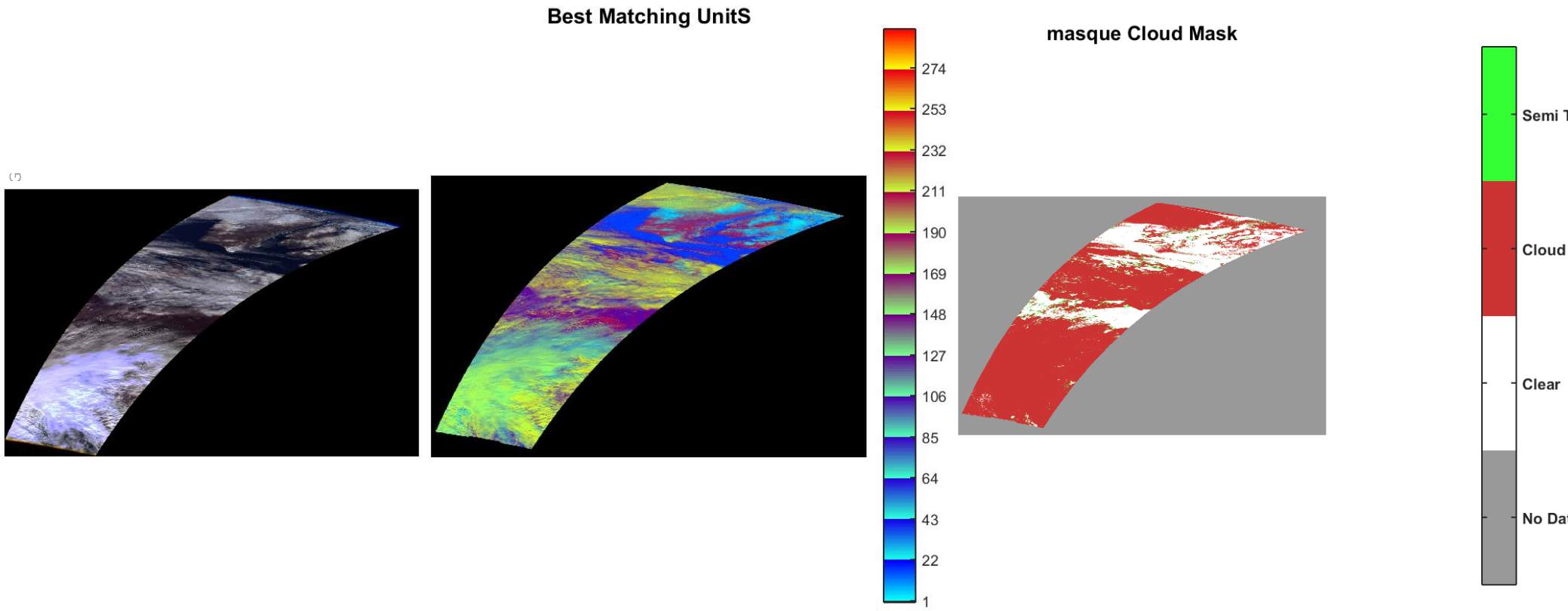






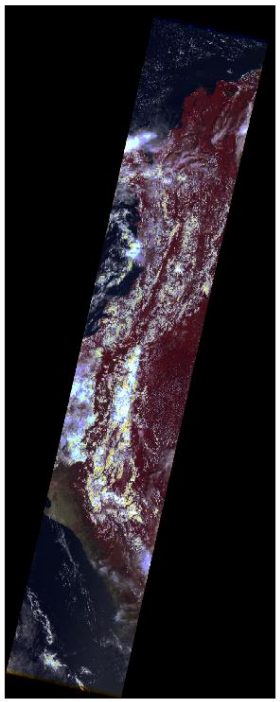
Assessment using the TDS

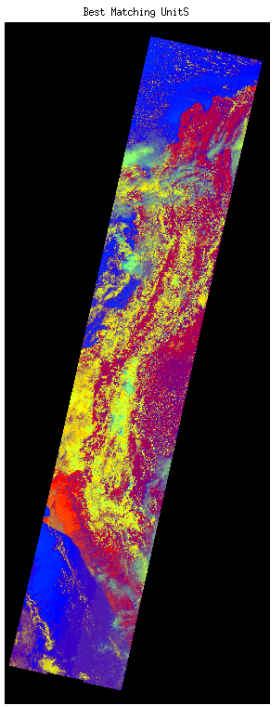
- **PROBAV_L2A_20140921_204654_2_333M_V01.hdf**

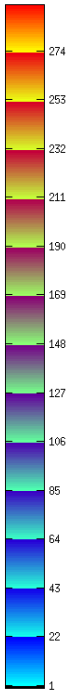


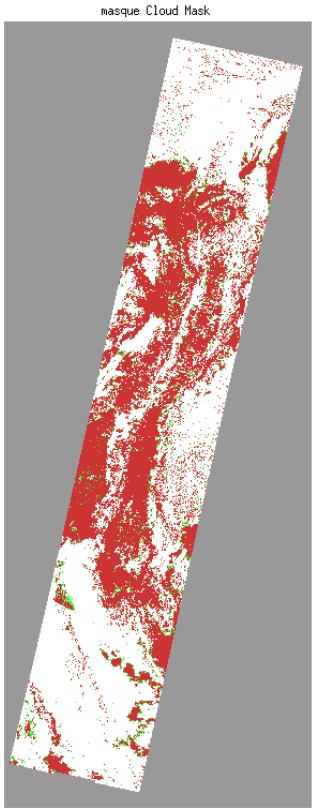
Assessment using the TDS

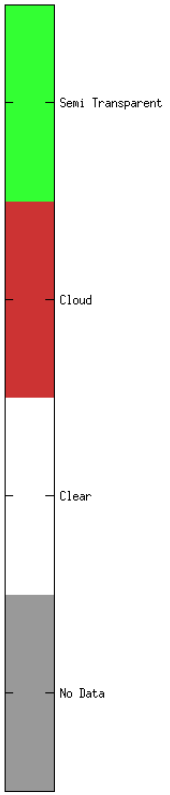
- **PROBAV_L2A_20141221_155703_2_333M_V001.hdf**











Comparison with the TDS

- Confusion Matrix cloud/clear

			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%**

Comparison with the TDS

- Confusion Matrix cloud/clear/semitransparent

			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%**
	Semi transparent	2	0	0.74%	0.81%	**1.26%**

Choice to be cloud conservative

Low mis detection (1.78 + 2.89%)

Potential for improvement

- Thanks to the validation report and the TDS, we have identified weakness of the algorithm.

- Several possibilities to improve the results
  - Just remind that the SOM has been built for land applications. So the number of pixels of the database located in the ocean and coastal waters is low.
    - Not difficult to change

Potential for improvement

- **Could the transitions 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.
  - Increase the number of classes ?

Implementation

- Developped in matlab
- Auxiliary data : Relative Spectral Response
- Library for SOM needed – Exists in many languages

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 SO 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.
  - >Learning using Proba-V directly, use the Status map for the labelling