Proba-V Cloud Detection Round Robin: Validation Results and Recommendations

R.Q. Iannone[1], F. Niro[1]
[1]Serco S.p.A.,
Via Sciadonna 24, 00044 Frascati, Italy
rosarioquirino.iannone@serco.com; fabrizio.niro@esa.int

P. Goryl[2], S. Dransfeld[2], B. Hoersch[2]
[2]ESA/ESRIN,
Via Galileo Galilei, 00044 Frascati, Italy

K. Stelzer[3], G. Kirches[3], M. Paperin[3], C. Brockmann[3]
[3]Brockmann Consult GmbH,
Max-Planck-Str. 2, 21502 Geesthacht, Germany

L. Gómez-Chova[4], G. Mateo-García[4]
[4]University of Valencia,
Catedrático José Beltrán 2, E-46980 Paterna, Spain

R. Preusker[5] , J. Fischer[5]
[5]Freie Universität Berlin
Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany

U. Amato[6]
[6]CNR,
Via Pietro Castellino 111, 80131 Napoli, Italy

C. Serio[7]
[7]Universita' degli studi della Basilicata,
Viale dell'Ateneo Lucano, 10 - 85100 Potenza, Italy

U. Gangkofner[8]
[8]eoConsultancy
Frax 516, A-6232 Münster, Austria

B. Berthelot[9]
[9]Magellium
Rue Hermes, 31520 Ramonville-Saint-Agne, France

M. -D. Iordache[10], L. Bertels[10], E. Wolters[10], W. Dierckx[10], I. Benhadj[10], E. Swinnen[10]
[10]Flemish Institute for Technological Research (VITO)
Boeretang 200, BE-2400 MOLBelgium

***Abstract*—This paper discusses results from 12 months of a Round Robin exercise aimed at the inter-comparison of different cloud detection algorithms for Proba-V. Clouds detection is a critical issue for satellite optical remote sensing, since potential errors in cloud masking directly translates into significant uncertainty in the retrieved downstream geophysical products. Cloud detection is particularly challenging for Proba-V due to the presence of a limited number of spectral bands and the lack of thermal infrared bands. The main objective of the project was the inter-comparison of several cloud detection algorithms for Proba-V over a wide range of surface types and environmental conditions. Proba-V Level 2a products have been distributed to six different algorithm providers representing companies and research institutes in several European countries.**

The considered cloud detection approaches are based on different strategies: Neural Network, Discriminant Analysis, Multi-spectral and Multi-textural Thresholding, Self-Organizing Feature Maps, Dynamic Thresholding, and Classification based on Cloud Optical Thickness. The results from all algorithms were analysed and compared against a reference dataset, consisting of a large number (more than fifty thousands) of visually classified pixels. The quality assessment was performed according to a uniform methodology and the results provide clear indication on the potential best-suited approach for next Proba-V cloud detection algorithm.

Keywords*—*Proba-V; Cloud Detection Algorithm; Round Robin.

I. INTRODUCTION

Proba-V provides global daily observations of Earth 'surface in the VNIR-SWIR region of the spectrum [1]. The quality of the Proba-V products is a key element to guarantee the achievement of the objectives of the mission, which are related to the monitoring of the land use and land cover and also to the understanding of the long-term behavior of the vegetation. The production of multi temporally composited cloud-free mosaics of land surfaces is crucial to yield information for the study of terrestrial vegetation structure and dynamics, and land cover mapping.

The cloud detection algorithms are highly dependent on the available spectral bands. The lack of TIR channels or dedicated cirrus band as the 1.38 micron band makes the cloud screening for Proba-V more challenging. The new operational Proba-V algorithm for cloud detection, which is an evolution of the legacy model used for SPOT-VGT [2], relies on the use of climatology of reflectances to define a dynamic threshold on the Blue band, which depends on the surface cover and on vegetation conditions. This method still presents some drawbacks, in particular a dependency on illumination and viewing geometry and some remaining misdetection at the edges of each status class (e.g., land/water).

