Algorithm Theoretical Basis Document (ATBD) for LAI, fAPAR, and fCover PROBA-V products at 300m resolution (GEOV3)

Frédéric Baret, Marie Weiss, Aleixandre Verger, Bruno Smets

IMPLEMENTING MULTI-SCALE AGRICULTURAL INDICATORS EXPLOITING SENTINELS

ATBD FOR LAI, FAPAR AND FCOVER FROM PROBA-V PRODUCTS AT 300M RESOLUTION (GEOV3)

IMAGINES_RP2.1_ATBD-LAI300M

ISSUE 1.73

EC Proposal Reference N° FP7-311766

Due date of deliverable: June 2013

Actual submission date: August 2013

Start date of project: 01.11.2012 Duration : 40 months

Name of lead partner for this deliverable: INRA

Book Captain: Fred Baret (INRA)
Contributing Authors: Marie Weiss (INRA)
Aleixandre Verger (CREAF)
B. Smets (VITO)

Project co-funded by the European Commission within the Seventh Framework Program (2007-2013)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document Release Sheet

Book Captain:	Fred Baret	Date: 31.05.2016	Sign. 
Approval:	R. Lacaze	Date: 03.06.2016	Sign. 
Endorsement:	M. Koleva	Date:	Sign.
Distribution:	Public		

CHANGE RECORD

Issue/Revision	Date	Page(s)	Description of Change	Release
	19.04.2013	All	First draft	D1.00
D1.00	26.06.2013		Improved StepB and EBF/No-EBF detection	I1.20
I1.20	01.07.2013		Add the list of parameters	I1.30
I1.30	21.07.2013		Improve the definition domain; Add quality flags; Detailed input description	I1.40
I1.40	24.07.2013		Step A completed	I1.50
I1.50	02.08.2013	All	Final version	I1.60
I1.60	19.09.2013	17-18 25-26	Update PROBA-V data descriptor Update meteo data descriptor	I1.61
I1.61	11.10.2013		Update after the Design Key Point: §3.1.3, Table 4, Table 7, Step A in §3.3.1.2 and §3.3.1.4, Step B in §3.3.2.1.2, §3.3.2.1.3, §3.3.2.1.4, §3.3.2.2.1, §3.3.2.2.5 and §3.3.2.3	I1.62
I1.62	08.07.2014		Clarifications and corrections (e.g. Figure 6) after Global Land Service Review 3	I1.70
I1.70	04.11.2015		Update to be compliant with operational implementation Revision of Section 2.4 and 3.3 after Global Land Review 4 Update Conclusion	I1.71
I1.71	24.11.2015	29	Update Table 6: quantitative quality indicators	I1.72
I1.72	31.05.2016	29 29 30 36	Table 4: 2 first line: change one parameter name, add another one (to describe the compositing window selection more clearly) Table 4: δ_{LAI} is set to 0.5 Table 6: maximum value of the number of observations has been corrected Rephrase section 3.3.2.1.1. Correct the	I1.73

			error $length_{no_EBF}^{max}$in place of $length_{EBF}^{max}$	
		37	Figure 8 has been modified to include land cover identification at the decadal date D of a given pixel	
		38	Rephrase the section. Correct the condition to decide the land cover of the considered pixel.	
		Erreur ! Signet non défini.	**Erreur ! Source du renvoi introuvable.** was modified to make it clearer	
		41	Rephrase to be hopefully clearer	
		46	Section 3.3.2.2.4 rewritten	
		44	Figure 13 has been modified to clarify the different conditions for gap filling	
		47	Add Figure 15 to illustrate the gap filling (step B3)	
		40,46	Equations describing the uncertainties have been reviewed	

TABLE OF CONTENTS

ACRONYMS

ATBD	Algorithm theoretical based Document
BRDF	Bidirectional Reflectance Distribution Function
DEM	Digital Elevation Model
EBF	Evergreen Broadleaf Forest
ECV	Essential Climate Variable
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
FCOVER	Fraction of vegetation cover
FIPAR	Fraction of Intercepted Photosynthetically Active Radiation
GAI	Green Area Index
GCOS	Global Climate Observation System
GEO	Group on Earth Observations
GEOV1	VGT bioGEOphysical product Version 1 (geoland2 project)
GEOV2	VGT bioGEOphysical product Version 2 (geoland2 project)
GEOV3	PROBA-V bioGEOphysical product Version 3 (ImagineS project)
GMES	Global Monitoring of Environment and Security
GTOS	Global Terrestrial Observation System
GUI	Graphic User Interface
LAI	Leaf Area Index
LDAS	Land Data Assimilation System
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Imaging Spectrometer
NIR	Near Infrared spectral domain
NNET	Neural Network
NNT	Neural Network Technique
SAA	Sun Azimuth Angle
SMAC	Simplified Method for the Atmospheric Correction
SZA	Sun Zenith Angle
SWIR	Short Wave Infrared
TBD	To be defined
TOA	Top of Atmosphere
TOC	Top of Canopy
TSGF	Temporal Smoothing and Gap Filling method

VAA	Viewing Azimuth Angle
VIS	Visible spectral domain
VZA	Viewing Zenith Angle
WGS84	World Geodetic System 1984

LIST OF FIGURES

LIST OF TABLES

1 BACKGROUND OF THE DOCUMENT

1.1 EXECUTIVE SUMMARY

The Copernicus program is the EU response to the increasing demand for reliable environmental data. The objective of the Copernicus Land Service is to continuously monitor and forecast the status of land territories and to supply reliable geo-information to decision makers, businesses and citizens to define environmental policies and take right actions. ImagineS intends to continue the innovation and development activities to support the operations of the Copernicus Global Land service, preparing the use of the new Earth Observation data, including Sentinels missions data, in an operational context.

The main objectives of IMAGINES are to (i) improve the retrieval of basic biophysical variables, mainly LAI, FAPAR and the surface albedo, identified as Terrestrial Essential Climate Variables, by merging the information coming from different sensors (PROBA-V and Landsat-8) in view to prepare the use of Sentinel missions data; (ii) develop qualified software able to process multi-sensor data at the global scale on a fully automatic basis; (iii) complement and contribute to the existing or future agricultural services by providing new data streams relying upon an original method to assess the above-ground biomass, based on the assimilation of satellite products in a Land Data Assimilation System (LDAS) in order to monitor the crop/fodder biomass production together with the carbon and water fluxes; (iv) demonstrate the added value of this contribution for a community of users acting at global, European, national, and regional scales.

This ATBD (Algorithm Theoretical Based Document) describes the proposed algorithm for the PROBA-V products at 300m resolution with justification of the choices made. It also proposes a real time approach for the product estimation with a delivery frequency of 10 days. The considered products are the following set of biophysical variables: LAI and FAPAR that are Essential Climate Variables (ECVs) as recognized by international organizations such as GCOS and GTOS. In addition, the FCOVER variable will be also generated since it corresponds to specific needs for some users.

This ATBD is an adaptation and a simplification of the GEOV2/VGT algorithm that was developed for the VEGETATION sensor. Because no PROBA-V data were available at the time of definition of the methodology, it was calibrated using VEGETATION2 sensor data at 1/112° spatial resolution. The proposed default values of the parameters will then need to be fine-tuned. Finally, when actual PROBA-V data will be available for at least few months, this algorithm will be run on a significant area to evaluate its performances and possibly adapt some of the parameters for improved performances.

1.2 SCOPE AND OBJECTIVES

One of the main objectives of ImagineS is to provide a continuity and improvement to geoland2 activities regarding the development of algorithms and associated products. The PROBA-V mission, launched in May 2013, constitutes a very important source of observations, not only because it ensures the continuity of VEGETATION sensors observations, but also because of its improved spatial resolution. The algorithm developed under the ImagineS project and run in operational processing chains provides to the scientific community as well as other stakeholders including policy makers, the proper information required for several applications. This has been detailed in the User Requirement Document (ImagineS_RP1.1).

The objective of this document is to provide a detailed description and justification of the algorithm proposed for version 1.0 of the algorithm based on daily TOA PROBA-V data at 300m resolution. Since no actual PROBA-V observations were available at the time of definition of this methodology, this ATBD heavily derives from the GEOV2/VGT algorithm (Baret et al, 2012) that was dedicated to provide historical and near real-time estimates of LAI, FAPAR and FCOVER variables from VEGETATION observations at 1/112° spatial resolution at a dekadal step.

1.3 CONTENT OF THE DOCUMENT

This ATBD document is split in 2 main sections:

1. **Algorithm overview.** This section contains:
   - A definition of the proposed products.
   - A brief description of the PROBA-V data from which the products are derived
   - The outline of the algorithm.
2. **Description of the algorithm**. This section contains:
   - The inputs required and outputs, including the quality indicators, provided by the algorithm.
   - The retrieval technique used. Neural network techniques constitute the core of the operational algorithm, completed with dedicated data compositing, filtering and smoothing processes.

The calibration of the neural networks is presented in details in Annex.

1.4 RELATED DOCUMENTS

1.4.1 Inputs

Overview of former deliverables acting as inputs to this document.

Document ID	Descriptor
ImagineS_RP1.1	Users Requirements Document
ImagineS_RP1.2	Service Specifications Document

1.4.2 Output

Overview of other deliverables for which this document is an input:

Document ID	Descriptor
ImagineS_RP5.1	Detailed Processing Model – GEOV3
ImagineS_RP5.3	Design Document – GEOV3
ImagineS_RP6.3	Product User Manual LAI, FAPAR, FCover
ImagineS_D6.4	300m LAI, FAPAR, FCover products
ImagineS_D6.6	PROBA-V LAI, FAPAR, FCover processing line
ImagineS_RP7.2	Scientific Validation Plan

1.4.3 External documents

Document ID	Descriptor
GIOGL1_ATBD_PROBA2VGT	Algorithm Theoretical Basis Document of the pre-processing module to convert PROBA-V data in VGT-like data

http://land.copernicus.eu/global/sites/default/files/products/GIOGL1_ATBD_PROBA2VGT_I1.10.pdf)

2 OVERVIEW

2.1 THE CONSIDERED PRODUCTS

The considered products correspond to actual vegetation biophysical variables that are defined below.

2.1.1 FAPAR

FAPAR corresponds to the fraction of photosynthetically active radiation absorbed by the canopy. The FAPAR value results directly from the radiative transfer in the canopy which is instantaneous. It depends on canopy structure, vegetation element optical properties and illumination conditions. FAPAR is very useful as input to a number of primary productivity models based on simple efficiency considerations (Prince, 1991) (McCallum et al., 2009). Most of the primary productivity models using this efficiency concept are running at the daily time step. Consequently, the product definition should correspond to the daily integrated FAPAR value that can be approached by computation of the clear sky daily integrated FAPAR values as well as the FAPAR value computed for diffuse conditions. To improve the consistency with other FAPAR products that are sometimes considering the instantaneous FAPAR value at the time of the satellite overpass under clear sky conditions (e.g. MODIS), a study investigated the differences between alternative FAPAR definitions (Baret et al., 2003). Results show that the instantaneous FAPAR value at 10:00 (or 14:00) solar time is very close to the daily integrated value under clear sky conditions. To keep a higher consistency with the FAPAR definition used in the CYCLOPES (Baret et al., 2007), and MODIS products, the instantaneous FAPAR value at 10:00 solar time under clear sky conditions (equivalent to black-sky conditions as defined also for albedo) was used.

FAPAR is relatively linearly related to reflectance values, and is little sensitive to scaling issues (Weiss et al., 2000; Hilker, 2010). Note also that the FAPAR refers only to the green parts of the canopy.

2.1.2 Cover fraction (FCover)

It corresponds to the gap fraction for nadir direction. FCOVER is used to separate vegetation and soil in energy balance processes, including temperature and evapotranspiration. It is computed from the leaf area index and other canopy structural variables and does not depend on variables such as the geometry of illumination as compared to FAPAR. For this reason, it is a very good candidate for the replacement of classical vegetation indices for the monitoring of green vegetation. Because of its quasi-linear relationship with reflectances, FCOVER is only marginally scale dependent (Weiss et al., 2000). Note that similarly to LAI and FAPAR, only the green elements are considered.

2.1.3 Leaf Area Index (LAI)

LAI is defined as half the developed area of photosynthetically active elements of the vegetation per unit horizontal ground area. It determines the size of the interface for exchange of energy (including radiation) and mass between the canopy and the atmosphere. This is an intrinsic canopy primary variable that should not depend on observation conditions. LAI is strongly non-linearly related to reflectance. Therefore, its estimation from remote sensing observations is scale dependent (Weiss et al., 2000; Garrigues et al., 2006). Note that vegetation LAI as estimated from remote sensing includes all the green contributors such as the understory when existing under forests canopies. However, except when using directional observations (Chen et al., 2005), LAI is not directly accessible from remote sensing observations due to the possible heterogeneity in leaf distribution within the canopy volume. Therefore, remote sensing observations are rather sensitive to the 'effective' leaf area index, i.e. the value that provides the same diffuse gap fraction while assuming a random distribution of leaves. The difference between the actual LAI and the effective LAI may be quantified by the clumping index (Chen et al., 2005) that roughly varies between 0.5 (very clumped canopies) and 1.0 (randomly distributed leaves). Note that similarly to the other variables, the retrieved LAI is mainly corresponding to the green elements: the correct term to be used would be GAI (Green Area Index) although we propose to still use LAI for the sake of simplicity.

2.2 PROBA-V INSTRUMENT AND DATA

The PROBA-V sensor has been launched in April 2013 onboard the PROBA platform (Figure 1). It weights around 160 kg with dimensions 0.77x0.73x0.84 m^3.

Figure 1: The PROBA platform with the PROBA-V instrument

The satellite flies at 820 km altitude on a circular helio-synchronous orbit with 11:00 equatorial crossing time. It provides a global observation on a daily basis for latitudes higher (lower) than 35° (-35°).

The instrument is viewing the surface under a 102.6° field of view, providing a swath of about 2500 km. An array of 6000x4 elements is used in the VIS-NIR (only 3 bands on the 4 potential ones are used) yielding to a ground sampling distance that varies across the swath from 100m up to 350m at the extremities of the swath (Figure 2, left). The SWIR domain is sampled using 3 arrays of 1024 elements, providing a ground sampling distance about twice as that in the VIS-NIR (Figure 2, right). This obviously poses a problem regarding the consistency of the radiometric information between the VIS-NIR and SWIR domains.

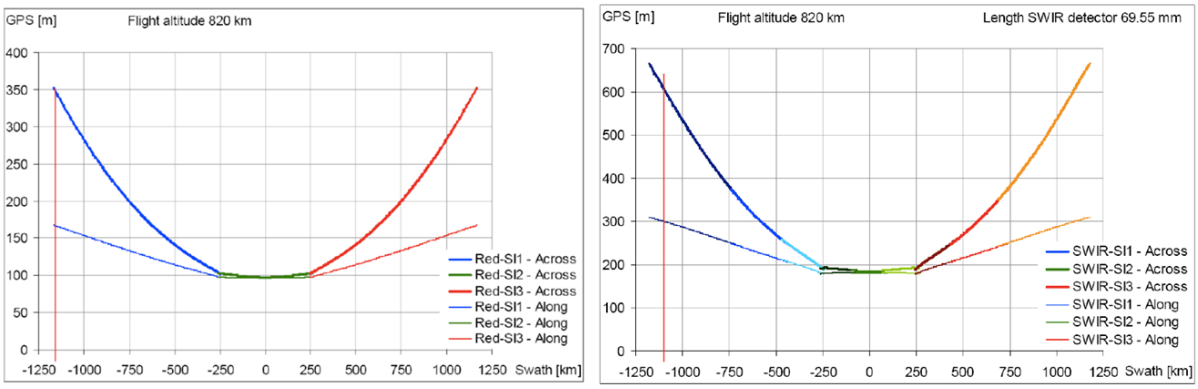

Figure 2: Ground sampling distance (GPS in m) as a function of the position on the swath (in km) for the VIS-NIR (left) and SWIR bands (right).

