ESA UNCLASSIFIED – For Internal Use Privileged

esa
esrin

Via Galileo Galilei Casella Postale 64 00044 Frascati Italy
T +39 06 9418 01 F +39 06 9418 0280 www.esa.int

DOCUMENT

Proceedings, Recommendations and Lessons Learnt from Proba-V Cloud Detection Round Robin Workshop, ESA/ESRIN, 1 March 2017

Prepared by F. Niro (EOP-GMQ)
Approved by P. Goryl (EOP-GMQ)
Reference TBD
Issue 1.1
Date of Issue 22/03/2017
Status Final
Document Type TN

European Space Agency
Agence spatiale européenne

Table of Contents

1 Introduction 3
1.1 Background 3
1.2 Acronyms and Abbreviations 4
1.3 Reference Documents 5
2 Proceedings 6
2.1 Welcome and Introduction 6
2.2 Algorithms Overview 6
2.3 Inter-comparison Results 10
2.4 Closing Session 11
3 Recommendations and Lessons Learnt 12
4 APPENDIX A: Participants 14

1 INTRODUCTION

During 1st March 2017 a one-day Workshop was held in ESA/ESRIN to discuss the results from an inter-comparison exercise on different cloud masking approaches for Proba-V. This document provides the Proceedings of the Workshop, the Recommendations and the list of participants. Presentations from the Workshop will be made available at the following web page:

https://earth.esa.int/web/sppa/meetings-workshops/expert-meetings/proba-v-cloud-detection-round-robin-meeting

1.1 Background

The Proba-V Cloud Detection Round Robin project started in Jan 2016 and was funded by ESA and BELSPO. The motivations were to investigate potential new algorithms for Proba-V cloud detection, analyse strengths and weaknesses of each method for various surface type and cloud cover conditions, collect ideas and recommendations for implementation.

The Round Robin was open to any interested scientist. Participants were gathered with an open call. A grant was allocated to each participant in order to partially cover the effort spent for algorithm development and testing. The main requirement for participation was to develop a novel algorithm and provide the relevant cloud mask for the given input data and the associated ATBD.

The Quality Assessment of the different algorithm was performed by Brockmann Consult using a tool (PixBox) [RD-1], which was already successfully used for such purpose in other ESA projects, in particular Land Cover CCI. PixBox validation is based on the collection of a large (more than 50 thousands) and statistically representative set of pixels, which are manually classified by a remote sensing expert by visual inspection. This set is used as "truth" to compare against during the validation.

Five groups participate to the Round Robin, providing different cloud detection approaches spanning a wide range of methodologies:

- CNR (Italy): Discriminant Analysis
- University of Valencia (Spain): Neural Network
- Free University of Berlin (Germany): Physical Retrieval of Cloud Properties (COT)
- Magellium (France): Self-Organizing Feature Mapping
- EoConsultancy (Austria): Multi-spectral and Multi-textural thresholding

In addition, the current operational Proba-V algorithm, developed by VITO (Belgium), was considered for the inter-comparison, it is based on a dynamic threshold method applied to the radiometry in the Blue band.

Additional information about the project, the delivered documentation and updates can be accessed at the following web page:

https://earth.esa.int/web/sppa/activities/instrument-characterization-studies/pv-cdrr/

1.2 Acronyms and Abbreviations

ATBD	Algorithm Theoretical Basis Document
CCI	Climate Change Initiative
EO	Earth Observation
ESA	European Space Agency
LTDN	Local Time of Descending Node
MEP	Mission Exploitation Platform
MERIS	MEdium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautical and Space Administration
NDVI	Normalized Difference Index
NDVI	Normalized Difference Vegetation Index
NDSI	Normalized Difference Snow Index
NDWI	Normalized Difference Water Index
NIR	Near-InfraRed
NN	Neural Network
OAA	OverAll Accuracy
OLCI	Ocean and Land Colour Instrument on board Sentinel-3
PA	Producer's Accuracy
PDF	Portable Document Format
PROBA-V	Project for on-board Autonomy-Vegetation
QA	Quality Assessment
QWG	Quality Working Group
SPOT	Satellite Pour l'Observation de la Terre
SZA	Sola Zenith Angle
SWIR	Short-Wave InfraRed
TOA	Top Of Atmosphere
TOC	Top Of Canopy
UA	User's Accuracy

