Proba-V Clouds Detection Round Robin Protocol
Discriminant Analysis

Umberto Amato*, Guido Masiello^, Carmine Serio^

*Istituto per le Applicazioni del Calcolo CNR, Napoli (Italy)

^Scuola di Ingegneria, Università della Basilicata, Potenza (Italy)

1. Introduction

The present work is part of the Round Robin exercise on Proba-V cloud detection (PV-CDRR) organized by the European Space Agency (ESA) in collaboration with the Belgian Science Policy Office (BELSPO).

Aim of the exercise is to intercompare some methodologies for cloud detection from Proba-V images, particularly to improve accuracy in particular conditions where issues still exist (e.g., illumination and viewing geometry, edges of typologies of surface, etc.).

As well known to detect clouds is not an easy task. On one side clouds themselves are not sharply defined; on the other side cloud detection is based on the contrast between clouds and underlying surface, therefore surface characterization is important for a successful cloud detection, but is itself a hard problem.

Practically all methodologies for cloud detection are based on the paradigm of (supervised) classification that goes through the following steps:

1) (training) to provide a set of radiances in *known* conditions of clear or cloudy sky;
2) (tuning or estimation) from this to tune parameters relevant to the two sky conditions relying on a classification methodology;
3) (test or prediction) finally to predict clear o cloudy condition on new radiances basing on the classification methodology and its parameters estimated at the tuning step.

Methodologies differ in several features that we can summarise in

a) how radiances at the taining step are obtained (simulated by a Radiative Transfer Model in several atmosphere conditions or real measurements by some sensor)
b) how sky conditions are estimated at the training step (naturally in the case of simulated radiances, by visual inspection of the image; by direct analysis of radiances, by some algorithm)
c) which methodology is adopted for the tuning and prediction step (physical thresholds, Statistical or Machine Learning methodologies (e.g., Discriminant Analysis, Support Vector Machine, K Nearest Neighbour)
d) which variables are directly used by the methodologies (radiances, reflectances, their linear and nonlinear combinations, Principal Components, physically based indicators, etc.)

e) choose some climatological regions and/or surface typologies within which radiances in clear or cloudy conditions can be considered homogeneous.

Finally we mention that an important aspect of a cloud detection algorithm is to properly balance its errors, intended as cloudy pixels classified as clear and clear pixels classified as cloudy. According to the approach they are called Type I error or False Positive or False Alarm and Type II error or False Negative or miss. Whereas the top (unreachable) objective is to have zero errors of both type, a significant question is how we want to balance the unavoidable mistakes: by application driven arguments (false cloudy are to absolutely be avoided otherwise a product is wrong), by global reliability (both clear and cloudy conditions are equally important), etc.

Our approach to the Proba-V exercise has mainly been to adapt the methods we had developed for other sensors to Proba-V. Actually all of them were thought for a high and moderately high number of spectral bands and with the presence of IR channels. Therefore we expected to drop all improvements based on transforming radiances and computing physical indicators.

2. Methodology

The general methodology that we have followed in the Proba-V exercise is classification: a labelled data set is known (i.e., scenes and the corresponding Clear/Cloudy status) from which parameters are estimated (training and tuning phase) and used to classify actual (unlabelled) scenes (test phase). In addition to make the training data set as homogeneous as possible some typologies of surface are defined on each one of which parameters depend.

2.1 Training data set (MODIS and SEVIRI)

A training data set whose labels are "true" is called a *gold* standard. This could be the case of the test data set provided at the end of the exercise or the 20000 blind radiances that have been considered for evaluating the methodologies. When the data set labels are assigned by another algorithm, then they are referred to as *silver* standard, to stress that they are not exempt from errors.

Two different cloud masks have been considered for the training set in the Proba-V exercise: SEVIRI and MODIS.

2.1.1 SEVIRI cloud mask

A cloud mask of one hemisphere is provided every 15 minutes based on SEVIRI sensor onboard METEOSAT Second Generation satellites. Spatial resolution is 3Km sub-satellite but degrades far from the equator and from the Greenwich meridian. As a consequence the grid is variable but fixed with respect to the time (the satellite is geostationary). For reliability reasons the cloud mask is provided only within a radius of about 60 degrees around the point at zero latitude and longitude. Of course no data are provided for the hemisphere including Americas, Oceania and most Asia.

The grid of the SEVIRI cloud mask is 3712x3712 pixels; due to the limit of 60 degree the number of valid cloud mask pixels is 11,953,264 (86.75%). To co-locate SEVIRI and PROBA scenes we resampled the SEVIRI grid to a uniform grid in latitude and longitude having the same number of pixels as the original SEVIRI images. This choice preserves the original space resolution close to the centre of the SEVIRI images. Of course far from the centre the finer resolution of the new grid is fake and the SEVIRI cloud mask is simply repeated within the coarser grid. Technically the procedure is equivalent to a Nearest Neighbour interpolation.

For the purpose of co-locating a Proba-V scene to SEVIRI we considered the closest SEVIRI scene with respect to time and, due to the coarser grid of SEVIRI than Proba-V, the closest pixel in space regardless of the amount of difference.

SEVIRI cloud mask provides four different conditions (clear over sea or land, cloudy, uncertain). We considered as valid SEVIRI pixels only the clear and cloudy ones.

Figure 1 shows an example of SEVIRI cloud mask.