The bands are relatively close to those of the VEGETATION instruments, with however some slight shifts (particularly for the blue and SWIR bands) as shown in Table 1.

Acronym	Center (nm)	Width (nm)	Potential Applications
B0 (blue)	470 (440)	46 (40)	Atmosphere, clouds
B2 (red)	650 (645)	80 (70)	Vegetation
B3 (NIR)	837 (835)	120 (110)	Vegetation, water bodies
SWIR	1610 (1665)	80 (170)	Vegetation, water bodies

Table 1: PROBA-V spectral characteristics: band center and width. The characteristics of VEGETATION instrument are in parenthesis.

PROBA-V S1 products are daily images, corrected for system errors (error registration of the different channels, calibration of all the detectors along the line-array detectors for each spectral band) and resampled to the Plate Carrée projection (lat/lon, WGS84). The pixel brightness count is the ground area's apparent reflectance as seen at the top of atmosphere

(TOA) in the four PROBA-V bands (Table 2). Some auxiliary data are required to process the PROBA-V data, as described in paragraph 3.1.

PROBA-V planes	Description
B0	B0 spectral band, Radiometry data
B2	B2 spectral band, Radiometry data
B3	B3 spectral band, Radiometry data
SWIR	SWIR spectral band, Radiometry data
NDVI	Normalized Difference Vegetation Index data
QC	Quality Control Bit NR 7: Radiometric quality for B0 coded as 0 if bad and 1 if good Bit NR 6: Radiometric quality for B2 coded as 0 if bad and 1 if good Bit NR 5: Radiometric quality for B3 coded as 0 if bad and 1 if good Bit NR 4: Radiometric quality for MIR coded as 0 if bad and 1 if good Bit NR 3: land code 1 or water code 0 Bit NR 2: ice/snow code 1, code 0 if there is no ice/snow Bit NR 1: 0 0 1 1 Bit NR 0: 0 1 0 1 Clear Shadow Uncertain Cloud
VZA-VNIR	view zenith angles for Visible and Near Infra Red channels
VAA-VNIR	view azimuth angles for Visible and Near Infra Red channels
VZA-SWIR	View zenith angles for SWIR channel
VAA-SWIR	View azimuth angles for SWIR channel
SZA	sun zenith angles
SAA	sun azimuth angles
TIME	Observation timing information

Table 2: PROBA-V S1 data descriptor

The PROBA-V S1 dataset contains a set of attributes linked to the product or to one of its datasets:

- Generic information
  - (1) Product reference
  - (2) Platform
  - (3) Instrument
  - (4) Processing information (date, time)
  - (5) Map projection information (name, family, units, reference, ..)
  - (6) Synthesis period information
- Radiometric related information
  - (1) Reference to the used tools for geo-modelling, radio-modelling, mapping, mosaicking, cloud/ice/snow detection, shadow detection, compositing
- Geometric related information
  - (1) Top/bottom left/right point positioning
  - (2) Latitude and longitude for center point of the geographic bounding box
- Quality information related to the Status Map
  - (1) Percentage cloud
  - (2) Percentage snow_ice
  - (3) Percentage land
  - (4) Percentage missing data
- Time information related to the Time dataset
  - (1) Observation start and end date/time

2.3 REQUIREMENTS FOR THE ALGORITHM SELECTION AND DESIGN

The objective is to develop an algorithm dedicated to the estimation of LAI, FAPAR and FCOVER from the PROBA-V series of observations at 300m ground sampling distance. The algorithm should provide high level of consistency with GEOV2/VGT when aggregated at the 1 km resolution. These LAI, FAPAR and FCOVER products, called here GEOV3, should have the same temporal sampling frequency of 10 days. Products should also be associated with quality assessment flags as well as quantified uncertainties. The algorithm runs at the pixel level without interactions with the surrounding pixels. The algorithm should provide real time estimation. This forces to perform short term projection of the product dynamics. Note that there are two major changes according to the GEOV2/VGT algorithm:

- The SWIR band is not used since it has not the desired geometrical characteristics due to the resolution and ground spatial sampling distance which are roughly twice that of the VIS-NIR domain. The theoretical performances of the neural networks without SWIR input over a subset of Belmanip2.1 data are very close than what are obtained with the SWIR input (0.20 instead of 0.14 for LAI; 0.044 instead of 0.039 for FAPAR; 0.040 instead of 0.039 for FCover). The slight advantage of using the SWIR

band for the Belmanip2.1 sites, which are relatively homogeneous, was expected to degrade quickly for heterogeneous landscapes. Further, the angular configuration of observations is different from that of VIS and NIR bands.

- It is not possible to use the climatology as a background information since
  - The landcover at the 300m pixel scale may change significantly from one year to another for a significant fraction of the pixels.
  - The building of the climatology would require at least few years, or, alternatively, should be derived from other observations including the MERIS or MODIS time series that would require a large effort.

2.4 ALGORITHM OUTLINE

The scheme proposed for GEOV3 is similar to that used for GEOV2/VGT and is sketched in Figure 3. The algorithm starts from the daily PROBA-V S1 top of the atmosphere reflectance products which are first transformed into instantaneous estimates of LAI, FAPAR, FCover (Step A in Figure 3). Then, smoothing and gap filling is achieved over a compositing temporal window that may be dissymmetric as in the case of the near-real time situation or at the beginning of the time series (Step B in Figure 3).

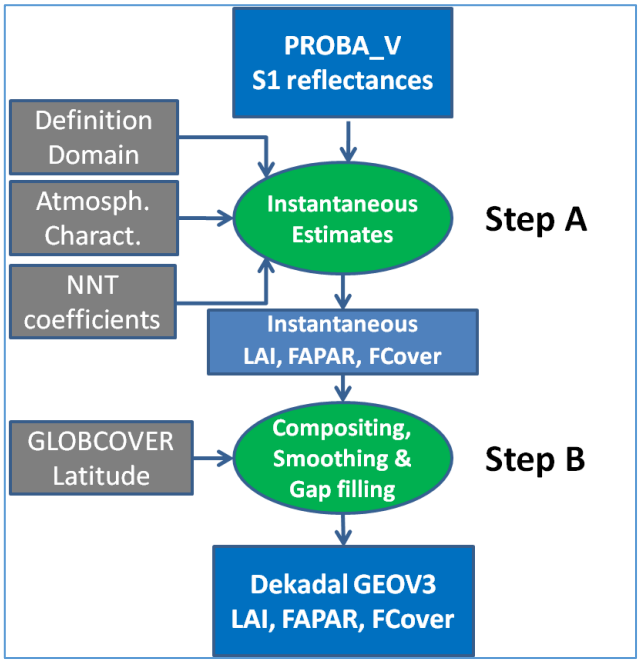

Figure 3: Flow chart showing the two processing steps

The compositing for real time estimates (Step B) is achieved according to the scheme described in Figure 4. We consider a particular dekadal date, *D*.

- In real time mode (when the actual date corresponds to *D*, i.e. the first line in Figure 4), the compositing is achieved using only the past observations, with a maximum compositing window spanning c_- days in the past. The value of the product at dekadal date *D* is therefore be relatively inaccurate.
- Then, the observations accumulate after dekadal date *D* when time passing. The value at dekadal date *D* is updated using the observations available after dekadal date *D* within the consolidation period. The accuracy of the product value progressively improves during this consolidation period.
- At the end of the consolidation period that lasts c_+ days, the value of the product value at dekadal date *D* has converged towards the 'historical' value, i.e. when there is at least c_- days (maximum length of the compositing window) after dekadal date *D*.

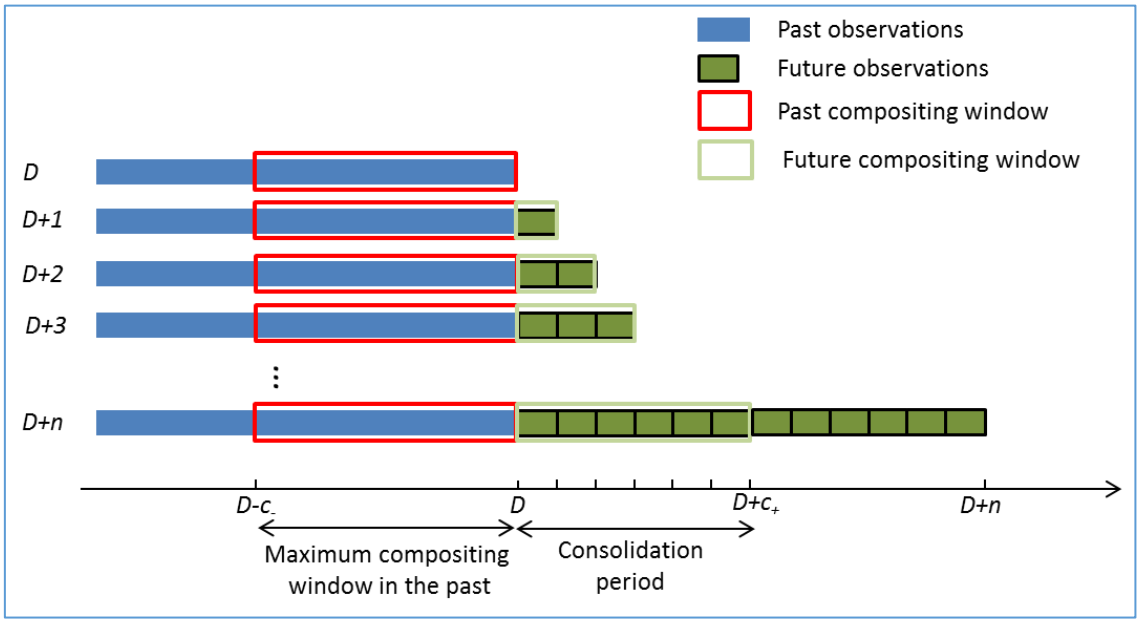

Figure 4: Scheme showing the compositing used for real time estimates

Note that when working in real time estimation and updating the values for each new dekad available, the compositing period is not symmetric: the maximum compositing window in the past is longer than the maximum compositing window in the future (after date *D*) that is limited by the length of the consolidation period. However, if the time series are reprocessed, the maximum length of the compositing period in the future should be equal to that in the past.

Once a new dekad is available after dekadal date D, the value of the product at dekadal date D is updated, and replaces the previous update. This allows saving significant space. However, it should be wise to save these intermediate values during a test phase to study the convergence during the consolidation phase.

The selection of the daily observations considered as valid in the compositing window (Step B) is made based on the LAI product. As a matter of fact, LAI is the most sensitive variables as compared to FCOVER and FAPAR.

3 ALGORITHM DESCRIPTION

In this section, the inputs and outputs are described, along with the quality flags considered. Then, the several steps of the algorithm are presented in details.

3.1 INPUTS

All these inputs are required for each considered pixel.

3.1.1 Top of Atmosphere daily reflectances

The daily synthesis (S1) of Top of Atmosphere reflectances in the three PROBA-V bands ($B0: blue$, $B2: red$, $B3: near\ infrared$) are required as inputs. They should be projected over the PROBA-V grid. Reflectances should be expressed in terms of reflectance factor, mainly varying between 0.0 and 0.7 for most land surfaces outside the hot-spot or the specular directions and cloud, snow or ice cover.

3.1.2 Geometry of acquisition

The geometry information is required as input to the atmospheric correction step as well as for the neural networks. It includes:

- the view zenith angle (VZA),
- the sun zenith angle (SZA),
- the relative azimuth angle ($SAA - VAA$)

3.1.3 Atmospheric conditions

Information on atmospheric conditions is provided as ancillary data. This information should be linked to the time-grid as included in the PROBA-V S1 Top Of Atmosphere products. It includes:

- the ozone content (OG)
    - global coverage low resolution climatology maps, one for each month with a reference date of 15th each month, represented by mean values over a period of 1 month.
    - a spatial resolution is 32/112 of a degree (0.285714286°) in the plate-carrée projection and with pixel coordinates always taken at the center of the pixel.
    - *alternative source*:
        - total column ozone (in Dobson) provided by ECMWF (TCO3)
        - regular grid structure (Gaussian N400 grid) with a resolution of 0.25° (in both longitude and latitude direction) covering the complete globe (90°N to 89.75°S and 180°W to 179.75°E) provided in FM-92 GRIB file format.

        - datasets are available with 6-hour intervals (00:00 UTC, 06:00 UTC, 12:00 UTC and 18:00 UTC times)
- the water content (WVG)
    - the vertically integrated water vapour from the ground level up to a level of 150 hPa, also called TCWV (Total Column of Water Vapour) and provided by MeteoServices in an ASCII file, expressed in g/m^2, ranging from 0 to 70000 with value -99999 as no data.
    - regular grid structure with a resolution of 0.25° (in both longitude and latitude direction) covering a limited region (75°N to 57°S and 180°W to 179.75°E). The data contains only significant values for the land pixels.
    - datasets are available with 6-hour intervals (00:00 UTC, 06:00 UTC, 12:00 UTC and 18:00 UTC times)
    - *fallback scenario*: a degraded file is created based on the most recent nominal data. For example, in case there is no nominal data available at 18:00, the 12:00 nominal file is used with a time step of 6 hours to create the degraded file.
    - *alternative source*: ECMWF WVG provided in a regular grid structure (Gaussian N400) with a resolution of 0.25° covering the complete globe (90°N to 89.75°S and 180°W to 179.75°E), in NETCDF format with the same 6-hour intervals expressed in unit kg/m^2.
- surface pressure derived
    - from the altitude of the pixel derived from a DEM at 90m spatial resolution. The equation used to derive pressure from the DEM is :

$$Pressure = 1013.25 * \left(1 - 0.0065\frac{z}{288.15}\right)^{5.255}$$

      This equation corresponds to the simple international barometric height formula for the standard atmosphere where a surface pressure is assumed to be 1013.25hPa, the surface temperature is 15°C and the temperature gradient is 0.65°C per 100 meter (Doerffer et al, 2011).

    - The Global Land Survey Digital Eleveation Model (GLSDEM) is used (http://www.glcf.umd.edu/data/glsdem/description.shtml). The GLSDEM is available in degree tiles with a resolution of Geographic 3 arc second (~90 m) ranging from 83°N to 56°S, and have been retiled to create 5° by 5° tiles.
    - *alternative source*: ECMWF surface pressure in a regular grid structure (Gaussian N400) with a resolution of 0.25° in both directions covering the complete globe (90°N to 89.75°S and 180°W to 179.75°E), in NETCDF format with a 6-hour intervals expressed in unit Pa.

The input data are interpolated in space to the resolution of the PROBA-V input image using bilinear interpolation, and if applicable the nearest input in time.

3.1.4 Algorithmic parameters

The two main steps (Step A and Step B) of the algorithm as showed in Figure 3 use a series of parameters listed in Table 3 and Table 4, respectively. Their roles and usages are further described in the various paragraphs of Section 3.3. The parameters used in the calibration of the neural networks and the elaboration of the definition domain are given in Annex: Neural Network Calibration.