VIS	Visible
VITO	Flemish institute for technological research
VNIR	Visible and Near-InfraRed
VIIRS	Visible Infrared Imaging Radiometer Suite

1.3 Reference Documents

[RD-1] Brockmann C., Paperin M., Danne O., Kirches, G., Bontemps, S., Stelzer, K., Ruescas, A. Cloud Screening and Pixel Characterisation: IdePix Approach and Validation Using PixBox, Sentinel-3 OLCI/SLSTR and MERIS/(A)ATSR workshop, ESA-ESRIN, Frascati, Italy, from 15 to 19 October 2012.

[RD-2] 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

[RD-3] 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

[RD-4] O.Chabiron, Détection de Nuages à Haute résolution spectrale, Rapport de stage, Master 2 Institut de Mathématiques Appliquées de Toulouse, Aout 2012.

2 PROCEEDINGS

The main points arising from the presentations and discussions are summarised below. Recommendations [*REC*] and Lessons Learnt [*LES*] are identified when they arose in the meeting, and collected in the final table in Section 3.

2.1 Welcome and Introduction

F. Niro (ESA/ESRIN) welcomed the participants and acknowledged their participation to the Round Robin; the success of the project, in terms of scientific value, was largely due to their contribution to the inter-comparison exercise. The objectives of the meeting were briefly recalled: to review the various algorithms' baseline definition, to discuss the inter-comparison results, to collect recommendations for cloud detection, which are relevant both for Proba-V as well as for other ESA optical sensors, mainly S2 and S3. One of the objectives of the project was already fulfilled; it is the increase of awareness on the mission data, since five different groups with a long lasting experience on optical sensors, approached Proba-V products for the first time.

R. Q. Iannone (ESA/ESRIN) reminds the overall project organization and motivations, the Round Robin Design, validation approach and expected deliverables. Serco managed the overall project in close collaboration with ESA and the Proba-V QWG. Brockmann Consult provided the scientific expertise and it was responsible for the algorithms validation and quality assessment. The Round Robin design was shown and discussed. The same set of blinded input data (no cloud or snow/ice flags), consisting of 4 days of Proba-V L2a 333m global products from 4 days in different 2014' seasons, were provided to all participants and to Brockmann Consult. The validation was based on pixel-by-pixel comparison of the manually classified mask with the one provided by the various participants. Visual inspection of selected images was used to complement the pixel-based validation, in order to analyse performances in cloud delineation and pattern recognition for critical conditions e.g., bright surfaces, turbid coastal water, and turbid atmosphere.

2.2 Algorithms Overview

L. G. Chova (University of Valencia, Spain) presented the algorithm developed with the support of G. M. Garcıa. L. G. Chova was strongly involved in the past in the development of cloud detection for ENVISAT/MERIS [RD-2], he's now also involved in the development of a cloud screening methodologies applied on Landsat images within the Google Earth Engine platform. For this Round Robin exercise, an advanced machine learning method is proposed. The approach consisted of two main steps: features extraction from the images and a supervised classification method trained on manually classified image's clusters. The extracted features include spectral and spatial physically based features, namely, brightness and whiteness in VIS and NIR bands and relevant spectral indices (NDSI and NDVI), as well as mean and standard deviation within moving windows of 3x3 and 5x5 size. In order to reduce complexity of the classifier and remove redundant information, the most relevant features were selected using filter approaches. Three classification methods were tested, the best performances being obtained with a Multilayer Perceptron Neural

Network (MLP) approach. For training the algorithm a large number of images have been manually classified into cloud/clear classes after an unsupervised clustering analysis. The proposed algorithm has several advantages for a potential implementation, in particular, it doesn't require any auxiliary data, it is extremely fast, and it is globally valid, as long as it is properly trained. Indeed, it was pointed out that, as for any NN algorithm, the availability of a comprehensive and accurately labelled training dataset, determines the quality of the results. The results of the proposed algorithm for Proba-V cloud detection will be presented at the next IGARSS2017 conference.