Recent years have seen an increasing interest in cloud detection from satellite imagery using different approaches. Cloud detection algorithms, indeed, can assume many forms, from single thresholding method to more elaborate multi-temporal, Bayesian and machine learning approaches [3, 4, 5].

With this in mind and in order to inter-compare the performances of different cloud detection algorithms for

Proba-V and to potentially identify the way to improve the current method, the European Space Agency (ESA) and the Belgian Science Policy Office (BELSPO) decided to organize a dedicated Round Robin exercise.
The main objectives of the Round Robin were twofold. The first objective is to provide better understanding of the advantages and drawbacks of the various techniques for various clouds and surface conditions.
The second objective is to collect lessons learnt on cloud detection in the VNIR and SWIR domain for land and coastal water remote sensing and reuse them in the frame of Sentinel-2 and Sentinel-3 missions.

Such an inter-comparison between six fundamentally different cloud detection approaches is of obvious great importance to assess the performance of the methods. More importantly, the implementation in the operational processing chain of a potential best candidates would enable to improve the overall accuracy and stability of geophysical products derived from Proba-V data.

II. ROUND ROBIN DESIGN

The design of the Round Robin was based on the definition of three types of datasets, which were relevant for the Proba-V intercomparison: the Input Reference Scenes, the Validation Dataset and the Test Dataset.

The Input Reference Scenes are the Proba-V Level 2a products, consisting of TOA reflectances in the 4 Proba-V bands, radiometrically and geometrically corrected, projected and resampled to the chosen spatial resolution. The products were provided with one single resolution, which is 333m. The available data set consists of 331 products acquired in four days covering the four seasons: 21/03/2014, 21/06/2014, 21/09/2014, and 21/12/2014.
An important aspect is that no training dataset was provided to the participants to avoid that the algorithms were tuned on this dataset.

For validation dataset, we intended the ensemble of pixels on which the quality assessment of the various algorithms had to be made. The validation dataset consisted on a relatively high number (53000) of carefully chosen pixels, extracted from the Input Reference Scenes.
This dataset was selected in order to be statistically representative of the different pixel classes, which were pertinent for our scope. The main pixel classes represented in the validation dataset were settled as clear sky, thick clouds, and semi-transparent clouds over different surfaces such as land, snow/ice and coastal water. Three different classes of semi-transparent clouds were identified: thick, average/semi-dense, and thin. The statistical distribution of the classes was representative of mean global clouds cover condition, having 60% of cloudy pixels (half of which semi-transparent clouds), and 40% of clear pixels. A large portion of thick clouds was saturated. They were marked as oversaturated.
The distribution of surface type was in-line with the typical observation scenario of the Proba-V sensor, acquiring largely over land (70%), with the remaining 30% equally distributed over coastal, inland water and snow/ice.
In addition of being statistically significant, the validation dataset needed to be global and representative of different seasons, meaning that the selected pixels needed to be equally spread over the globe for the four seasons, allowing correctly representing the different climatological conditions and land cover types. Finally, different geometries of observation were included (sun and viewing geometries) in order to represent different illumination condition and atmospheric path radiances.
The validation dataset was hidden to the Round Robin participants, in order to avoid that algorithms were adjusted to the pixels, where the actual quality assessment was made.

The test dataset represented a small but statistically representative subset of the validation dataset. The purpose of the test dataset was to provide to the participants example of pixel classification criteria and nomenclature. The test dataset was therefore limited, but still provided a representative example of all the pixel classes.

III. CLOUD DETECTION ALGORITMS

Various cloud detection algorithms were developed in the framework of this Round Robin. Table 1 summarizes the cloud detection algorithms and the auxiliary data, which is deemed necessary to constraint a specific algorithm (e.g., land cover maps, surface albedo). The difference between the algorithms lies in the different approaches used to select and separate cloudy pixels from clear pixels.