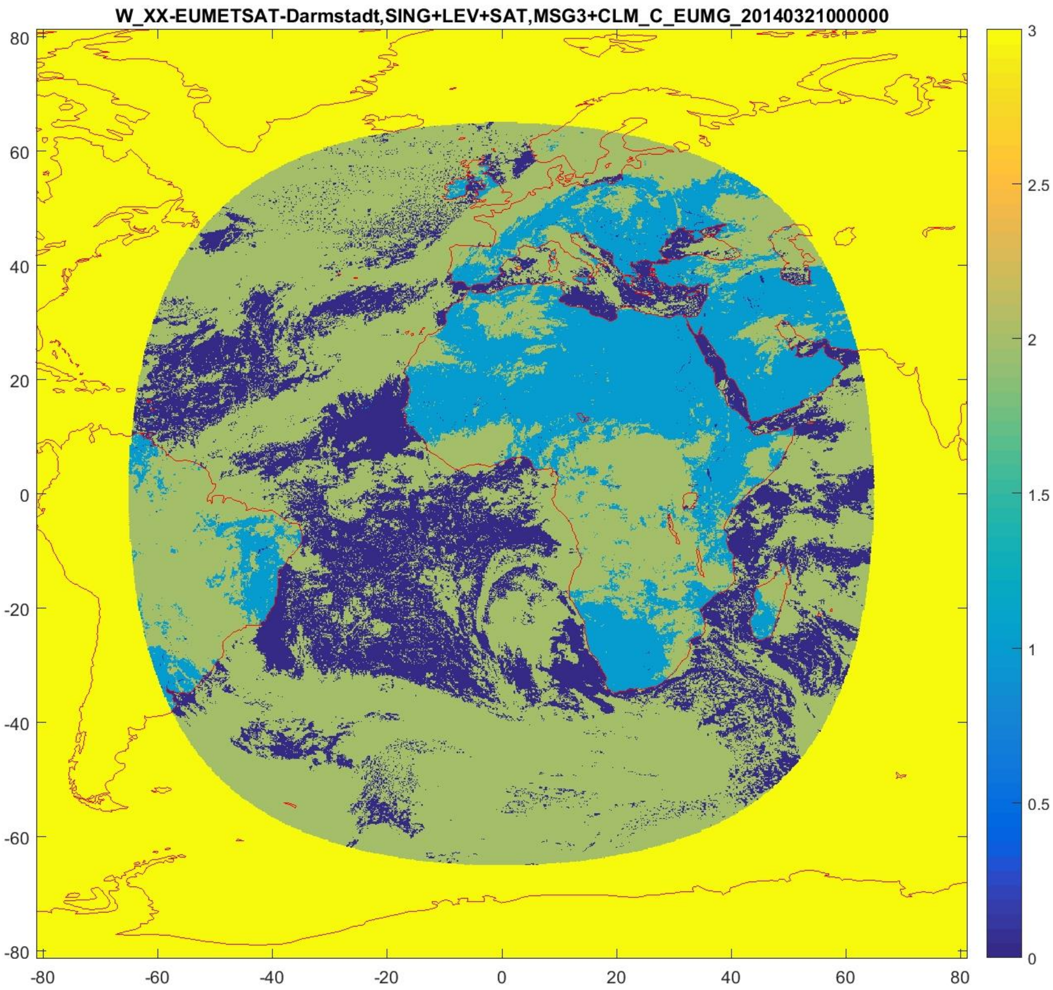

Figure 1 Example of SEVIRI cloud mask.

2.1.2 MODIS cloud mask

MODIS sensor flies onboard EOS Terra and Aqua satellites on polar orbits. We considered all the MODIS images for the four days of the Proba-V exercise (total of 2276 files) at 1Km spatial resolution (products MOD35 and MYD35). The grid of latitude and longitude is not uniform and it is given at a coarser grid than the cloud mask. For this reason the latitude and longitude grids are first interpolated to the full resolution of the cloud mask. Since the grid of the MODIS files depends on the granule and therefore is not fixed with time or space, for practical computational reasons we could not co-locate the Proba-V grid into the MODIS one as made with SEVIRI. Therefore we adopted a simpler reverse solution where the MODIS grid is co-located into the Proba-V uniform grid. This is possible because the MODIS grid is coarser than the PROBA-V one. Actually despite of this each MODIS grid point is mapped into only one Proba-V point; as a consequence a high number of Proba-V pixels have no associated MODIS cloud mask. This is not a problem considering the very high number of available Proba-V scenes. Since MODIS orbits are polar, it is not guaranteed that there is a good match in time with a Proba-V scene; we considered as simultaneous only scenes that were far at most 30 minutes between each other.

MODIS considers the following conditions of cloudiness: cloudy, uncertain clear, probably clear, confident clear. We consider as valid pixels only cloudy and confident clear.

Figure 2 shows an example of SEVIRI cloud mask.

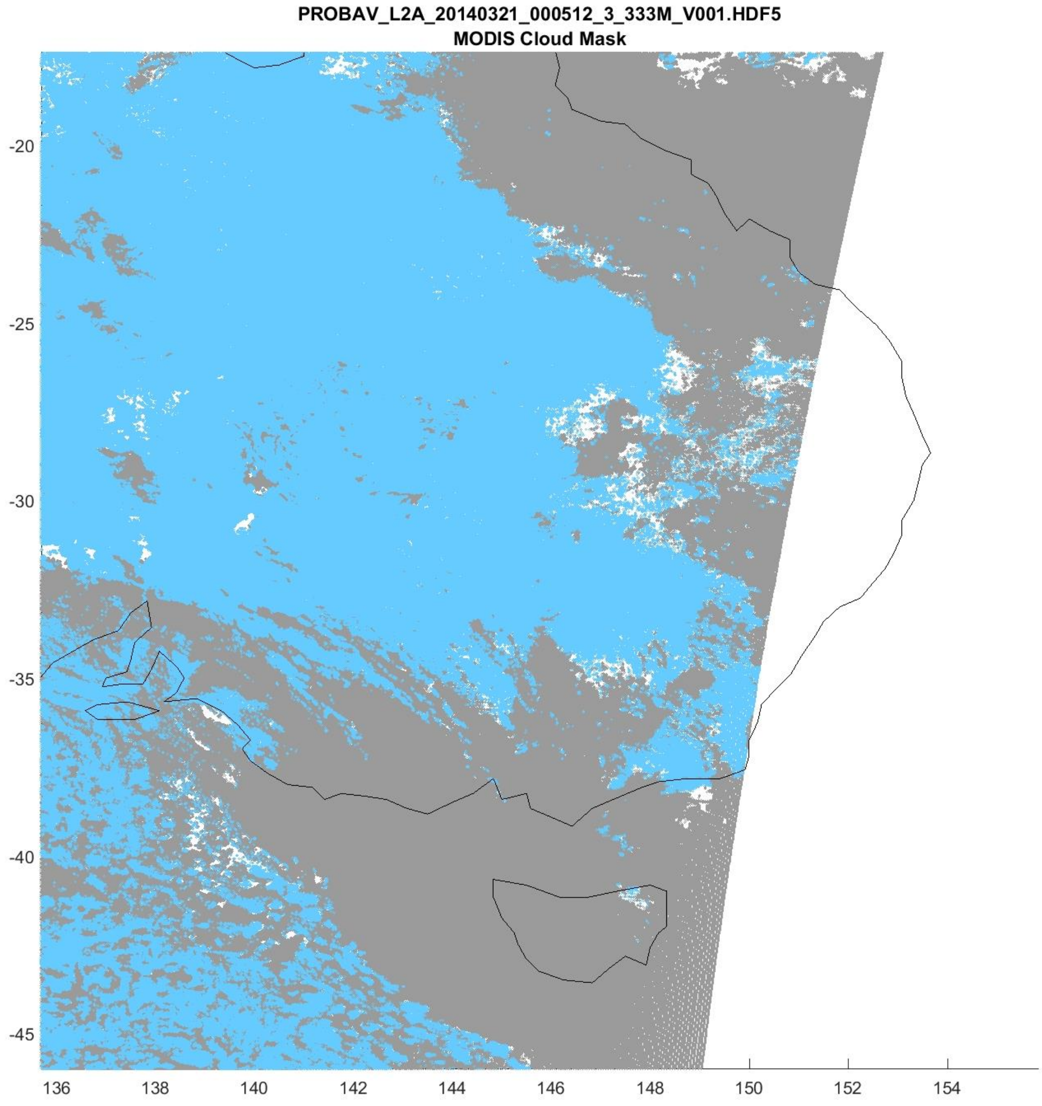

Figure 2 Example of MODIS cloud mask co-located to Proba-V (light blue: clear; gray: cloudy; white: undefined)