Parameter	Descriptive Name	Type	Range	Reference
$P^{tol}_{min\_LAI}$	Tolerance limit minimum for LAI	Float	-0.5-0 (-0.2)	3.3.1.6
$P^{tol}_{max\_LAI}$	Tolerance limit maximum for LAI	Float	7-10 (10)	3.3.1.6
$P^{tol}_{min\_FAPAR}$	Tolerance limit minimum for FAPAR	Float	-0.2-0 (-0.1)	3.3.1.6
$P^{tol}_{max\_FAPAR}$	Tolerance limit maximum for FAPAR		0.94-1.14 (1.04)	3.3.1.6
$P^{tol}_{min\_FCOVER}$	Tolerance limit minimum for FCOVER	Float	-0.2-0 (-0.1)	3.3.1.6
$P^{tol}_{max\_FCOVER}$	Tolerance limit maximum for FCOVER	Float	1-1.2 (1.1)	3.3.1.6
α_{B_0}	Slope of the spectral conversion for B_0 input reflectance	Float	0.997121	3.3.1.2
α_{B_2}	Slope of the spectral conversion for B_2 input reflectance	Float	0.998302	3.3.1.2
α_{B_3}	Slope of the spectral conversion for B_3 input reflectance	Float	1.000472	3.3.1.2
β_{B_0}	Offset of the spectral conversion for B_0 input reflectance	Float	0.00344	3.3.1.2
β_{B_2}	Offset of the spectral conversion for B_2 input reflectance	Float	0.002937	3.3.1.2
β_{B_3}	Offset of the spectral conversion for B_3 input reflectance	Float	0.00236	3.3.1.2

Table 3: The algorithmic parameters used in Step A. Values between brackets are used in the processing chain. The slope and offset of the spectral conversion have been established by VITO in the framework of the Copernicus Global Land Service [GIOGL1_ATBD_PROBA2VGT].

Parameter	Descriptive Name	Type	Units	Range	Step
$length_{EBF}^{\max\_bef}$	Maximum length of the half compositing window for EBF case	Int.	days	180-720 (210)	3.3.2.1.1
$length_{EBF}^{\max\_aft}$	Maximum length of the half compositing window for EBF case	Int	days	20-120 (60)	
N_{EBF}^{min}	Number of valid observations required to compute LAI for EBF case	Int.	-	5-20 (20)	3.3.2.1.2
$f_x(N_{tot})$	Frequency used to consider an observation as valid	Float	%	50-99 (90)	3.3.2.1.2
N_{EBF}^{max}	Number of decads used to compute the fraction of EBF cases	Int.	dekad	18-72 (36)	3.3.2.1.3
Lat_{EBF}^{max}	Latitude max at which EBF can be found (absolute value)	Float.	°	20-35 (28.5)	3.3.2.1.3
f_{EBF}^{min}	Ratio of number of dates at which the pixel is identified as EBF over the N_{EBF}^{max} decads	Float	-	0.1-0.5 (0.8)	3.3.2.1.3
LAI_{EBF}^{min}	Threshold LAI value required to detect EBF (EBF should have LAI> LAI_{EBF}^{min}	Float	-	2-6 (4)	3.3.2.1.3
$Diff_{EBF}^{thres}$	Used to apply a threshold on the difference δ_{LAI_Valid} between consecutive valid observations to detect EBF	Float	-	0.5-3 (0.9)	3.3.2.1.3
$Percent_{EBF}$	Value of the percentile of difference between consecutive observations used to detect EBF	Float	%	50-90 (80)	3.3.2.1.3
$N_{outlier}^{min}$	Minimum number of observations required for outlier identification	Int	-	3-10 (5)	3.3.2.2.1
$tol_{outlier}^{abs}$	Value of outlier threshold (absolute value). used to detect outliers	Float	-	0.05-0.5 (0.1)	3.3.2.2.1
$tol_{outlier}^{rel}$	Value of outlier threshold (relative value) used to detect outliers	Float	-	0.2-1.0 (0.6)	3.3.2.2.1
$length_{no\_EBF}^{outlierabs}$	Length of the half window used for outlier rejection for no_EBF case	Int.	days	10-25 (20)	3.3.2.2.1
$length_{no\_EBF}^{min}$	Minimum length of the half compositing window for no_EBF case	Int.	days	10-60 (20)	3.3.2.2.2
$length_{no\_EBF}^{max}$	Maximum length of the half compositing window for no_EBF case	Int.	days	20-120 (60)	3.3.2.2.2

$N_{no\_EBF}^{max}$	Maximum number of valid observations for non_EBF case in each half window	Int.	-	5-20 (10)	3.3.2.2.2
N_{linear}	Minimum number of valid observations for degree2 polynomial fit, non EBF	Int.	-	3-7 (5)	3.3.2.2.3
k	value of coefficient k in the weighing of observations for the polynomial fitting for non EBF case	Float	-	1-20 (2)	3.3.2.2.3
N_{miss}	Minimum number of observations for non_EBF case	Int	-	2-7 (3)	3.3.2.2.3
N_{gap}^{max}	Length in decades of period with missing data that can be filled	Int.	dekad	3-10 (6)	3.3.2.2.3
$\Delta_{no\_EBF}^{max}$	maximum distance between the date of the dekad and the nearest observations required to compute the product in the no EBF case	Int.	days	5-30 (15)	3.3.2.2.3
$length_{no\_EBF}^{interp}$	Length of the window used to interpolate between the 2 nearest data before and after decade when no_EBF algorithm fails	Int	days	10-20 (15)	3.3.2.2.3
$length_{no\_EBF}^{nearest}$	Length of the window used to select the nearest data before or after decade when no_EBF algorithm fails and interpolation is not possible	Iint	days	5-10 (5)	3.3.2.2.3
tol_{FAPAR}^{min}	Tolerance minimum on FAPAR used to reject estimated values outside the expected range of variation	Float	-	-0.2-0 (-0.1)	3.3.2.2.4
tol_{FCOVER}^{min}	Tolerance minimum on FCOVER used to reject estimated values outside the expected range of variation	Float	-	-0.2-0 (-0.1)	3.3.2.2.4
tol_{LAI}^{min}	Tolerance minimum on LAI used to reject estimated values outside the expected range of variation	Float	-	-0.5-0 (-0.2)	3.3.2.2.4
δ_{LAI}	Threshold value for the confidence interval on the estimated value	Float	-	0.2-1.0 (0.5)	3.3.2.2.4
tol_{LAI}^{max}	Tolerance maximum on LAI used to reject estimated values outside the expected range of variation	Float	-	7-10 (10)	3.3.2.2.4
tol_{FAPAR}^{max}	Tolerance maximum on FAPAR used to reject estimated values outside the expected range of variation	Float	-	0.94-1.14 (1.04)	3.3.2.2.4

tol_{FCOVER}^{max}	Tolerance maximum on FCOVER used to reject estimated values outside the expected range of variation	Float	-	1-1.2 (1.1)	3.3.2.2.4

Table 4: The algorithmic parameters for Step B. Values between brackets are used in the processing chain.

3.2 OUTPUTS

Three types of outputs are expected:

- The dekadal values of LAI, FAPAR and FCOVER
- Quantitative quality assessment (QA) indicators of the products
- Qualitative quality indicators (QC)

3.2.1 The LAI, FAPAR, FCover products

The outputs are computed by application of the algorithm over each pixel at each dekadal date. They include the LAI, FAPAR and FCOVER values as described previously. The range of variation and resolution are presented in Table 5. Note that the same conventions as for the previous versions of products, known as GEOV1 and GEOV2, are used here.

P	Products	Physical Minimum	Physical Maximum	Max DN value	Scaling factor	Offset
1	LAI	0.0	7.0	210	30	0
2	FAPAR	0.0	0.94	235	250	0
3	FCOVER	0.0	1.0	250	250	0

Table 5: Minimum, maximum values and associated resolution for LAI, FAPAR and FCOVER products.

3.2.2 Quality Indicators

In addition to the LAI, FAPAR and FCover values, quantitative (QA) and qualitative (QC) quality indicators are also generated. They are listed in Table 6 and Table 7, respectively.

QA	Quantitative quality indicator	Physical Minimum	Physical Maximum	Max DN value	Scaling factor	Offset	Missing Value
1	Number of available valid instantaneous estimates in the compositing window	0.0	40	40	1	0	-
2	Length in days of semi-period before D	15	210	210	1	0	-
3	Length in days of semi-period after D	0	60	60	1	0	-
4	Uncertainty of LAI products	0.0	7.0	210	30	0	255
5	Uncertainty of FAPAR products	0.0	0.94	235	250	0	255
6	Uncertainty of FCOVER products	0.0	1.0	250	250	0	255

Table 6: Minimum, maximum values and associated resolution for the quantitative quality indicators of LAI, FAPAR and FCOVER. D means dekadal date.

QC	Qualitative quality indicator	Meaning	Reference
1	Land/Water	QC(1)=0: Water (heritage from PROBA-V QC) QC(1)=1: Land	Table 2
2	Vegetation class identified	QC(2)=0: no_EBF case QC(2)=1: EBF case	3.3.2.1.3
3	Missing dekad interpolated	QC(3)=0: Not interpolated QC(3)=1: Interpolated	3.3.2.3
4	Not used	Not used	
5	Method used for product computation EBF case	EBF case: QC(5)=0: Based on daily observations QC(5)=1: Based on previous dekadal product	3.3.2.1.2
6:7	Method used for product computation Non-EBF case	No_EBF case: QC(6:7)=00: Second degree polynomials fit QC(6:7)=01: Linear fit QC(6:7)=10: Interpolation between the two nearest dates within $length_{no\_EBF}^{interp}$ days QC(6:7)=11: Nearest data within $length_{no\_EBF}^{nearest}$ days or missing value (in this case the product value is set to missing value)	3.3.2.2.3
8	Instantaneous EBF classification flag	QC(8)=0: no_EBF case instantaneous value QC(8)=1: EBF case instantaneous value	3.3.2.1.3

Table 7: Qualitative quality indicators

3.3 DETAILED DESCRIPTION

As shown on Figure 3, the algorithm relies on 2 main steps:

- Step A: instantaneous estimates of LAI, FAPAR, FCover (Figure 5)
- Step B: compositing, smoothing and gap filling (Figure 6)

3.3.1 Instantaneous LAI, FAPAR, FCover estimates (Step A)

It yields a first estimate of instantaneous products from the PROBA-V S1 Top-of-Atmosphere reflectances (Figure 5).

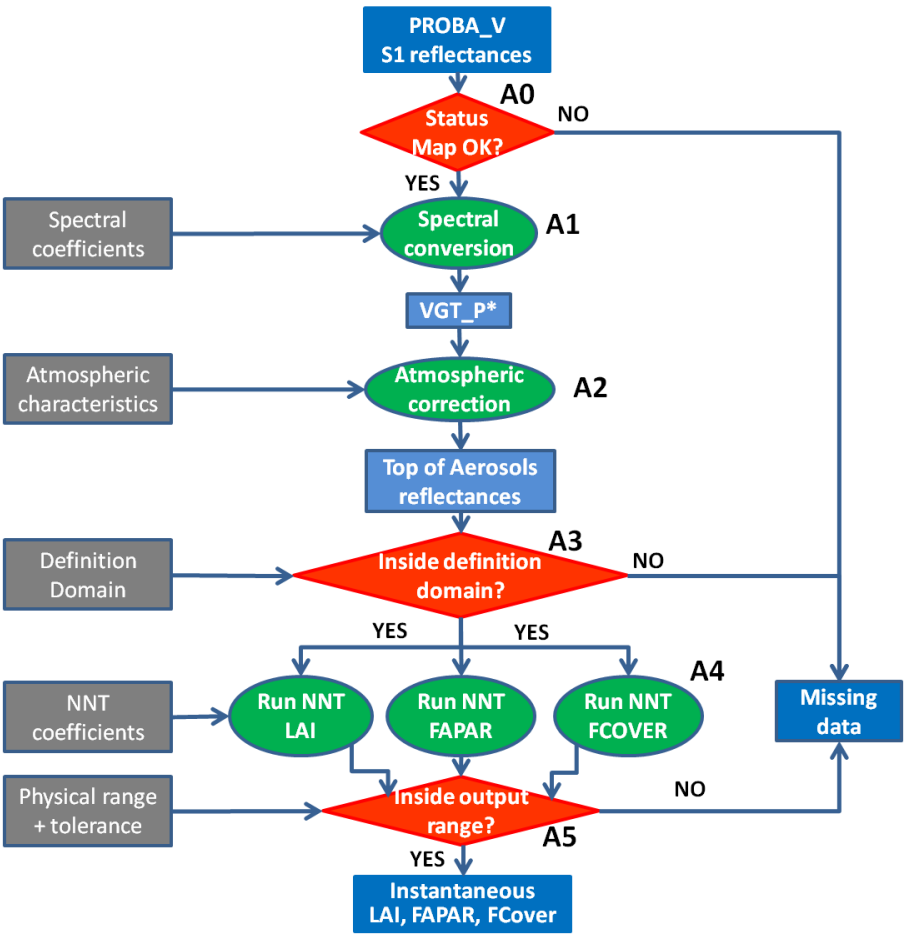

Figure 5: Flow chart describing the instantaneous product estimation (Step A)

3.3.1.1 Rejection of input data based upon their quality status (Step A0)

The PROBA-V QC plane (Table 2) is first used to keep only valid pixels i.e. clear pixels with all the 3 bands have good radiometric quality, located over land, not covered by ice, cloud or snow.

3.3.1.2 Spectral conversion of TOA reflectances (Step A1)

Since there were any existing PROBA-V data available at the moment of definition of this methodology to be used for training the neural networks, the training was based on SPOT/VEGETATION 1km data similarly to what was done for GEOV2/VGT (See Annex). However, only the 3 VIS-NIR bands were used since the SWIR band of PROBA-V sensor is associated to a degraded spatial resolution.

Because the spectral characteristics of PROBA-V sensor are slightly different from those of VEGETATION (Table 1), a spectral conversion was applied on the actual PROBA-V TOA reflectances as:

$$\widehat{\rho_{VGT(Bx)}} = \alpha_{Bx}.\rho_{PROBA}(Bx) + \beta_{Bx} \quad [1]$$

where $\widehat{\rho_{VGT(Bx)}}$ is the converted PROBA-V TOA reflectance (called VGT_P* in Figure 5) and $\rho_{PROBA}(Bx)$ is the TOA PROBA-V reflectance and α_{Bx} is the conversion coefficient for band Bx.

	B0	B2	B3
α_{Bx}	0.997121	0.998302	1.000472
β_{Bx}	0.00344	0.002937	0.00236

Table 8: Spectral conversion coefficients (α_{Bx} and β_{Bx}) between VEGETATION and PROBA-V sensors [GIOGL1_ATBD_PROBA2VGT]

3.3.1.3 Atmospheric correction (Step A2)

The atmospheric correction is achieved using the SMAC4.0 model (Rahman and Dedieu, 1994). It uses as inputs:

- The top of atmosphere reflectance in the three bands $B0$, $B2$ and $B3$.
- The ozone content (see description in 3.1.3)

- The water vapor content (see description in 3.1.3)
- The surface pressure (derived from the pixel altitude) (see description in 3.1.3)
- Aerosol optical thickness at 550nm is set to zero (no aerosol)
- the view zenith angle (VZA),
- the sun zenith angle (SZA),
- the relative azimuth angle ($SAA - VAA$),

The output are the top of aerosol reflectances in the three bands. Note that the coupling between aerosol scattering and Rayleigh scattering as well as gas absorption is here neglected. Simulations show that this coupling term is relatively weak except in the $B0$ band (Figure 6).

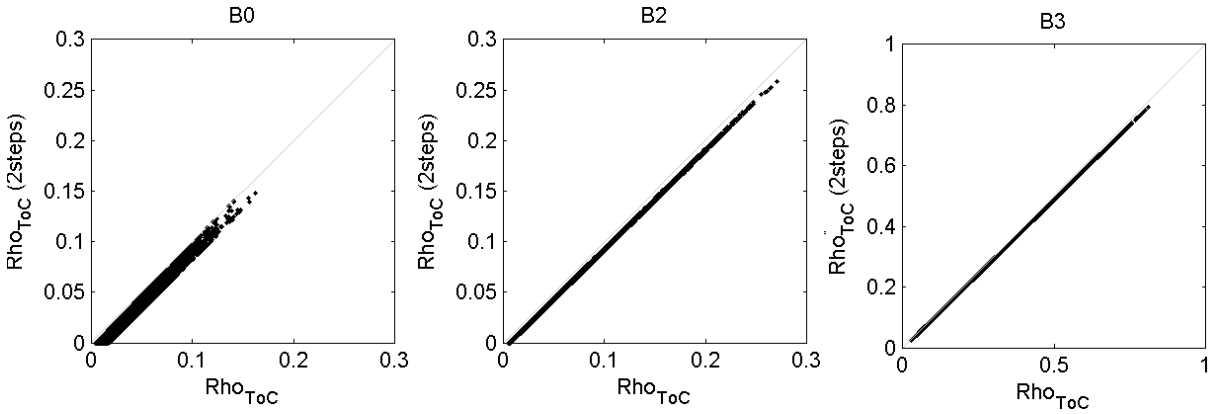

Figure 6: Comparison between the top of canopy reflectance derived from atmospheric correction achieved in one single step (Rho_{TOC}, i.e. running SMAC using the known atmospheric characteristics) and the value derived in two steps (Rho^*_{TOC}(2steps), i.e. running SMAC using the known atmospheric characteristics but with no aerosol, then running SMAC a second time using the actual aerosol characteristics but with pressure, ozone and water vapor set to 0).