TABLE I: Cloud Detection Algorithms developed in the framework of the Round Robin and list of the auxiliary data needed for each algorithm.

Algorithm Name	**Methods**	**Auxiliary Files**
ALGO 1	Cumulative Discriminant Analysis	Seviri and Modis cloud mask; Globcover mask
ALGO 2	Cumulative Discriminant Analysis	Seviri cloud mask; Globcover mask
ALGO 3	Cumulative Discriminant Analysis	Modis cloud mask; Globcover mask
ALGO 4	Multi-spectral and multi-textural thresholding	Land Cover data of the ESA CCI
ALGO 5	Multilayer Perceptron (MLP) Neural Network	None
ALGO 6	Kohonen Self-Organizing Maps	Modis cloud mask
ALGO 7	Dynamic Thresholding	Land Cover data of the ESA CCI; GlobAlbedo surface reflectance
ALGO 8	Classification based on Cloud Optical Thickness	ERA-interim; DEM (GTOPO 30); GlobAlbedo surface reflectance

The approaches considered in this contest are the following ones:

ALGO1-3: The main methodology adopted for discriminating clouds is Cumulative Discriminant Analysis

(CDA), described in detail by [6]. In a one-dimensional problem (1 spectral band) it considers empirical estimates of the cumulative functions of reflectance, *x*, in clear and cloudy conditions, $\dot{F}_{Clear}(x)$ and $\dot{F}_{Cloudy}(x)$, respectively. The Discriminant Analysis produce a decision rule $\Gamma(x;\mathbf{X})$, with $\mathbf{X}$ being the training data set. The methodology relies therefore on a classification of binary variables (1=clear; 2=cloudy) depending on a threshold Θ, needed to classify a given pixel as clear or cloudy. The system needs a training dataset and a reference cloud mask.

The training data set in this case was a subset of the full Proba-V scenes for which a cloud mask is estimated by a consolidated algorithm (*silver* standard). The following three cloud masks have been considered as *silver* standard:

a) Joint SEVIRI and MODIS cloud mask. It is obtained retaining only scenes where the SEVIRI and MODIS cloud masks were agreeing (ALGO1).
b) SEVIRI cloud mask. It is obtained starting from radiance measured from the SEVIRI sensor on-board geostationary METEOSAT Second Generation (MSG) satellites every 15 minutes within a radius of about 60 degrees around the point at zero latitude and longitude at 3km spatial resolution (ALGO2).
c) MODIS cloud mask. It is obtained from radiance measured by MODIS sensor on-board Terra and Aqua EOS satellites at 1km spatial resolution and with a time difference of maximum 30 minutes between corresponding Proba-V and MODIS scenes (ALGO3).

In order to have as homogeneous reflectances as possible, different training sets have been setup according to different types of surface underlying scenes. Surface type has been estimated from the GlobCover map [7], based on land cover maps from the 300m MERIS sensor on-board the ENVISAT satellite. Five surface typologies have been defined, such as water, vegetation, bare land, urban, and snow/ice.

ALGO4: The methodology of the cloud detection algorithm consists of thresholding image derived parameters, which consists of multispectral indices (band ratios, band differences, NDIs), texture parameters, and local parameters derived from intermediate results with moving windows of various sizes. As auxiliary data, the water surfaces of the Climate Change Initiative (CCI) Land Cover data from 2010 are used.

The developed methodology, as it is based on thresholding image parameters, belongs to the most commonly used cloud screening approaches. However, the procedure presented is a completely new adaptation of such methods to Proba-V data, and does not incorporate known thresholds or parameter definitions from the literature. It has been developed empirically, step by step, with a trial and error procedure, selecting and interlinking finally the best working and feasible single steps, parameters, thresholds and their combinations. The algorithm is structured into three major processing steps with clearly defined substructures each. The processing has been implemented in the Spatial Model Editor (2016) of ERDAS Imagine, a graphical tool for performing interlinked sequences of image and GIS data processing. The procedure consists of three models that have the following functions:

1. Cloud/haze/snow/ice retrieval producing 66 spectral classes without thematic assignment to either one of the target classes yet.
2. Generation of image parameters for the subsequent filtering and thematic assignment.
3. Filtering and thematic assignment of the 66 classes derived in (1) to Clouds, Haze, Snow/Ice, and Thin/Partial Snow/Ice, and addition of some cloud pixels in gaps, buffers and water areas.