2.2 Surface typology

To well characterize the surface is important for an accurate cloud mask detection because clouds have to be discriminated against the underlying surface. An important step in cloud detection is to sort pixels into regions that are homogeneous in their spectral behaviour as far as typology of surface is considered.

Proba-V scenes are endowed with a sea/land mask and an algorithm for snow/ice detection. In order to have a more detailed classification of the surface we considered one ancillary surface classification.

2.2.1 GlobCover is a 2005 ESA initiative jointly with JRC, EEA, FAO, UNEP, GOFC-GOLD and IGBP. The aim of the project was to provide land cover maps from the 300m MERIS sensor onboard the ENVISAT satellite mission. We considered the map that covers the period January - December 2009. The GlobCover map is provided as an image 55,800x129,600 pixels in an equispaced grid with range $[-65^o, 80^o]$ for latitude and $[-180^o, 180^o]$ for longitude. The map classifies surface into 22 different classes having proper codes. For the purpose of the present Proba-V esercise we considered the surface typologies described in

2.2.2 Table *1*. A representative image of the surface mask obtained from GlobCover is shown in Figure 3.

Surface typology	GlobCover code	Occurrence in the dataset
Water	210	40.2%
Vegetation	≤ 180	50.1%
Bare Land	200	8.1%
Urban	190	0.2%
Snow/Ice	220	1.4%

Table 1: Surface typology estimated by GlobCover and percentage of occurrence in the full Proba-V dataset.

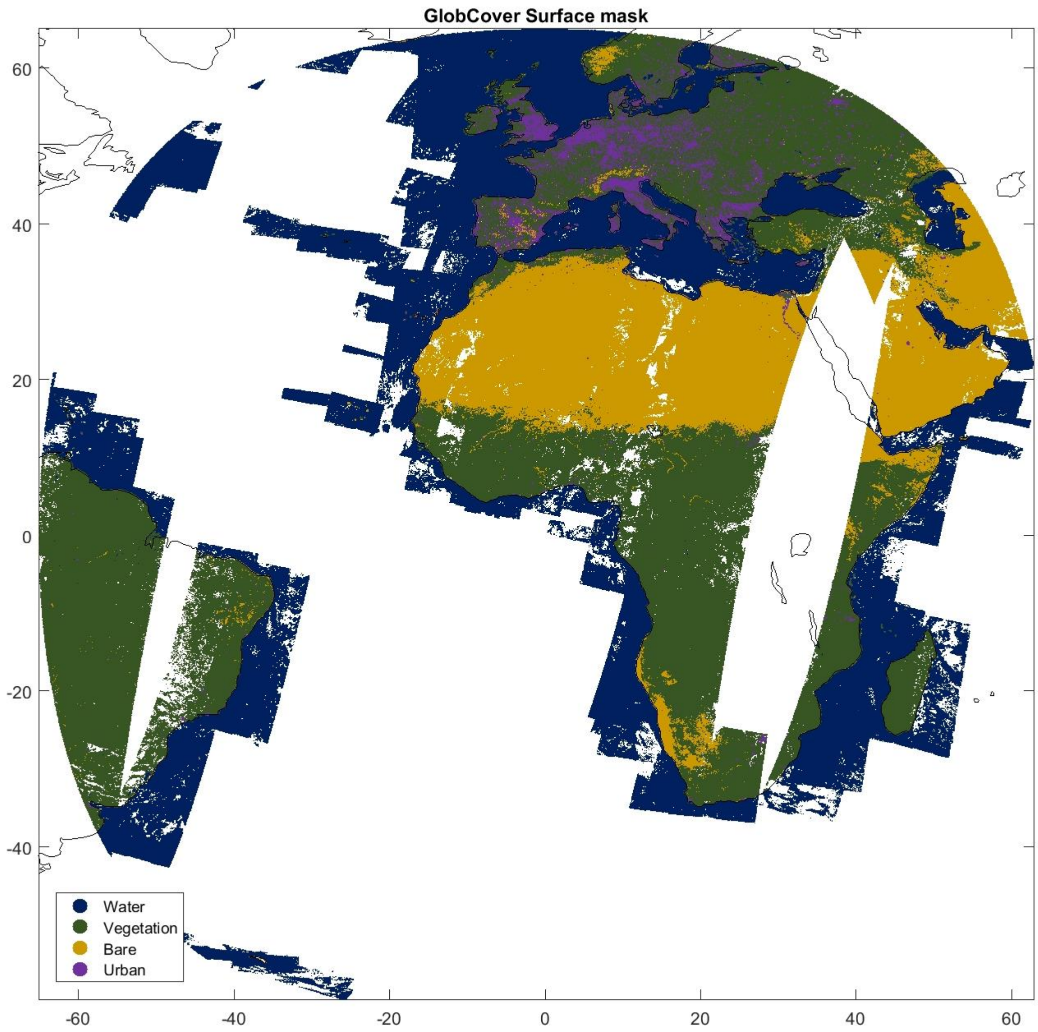

Figure 3: Surface mask obtained from GlobCover

2.3 Statistical methodology

The main methodology adopted for discriminating clouds is Cumulative Discriminant Analysis (CDA), introduced in detail in Amato et al. (2014) and here summarized.