During the discussion, it was stressed the interest in accessing the validation dataset of more than 50,000 pixels visually classified by Brockmann Consult. Most of the participants agreed that their algorithms would benefit from this dataset. On the other hand, K. Stelzer and C. Brockmann pointed out that if we release to the participants this validation dataset, we would loose the possibility to reuse it for validation of future cloud detection algorithm.

[REC-1: Considering the effort spent in this validation dataset generation and the need of reuse it in the frame of future Proba-V QWG validation activities, it is recommended not to release the complete Round Robin Validation Dataset to the participants].

A lesson learnt was additionally highlighted on the use of the Mission Exploitation Platform (MEP), which was offered as optional tool for the Round Robin. Most of the participants had issues in using commercial licenses on the MEP (e.g., Matlab) and owing to the limited size of input data and the limited time for the exercise, all prefer to work on their platforms for developing and running their algorithms.

[LES-1: The usage of the MEP virtual environment for algorithm development requires non-negligible effort for code adaptation and SW environment porting, this is not ideal for short-time projects and when the size of input data is limited].

U. Amato (National Research Center, CNR, Italy) presented the algorithm, which was implemented in collaboration with C. Serio. The algorithm already developed for SEVIRI [RD-3], was here adapted to Proba-V. The methodology is based on Cumulative Discriminant Analysis (CDA), which is a supervised pixel classification method allowing to estimate statistical properties of the classes. As any supervised method, the accuracy of the pixel classification is strongly dependent on the representativeness and accuracy of the used training dataset. The method was tested with three different training datasets derived from space and time collocated cloud mask from MODIS, SEVIRI and joint MODIS+SEVIRI mask (retaining only pixels where cloudy or clear labels of SEVIRI and MODIS agree). The results for the three datasets were all provided to the Round Robin and will be considered as "separate" algorithms in the final quality assessment. A global land cover map @300m, from the ESA GlobCover project, was used as ancillary information within the algorithm to better discriminate cloud spectral features from the underlying surface. The provided cloud mask strongly depends on the definition of the cost function to be minimised within the CDA scheme, this choice drives the conservative nature of the mask, being either clear-sky or cloud conservative. Different spectral dependence of clear and cloudy pixels with respect to SZA is also shown; this would need to be taken into account to improve cloud masking.

During the discussion, the accuracy of the training dataset was debated, in particular with respect to the significant difference in observation time with MODIS (30min) and in spatial resolution with SEVIRI (3km). The validation results showed that the joint MODIS+SEVIRI mask has the best performances. The availability of the Round Robin validation dataset, to be used as "gold standard" for training was further requested. [*See REC-1 within L. G. Chova presentation*].

U. Gangkofner (EOConsultancy, Austria) introduced her algorithm, which was expanded to further classify the pixel into additional classes, besides the required cloud/clear, i.e., semi-transparent, ice/snow and thin snow. The method can be defined as Multi-spectral and Multi-textural thresholding method, in which a large set of spectral and spatial features, including several spectral indices (such as NDVI, NDWI and a dedicated NDI), and various statistical indicators computed in moving windows of various sizes, are used to discriminate the various classes. Thresholds are properly tuned by inspecting segmented images in relevant Region of Interest. The method is relatively new, though it belongs to the general family of thresholding methods, and it allows extremely accurate and detailed classification, including treatment of the spatial shift between SWIR and VNIR focal planes. As ancillary data, the Land Cover CCI map is used to improve cloud discrimination for different surface types. The algorithm was fully developed in ERDAS IMAGINE (2016) Spatial Model Editor. A relatively long computation time was anticipated.

During the discussion, several drawbacks were highlighted regarding the potential operational implementation, in particular a relatively long computation time, the need for complete porting of the code to an open-source language, as well as the very complex algorithm processing flow. On the other hand, this approach showed that in case of mono-temporal methods, spatial texture features are essential for pixel class discrimination, especially over critical surface areas, such as salt lakes, wetlands, dry surfaces or agricultural fields. Investigation over the mountainous areas shows some remaining issues in cloud/snow classification, which could be potentially improved by using a DEM.

[LES-2: Appropriate definition of multi-textural features are essential for improving cloud discrimination over complex landscape and critical surface conditions, this can be an alternative to multi-temporal approaches. Furthermore, for cloud/snow discrimination over mountainous areas, the use of the DEM can be an interesting option].