ALGO5: The proposed cloud masking process relies on the extraction of meaningful physical features (e.g. brightness and whiteness) that are combined with spatial features to increase the cloud detection accuracy [8]. In short, the four Proba-V spectral channels (4 TOA reflectance bands), ten physically-based spectral features, and the mean and standard deviation at two different scales for each pixel-based feature were considered. In order to train statistical machine learning models from real data, a representative number of samples had to be labeled as cloud-contaminated or cloud-free samples. To label in a semi-automatic way a sufficient number of pixels from the Proba-V images, the user-driven methodology proposed for MERIS in [9] to the Proba-V images was adapted, where the labeling of cloud clusters found in the image is done by an expert. Then, a supervised pixel-based classification, based on the TOA reflectance and on the manually labeled training set, is applied to these features providing the pixel label ('cloud' or 'cloud free'). Although several advanced supervised classification algorithms have been tested, the multilayer perceptron (MLP) neural network [10] has been selected due to its good performance and capability of ingesting a large number of training samples.

ALGO6: The Proba-V Cloud detection algorithm has been performed using a generic tool based on a classification with unsupervised Kohonen self-organizing maps (SOM) [11]. The SOM performance for cloud classification has been assessed and trained on MODIS and Landsat dataset and the results have been validated against MODIS CLOUD masks. 294 classes have been gathered in three groups: clear, cloudy, partially cloudy. In a second step, the topological map has been derived to fit PROBA V spectral bands and applied automatically to the PROBA-V dataset.

ALGO7: This cloud detection algorithm is based on adjusted thresholds on the reflectances in the Blue and SWIR channels. The approach is based on the usage of climatological maps of albedo (monthly averages obtained from the GlobAlbedo dataset). In addition to the GlobAlbedo dataset, also a 10 years global climatology of monthly means of MERIS reflectances in the blue band was used. This auxiliary information is used to design a “dynamic threshold” algorithm, with cloud tests customized for each status class (land, water, snow/ice, unknown land cover) [12].

ALGO8: This cloud detection algorithm relies on the estimation of the cloud optical thickness (COT) in the NIR and RED bands. The COT algorithm is based on look

up tables of coefficients derived from radiative transfer calculations using the radiative transfer code MOMO [13, 14], respecting the full range of observation and sun geometries, variations of the underlying surface (Lambertian above land, or wind roughened water surface above water [15]), and variations of cloud optical thickness and height.
For the determination of the COT, auxiliary surface albedo (utilizing MERIS Albedomap data [16]) and wind speed (from the meteorological re-analysis ERA-Interim) are used. Eventually, if the COT exceeds a particular threshold (here 1), a pixel is considered to be cloudy. Above land the RED-COT and above sea the NIR-COT are used, since typically, the respective surfaces are darker in these bands and the information gaining contrast is increased. A cloudy pixel is restored as snow covered, if the NDSI exceeds 0.45 and the 2m surface temperature (from ERA-Interim) is below 5°C.