Let us first consider a univariate model (e.g., one spectral band) with variable x. Let $f_{Clear}(x)$ and $f_{Cloudy}(x)$ be its density functions in clear and cloudy conditions, respectively, and $F_{Clear}(x)$ and $F_{Cloudy}(x)$ the cumulative functions that we estimate by the empirical counterparts $\hat{F}_{Clear}(x)$ and $\hat{F}_{Cloudy}(x)$. The Discriminant Analysis will produce a decision rule $\Gamma(x;\mathbf{X})$, with $\mathbf{X}$ being the training data set depending on a threshold θ, given by

$$\Gamma(x;\mathbf{X}) = \begin{cases} 1\ (Clear) if\ x \le \theta \\ 2\ (Cloudy) if\ x > \theta \end{cases} \text{ or } \Gamma(x;\mathbf{X}) = \begin{cases} 1\ (Clear) if\ x > \theta \\ 2\ (Cloudy) if\ x \le \theta \end{cases}$$

where the threshold and the direction of the threshold (i.e., left or right rule) are chosen basing on the training data set $\mathbf{X}$.

The estimate of the threshold and of the direction are obtained by minimizing the Cost function $C(\mathbf{X};\theta)$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, C(\mathbf{X};\theta),$$

with

$$C(\mathbf{X};\theta) = \min\left(\max\left(1-\hat{F}_{Clear}(\theta),\hat{F}_{Cloudy}(\theta)\right),\max\left(\hat{F}_{Clear}(\theta),1-\hat{F}_{Cloudy}(\theta)\right)\right)$$

depending on the threshold θ and on the direction (arguments of the first or second maximum).

In practice the arguments of the maximum are the Type I and Type II errors of the classification in both directions of the threshold (see Fig. 1 for the density functions and Fig. 2 for the cumulative ones for one direction of the threshold).

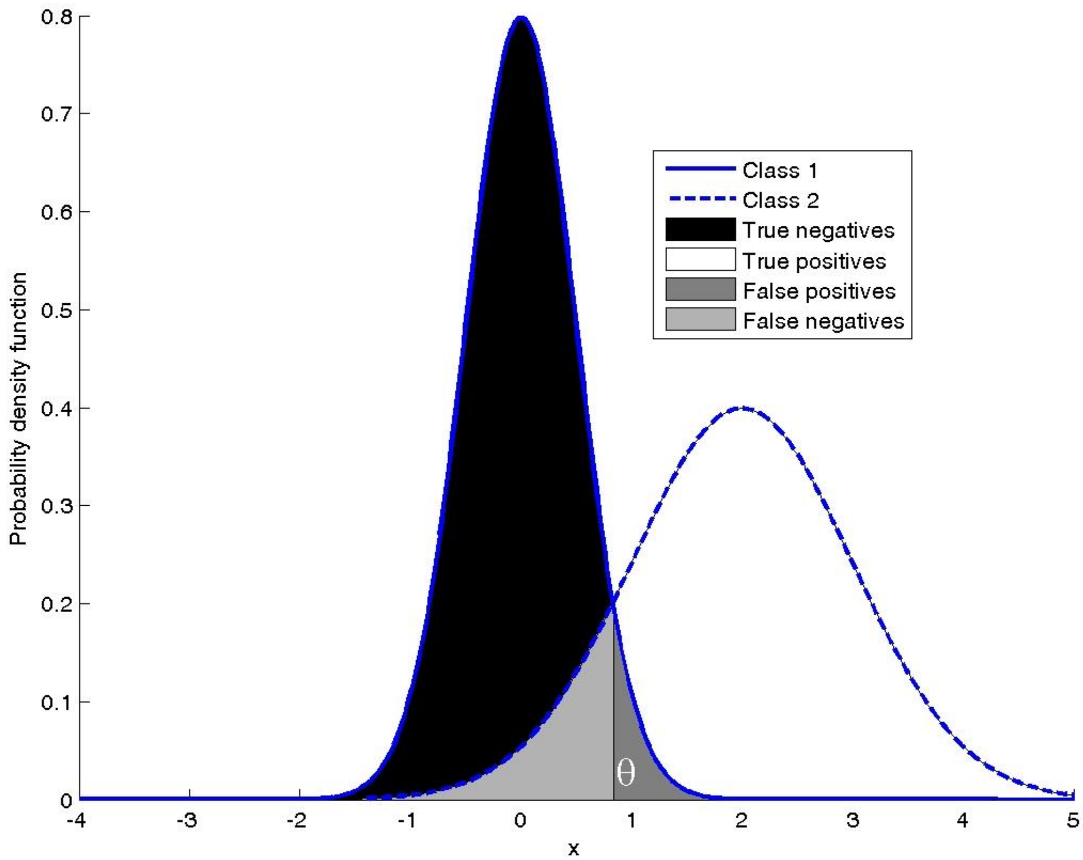

Figure 4: Density functions of the two classes between which to discriminate.

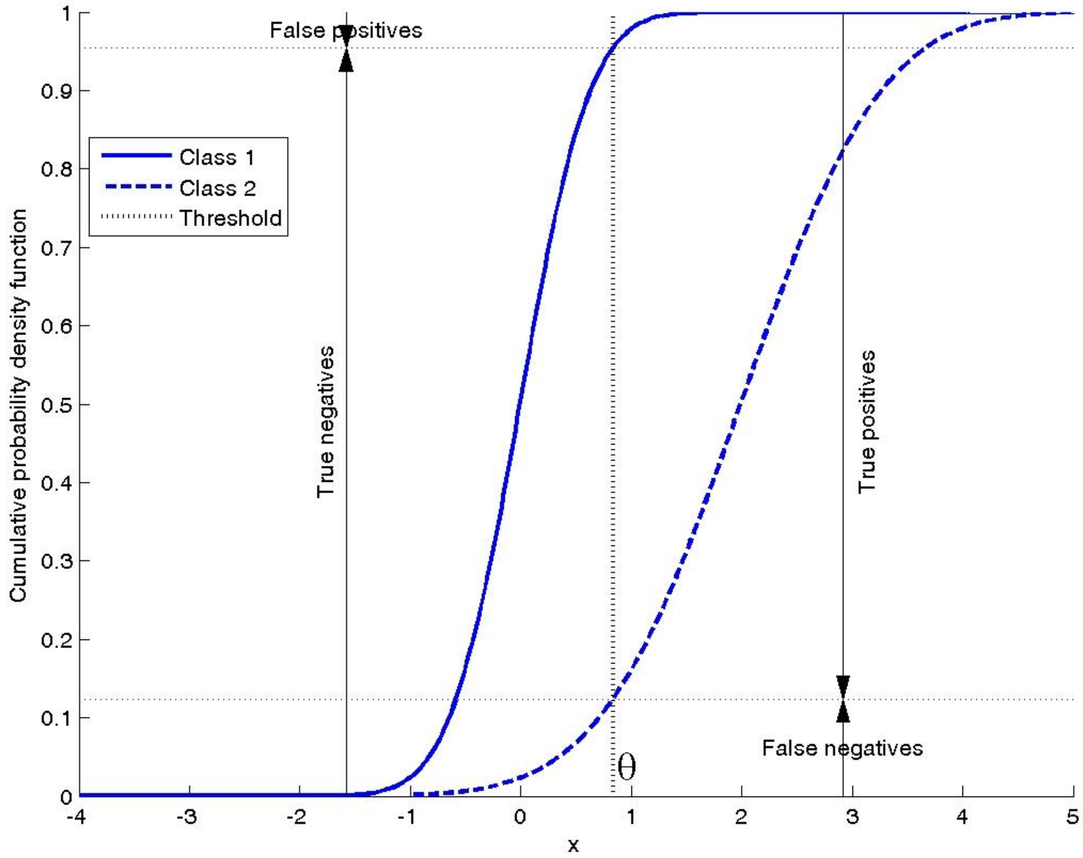

Figure 5: Cumulative distribution functions of the two classes between which to discriminate.