3.3.1.4 First outlier rejection (Step A3)

The top of aerosols reflectances outside the definition domain as described in Annex (Section 3) are removed. The corresponding pixels are considered outliers. This should allow rejecting cloud/snow/water contaminated values.

3.3.1.5 Deriving instantaneous estimates using neural networks (Step A4)

One specific neural network was previously calibrated for each of the 3 variables considered (LAI, FAPAR, and FCover) (see Annex). They are then applied to each individual observation (one pixel at a given date) using the top-of-aerosols reflectances in the 3 VIS-NIR bands and to 3 angles characterizing the sun and view directions of the observation to estimate the corresponding instantaneous LAI, FAPAR and FCover values.

3.3.1.6 Output outlier rejection (Step A5)

The instantaneous values lying outside the physical range of variation of the variables extended by the tolerance limits (Table 9) are rejected. Three cases of the variable values are possible:

1. The value is within expected physical range of variation (Table 5): It is considered valid.
2. The value is within the tolerance limits (Table 9) but higher (lower) than the physical maximum (minimum) (Table 5). The value is considered valid but set to the physical maximum (minimum).
3. The value is outside the tolerance limits: it is considered invalid.

If one of the variables is considered invalid, all the three variables are rejected. This allows keeping a high level of consistency between the three products (LAI, FAPAR and FCover).

	LAI	FAPAR	FCOVER
P^{tol}_{min}	-0.20	-0.1	-0.1
P^{tol}_{max}	10.0	1.04	1.1
Physical Minimum	0	0	0
Physical Maximum	7	0.94	1

Table 9: Physical range of variation and tolerance limits (Minimum (P^{tol}_{min}) and maximum (P^{tol}_{max})) used to reject the output.

3.3.2 Compositing, smoothing and gap filling (Step B)

The compositing, filtering and gap filling module is run for each dekadal date *D*. Two cases are considered: Evergreen Broadleaf Forest (EBF) or not EBF (no_EBF). Each pixel is first processed as an EBF. This provides the required criterions for deciding whether the pixel is actually an EBF or not. In the latter case, the no_EBF algorithm is run. Finally, if the previous dekadal product is missing, a gap filling algorithm is run to fill the preceding small gaps. This is sketched in Figure 7.

Note that similarly to Step A, the algorithm produces in parallel the three variables (LAI, FAPAR and FCover) since the final dekadal products are strongly inter-related.

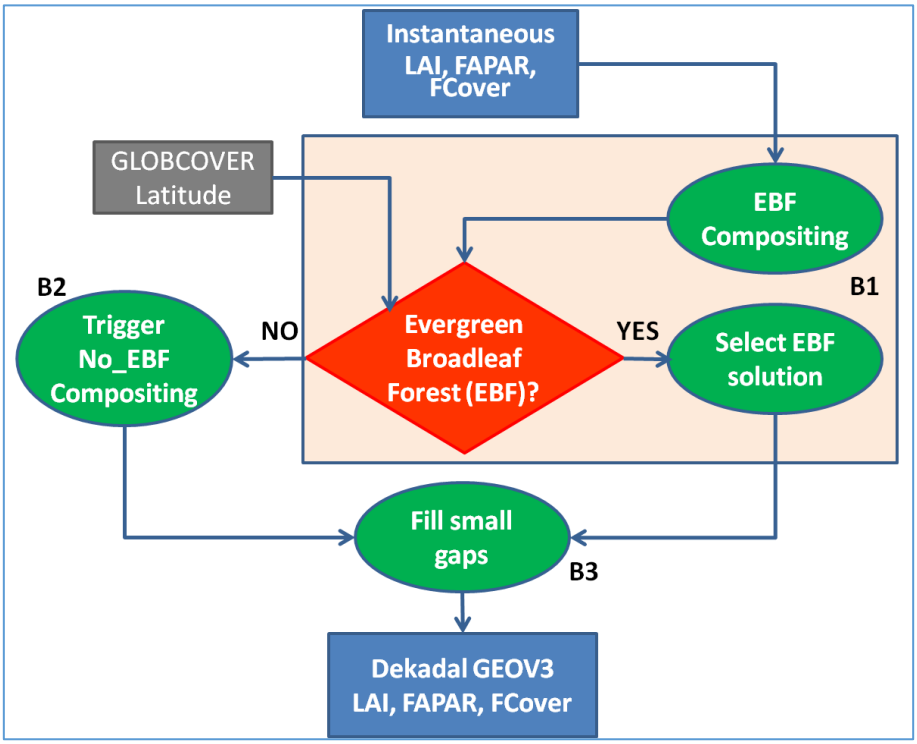

Figure 7: Flow chart describing Step B: compositing, smoothing and gap filling.

3.3.2.1 Run Evergreen Broadleaf Forest case (Step B1)

Evergreen Broadleaf Forests are characterized by:

- A geographic location within the tropical area
- A limited seasonality
- A large amount of cloud contaminated observations.

The processing of such situations needs to take into account explicitly these peculiarities. The algorithm includes 3 steps for each dekadal date:

1. Adjusting the length of the compositing window (Step B1.1)
2. Computing the variables estimate for the dekadal date (Step B1.2)
3. Deciding whether the pixel belongs actually to the EBF or no_EBF class (Step B1.3)

3.3.2.1.1 Adjusting the length of the compositing window (Step B1.1)

First, a large compositing window is considered. It is defined as the maximum length (in days) of the period before the decadal date ($length_{EBF}^{\max\_bef}$), and the maximum length of the period after this date ($length_{EBF}^{\max\_aft}$). $length_{EBF}^{\max\_aft}$ is lower than $length_{EBF}^{\max\_bef}$ to prevent from having a too long consolidation period while being consistent between historical processing and near real time situations. We therefore chose the same value as for the non evergreen broadleaf forest case (section 3.3.2.2). $length_{EBF}^{\max\_bef}$ is chosen in order to avoid missing values for the first dates of the time series. As default values, we propose to use $length_{EBF}^{\max\_bef} = 210\ days$ and $length_{EBF}^{\max\_aft} = 60\ days$.

Then, the N_{max} closest available observations to dekadal date D are selected, with $N_{max} = 20$ as default value.

3.3.2.1.2 Computing the decadal estimates (Step B1.2)

Two cases are considered depending on the available number of observations, $N_{tot} = N_{before} + N_{after}$ in the compositing window (Figure 8):

1. ***If $N_{tot} \geq N_{EBF}^{min}$, computing the mean of 'valid' daily observations.*** If the number of available observations is larger than the minimum allowed number of observations, N_{EBF}^{min}, a frequential filter is first applied to discard the low values corresponding to cloud contamination: date t of a daily observation is considered valid if $LAI(t) \geq f_x(N_{tot})$, where $f_x(N_{tot})$ is the percentile x value computed on the available total daily observations over the compositing window, N_{tot}. Once enough PROBAV data will be available, the value of x could be fine-tuned, but as a default value, x was set to $x = 90\%$. Then, the mean computed over the valid observations and assigned to the dekadal date D:

$$If\ N_{tot} \geq N_{EBF}^{min}, P(D) = mean(P(t), with\ LAI(t) \geq f_{90}(N_{tot}))$$

2. ***If $N_{tot} < N_{EBF}^{min}$*** within the compositing window, the product value of dekadal date D is set to the product value corresponding to the previous dekad. If there was missing data for the previous dekad, then a missing value is set to the dekadal date D.

$$If\ N_{tot} < N_{EBF}^{min}, P(D) = P(D-1)$$

A quality indicator (QC(5) in Table 7) specifies if the product value is actually computed based on the daily observations considered for dekadal date D (case 1) or if it is estimated using the previous dekadal value (case 2).

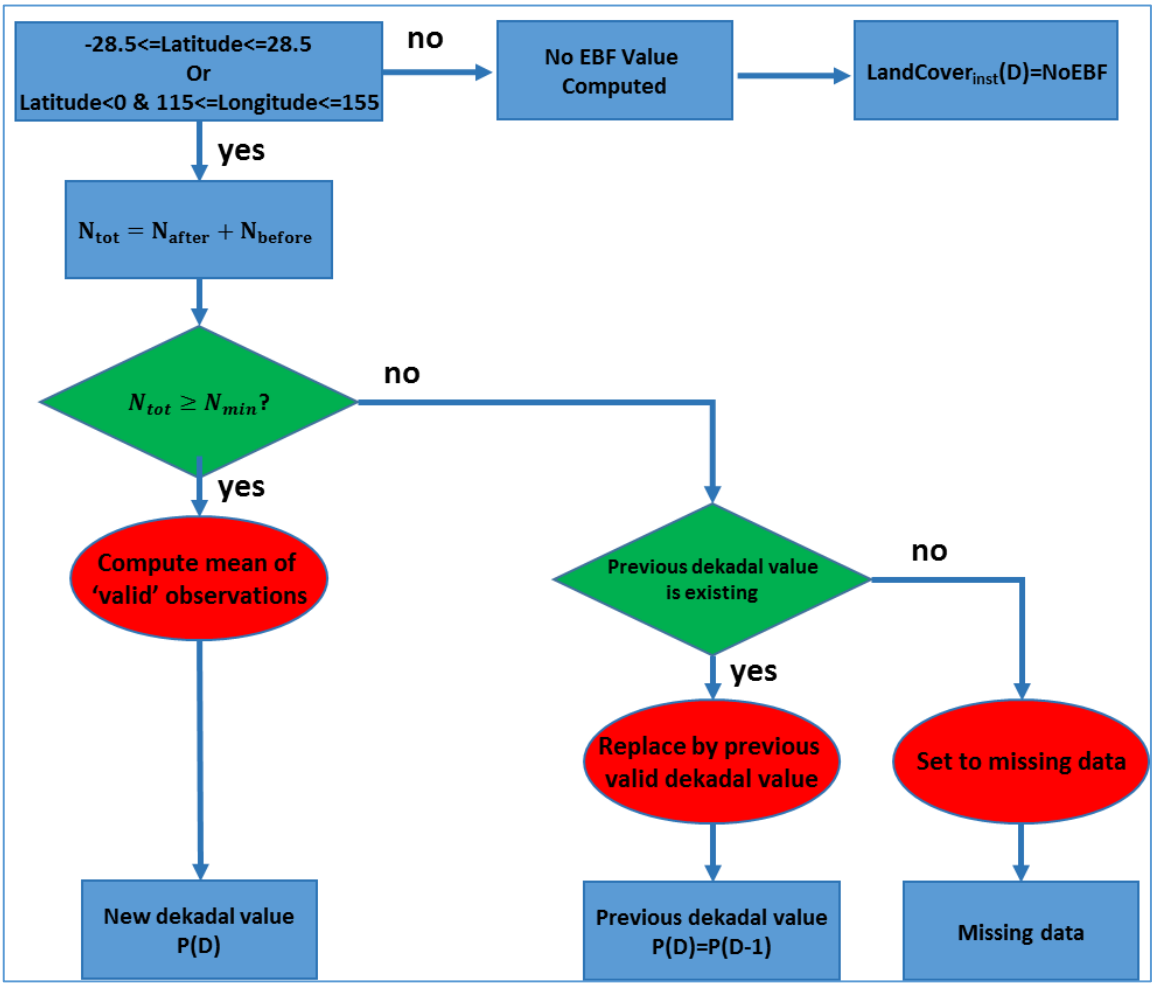

Figure 8: Flow chart showing the different cases for smoothing, and gap filling at the decadal date D for Evergreen Broadleaf Forest situations. D indicates the current decade at which the product is computed (D-1 being the previous decade).

3.3.2.1.3 Checking the landcover (Step B1.3)

The use of an ancillary classification, such as GLOBCOVER, was a possible solution. However, it poses at least three problems:

1. The grid used for GLOBCOVER (1/360°) is different from that used for PROBA-V (1/336°).

2. The classification map is fixed, although some pixels may change due to deforestation or reforestation,
3. The classification map may incorporate some uncertainties

For this reason, we preferred using the PROBA-V observations directly to identify the EBF cases when possible. However, at the beginning of the time series, when very few observations are available, or when the data are either too few or too scattered, the GLOBCOVER classification will be used pertinently. It should be projected over the same grid as that used for PROBA-V using the closest neighbor interpolation

We first estimate an instantaneous land cover class $LC_D(D)$ of a given pixel as an evergreen broadleaf forest or not using three criterions that must be simultaneously fulfilled (Figure 9Figure 8):

1. The pixel should be at a latitude between -Lat_{EBF}^{max} and $Lat_{EBF}^{max} = 28.5°$ being the default value and corresponding to the range where most of EBFs are expected. As some EBF can also be found in Australia, a second test is performed on both latitude (<0) and longitude (between 115°E and 155°E). This criterion prevents from problems observed over needleleaf forests with high LAI values and a high probability of cloud cover inducing strong scattering between observations.
2. The estimated LAI value should be higher than LAI_{EBF}^{min} with default value $LAI_{EBF}^{min} = 4.0$. As a matter of fact, EBF are characterized by high LAI values.
3. There is a significant level of noise in the daily LAI temporal profile. To evaluate the level of noise, the distribution of the absolute value of the difference between two consecutive valid observations (δ_{LAI_Valid}) is computed over the temporal window used for EBF case. Indeed, high differences between two close dates correspond to situations with high probability of cloud contamination and induce scattering in the data. The LAI value corresponding to the percentile, $Percent_{EBF}$, (with default value $Percent_{EBF} = 80\%$) is also computed. Then, if δ_{LAI_Valid} is higher than the threshold value $Diff_{EBF}^{thres}$ *$LAI(Percent_{EBF}$), $Diff_{EBF}^{thres}$ = 0.9 being the default value, and condition 1 and 2 are fulfilled, the pixel is classified as an EBF.

However, because of the generally high cloud cover and contamination, the instantaneous estimates of landcover derived from the 3 previous criterions are regularized using the estimates from the previous dekads. As a matter of fact, the landcover class of a pixel is not likely to change often with time. The previous period of $(N_{EBF}^{max} - 1)$ dekads is considered. The frequency of dekads assigned to EBF in the previous period of N dekads is computed (f_{EBF}, Figure 10**Erreur ! Source du renvoi introuvable.**).

- If the frequency is:
  - $f_{EBF} \leq 1 - f_{EBF}^{min}$ then the dekad is considered as non_EBF
  - $f_{EBF} \geq f_{EBF}^{min}$ then the dekad is considered as EBF.

- If this frequency is outside the range $1 - f_{EBF}^{min} < f_{EBF} < f_{EBF}^{min}$, the estimates from the observations are considered unreliable and the GLOBCOVER classification is used. The default value of N_{EBF}^{max} was set to $N_{EBF}^{max} = 36\, dekads$ (≈360 days) and f_{EBF}^{min} was set to $f_{EBF}^{min} = 0.8$.

Note that, at the beginning of the time series, when $N_{available} < N_{EBF}^{max}$, the GLOBCOVER classification is used to fill the $N_{EBF}^{max} - N_{available}$ missing dekads.