B. Berthelot (Magellium, France) presented her method for cloud detection. The algorithm, based on Neural Network Kohonen classification, uses references topological maps created from MODIS L1c data, which can be in principle adapted to any sensor, after spectral adjustment to the sensors' specific bands. The method was successfully used to produce Landsat cloud mask [RD-4] and it was here tested for Proba-V data. Aqua MODIS TOA reflectances in the seven bands used for land applications (up to 2.1 micron) were used as reference dataset. MODIS cloud mask were used for a supervised statistical learning with the objective to assign the pixel's label: cloud/clear/semi-transparent. The MODIS referent vectors were then resampled to Proba-V channels using the relevant SRF. The spectral interpolation at TOA is appropriate over cloudy areas with smooth spectral signature, while over land it introduces artefacts. The maps obtained are then used for the classification and the quality assessment is performed over the provided test dataset. The algorithm can be easily tuned to be more or less clear-sky conservative. The current version was optimized over land surfaces, but can be easily expanded to coastal water by extending the reference maps. No ancillary data is used for the classification, beside the SRF used for spectral adjustment. The code was developed in Matlab and several ways for further improving/extending the algorithm were already identified, including the use of Proba-V data for the learning phase as well as specialized maps for snow detection.

During the discussion, the issue of spectral interpolation was raised, in particular when dealing with semi-transparent clouds, where spectral signature may include some remaining land spectral features.

R. Preusker (Free University of Berlin, Germany) presented his idea for cloud detection. It is a general physical approach that applies to any sensor. The idea is not to perform a pixel classification, rather to retrieve the Cloud Optical Thickness (COT) from Proba-V measurements. When COT is retrieved, cloud flagging will be only a matter of tuning the threshold based on the applications' needs. The retrieval of COT is based on the computation of a set of forward model simulations, using the MoMo radiative transfer code developed at FUB. The simulations should represent a wide range of atmospheric and surface conditions, i.e., different COT, cloud height, surface brightness and roughness, geometry of observation. Ozone and Water vapour are kept constant in the simulations. The inverse modelling consisted of a standard Artificial Neural Network (ANN) approach. The COT threshold for cloud flagging is set dependent on the surface type (COT in NIR over clear water, COT in Red over Land), COT in the Blue is not used (ambiguity with Rayleigh), while the SWIR is used for cloud/snow discrimination. Auxiliary data include DEM, Land/Sea mask, wind speed (ERA interim), Albedo Maps climatology in Red and NIR (derived from MERIS) and 2m air Temperature (ERA interim), the latter being used for snow pixel restoration. The algorithm is extremely fast and was developed entirely on Python with a modular architecture; most of the computing time is actually taken from interpolation of ERA interim values to Proba-V swath and time. Several potential improvements were already identified, including use of better albedo maps from MODIS/VIIRS, introduction of semi-transparent cloud flag and cloud shadow estimation based on geometry and radiometry; estimation of cloud probability is also possible, though it will require much more work. The strength of the algorithm is that it is not based on a subjective or manual detection of cloud, rather it retrieves a physical cloud property (COT), which can be rigorously validated. The performances are strongly dependent on the used albedo maps, and the clear message is to use MODIS/VIIRS maps for this purpose.

During the discussion it was highlighted that this is the first attempt to retrieve cloud optical properties (COT) from Proba-V sensor, despite the limited information available in the four Proba-V spectral bands. The first maps of COT from Proba-V look reasonable. The same approach will be tested in the future on Sentinel-3 data.

Concerning the Proba-V Round Robin, a first rough estimate of threshold was used for detecting clouds/clear pixels for the provided input data. Cloud masks were provided to Brockmann Consult and were considered to be part of the final Quality Assessment. The ATBD needs still to be provided by FUB.