At the threshold $\hat{\theta}$ that maximizes the Cost function both Type I and Type II errors coincide.

The choice of this Cost function is least noncommittal with respect to the average conditions of the sky, cloudy or clear. The rationale behind is that we want to simultaneously minimize both Type I and II errors with an equal balancing between the twos.

We point out that the quantity

$$R(\mathbf{X}; x) = \min\left(1 - \hat{F}_{Clear}(x), \hat{F}_{Cloudy}(x), \hat{F}_{Clear}(x), 1 - \hat{F}_{Cloudy}(x)\right)$$

for a generic radiance x can be considered as an indicator of the reliability of the Clear/Cloudy label assigned to a pixel.

The multivariate CDA is straightforwardly obtained from the univariate one by assuming independence of the joint density function (and consequently of the corresponding cumulative one). Therefore the multivariate cumulative function is the product of the univariate ones.

The algorithm for estimating thresholds and direction goes through the following steps:

- Compute unidimensional cumulative distribution functions for each spectral band $\hat{F}^{d}_{Clear}(x_d)$, $\hat{F}^{d}_{Cloudy}(x_d)$.
- Compute directional cumulative distributions functions according to the selected direction of threshold inequalities

$$\tilde{F}^{d}_{Clear}(x_d) = \begin{cases} \hat{F}^{d}_{Clear}(x_d), if\ x_d \le \theta_d \Rightarrow Clear \\ 1 - \hat{F}^{d}_{Clear}(x_d), if\ x_d \ge \theta_d \Rightarrow Clear \end{cases}$$

$$\tilde{F}^{d}_{Cloudy}(x_d) = \begin{cases} \hat{F}^{d}_{Cloudy}(x_d), if\ x_d \le \theta_d \Rightarrow Clear \\ 1 - \hat{F}^{d}_{Cloudy}(x_d), if\ x_d \ge \theta_d \Rightarrow Clear \end{cases}$$

- Compute Type I and Type II Errors as

$$E^{I}(x_1, \dots, x_D) = 1 - \prod_{d=1}^{D} \tilde{F}^{d}_{Clear}(x_d)$$

$$E^{II}(x_1, \dots, x_D) = \prod_{d=1}^{D} \tilde{F}^{d}_{Cloudy}(x_d)$$

- Compute the cost function as

$$C(\theta_1, \dots, \theta_D) = \max\left(E^{I}(\theta_1, \dots, \theta_D), E^{II}(\theta_1, \dots, \theta_D)\right)$$

- Minimize the Cost function yielding thresholds and their directions

$$\left(\hat{\theta}_1, \dots, \hat{\theta}_D\right) = \min_{\theta_1,\dots,\theta_D} C(\theta_1, \dots, \theta_D)$$

3. Results

The total number of scenes available in the 331 files of the full data set is 50,532,546,354, of which 7,731,538,861 (15.2%) are good scenes, the remaining ones being off sensor view, sun glint or missing radiance. To all scenes it is possible to assign a surface typology according to GlobCover. Coverage for the full Proba-V dataset is shown in

Table *1*.

On the contrary not all 7.7 Giga pixels could be co-located with the SEVIRI and MODIS cloud masks.

Table *2* shows the useful scenes with both cloud masks, together with their status (clear/cloudy).

	SEVIRI Cloud mask	MODIS Cloud mask
Clear	46,432,741 (50.6%)	120,640,955 (40.1%)
Cloudy	45,311,833 (49.4%)	187,331,229 (59.9%)
Total	91,744,574	300,972,184

Table 2: Number of valid Proba-V scenes co-located with the SEVIRI and MODIS cloud mask.

In practice there are over 300M scenes for the MODIS cloud mask and over 90M for the SEVIRI one. The exercise was worked in a classic training-test paradigm according to the following steps:

1) Once the training dataset has been chosen, parameters of the chosen classification method (CDA) are computed and stored.
2) Performance of the methodology is estimated in terms of some indicator.

As far as the training phase is concerned, we considered three different scenarios according to the chosen *silver* cloud mask:

a) SEVIRI cloud mask
b) MODIS cloud mask
c) Joint SEVIRI&MODIS cloud mask. This is obtained from the SEVIRI and MODIS cloud masks by retaining only pixels for which there was an agreement between the two independent cloud masks.

In the test phase in principle a different test dataset from the training one should be used. In practice this can be easily obtained by splitting the original dataset into two parts with standard proportions (typically 2/3 for the training set and 1/3 for the test one). The large size of the sample allows to apply a similar procedure. However for the same reason the full statistical properties (i.e., cumulative distribution) of reflectance in both clear and cloudy conditions are well estimated with such large numbers, therefore there is no significant change of performance when using a different dataset for the test phase or the same dataset. For the same reason, to speed up calculations and to solve some troubles due to memory usage we limited the size of the training set (and therefore of the test one) to 3 Million.

Actually the Proba-V exercise is provided with a limited annotated dataset of 1350 scenes for which the status of the sky (clear or cloudy) is given according to visual inspection (*gold* standard). This dataset is a subset of the final (unknown) evaluation one. Therefore the classification procedure can be endowed with the following phase

3) Performance of the methodology is estimated on the sample *gold* dataset.

We stress that parameters of the methodology are obtained at the end of Phase 1), therefore the *gold* standard data set is no way involved in their estimation.

Results of the analysis are shown in the following Table 3–Table 5 for the SEVIRI, MODIS and SEVIRI&MODIS cloud masks, respectively.