IV. QUALITY ASSESSMENT AND RESULTS

In order to perform the Quality Assessment of the different algorithms, a tool, named PixBox [17], has been used. The PixBox tool has been populated by the pixel classified by visual inspection, which were used to create the validation dataset. The PixBox tool ensures a comprehensive set of statistical quality metrics in order to finally inter-compare the performances of the various algorithms. These metrics are based on confusion matrices, which allow assessing the classification accuracy by pixel-wise comparison of the considered mask with respect to the reference PixBox classification.
Performance metrics include the overall accuracy (OAA) of the algorithm, defined as the total number of correct classifications divided by the total number of sample points, the producer's accuracy (PA), which reflects the proportion of sample points correctly classified as X over the number of points observed to be X and the user's accuracy (UA), i.e., the proportion of sample points correctly classified as X over the number of points predicted to be X. The difference between producer's and user's accuracy is the difference between defining accuracy in terms of how well the cloudy or clear pixels can be mapped (producer's accuracy) versus how reliable the classification is to the user (user's accuracy). The OAA, the PA and the UA indices have been used to study the performance of the different algorithms for cloud detection under different conditions.

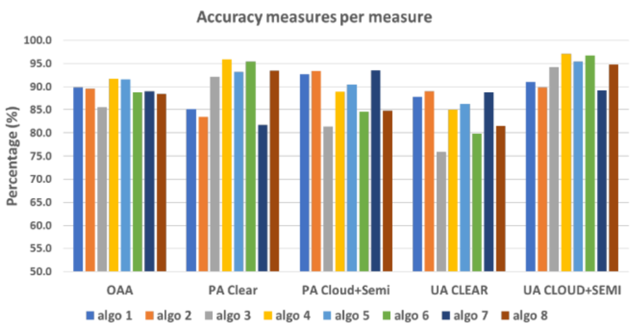

Fig.1: OOA, PA and UA indices characterizing the eight participating algorithms.

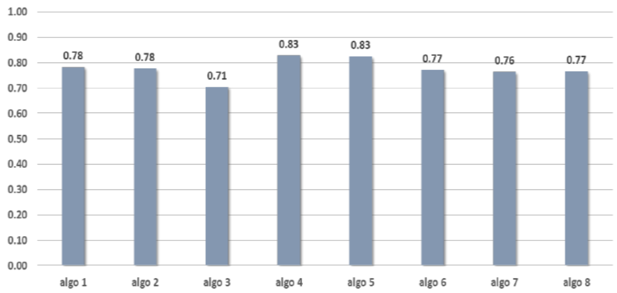

Fig. 2: Reliability of the eight participating algorithms as measured by the Krippendorff's *Alpha* coefficient.

In particular, the behavior of the algorithms for different surfaces, such as land, bright surfaces, coastline and water has been studied. Furthermore, it has been also investigated how all algorithms behave with respect to semi-transparent clouds.

Figure 1 shows the OAA, the PA and the UA for all algorithms per measure. All overall accuracies are larger than 85% and two algorithms have an OAA > 90%. Algo 3, 6 and 8 show high values in the UA CLOUD, but compared to Algo 4 and 5, they have low values in PA for clouds indicating that they are omitting clouds.
Algo 1, 2 and 7 show highest producer's accuracy for clouds, indicating that the clouds are detected with high probability but that also clear cases are classified as cloud, as indicated by the lower producer's accuracy for clear.
The Krippendorff's *Alpha* coefficient has been calculated in order to have a good measure of reliability. *Alpha* assumes the value 1, if the agreement between what it is identifies as cloud by the algorithm and the reference observations is perfectly matched, whereas *Alpha* is 0 when there is a systematic disagreement. Figure 2 shows the Krippendorff's *Alpha* coefficient for each participating algorithm. Overall all algorithms are of very good quality. The Quality Assessment was further complemented by visual inspection of several RGB images to identify issues in cloud borders delineation.

V. CONCLUSION

Accurate cloud detection is an important step for building a stable retrieval process for geophysical products from Proba-V data. Throughout this project an inter-comparison of the performances of different cloud detection algorithms for Proba-V was carried out.
Each algorithm has strong and weak points. Algo 4 and Algo 5 are in a very good position. Algo 1 and Algo 2 are very similar, but Algo 2 is more clear-sky conservative, flagging clear surfaces as cloud while having a good detection of semi-transparent clouds. Algo 3, 6 and 8 are the ones detecting less clouds, but they have very little commission errors, e.g. at coastlines or in inland waters or over bright surfaces. Algo 7 is the most conservative and therefore detects small clouds, cloud borders and semi-transparent clouds but is flagging many clear pixels.