2.3 Inter-comparison Results

K. Stelzer (Brockmann Consult, Germany) presented the final validation results, based on statistical comparison of each cloud mask (including the operational one) with respect to the reference dataset of more than 50,000 manually classified pixels. The analysis was further broken down into different surface types, in particular land and water pixels, as well as for bright and semi-transparent clouds. RGB images were also presented showing the capability of each algorithm to correctly delineate cloud textural features over complex surface or atmospheric conditions. The detailed analysis can be found in the Workshop's presentation as well as in the Validation Report of the project, they will be both available on the relevant SPPA web pages:

https://earth.esa.int/web/sppa/activities/instrument-characterization-studies/pv-cdrr/

Overall, all algorithms show to be of very good quality, though strengths and weaknesses of each method are clearly highlighted. Two algorithms were eventually found to have the best performances: the one from EOConsultancy and the one from University of Valencia. The current operational cloud mask (the one used for the Collection 1 reprocessing) has good statistical performances, though the inspection of RGB images highlighted some clear issues, namely a general tendency to commission errors (over-detection of clouds), and an issue (omission error) in bright cloud delineation over water. While the clear-sky conservative nature of the algorithm was already know and it is in-line with the requirements of the land community, the issue over water needs to be further investigated within the QWG.

[REC-2: To bring the results of the Round Robin exercise to next Proba-V QWG: in particular, to evaluate the potential use of the best performing algorithms for implementation in the future processor baseline, and to investigate the issue of the current operational cloud mask in delineating cloud over water].

During the discussion it was highlighted that all algorithms seem to be better performing over land than water, this is not expected, since cloud flagging is in general easier over dark and homogeneous water surfaces. This could be due to the fact that in general all algorithms were tuned over land surfaces, and also to the selection of coastal water pixels, often over turbid water, within the reference database.

One idea was raised during the discussion: to expand the inter-comparison by studying the "distance matrix" of the different algorithms, this means to derive a NxN symmetric matrix, where N is the number of algorithms. The elements of the matrix will be the averaged absolute differences in cloud flagging (considering only 0/1 for cloud/clear) for each pair of algorithms. Ideally we could derive a distance matrix for each statistically representative subset of validation pixels, e.g., land, water, semi-transparent clouds. This comparison strategy, which is also adopted in the frame of S2 ACIX, is valuable in highlighting the agreement/deviations of each pair of algorithms and to identify outliers and systematic features in cloud flagging.

[REC-3: To expand the inter-comparison results by computing distance matrices for specific set of validation pixels, e.g., land, water].

2.4 Closing Session

During the closing session, the scientific value of the Round Robin exercise was further stressed. As confirmed by the positive feedbacks received after the meeting, it was clear that similar exercises should be repeated in the future, with similar or different targets (e.g., snow/ice flagging, cloud shadows). The open call formula is also the best approach for stimulating and gathering new ideas, allowing to reach the broad scientific community. The Round Robin design, which was based on a validation dataset hidden to the participants, was recognized as being fair, providing a solid and unbiased protocol for quality assessment.

[REC-4: To consider similar Round Robin exercise with similar or different targets for fostering new ideas, following similar open call approach and a solid and well consolidated quality assessment protocol].

A more general discussion on the cloud flagging requirements for Proba-V (and for similar optical sensors) was further initiated. Ice/snow and shadow flags are also required (they are part of the operational processor), though they were not considered in this exercise, which was focused only on cloud/clear. With respect to the recurrent debate on clear-sky versus cloud conservative masks, the focus should be moved to what it is really needed by the end-users. For Proba-V, and in general for mission with focus on land applications, the main driver should be to provide to the end-users the largest possible amount of clear pixels with the maximum reliability on their clear status. At a first stage, therefore, clear-sky conservative cloud mask should be preferred, since Proba-V was not designed for cloud properties retrieval. On the other hand, clear-sky conservative cloud masks have a more or less pronounced tendency to commission errors, with resulting decrease of clear pixels' availability within the image, this is not ideal for multi-temporal composites generation. We should finally stress that different applications may have different sensitivity to remaining undetected clouds (e.g., cirrus); for instance, phenology studies can use multi-temporal information to remove or smooth-out cloud contaminated pixels in the long-term series, some ocean applications (e.g., OC Polymer) are also resistant to cirrus clouds.

[LES-3: The choice of a cloud mask, in particular with respect to clear-sky versus cloud conservative masks, remain challenging, being dependent on the requirements of the downstream applications].