Tables can be read in this way. Results are separately shown for the five surface typologies (Water, Vegetation, Bare land, Urban, Snow/Ice). For each one the following quantities are shown:

- Size of the training dataset
- Size of the test dataset (according to discussion above they coincide)
- For each of the 15 possible k-combinations of reflectance the success percentage in all sky conditions and separately for Clear and Cloudy conditions. They are simply obtained as the fraction of training data set scenes correctly classified by the methodology.
- Then we considered a preliminar transform of reflectances into Principal Components and 4 cases corresponding to the number of retained Principal Components. For each one the same performance parameters as Reflectances are given
- Finally in the last row we give the best case (among the combinations of radiance and Principal Components) according to the same Minimum Cost criterion applied for CDA and the corresponding performance indexes.

For a better visual inspection in Table 3–Table 5 we colored by light green, orange and light red the cases yielding the best performance under Clear, Cloudy and global conditions, respectively.

We observe that by the very nature of CDA (in particular of the Cost criterion used to estimate parameters) and by the choice of using the same datasets for the training and test Phases the performance in Clear and Cloudy conditions (and therefore the global one) practically coincide.

Table 6 shows the results of classification on the *gold* reduced test dataset sorted by surface typology when SEVIRI and MODIS cloud masks are considered for the training dataset. We observe that globally the success percentages are similar (80.1% and 79.0% for SEVIRI and MODIS cloud masks, respectively). However their values for Clear and Cloudy conditions are very different, with MODIS being much more balanced between the two conditions. As before Bare land typology gives the worst results.

Finally we mention that for 43 scenes out of 1350 *gold* ones a MODIS cloud mask is available, and that the twos match 39 times (83.3% success); for the SEVIRI cloud mask the number of co-locations is much smaller (only 6, 83.3% success percentage). These digits can be considered as an estimate of the top performance reachable by methods based on the SEVIRI and MODIS cloud masks. Full results are shown in Table 6.

CDA SEVIRI	Water			Vegetation			Bare land			Urban			Snow/Ice		
	Global	**Clear**	**Cloudy**	**Global**	**Clear**	**Cloudy**	**Global**	**Clear**	**Cloudy**	**Global**	**Clear**	**Cloudy**	**Global**	**Clear**	**Cloudy**
Training size	3.000.000	1.012.707	1.987.293	3.000.000	1.149.659	1.850.341	3.000.000	2.624.951	375.049	300.645	83.887	216.758	60.291	24.038	36.253
Test size	3.000.000	1.012.707	1.987.293	3.000.000	1.149.659	1.850.341	3.000.000	2.624.951	375.049	300.645	83.887	216.758	60.291	24.038	36.253
BLUE	80,0	79,7	80,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,4	81,7	79,0	78,2	79,6
RED	80,8	80,4	81,0	75,8	75,7	75,8	59,3	59,2	59,8	78,2	78,0	78,3	62,7	62,1	63,1
NIR	81,1	81,0	81,2	76,8	76,6	76,9	55,9	55,9	56,3	77,5	77,4	77,6	80,1	80,0	80,2
SWIR	80,9	80,7	81,0	62,1	61,9	62,2	62,4	62,5	62,1	67,2	67,1	67,3	62,2	61,5	62,7
BLUE-RED	80,7	80,7	80,8	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	79,0	78,8	79,2
BLUE-NIR	81,1	81,0	81,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	80,1	80,0	80,2
BLUE-SWIR	80,9	80,7	81,0	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	79,0	78,8	79,2
RED-NIR	81,1	81,0	81,2	76,8	76,8	76,8	59,5	59,5	59,6	78,2	78,1	78,2	80,1	80,0	80,2
RED-SWIR	80,9	80,7	81,0	75,8	75,7	75,8	62,3	62,3	62,3	78,2	78,1	78,2	62,7	62,1	63,1
NIR-SWIR	81,1	81,0	81,2	76,8	76,8	76,8	62,3	62,3	62,3	77,5	77,4	77,6	80,1	80,0	80,2
BLUE-RED-NIR	81,1	81,0	81,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	80,1	80,0	80,2
BLUE-RED-SWIR	80,9	80,7	81,0	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	79,0	78,8	79,2
BLUE-NIR-SWIR	81,1	81,0	81,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	80,1	80,0	80,2
RED-NIR-SWIR	81,1	81,0	81,2	76,8	76,8	76,8	62,3	62,3	62,3	78,2	78,1	78,2	80,1	80,0	80,2
BLUE-RED-NIR-SWIR	81,1	81,0	81,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	80,1	80,0	80,2
P.C. 1	80,9	80,9	80,9	76,9	76,9	76,9	58,7	58,7	58,7	78,9	78,9	78,9	77,6	77,6	77,6
P.C. 1-2	80,9	80,9	80,9	76,9	76,9	76,9	70,7	71,4	65,8	79,7	81,9	78,8	77,6	77,6	77,6
P.C. 1-2-3	80,9	80,9	80,9	76,9	76,9	76,9	70,7	71,2	67,1	78,9	78,9	78,9	77,6	77,6	77,6
P.C. 1-2-3-4	80,9	80,9	80,9	76,9	76,9	76,9	74,1	75,4	64,4	78,9	78,9	78,9	77,6	77,6	77,6
MAX-MIN	**81,0**	**NIR**		**81,3**	**BLUE**		**67,1**	**P.C. 1-2-3**		**81,6**	**BLUE-RED**		**80,0**	**NIR**	

Table 3: Results of the exercise on the SEVIRI training dataset for the SEVIRI cloud mask.