REFERENCES

[1] W. Dierckx, et al., "PROBA-V mission for global vegetation monitoring: standard products and image quality." International Journal of Remote Sensing 35.7 (2014): 2589-2614.

[2] G. Lisens, P. Kempeneers, F. Fierens, and J. Van Rensbergen, Development of Cloud, Snow, and Shadow Masking Algorithms for VEGETATION Imagery. Proceedings of Geoscience and Remote Sensing Symposium, IGARSS 2000, Honolulu, HI 2: 834–836.

[3] O. Hagolle, M. Huc, D.V. Pascual, G. Dedieu, A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENµS, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment,* 2010, 114(8), 1747-1755.

[4] A. Hollstein, J. Fischer, C. Carbajal Henken, R. Preusker, Bayesian cloud detection for MERIS, AATSR, and their combination. *Atmospheric Measurement Techniques Discussions,* 2014, 7(11), 11045-11085.

[5] J. D. Jang, A. A Viau, F.Anctil, E. Bartholomé, (2006). Neural network application for cloud detection in SPOT VEGETATION images. *International Journal of Remote Sensing,* 2006, 27(4), 719-736.

[6] U. Amato, et al., Cloud mask via cumulative discriminant analysis applied to satellite infrared observations: scientific basis and initial evaluation. *Atmos. Meas. Tech.* **7**, 2014, 3355–3372

[7] T. R. Loveland,et al.: Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data: *Int. Journal of Remote Sensing* **21**(6/7), 2000, 1303–1330

[8] Linlu Mei, et al.: “A cloud masking algorithm for the XBAER aerosol retrieval using MERIS data,” Remote Sensing of Environment, vol. In Press, 2017.

[9] L. Gómez-Chova, et al.:, “Cloud-screening algorithm for ENVISAT/MERIS multispectral images,” IEEE Trans. on Geoscience and Remote Sensing, vol. 45, no. 12, 2007 pp. 4105–4118,

[10] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, Englewood Cliffs, NJ, USA, 1999

[11] O.Chabiron, Détection de Nuages à Haute résolution spectrale, Rapport de stage, Master 2 Institut de Mathématiques Appliquées de Toulouse, Aout 2012.

[12] E. L. A.Wolters, E.Swinnen, I. Benhadj, W. Dierckx, PROBA-V cloud detection evaluation and proposed modification, QWG Technical Note, 17/7/2015.

[13] G. N. Plass, G. W. Kattawar, and F. E. Catchings. Matrix operator theory of radiative transfer. 1: Rayleigh scattering. Applied Optics, 12(2):314–329, 1973.

[14] F. Fell and J. Fischer. Numerical simulation of the light field in the atmosphere ocean system using the matrix-operator method. Journal of Quantitative Spectroscopy and Radiative Transfer, 69(3):351 – 388, 2001.

[15] C. Cox and Munk, W.: Measurement of the Roughness of the Sea Surface from Photographs of the Sun‘s Glitter, J. Opt. Soc. Am.,44, 838–850, 1954.

[16] Muller, J.-P., Preusker, R., Fischer, J., Zuhlke, M., Brockmann, C., and Regner, P.: ALBEDOMAP: MERIS land surface albedo retrieval using data fusion with MODIS BRDF and its validation using contemporaneous EO and in situ data products, in: Geoscience and Remote Sensing Symposium, 2007, IGARSS, 2007,IEEE International, 2404–2407, 2007

[17] C. Brockmann, et al.: A Cloud Screening and Pixel Characterization: IdePix Approach and Validation Using PixBox, Sentinel-3 OLCI/SLSTR and MERIS/(A)ATSR workshop, ESA-ESRIN, Frascati, Italy, 15-19 Oct 2012.