One flag (QC(2) in Table 7) is then set to report the landcover $(LandCover_{est}(D))$ that is finally assigned, when considering the f_{EBF} value computed over the *N* dekads. Another flag (QC(8) in Table 7) corresponds to $LC_D(D)$: QC(8)=1 if the site is detected as EBF.

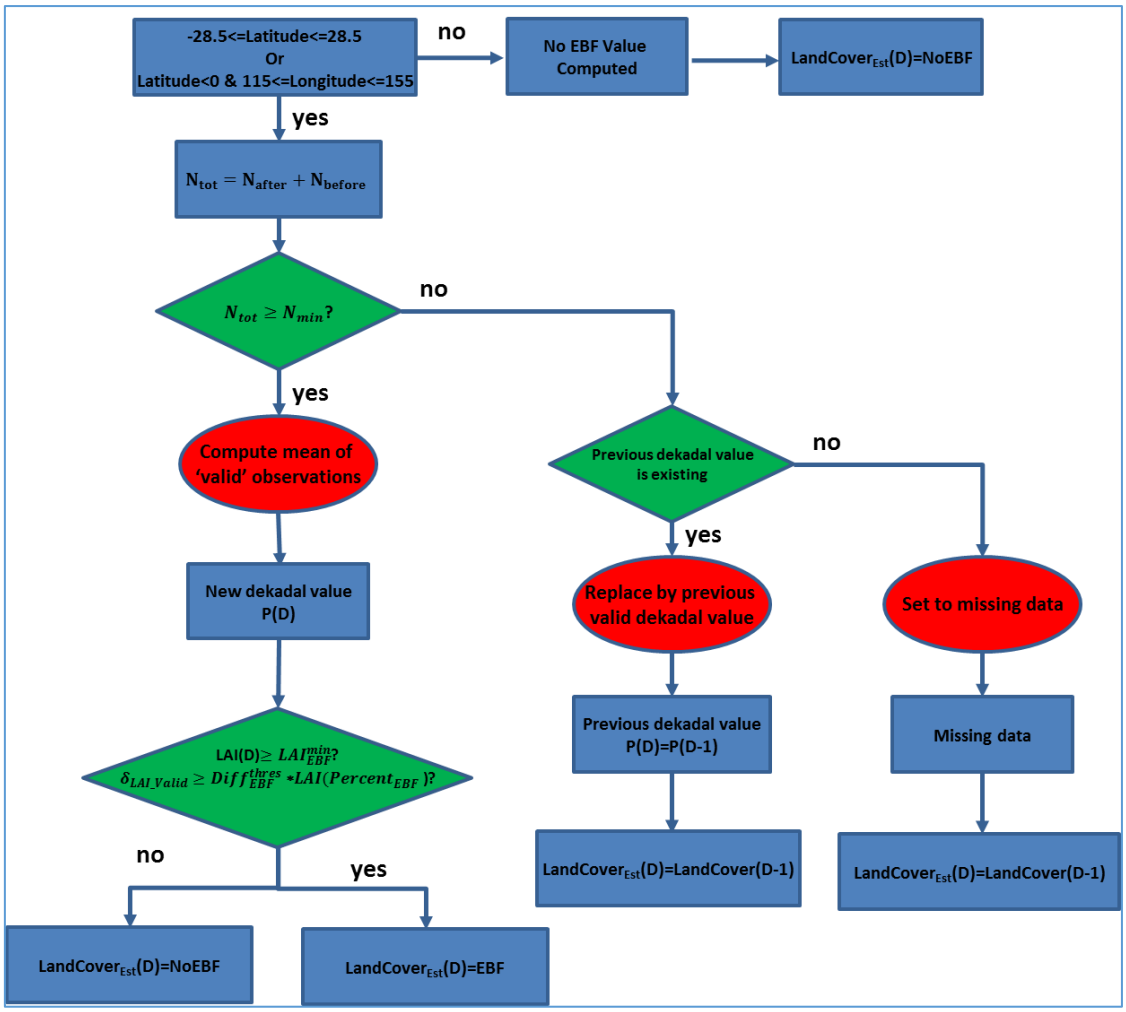

Figure 9. Flow chart showing the land cover class identification after applying the EBF compositing algorithm. The computation of f_{EBF} is illustrated in Figure 10 Erreur ! Source du

renvoi introuvable.**. D indicates the current decade at which the product is computed (D-1 being the previous decade).**

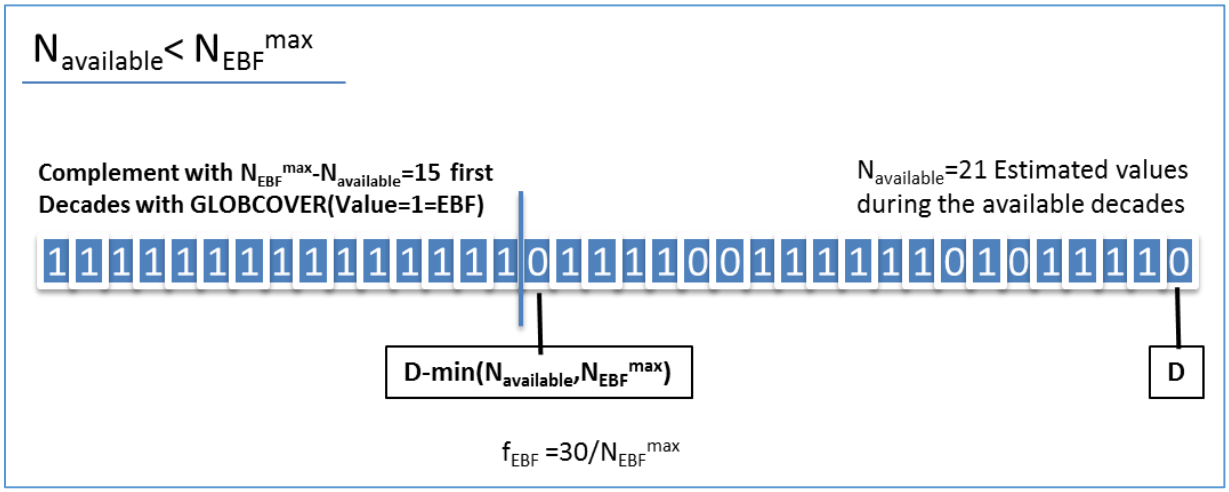

Figure 10: Example of landcover class determination criterion: only the 21st decades are available. The instantaneous land cover at decadal date D. $LC_D(D)$ is 0 (Non EBF). The final class $(LandCover_{Est}(D))$ attributed to this pixel is EBF ($f_{EBF} = \frac{30}{36} > f_{EBF}^{min}$)

3.3.2.1.4 Computing the uncertainties for EBF (Step B1.4)

- If $N_{tot} \geq N_{EBF}^{min}$. The uncertainties are computed as the Root Mean Square Error (RMSE) between the daily valid observations of the product $P(t)$ and the product value at decade D, $P(D)$:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n_{valid}} (P(t) - P(D))^2}{n_{valid}}}$$

  Where n_{valid} is the number of valid observations used to compute $P(D)$. A valid observation corresponds to the daily values higher than LAI_{EBF}^{min}

- If $N_{tot} < N_{EBF}^{min}$, two cases are considered:
  - If the uncertainty value at the previous dekad D-1 is available, the uncertainties are set to the ones at D-1.
  - If there is no previous uncertainty value (missing value), then the uncertainty at D is set to missing value, e.g. 255.

3.3.2.2 Non-Evergreen Broadleaf Forest case (Step B2)

Four steps are considered:

1. First, outliers are rejected based on the smoothness of the temporal profiles of instantaneous estimates (Step B2.1).
2. Second, the compositing window is adjusted (Step B2.2).
3. Then, the valid daily observations within the compositing window are used to fit a linear or second degree polynomial depending on the number of available daily observations (Step B2.3).
4. Finally, the uncertainties are calculated (Step B2.4).

3.3.2.2.1 Rejecting outliers based on expected smoothness of instantaneous estimates time series (Step B2.1)

The consistency of each instantaneous estimate with regards to the before and after observations is evaluated using a temporal window of half-length equal to $length_{no\_EBF}^{outlier}$ and centered on the considered instantaneous estimate. The default value for $length_{no\_EBF}^{outlier}$ is set to $length_{no\_EBF}^{outlier} = 20\ days$. However, if there are less than $N_{outlier}^{min}$ instantaneous estimates within the temporal window, no outlier is considered. The default value of $N_{outlier}^{min}$ is set to 5. The maximum LAI values and associated dates observed in the temporal window before (LAI_{before}^{max}) and after (LAI_{after}^{max}) are identified. The maximum LAI value is used here since it is the less affected by possible cloud contamination. Then, the value of the considered observation,LAI, is compared to the value interpolated, LAI_{interp}, at the same date between LAI_{before}^{max} and after $LAI_{after}^{max}$to detect positive (negative) peaks:

- Positive peak: $LAI \geq LAI_{interp}, +\max\left(tol_{outlier}^{abs}, tol_{outlier}^{rel}.LAI_{interp}\right)$
- Negative peak: $LAI \leq LAI_{interp}, -\max\left(tol_{outlier}^{abs}, tol_{outlier}^{rel}.LAI_{interp}\right)$

where $tol_{outlier}^{abs}, tol_{outlier}^{rel}$ are respectively the absolute and relative values of the tolerance, with $\left[tol_{outlier}^{abs}, tol_{outlier}^{rel}\right] = [0.1, 0.6]$ as default values. If a peak is detected for the considered observation, it is rejected. Note that the actual number of days used before the dekadal date of the product is therefore not strictly $\mathrm{length}_{\mathrm{no\_EBF}}^{\mathrm{max}}$ but $\mathrm{length}_{\mathrm{no\_EBF}}^{\mathrm{max}} + \mathrm{length}_{\mathrm{no\_EBFoutlier}}^{\mathrm{outlierabs}}$ to account for the window required for local outlier rejection ($length_{no\_EBFoutlier}^{outlierabs}$).

3.3.2.2.2 Adjusting the length of the compositing window (Step B2.2)

The compositing window for a given dekadal date D is made of two parts: the sub-window before date D and the sub-window after date D. The maximum length of each sub-window is set to $length_{no\_EBF}^{max}$ (in days). The value of $length_{no\_EBF}^{max}$ will be tuned, but as a default value, it will be set to $length_{no\_EBF}^{max} = 60\ days$ similarly to what is achieved with GEOV2/VGT.

The actual length of each sub-window and the corresponding number of available instantaneous estimates are defined by:

- ***The sub*-window before dekadal date D.** The length of the sub-window before dekadal date D $length_{before}$ (in days), is defined by the period that contains the $N_{no\_EBF}^{max}$ observations which are the closest to dekadal date D. The value of $N_{no\_EBF}^{max}$ will be tuned, but it is set to $N_{no\_EBF}^{max} = 10$ as default. However, a minimum length of the sub-window, $length_{no\_EBF}^{min}$, is considered to increase the robustness of the fitting. The value of $length_{no\_EBF}^{min}$ is set to $length_{no\_EBF}^{min} = 20\ days$. The actual number of observations will therefore depend on the considered cases (Figure 11) and will be $0 \leq N_{before} \leq length_{no\_EBF}^{min}$. The actual length of the sub-window is also variable, with $length_{no\_EBF}^{min} \leq length_{before} \leq length_{no\_EBF}^{max}$, depending on the available observations in the vicinity of the dekadal date considered.

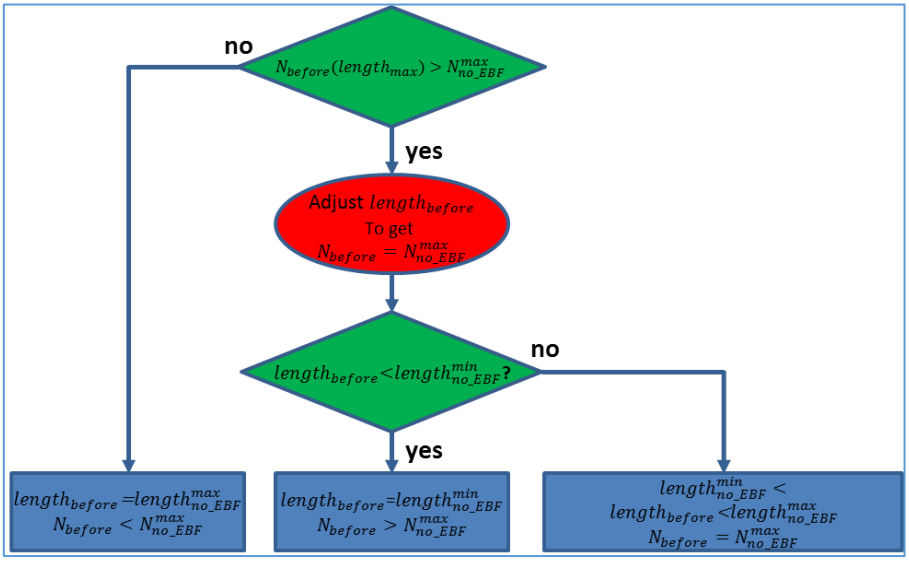

Figure 11: Scheme describing how to set the values $length_{before}$ and N_{before}.

- ***The sub*-window after dekadal date D.** Because we are considering here near-real time products, the length of the sub-window after decadal date D $length_{after}$ (in days) will depend on the considered date within the consolidation period. $length_{after}$ is defined as the period within $length_{no\_EBF}^{max}$ which contains the closest $N_{no\_EBF}^{max}$ daily estimates to dekadal date D. The value of $N_{no\_EBF}^{max}$ is the same as for the temporal window used before the dekadal date D. A minimum length of the sub-window, $length_{no\_EBF}^{min}$, is also considered to increase the robustness of the fitting. The value of $length_{no\_EBF}^{min}$ is set to the same value as the one used for the temporal window before dekadal date D.

Obviously, when considering real time products, no data will be available after dekadal date D and thus $length_{after} = 0\ days$. During the consolidation period, when only one dekad after date D is available, $length_{after} = 10\ days$. The actual value of $length_{after}$ will therefore be set to $length_{after} = \min(t - D, length_{no_{EBF}}^{min})$, where t corresponds to the latest daily product value available. The actual length of the sub-window after is therefore fluctuating, with $0 \leq length_{after} \leq length_{no\_EBF}^{max}$, depending on the available daily estimates in the vicinity of the considered dekadal date (Figure 12).

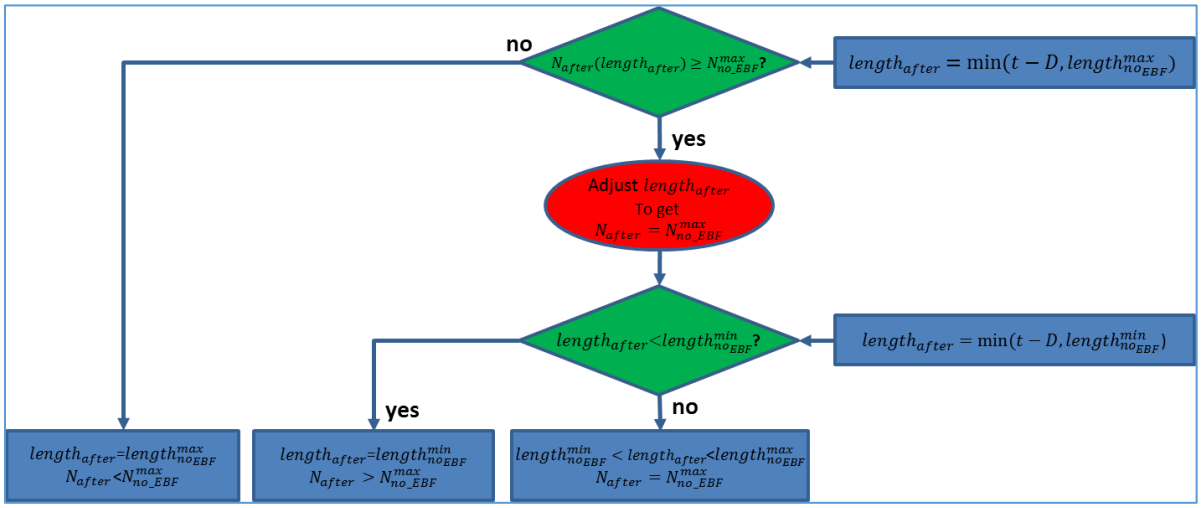

Figure 12: Scheme describing how to set the values $length_{after}$ and N_{after}. t represents the latest date at which a daily product is available. D indicates the current decade at which the product is computed.