[REC-5: In order to address the requirements of the various applications, leaving to the user the freedom to decide the level of cloud contamination that can be acceptable for its application, the best approach is to move toward a cloud probability mask, or at least to provide additional classes, e.g. thick clouds, semi-transparent clouds].

As a conclusive remark for the meeting, it was recalled that the results will be presented at the MultiTemp2017 Conference, to be held from 27 to 29 June in Bruges. All participants agreed to be co-author of a paper for the conference (possibly oral), furthermore, it was stressed that the results may also deserve publication in a peer-reviewed journal as long as some the analysis and discussion is expanded (e.g., distance matrix, impact of SZA). Considering that MultiTemp2017 requires an extended peer-reviewed paper, this will be the first target, in a second step and taking into account the feedback from the conference the preparation of a collective scientific paper will be considered.

[REC-6: To expand the analysis and discussion of results with the goal of first submitting a contribution to MultiTemp2017 proceeding and possibly to a peer-reviewed journal].

3 RECOMMENDATIONS AND LESSONS LEARNT

The key recommendations and lessons learnt arising from the meeting are summarised in the table below. The forum where this inputs will be discussed is also provided.

Ref	Summary	Forum
[REC-1]	*Considering the effort spent in this validation dataset generation and the need of reuse it in the frame of future Proba-V QWG validation activities, it is recommended not to release the complete Round Robin Validation Dataset to the participants.*	*Proba-V QWG*
[LES-1]	*The usage of the MEP virtual environment for algorithm development requires non-negligible effort for code adaptation and SW environment porting, this is not ideal for short-time projects and when the size of input data is limited.*	*ESA-EOP-GMQ*
[LES-2]	*Appropriate definition of multi-textural features are essential for improving cloud discrimination over complex landscape and critical surface conditions, this can be an alternative to multi-temporal approaches. Furthermore, for cloud/snow discrimination over mountainous areas, the use of the DEM can be an interesting option.*	*Proba-V QWG*
[REC-2]	*To bring the results of the Round Robin exercise to next Proba-V QWG: in particular, to evaluate the potential use of the best performing algorithms for implementation in the future processor baseline, and to investigate the issue of the current operational cloud mask in delineating cloud over water.*	*Proba-V QWG*

Ref	**Summary**	**Forum**
[REC-3]	*To expand the inter-comparison results by computing distance matrices for specific set of validation pixels, e.g., land, water.*	*Round Robin participants*
[REC-4]	*To consider similar Round Robin exercise with similar or different targets for fostering new ideas, following similar open call approach and a solid and well consolidated quality assessment protocol.*	*ESA-EOP-GMQ*
[LES-3]	*The choice of a cloud mask, in particular with respect to clear-sky versus cloud conservative masks, remain challenging, being dependent on the requirements of the downstream applications.*	*Proba-V QWG*
[REC-5]	*In order to address the requirements of the various applications, leaving to the user the freedom to decide the level of cloud contamination that can be acceptable for its application, the best approach is to move toward a cloud probability mask, or at least to provide additional classes, e.g. thick clouds, semi-transparent clouds*	*ESA-EOP-GMQ*
[REC-7]	*To expand the analysis and discussion of results with the goal of first submitting a contribution to MultiTemp2017 proceeding and possibly to a peer-reviewed journal.*	*Round Robin participants*

4 APPENDIX A: PARTICIPANTS

The list of Meeting's participants is provided in the following table

Name	Affiliation	Country
Philippe Goryl	ESA/ESRIN	Italy
Fabrizio Niro	ESA/ESRIN	Italy
Rosario Q. Iannone	ESA/ESRIN	Italy
Erminia De Grandis	ESA/ESRIN	Italy
Daniele Casella	ESA/ESRIN	Italy
Georgia Doxani	ESA/ESRIN	Italy
Kerstin Stelzer	Brockmann Consult	Germany
Carsten Brockmann (Webex)	Brockmann Consult	Germany
Rene Preusker	Free University of Berlin	Germany
Ute Gangkofner	EOConsultancy	Austria
Carmine Serio	U. degli Studi della Basilicata	Italy
Umberto Amato	CNR	Italy
Béatrice Berthelot	Magellium	France
Luis Gómez-Chova	University of Valencia	Spain
G. Mateo-García	University of Valencia	Spain