CDA MODIS	Water			Vegetation			Bare land			Urban			Snow/Ice		
	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy
Training size	3.000.000	531.745	2.468.255	3.000.000	1.107.612	1.892.388	3.000.000	2.232.046	767.954	452.712	136.498	316.214	3.000.000	1.603.830	1.396.170
Test size	3.000.000	531.745	2.468.255	3.000.000	1.107.612	1.892.388	3.000.000	2.232.046	767.954	452.712	136.498	316.214	3.000.000	1.603.830	1.396.170
BLUE	67,5	67,4	67,5	78,3	78,3	78,4	83,8	83,8	83,9	82,3	82,1	82,3	56,9	63,3	49,5
RED	67,7	67,7	67,7	77,2	77,1	77,2	66,6	66,6	66,7	80,2	80,2	80,3	59,8	60,0	59,6
NIR	68,3	68,2	68,3	76,2	76,1	76,3	62,5	62,5	62,8	79,1	79,0	79,2	62,1	62,3	61,9
SWIR	80,2	80,0	80,2	65,4	65,2	65,5	67,3	67,4	67,1	68,7	68,5	68,8	79,4	79,3	79,6
BLUE-RED	67,7	67,7	67,7	78,3	78,3	78,4	83,8	83,8	83,9	82,2	82,2	82,2	59,8	59,8	59,7
BLUE-NIR	68,2	68,2	68,3	78,3	78,3	78,4	83,8	83,8	83,9	82,2	82,2	82,2	62,1	62,1	62,1
BLUE-SWIR	80,1	80,1	80,1	78,3	78,3	78,4	83,8	83,8	83,9	82,2	82,2	82,2	79,5	79,5	79,5
RED-NIR	68,2	68,2	68,3	77,1	77,1	77,1	66,6	66,6	66,7	80,2	80,2	80,3	62,1	62,3	62,0
RED-SWIR	80,1	80,1	80,1	77,1	77,1	77,1	67,2	67,3	67,2	80,2	80,2	80,3	79,5	79,5	79,5
NIR-SWIR	80,1	80,1	80,1	76,2	76,2	76,2	67,2	67,3	67,2	79,1	79,1	79,1	79,5	79,5	79,5
BLUE-RED-NIR	68,2	68,2	68,3	78,3	78,3	78,4	83,8	83,8	83,9	82,2	82,2	82,2	62,1	62,1	62,1
BLUE-RED-SWIR	80,1	80,1	80,1	78,3	78,3	78,4	83,8	83,8	83,9	82,2	82,2	82,2	79,5	79,5	79,5
BLUE-NIR-SWIR	80,1	80,1	80,1	78,3	78,3	78,4	83,8	83,8	83,9	82,2	82,2	82,2	79,5	79,5	79,5
RED-NIR-SWIR	80,1	80,1	80,1	77,1	77,1	77,1	67,2	67,3	67,2	80,2	80,2	80,3	79,5	79,5	79,5
BLUE-RED-NIR-SWIR	80,1	80,1	80,1	78,3	78,3	78,4	83,8	83,8	83,9	82,2	82,2	82,2	79,5	79,6	79,4
P.C. 1	67,5	67,5	67,5	78,3	78,3	78,3	83,6	83,6	83,6	82,2	82,2	82,2	58,5	58,5	58,5
P.C. 1-2	67,5	67,5	67,5	78,3	78,3	78,3	83,6	83,6	83,6	82,2	82,2	82,2	58,5	58,5	58,5
P.C. 1-2-3	67,5	67,5	67,5	78,3	78,3	78,3	83,6	83,6	83,6	82,2	82,2	82,2	58,5	58,5	58,5
P.C. 1-2-3-4	67,5	67,5	67,5	78,3	78,3	78,3	83,6	83,6	83,6	82,2	82,2	82,2	58,5	58,5	58,5
MAX-MIN	**80,1**	**BLUE-SWIR**		**78,3**	**P.C. 1**		**83,8**	**BLUE**		**82,2**	**P.C. 1-2-3**		**79,5**	**BLUE-SWIR**	

Table 4: Results of the exercise on the training dataset for the MODIS cloud mask.

CDA SEVIRI&MODIS	Water			Vegetation			Bare land			Urban			Snow/Ice		
	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy
Training size	3.000.000	569.362	2.430.638	3.000.000	1.104.020	1.895.980	3.000.000	2.769.386	230.614	235.546	55.957	179.589	41.570	16.344	25.226
Test size	3.000.000	569.362	2.430.638	3.000.000	1.104.020	1.895.980	3.000.000	2.769.386	230.614	235.546	55.957	179.589	41.570	16.344	25.226
BLUE	86,3	86,0	86,4	88,8	88,6	88,9	82,4	82,3	82,9	87,7	87,5	87,7	86,1	85,6	86,5
RED	87,7	87,3	87,7	81,9	81,8	82,0	66,5	66,5	67,0	84,0	84,0	84,1	70,9	70,7	71,1
NIR	88,7	88,5	88,8	82,5	82,4	82,5	62,0	62,0	62,2	83,4	83,3	83,4	86,5	86,4	86,5
SWIR	89,1	88,8	89,2	66,3	66,1	66,4	70,0	70,0	69,7	72,1	71,9	72,1	66,3	66,0	66,4
BLUE-RED	87,6	87,5	87,6	88,8	88,8	88,8	82,7	82,6	82,8	87,7	87,5	87,7	86,0	85,8	86,2
BLUE-NIR	88,7	88,5	88,8	88,8	88,8	88,8	82,7	82,6	82,8	87,7	87,5	87,7	86,5	86,4	86,5
BLUE-SWIR	89,0	89,0	89,0	88,8	88,8	88,8	82,7	82,6	82,8	87,7	87,5	87,7	86,0	85,8	86,2
RED-NIR	88,7	88,5	88,8	82,5	82,4	82,5	66,8	66,8	66,8	84,0	84,0	84,1	86,5	86,4	86,5
RED-SWIR	89,0	89,0	89,0	81,9	81,9	81,9	69,9	69,9	69,9	84,0	84,0	84,1	70,9	70,7	71,1
NIR-SWIR	89,0	89,0	89,0	82,5	82,4	82,5	69,9	69,9	69,9	83,4	83,3	83,4	86,5	86,4	86,5
BLUE-RED-NIR	88,7	88,5	88,8	88,8	88,8	88,8	82,7	82,6	82,8	87,7	87,5	87,7	86,5	86,4	86,5
BLUE-RED-SWIR	89,0	89,0	89,0	88,8	88,8	88,8	82,7	82,6	82,8	87,7	87,5	87,7	86,0	85,8	86,2
BLUE-NIR-SWIR	89,0	89,0	89,0	88,8	88,8	88,8	82,7	82,6	82,8	87,7	87,5	87,7	86,5	86,4	86,5
RED-NIR-SWIR	89,0	89,0	89,0	82,5	82,4	82,5	69,9	69,9	69,9	84,0	84,0	84,1	86,5	86,4	86,5
BLUE-RED-NIR-SWIR	89,0	89,0	89,0	88,8	88,8	88,8	82,7	82,6	82,8	87,7	87,5	87,7	86,5	86,4	86,5
P.C. 1	87,6	87,6	87,6	83,6	83,6	83,6	67,9	67,9	67,9	85,0	85,0	85,0	83,8	83,8	83,8
P.C. 1-2	87,6	87,6	87,6	83,6	83,6	83,6	84,4	84,8	79,6	85,5	86,1	85,2	83,8	83,8	83,8
P.C. 1-2-3	87,6	87,6	87,6	83,6	83,6	83,6	78,8	78,8	78,8	85,3	86,1	85,1	83,8	83,8	83,8
P.C. 1-2-3-4	87,6	87,6	87,6	83,6	83,6	83,6	78,8	78,8	78,8	85,0	85,0	85,0	83,8	83,8	83,8
MAX-MIN	**89,0**	**BLUE-SWIR**		**88,8**	**BLUE-RED**		**82,6**	**BLUE-RED**		**87,5**	**BLUE**		**86,4**	**NIR**	

Table 5: Results of the exercise on the training dataset for the joint SEVIRI&MODIS cloud mask.