This general rule applies to all the 4 cases encountered, i.e. for:

1. ***The start of the time series***, when no data is available before dekadal date D. In this case the processing should start only when $length_{no\_EBF}^{max}$ days have been accumulated.
2. ***Real time estimates***, when no data is available after dekadal date D
3. ***The consolidation period***, when only few observations are available after dekadal date D. The consolidation period lasts $length_{no\_EBF}^{max}$ days.
4. ***Historical processing***, when dekadal date D is followed by observations available for the next $length_{no\_EBF}^{max}$ days.

3.3.2.2.3 Computing the dekadal estimates (Step B2.3)

The smoothing, gap filling and projection is achieved based on a low degree polynomial fit. The degree of the polynomials used over the compositing window is modulated depending on the available number of observations, $N_{tot} = N_{before} + N_{after}$.

A first test is made to detect if the nearest observations within the compositing window is sufficiently close to the dekad date (dekad-nearest_observation <$\Delta_{no\_EBF}^{max}$). $\Delta_{no\_EBF}^{max}$ is set to 15. If this is not the case, no value is computed for this dekad; otherwise three cases are considered (Figure 13):

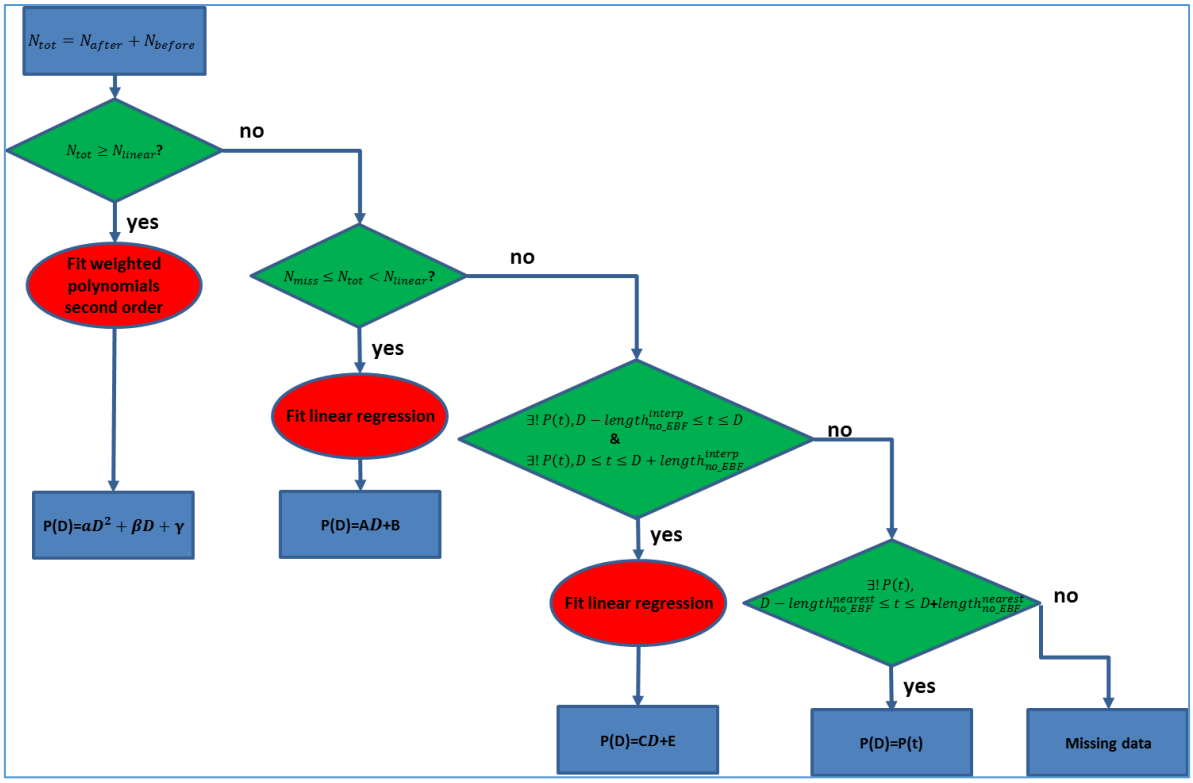

Figure 13: Flow chart showing the three cases considered for smoothing, gap filling and projection for non-Evergreen Broadleaf Forest situations. t represents the dates at which daily products P are available. D indicates the current decade at which the product is computed.

- ***If $N_{tot} \geq N_{linear}$, fitting a second degree polynomial.*** If the number of available instantaneous estimates is higher than the minimum allowed number of observations for fitting the polynomials, N_{linear}, a weighted second degree polynomial fitting is applied. The weights are estimated in two iterations. The value of N_{linear} will be fine tuned, but as a default value, we will set $N_{linear} = 5$.

- For the first iteration, weights of available instantaneous estimates, w, are fixed to $w = 1.0$.
- For the second iteration, the weights, w, are computed according to the distance of the instantaneous estimates from the result of the first iteration. A sigmoidal function was considered for computing the weights w (Figure 14):

$$w = \frac{2}{1+\exp(-k*\delta)}$$

where δ is the difference between the instantaneous estimates and the first polynomial estimates, and k a slope parameter ($k = 2$, but this value could be refined). Smaller weight values are considered when the instantaneous estimates are lower than the first polynomial fit: residual clouds or snow contamination generally lowers the value of the daily product. Conversely, higher weights are considered for the larger δ values to better fit to the maximum values that are considered more reliable. This minimizes the 'possible underestimation due to the smoothing procedure observed over peaks in the first iteration, when all the weights w are set to 1.0.

The corresponding flag QC(6:7) in Table 7 is set to 00.

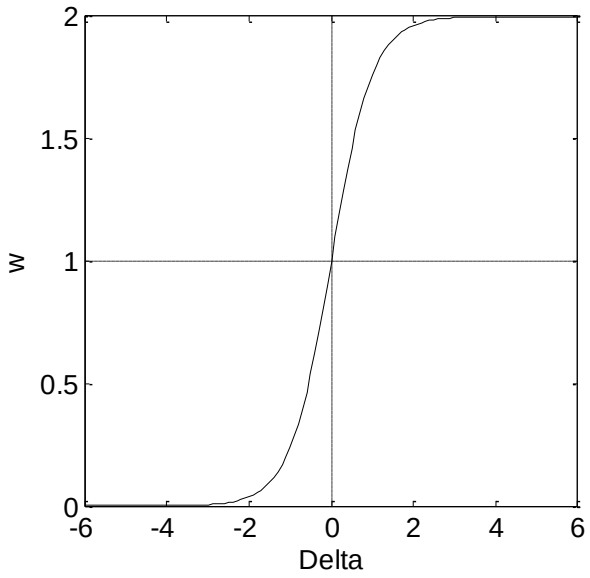

Figure 14: The weighting function used in the polynomial fitting. Delta represents the difference between the actual instantaneous data and the estimates from the first iteration of the polynomial. The slope k was set here to $k = 2$.

- ***If $N_{miss} \leq N_{tot} < N_{linear}$, adjusting a linear fit.*** When less than N_{linear} instantaneous estimates are available in the compositing window, but more than N_{miss} observations

are available within the compositing window, a simple linear fit is applied to provide a value for dekad D. The same principles are used as in the previous case, with one fit with $w = 1$ followed by one fit with the weights computed as previously using the function presented in Figure 14. The corresponding flag QC(6:7) in Table 7 is set to 01.

- ***If $N_{tot} < N_{miss}$, low amount of data.*** If less than N_{miss} instantaneous estimates are available in the compositing window (N_{miss} default value is set to 3.0), then 3 cases must be considered:
    - Two observations, one before and one after the dekad are available within $\pm length_{no\_EBF}^{interp}$ days (set to 15): then, a simple linear interpolation between those two instantaneous estimates is performed at the dekad date. The corresponding flag QC(6:7) in Table 7 is set to 10.
    - A single instantaneous estimate is available within $\pm length_{no\_EBF}^{nearest}$ days (set to 5) around the dekad: then, the corresponding value is attributed to the dekad. The corresponding flag QC(6:7) in Table 7 is set to 10.
    - For the other cases, it results in missing data for that particular dekad. The corresponding flag QC(6:7) in Table 7 is set to 11 and the product value is set to 'missing' data (=255).
- If the estimates are outside the variable range ± a tolerance value (tol_{LAI}, $tol_{FAPAR}, tol_{FCOVER}$), the value is set to missing value

However, when used in projection mode for cases where little observations are available, the polynomial fit may lead to inconsistent estimates. These situations are identified by computing the confidence interval Δ_{LAI} at 95% of the student distribution of LAI. If Δ_{LAI} is larger than a threshold value, defined as $\delta LAI.median(P_{no\_EBF}(t))$, the uncertainty is not computed and the product set to missing data (255)

3.3.2.2.4 Computing the uncertainties (Step B2.4)

If the product value does not correspond to a missing data, the uncertainties are calculated as the root mean square error between the actual daily product estimates $P_{noEBF}(t)$ and the estimated product value from the polynomial fit, $P_{noEBF}^{est}(t)$:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n_{valid}}(P_{no_{EBF}}(t) - P_{noEBF}^{est}(t))^2}{n_{valid}}}$$

In accordance with section 3.3.2.2.3, when used in projection mode for cases where little observations are available, the polynomial fit may lead to inconsistent estimates. When Δ_{LAI} is larger than $\delta LAI.median(P_{no\_EBF}(t))$, the uncertainty is not computed (missing value=255). If $N_{tot} < N_{miss}$, the uncertainties values are set to 255.

3.3.2.3 Filling Gaps (Step B3)

This is a post-processing step. Its completion does not change the values of the dates where data are already available.

If the product is computed at 2 consecutive dekads D and D-1 and the value of the previous dekad, $(D-2)$, is missing, a gap filling method is applied to fill small gaps (Figure 15). In this case, the algorithm evaluates if there are valid products between D and $(D-N_{gap}^{max})$, with N_{gap}^{max} corresponding to the maximum length of a gap to be filled. Default value for N_{gap}^{max} is set to $N_{gap}^{max} = 6.dekads$ If there is no value, then the gap is not filled and QC(3)=0. If there are existing products between dekads D and $(D-N_{gap}^{max})$, a linear interpolation is applied. A flag (QC(3)=1 in Table 7) is raised to indicate that the product corresponds to a linear interpolation.

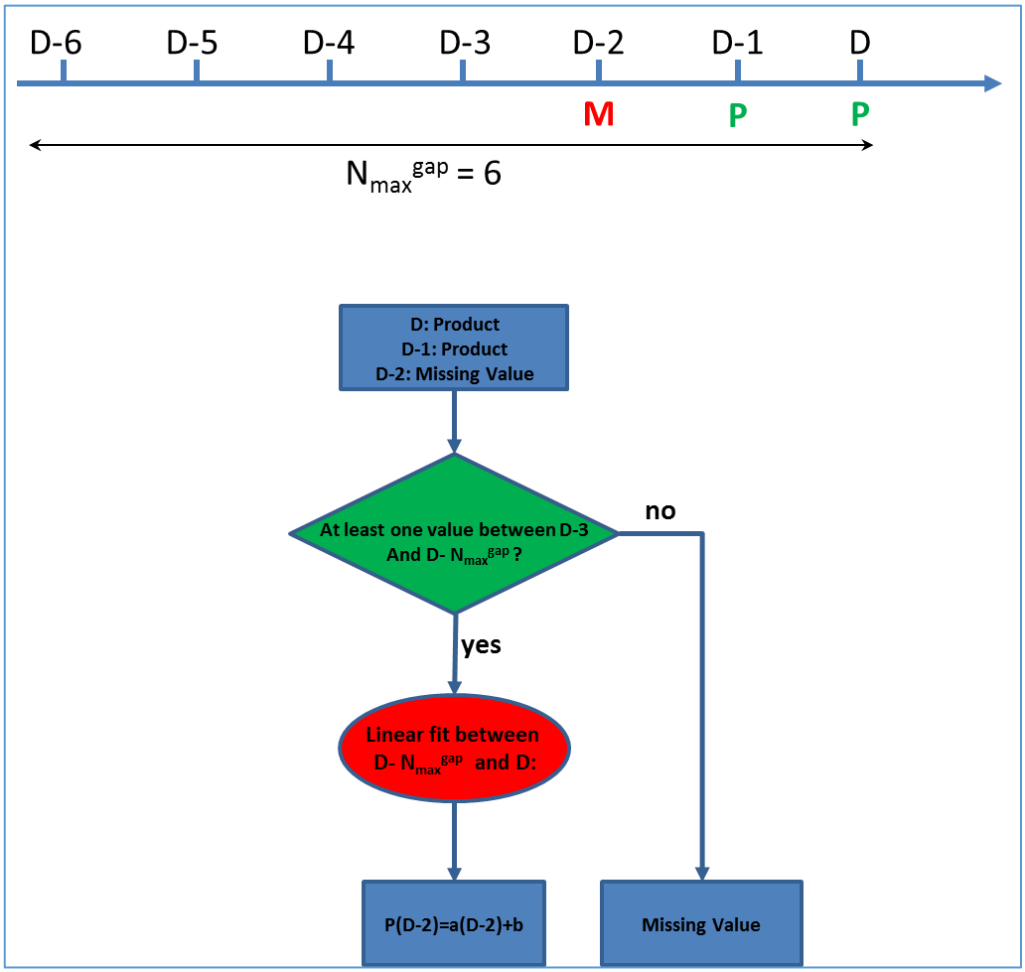

Figure 15. Scheme describing step B3 to fill gaps (no more than $N_{max}^{gap} = 6$ decades)

4 CONCLUSION

This GEOV3 algorithm is an adaptation and a simplification of GEOV2/VGT near real time algorithm. It does not require the use of a climatology since this climatology does not exist at this spatial resolution.

The performance of the algorithm is assessed at the global scale following the validation protocols defined by the LPV group of CEOS. This is performed in the context of the Copernicus Global Land service which disseminates the GEOV3 products (http://land.copernicus.eu/global/products) associated with a Product User Manual and a Quality Assessment report.

A way of improvement could be to fine-tune the parameters of the algorithm with actual PROBA-V data.

5 REFERENCES

Baret , F., Verger, A. and Weiss, M., 2012. GIO Global Land Component – Algorithm Theoretical Basis Document - LAI, FAPAR and FCOVER Version 2.0 (GEOV2), Issue I1.10, in pp. 72, INRA-EMMAH, Avignon.

Baret , F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Weiss, M., Samain, O., Roujean, J.L., Leroy, M., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: Principles of the algorithm. *Remote Sensing of Enviroment 110, 275-286*.

Baret, F., Morissette, J., Fernandes, R., Champeaux, J.L., Myneni, R., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigue, S., Nickeson, J., 2006. Evaluation of the representativeness of networks of sites for the global validation and inter-comparison of land biophysical products. Proposition of the CEOS-BELMANIP. *IEEE Transactions on Geoscience and Remote Sensing 44, 1794-1803.*

Baret, F., Leroy, M., Roujean, J.L., Knorr, W., Lambin, E., Linderman, M., 2003. CYCLOPES User Requirement Document. *INRA-CSE, Avignon*.

Chen, J.M., Menges, C.H., Leblanc, S.G., 2005. Global mapping of foliage clumping index using multi-angular satellite data. *Remote Sensing of Environment 97, 447-457*.

Defourny, P., Schouten, L., Bartalev, S., Bontemps, S., Cacetta, P., de Wit, A.J.W., di Bella, C., Gerard, B., Giri, C., Gond, V., Hazeu, G.W., Heinimann, A., Herold, M., Knoops, J., Jaffrain, G., Latifovic, R., Lin, H., Mayaux, P., Mucher, S.A., Nonguierma, A., Stibig, H.J., Van Bogaert, E., Vancutsem, C., Bicheron, P., Leroy, M. and Arino, O., . , 2009. Accuracy Assessment of a 300 m Global Land Cover Map: the GlobCover Experience., *Proceedings of the 33rd International Symposium on Remote Sensing of Environment, Stresa (Italy), May 2009*.

Doerffer, R., Bourg, L., Goryl, P. and Bouvert, M. 2011. Alternative atmospheric correction procedure for case 2 water remote sensing using MERIS. ATBD V1.0. 27pp. *https://earth.esa.int/instruments/meris/atbd/abtd_2.25_v1.0.pdf*

Garrigues, S., Allard, D., Baret, F., Weiss, M., 2006. Influence landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. *Remote Sensing of Enviroment 105, 286-298*.

Hilker, T., Hall, F.G., Coops, N.C., Lyapustin, A., Wang, Y., Nesic, Z., Grant, N., Black, T.A., Wulder, M.A., Kljun, N., Hopkinson, C., & Chasmer, L., 2010. Remote sensing of photosynthetic light-use efficiency across two forested biomes: Spatial scaling. *Remote Sensing of Environment, 114, 2863.*

McCallum, JL, Wagner W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S. et al., 2009. Satellite-based terestrial production efficiency modeling. *Carbon Balance and Management, 4, 8*.

Prince, S.D., 1991. A model of regional primary production for use with coarse resolution satellite data. *International Journal of Remote Sensing*.

Rahman, H., Dedieu, G., 1994. SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *International Journal of Remote Sensing 15, 123-143*.

Verger, A., 2008. Analisi comparativa d'algorismes operacionals d'estimacio de parametres biofisics de la coberta vegetal amb teledeteccio. *Departament de fisica de la terra i termodinamica (pp. 277).* Valencia, Spain : Universitat de Valencia.

Weiss, M., Baret, F., Myneni, R., Pragnère, A., Knyazikhin, Y., 2000. Investigation of a model inversion technique for the estimation of crop charcteristics from spectral and directional reflectance data. *Agronomie 20, 3-22*.

6 ANNEX: NEURAL NETWORK CALIBRATION

1. PREPARATION OF THE TRAINING DATA SET

The neural networks were calibrated using the BELMANIP2 (Baret et al., 2006) sites extracts over the 2005-2008 period (therefore based on VEGETATION2 sensor). The top of atmosphere reflectance are first corrected from atmospheric effects except aerosol using the actual water vapor, ozone, pressure (derived from altitude) based on the SMAC4.0 model (Rahman and Dedieu, 1994). The derived corrected top of aerosol reflectance is then used as inputs to the neural network along with the observational configuration:

- Atmospherically corrected reflectance in the three VEGETATION bands (B0, B2, B3),
- the cosine of the view zenith angle (cos(VZA)),
- the cosine of the sun zenith angle (cos(SZA)),
- the cosine of the relative azimuth angle (cos(SAA-VAA)),

Note that the use of the blue band allows implicitly correcting for the actual aerosol optical thickness and type. The output is the corresponding instantaneous value of the biophysical variable (LAI, FAPAR and FCover). To be consistent with GEOV1/VGT algorithm, this output is computed similarly by fusing CYCLOPES version 3.1 and MODIS collection 5 products. It consists in a weighted average of both products. The weighting, w, is designed to enhance the specific advantage of each product while limiting their deficiencies. As compared to GEOV1 (Figure 16), this function is smoother, limiting the brutal change in the weight observed for $LAI_{CYCV31} = 4$. Similarly to GEOV2, $w = 0.5$ when $LAI_{CYCV31} = 2$.

$$w = \frac{1}{0.982}\left(1 - \frac{1}{(1 + \exp(-2.LAI_{CYCV31} + 4))}\right)$$

$$\begin{cases} LAI_{fused} = LAI_{MODC5}.(1-w) + LAI_{CYCV31} \cdot w \\ fAPAR_{fused} = fAPAR_{MODC5} \cdot (1-w) + fAPAR_{CYCV31} \cdot w \end{cases} \quad (1)$$

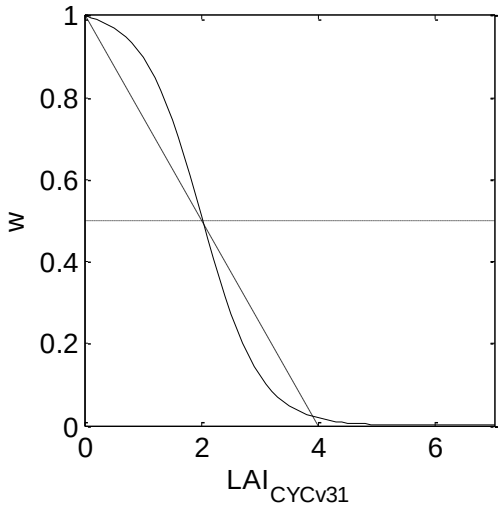

Figure 16: The weighting function used in GEOV2 for the fusion between CYCLOPES and MODIS LAI and FAPAR products. The dashed line corresponds to the weight used for generating GEOV1 products. The dotted line corresponds to $w = 0.5$.

Note that for FCover, no fusion was completed since CYCLOPES was the only existing one. However, several evaluations have shown that CYCLOPES FCOVER products were suffering from a significant systematic underestimation (Verger, 2008). As for GEOV1 and GEOV2 FCover product, this was corrected for by applying a scaling factor to the CYCLOPES V3.1 products ($FCOVER_{CYCV31}$). This factor ($\frac{1}{0.6872}$) was corresponding to the inverse of the FCover value for the 99% cumulated frequency that should be very close to 1.0 since it should correspond to very dense canopies:

$$FCOVER_{best} = \frac{1}{0.6872} \cdot FCOVER_{CYCV31} \quad (3)$$

Where $FCOVER_{best}$ is the value that used for training the neural network. As both LAI and FAPAR are issued from the fusion of two products while FCover is only derived from one of those products, some inconsistency in FCover-FAPAR or FCover-LAI relationships might be observed, especially for high values. Indeed, FCover (gap fraction at nadir) should always be lower than FAPAR that can be approximated by the FIPAR (gap fraction in the sun direction). Figure 17 shows that this is not the case for a significant number of points (33%) for FAPAR higher than 0.5 (33%). However, as high product values must be represented within the neural network training data set, we decided not to discard them. This could be further improved by using more years of data or adding new BELMANIP sites corresponding to dense canopies.

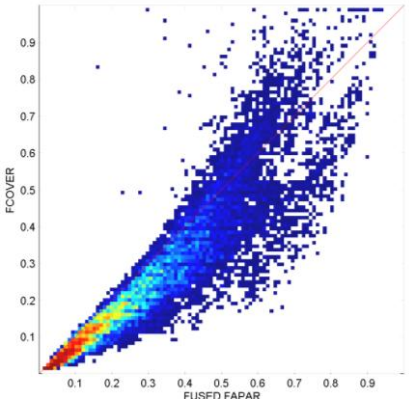

Figure 17: Relationship between fused FAPAR and FCOVER used for the neural network training. The 1:1 line is shown in red.

To facilitate the training process, individual MODIS and CYCLOPES products used in the learning data base were first carefully filtered based on their expected temporal smoothness. For this purpose, a specific tool was developed allowing to visually discarding data that appear to be outliers (Figure 18). Several manipulations were completing this first filtering process to get more robust results.

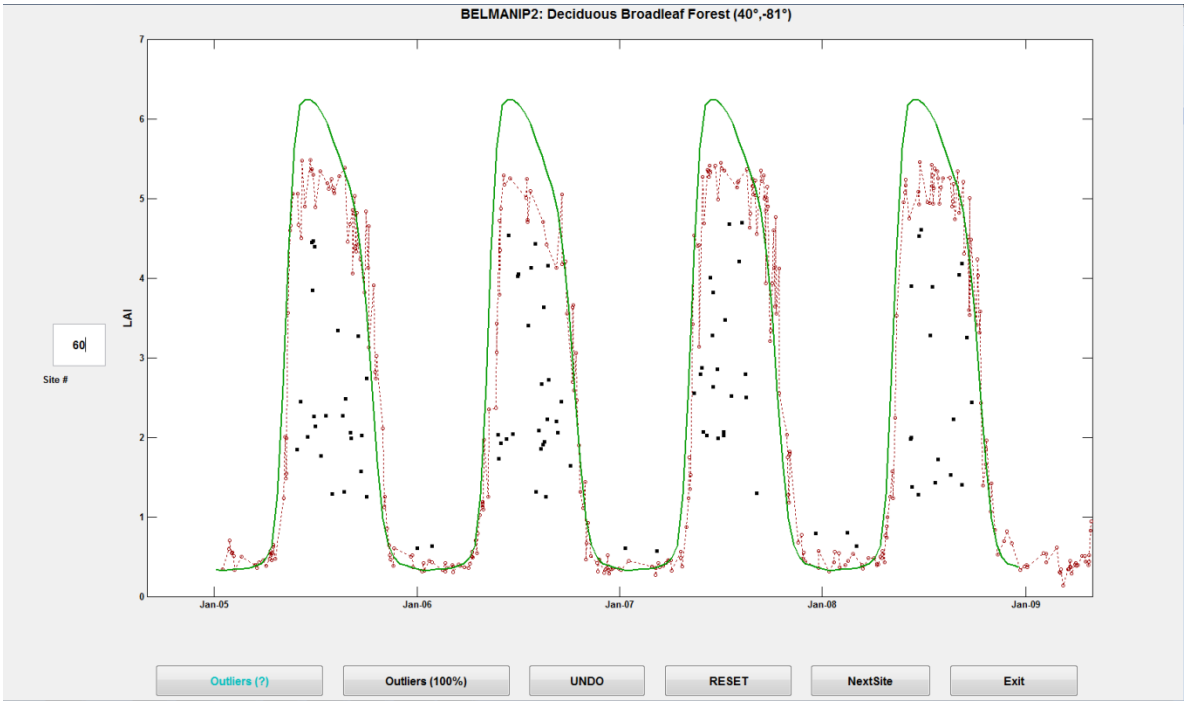

Figure 18: Manual filtering of the outliers. The green line corresponds to the GEOV1 climatology product, red circles to VGTP-V0 (first neural net) valid data, black squares are outliers.

To generate the training data base, all the inputs (reflectance) and outputs (products) should share the same spatial and temporal support. For this purpose, MODIS products were re-projected into the CYCLOPES system (Lat/Lon on WGS-84) using the MODIS re-projection tool. Then, the CYCLOPES dates (dekadal frequency) were used. To minimize the effects of geometrical uncertainties and differences between products, a 3x3 pixel area centered over the BELMANIP2 sites was considered. A dedicated selection process was then applied to get as consistent as possible products regarding the possible contamination by clouds. The selection process assumes that the 3x3 sites are homogeneous, what is mainly the case by construction of the ensemble of 445 BELMANIP2 sites. It is achieved for each date where a daily VGT-P product is available through three steps:

1. Selection of the MODIS and CYCLOPES products: all the MODIS and CYCLOPES products available within ±10 days around the VGT-P daily product date are gathered over the 3x3 pixels. For MODIS, only the main and main + saturation LAI and FAPAR products are considered. This will result in nMOD (0<nMOD<.36) and nCYC (0<nCYC<.18) available products.
2. If there are at least 5 valid products for MODIS and CYCLOPES (nMOD>4 and nCYC>4), then the difference ΔMOD and ΔCYC between the 70% and 90% percentiles within respectively the nMOD and nCYC LAI values available is computed.
3. If ΔMOD <0.2 and ΔCYC <0.2 then the 70% percentile value of LAI and FAPAR products is computed over the nMOD and nCYC products available. This step is applied only when the site is not considered as a bare soil. A bare soil is identified using the LAI climatology computed on GEOV1_VGT if the 90% percentile of the data is less than 0.05. For these bare soil sites (a total of 58), the variables were set to 0.

If steps 2 or 3 are not valid, this results in a missing value. The several threshold values used in the above steps were defined after trial and error tests to reduce the large variability observed over the individual MODIS LAI and FAPAR values and get more consistency between MODIS and CYCLOPES products. The first condition on ΔMOD and ΔCYC over LAI products prevents from using too unstable values, while the lower values may show higher variability because of possible cloud contamination or atmospheric residual effects. Similarly, the 70% percentile value selected for LAI and FAPAR reduces the occurrence of cloud and atmosphere artifacts.

The consistency between the resulting CYCLOPES and MODIS LAI and FAPAR composited values was further checked. Cases showing strong discrepancies were eliminated (Figure 19). The resulting filtered MODIS and CYCLOPES products show a very good consistency with the fused products derived from the application of equation (1) on the composited MODIS and CYCLOPES products (Figure 20). As expected, for LAI<2 and FAPAR<0.5, CYCLOPES contribution to the fused products is the largest. Conversely, for LAI>2 and FAPAR>0.5, MODIS contributes the more to the fused product.

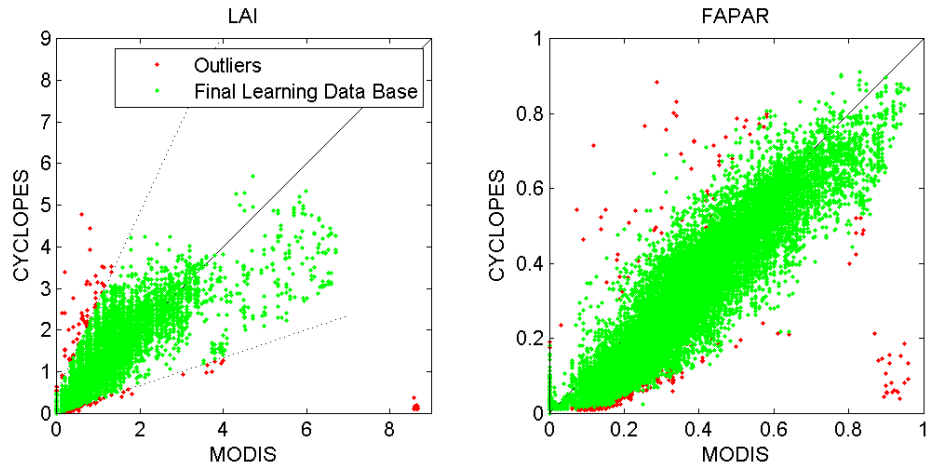

Figure 19: Consistency check between MODIS, CYCLOPES and the fused LAI (left) and FAPAR (right) product. The green points are those considered in the further processing steps. The red points are eliminated.

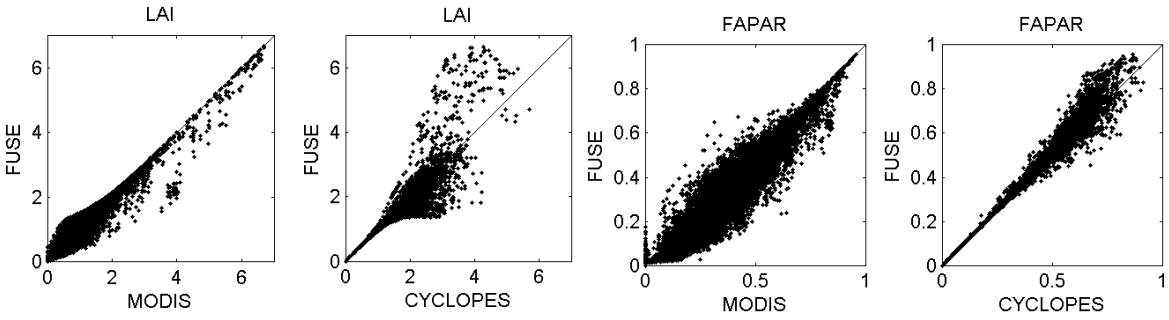

Figure 20: Relationships between the values resulting from the fusion of MODIS and CYCLOPES products according to equation (1) as a function of composited MODIS and CYCLOPES products for LAI (left) and FAPAR (right) variables.