Surface typology	SEVIRI			MODIS			SEVIRI&MODIS		
	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy
Water (326 scenes)	71.8	61.0	88.9	81.6	82.5	80.2	78.2	73.0	86.5
Vegetation (921 scenes)	83.3	62.8	91.4	78.3	72.4	80.6	82.7	63.6	90.3
Bare land (83 scenes)	73.5	78.6	68.3	73.5	81.0	65.9	73.5	81.0	65.9
Urban (9 scenes)	100	100	-	100	100	-	100	100	-
Ice/snow (10 scenes	100	-	100	90	-	90	100	-	100
Global (1350 scenes)	80.1	64.1	89.9	79.0	77.5	79.8	81.3	69.3	88.5

Table 6: Success percentages on the reduced test *gold* dataset based on the SEVIRI and MODIS training cloud mask.

4. Discussion

The results presented in the previous Section are only a subset of extensive experiments worked on the Proba-V dataset. In particular we considered further surface classifications (PROBA, MODIS, GLCC) and a bunch of classification methods (Linear Discriminant Analysis, Quadratic Discriminant Analysis, Principal Component Discriminant Analysis, Support Vector Machine, K-NN classification). We also considered a classification methodology based on regression, with the twofold aim of classification itself and use of more ancillary data into the problem (zenith angle, introduction of nonlinear dependence on reflectance and/or Principal Components) in an ANOVA framework. They are not reported here because we considered them not reliable enough to be used on an operative basis or because of no improvement of the results according to our measure of performance. Full details will be given in forthcoming papers.

Analysis of Table 3–Table 5 shows that performance of SEVIRI and MODIS cloud masks on the training data set is around 80% but somewhat depending on the surface type. Performance of the joint SEVIRI&MODIS cloud mask is significantly higher (up to 89% over Water). This has to be expected because the quality of the training dataset is higher (due to the agreement between two independent cloud masks, SEVIRI and MODIS). However this cannot be taken as a conclusive argument because reasonably the pixels for which the two cloud masks agree are the “easiest” to classify, therefore performance index is biased; moreover the training datasets do not adequately represent the conditions where classification is more uncertain (mainly the ones with mixed conditions or at the boundary between clear and cloudy conditions), therefore the estimated parameters are not tailored on those conditions.

There is no strong evidence of the role of multispectrality. This could have been expected considering the low number of spectral channels. First of all we observe that a single channel, yet depending on the surface type, is able to yield performance equal to that using the full set of channels. In particular the best channels are SWIR for Water surface, BLUE for Vegetation, Bare land and Urban surfaces, NIR for Snow/Ice. In practice multispectrality plays a significant role with respect to different surface typologies but only in a less extent within a typology. Related to these arguments, there is no advantage in computing Principal Components of reflectances.

For ease of presentation, the final results provided for the Proba-V exercise refer to the case of full set of spectral bands (row BLUE-RED-NIR-SWIR in Table 3–Table 5). Moreover three different solutions were provided according to the cloud mask used for the training phase: CNR-UNIBAS1, CNR-UNIBAS2 and CNR-UNIBAS3 according to SEVIRI&MODIS, SEVIRI and MODIS cloud masks, respectively.

Finally we want to point some issues resulting from the Exercise that we believe interesting for a general discussion:

- *Comparison of operational land masks*. It is generally acknowledged how mandatory is to tailor estimate of parameters of a classification methodology to the underlying surface type of a pixel due to the importance of the contrast of the surface with clouds. We found a significant mismatch among the different surface typologies we considered (GlobCover, GLCC, Proba-V). In the present exercise we relied on the GlobCover one because of its native spatial resolution similar to Proba-V (300mt) and because more recent.
- *Optimal surface mask*. To choose specific surface typologies for cloud detection is a trade-off between homogeneity of the classes and its number. Therefore it is interesting to find an optimal sorting of the different typologies specific to the cloud detection, possibly also including production of surface masks depending on season.
- *Operational cloud masks (silver standard)*. Table 3–Table 5 show that the MODIS and SEVIRI operational cloud masks have significant mismatch in the classification (see for example the percentage of Clear and Cloudy pixels in each of them). It is interesting to assess such mismatches with respect to, e.g., spatial and time resolution, surface typology. We observe for example that MODIS spatial resolution (1Km) is better than SEVIRI one (3Km ssp), but the reverse occurs for time resolution.

- *Cost function*. The choice of the Cost function strongly affects the retrieval of the cloud mask. Its main effect is to foster retrieval of one of the two sky conditions (Clear or Cloudy). This can have motivations driven from applications, e.g., to be more conservative with respect to the cloudy conditions. Practically this can mean to use other performance indicators than Type I and Type II error, also acting on the priori probabilities involved in the Discriminant Analysis, maybe depending on the global scene under analysis.
- *Solar zenith angle*. Considering that solar zenith angle significantly affects radiance, the best way to incorporate this dependence into the classification methodology has to be investigated.
- *Nonlinear models*. Most classification methodologies for cloud detection are linear with respect to radiances/reflectances. Nonlinearity can be introduced in several ways. In the present Proba-V exercise we augmented the variables with new variables nonlinearly depending on reflectance. Results (not shown in the present report for the sake of brevity) indicate that even though the considered nonlinear effects had a significant impact on the reflectances, however they did not result in a straightforward improvement of the performance in detecting clouds.
- *Full Proba-V test dataset*. Finally all analyses performed in the exercise should be assessed with the full 20K size Proba-V test dataset that is a unique example of extensive *gold* standard dataset. It could even be used for training the methodologies for cloud detection. For this reason it would be very important to put this database at disposal at least of the participants in the project.

5. References

U. Amato, L. Lavanant, G. Liuzzi, G. Masiello, C. Serio, R. Stuhlmann, and S. A. Tjemkes: Cloud mask via cumulative discriminant analysis applied to satellite infrared observations: scientific basis and initial evaluation. *Atmos. Meas. Tech.* **7**, 3355–3372 (2014)

T.R. Loveland, B.C. Reed, J.F. Brown, D.O. Ohlen, J. Zhu, L. Yang, J.W. Merchant: Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data: *International Journal of Remote Sensing* **21**(6/7), 1303–1,330 (2000)