To eliminate outliers in the input reflectances, they were tentatively transformed into LAI products by training a specific neural network that relates the reflectances with the corresponding fused MODIS and CYCLOPES LAI products. For this purpose, the input reflectances were used to calibrate a back-propagation neural network in order to generate a first estimate (Product_0) of the daily values of the biophysical products. A typical 2 layers architecture as that used to generate GEOV1/VGT products was selected with one hidden layer composed of 5 neurons characterized by a tangent sigmoid function and one output layer composed of one linear neuron (Figure 21). Inputs and output were normalized.

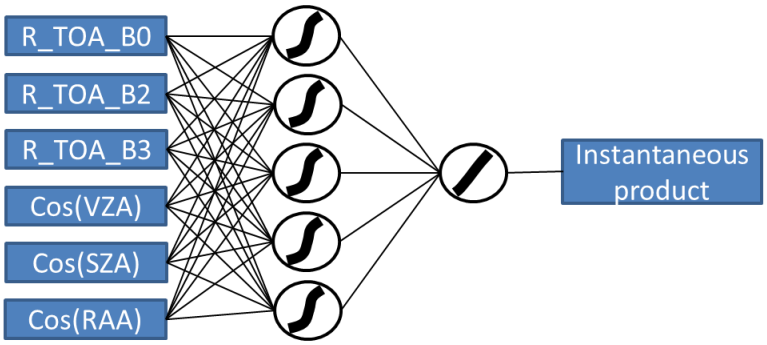

Figure 21: Architecture of the neural network used to compute the instantaneous products.

The neural networks (NNTs) were applied over the 2005-2008 period and then the outliers in the instantaneous estimates were manually filtered out considering the GEOV1/VGT climatology as background information and using a GUI interface specially developed. The expected temporal consistency was used here as the main criterion. The filtering was achieved by several persons to prevent serious bias in the selection process. An example of filtering is shown in Figure 18.

The data were further filtered using three additional criterions whose parameters are summarized in Table 10:

- ***Total air mass.*** The total air mass, m, computed as $m = \frac{1}{\cos \mathrm{SZA}} + \frac{1}{\cos \mathrm{VZA}}$ where SZA and VZA are respectively the sun and view zenith angles. The values for which $m > m_{max}$ were considered as unreliable because of too large atmospheric BRDF effects. They were thus flagged as outliers. The values $m_{max} = 4$ is used as a default value, but will be further refined.
- ***Sun zenith angle.*** The sun zenith angle should be lower than $\mathrm{SZA} < SZA_{max}$ i.e. excluding winter period in the northern latitudes which are prone to cloud cover, snow cover as well as poor illumination conditions. The values $SZA_{max} = 75°$ is used as a default value, but will be further refined.
- ***Soil line.*** Points lying below the soil line in the $B2$ and $B3$ bands were considered with a high probability of being contaminated by significant fraction of water-bodies or clouds. They correspond to the following condition: $\rho_{B3} < 0.54 \frac{\rho_{B2} - 0.04}{0.5 - 0.04}$ where ρ_{B2}, and ρ_{B3} represent respectively the TOA reflectance for bands $B2$ and $B3$. These data points were flagged as outliers.

Note that if these cases are not included in the training database, they will be outside the definition domain. The use of the definition domain to reject outliers will thus result in eliminating the cases considered above.

Parameter	Descriptive Name	Type	Units	Range
m_{max}	Maximum air mass	Float		4
SZA_{max}	Maximum sun zenith Angle	Float	degree	75°
SL_1	Soil Line coefficient 1	Float		0.54
SL_2	Soil Line coefficient 2	Float		0.04
SL_3	Soil Line coefficient 3	Float		0.5

Table 10: Values of parameters used for the preparation of the training dataset.

2. TRAINING THE NETWORKS FOR OPERATIONAL APPLICATIONS AND SCALING THE OUTPUTS

A second neural network with the same architecture as the first one was trained over the filtered reflectance inputs. This network is the one implemented in the processing chain. In addition to the simple training, the outputs were scaled when necessary so that their range of variation better matches the expected one. The scaling was incorporated in the 'denormalization' step of the raw network outputs.

The theoretical performances show that the training process was successful for the three variables (Figure 22): high correlation coefficient and very little bias observed.

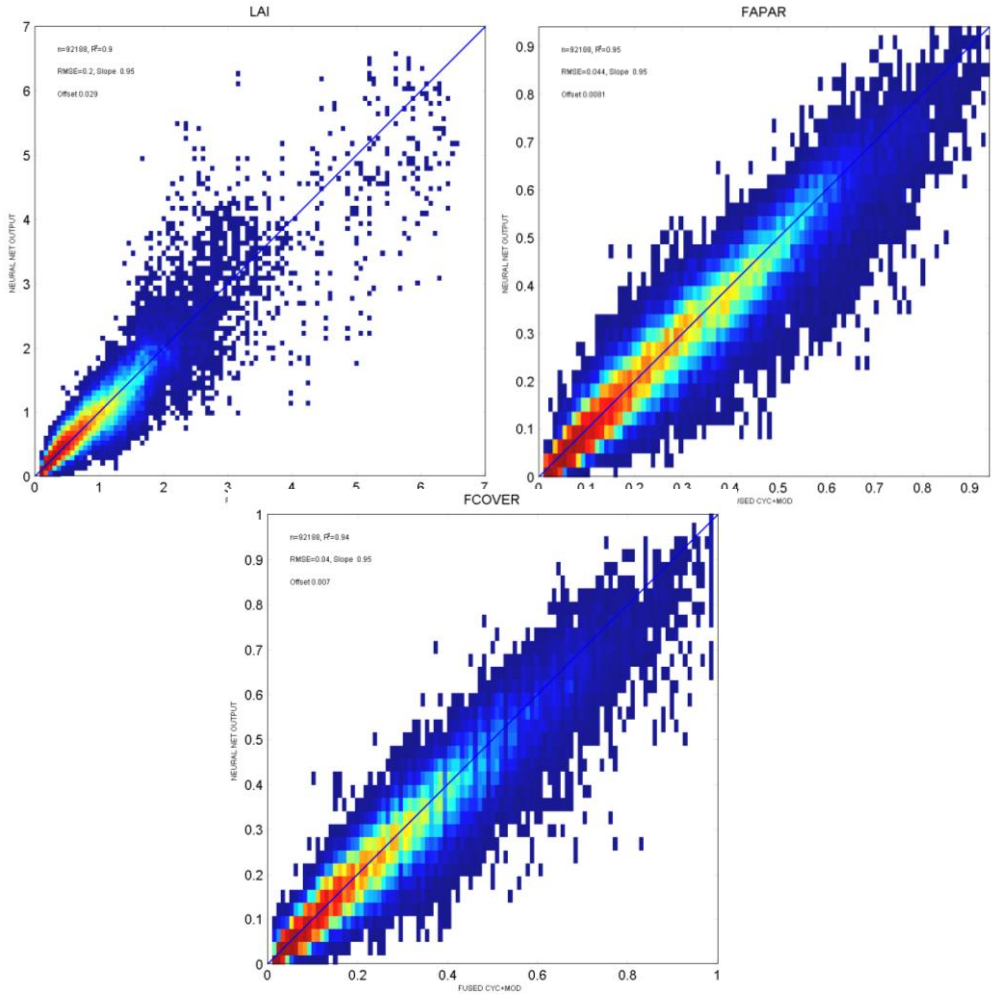

Figure 22: Theoretical performances of the neural networks used for LAI, FAPAR and FCOVER products. Each scatterplot between neural network outputs and the input fused MODIS and CYCLOPES products is displayed as a density plot: the more red, the denser the points are.

To apply the neural network, the following steps must be completed:

- **Normalization of the inputs**: for all the inputs X, the following normalization equation must be applied:

$$Xnorm = 2.(X - Xmin)/(Xmax - Xmin) - 1$$

where $Xnorm$ is the normalized input value, and $Xmin$ and $Xmax$ are computed over the neural network training data set.

- **Run the neural network.** The neural network is described by its architecture, i.e. the number of hidden layers and the output layer. Each layer is described by its number

of neurons, associated weight and biases and transfer function. For the neurons of the hidden layers, the transfer function is a tangent sigmoid function given by: $y = Tansig(x) = 2/(1 + exp(-2x)) - 1$, while for the output layer the transfer function is linear ($y = x$) .

- **Denormalization of the output.** It simply consists in applying the inverse function used for input normalization:

$$Y = 0.5.(Ynorm + 1) * (Ymax^* - Ymin^*) + Ymin^*$$

where $Ynorm$ is the normalized output value issued from the NNT, and $Ymin^*$ and $Ymax^*$ are computed over the neural network training data set.

- **Scaling the outputs.** To better match the expected range of variation of the outputs with the theoretical one, the outputs were multiplied by a scaling coefficient, $Scaling$. This was achieved by dividing the theoretical maximum value for FAPAR (0.94) and FCOVER (1.00) by the 99% percentile value computed over the cumulated distribution frequency of the output Y values of the neural network (Figure 23). Note that no scaling was applied to the LAI product.

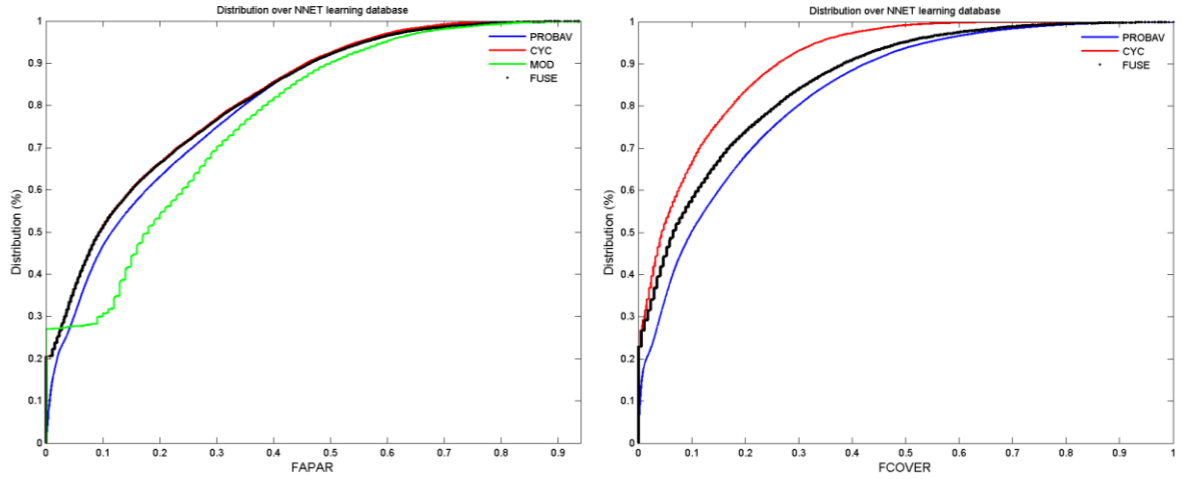

Figure 23: The cumulated distribution frequency of FAPAR (left) and FCOVER (right) used of CYCLOPES, MODIS, fused and GEOV2 products.

Finally, the normalization integrating the scaling applied to each coefficient writes:

$$Ymax = Scaling.Ymax^*$$
$$Ymin = Scaling.Ymin^*$$

Note that the scaling values are integrated in the coefficients used for the denormalization.

3. GENERATION OF THE DEFINITION DOMAIN

The convex hull made by the data retained after this manual selection process and additional filtering based on the air-mass, SZA value and soil line was formed in the B0, B2 and B3 feature space to generate the definition domain that is used to check if the inputs are valid or not before applying the neural network. Note that conversely to the NNT training data set for which we selected VEGETATION data matching MODIS and CYCLOPES valid data, the definition domain is built using all the available valid VEGETATION data during the 2005-2008 period over the BELMANIP2 set of sites. The definition domain would need to be refined once PROBA-V actual observations at 300m resolution will be released. As a matter of facts, when the spatial resolution increases, the definition domain slightly expands because of the inclusion of more pure cases.

The definition domain was gridded using 30^3 cells by dividing each dimension into 30 equally spaced steps between the minimum (B^{min}) and maximum (B^{max}) values for the bands. This corresponds to 27 000 combinations (cells). Note that the B^{max} value is computed on the training data set, by allowing an absolute and relative tolerance of 0.01. The definition domain was finally described by the cells containing no data flagged as outliers. To get a more compact definition domain, closing morphological operations were applied in the 3D space. The Table 11 summarizes the values of parameters used to set-up the definition domain while the Figure 24 shows the final definition domain for all the band combinations (independently on the observational configuration).

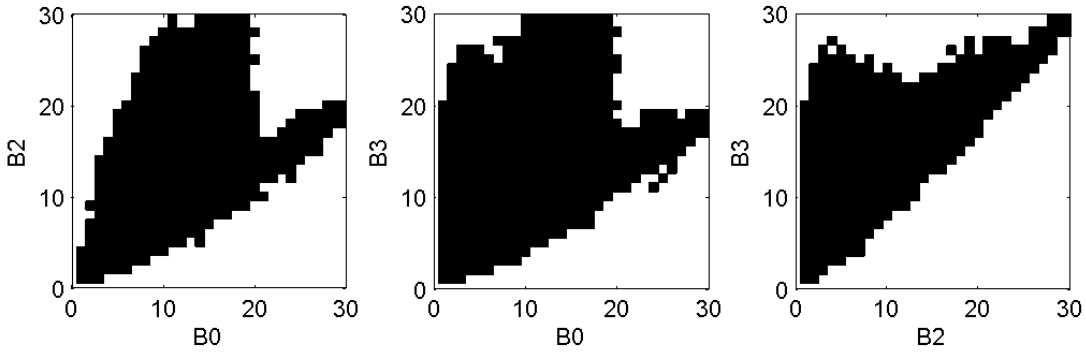

Figure 24: The convex hull that corresponds to the definition domain using the manually filtered outliers. Pixel will be declared as valid if they are within the area defined by the black areas. The 30 cells are distributed equally over the different reflectance ranges (between minimum and maximum values).

Finally, to determine if a set of inputs (reflectance in the 3 bands, cosine of the view zenith, sun zenith and relative azimuth angle) is valid for the neural networks, the following steps are followed:

- Air mass test: $\frac{1}{\cos \text{SZA}} + \frac{1}{\cos \text{VZA}} \leq 4$
- Sun zenith Angle test $SZA \leq 75°$
- Test if the set of reflectance is within or out of the definition domain
    - Compute the corresponding coordinates in the definition domain grid

      $$\frac{\rho_{B0} - B_0^{min}}{B_0^{max} - B_0^{min}} * 30$$

    - Check if this set of 3 coordinates is within the definition domain (i.e. corresponds to the coordinates of a black square in Figure 24).

Parameter	Descriptive Name	Type	Range
B_0^{min}	Minimum reflectance value (Blue)	Float	0
B_0^{max}	Maximum reflectance value (Blue)	Float	0.25
B_2^{min}	Minimum reflectance value (Red)	Float	0
B_2^{max}	Maximum reflectance value (Red)	Float	0.58
B_3^{min}	Minimum reflectance value (NIR)	Float	0
B_3^{max}	Maximum reflectance value (NIR)	Float	0.7
$NCell(B_0)$	Number of cells used to build the definition domain with B_0	Int.	10-50 (30)
$NCell(B_2)$	Number of cells used to build the definition domain with B_2	Int.	10-50 (30)
$NCell(B_3)$	Number of cells used to build the definition domain with B_3	Int.	10-50 (30)

Table 11: Values of parameters used to set-up the definition domain.