Issue: I1.41



Copernicus Global Land Operations "Vegetation and Energy"

"CGLOPS-1"

Framework Service Contract N° 199494 (JRC)

ALGORITHM THEORETHICAL BASIS DOCUMENT

Leaf Area Index (LAI)

Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) Fraction of green Vegetation Cover (FCover)

Collection 1km
Version 2

Issue I1.41

Organization name of lead contractor for this deliverable: CREAF

Book Captain: Aleixandre Verger (CREAF)

Contributing Authors: Frédéric Baret (INRA)

Marie Weiss (INRA)

Date Issued: 11.03.2019

Issue: I1.41



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)	
СО	Confidential, only for members of the consortium (including the Commission Services)	

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **2 of 93**

Date Issued: 11.03.2019

Issue: I1.41



Document Release Sheet

Book captain:	Aleixandre Verger	Sign	Date 11.03.2019
Approval:	Roselyne Lacaze	Sign Racy	Date 18.03.2019
Endorsement:	Michael Cherlet	Sign	
Distribution:	Public		

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **3 of 93**

Date Issued: 11.03.2019

Issue: I1.41



Change Record

Issue/Rev	Date	Page(s)	Description of Change	Release
	26.03.2014	All	First Issue	I1.00
l1.00	10.09.2014	All	Update after external review	I1.10
I1.10	04.09.2015	All	Update with new daily Input Data (VGT reprocess, S1 TOC)	l1.20
l1.20	03.11.2015	All	Improvements of EBF estimates (specific neural net in Step A), improvements in Step B	l1.30
I1.30	12.04.2016	32 49-50	Add one parameter in Table 7 Clarify the meaning of quality indicators (4.3.5)	l1.31
l1.31	04.07.2016	All	Update after external review Improvement of Step A to ensure consistency between PROBA-V and VGT estimates	l1.32
l1.32	03.01.2017	27-29 34 89-92	Add description of input PROBA-V data Revised quality flag description (Table 11) Additional graphs in Annex 3	I1.40
I1.40	11.03.2019	4.4 4.2.2	More details about algorithm limitations More details about quality flag	l1.41

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **4 of 93**

Issue: I1.41



TABLE OF CONTENTS

Eχ	ecutiv	re Summary	13
1	Вас	kground of the document	14
	1.1	Scope and Objectives	14
	1.2	Content of the document	
	1.3	Related documents	
	1.3.1		
	1.3.2	•	
	1.3.3		
	1.3.4		
2	Rev	iew of Users Requirements	17
3	Ove	erview	20
	3.1	The considered products	20
	3.1.1	L FAPAR	20
	3.1.2	2 Cover fraction (FCover)	20
	3.1.3	3 Leaf Area Index (LAI)	20
	3.2	VGT and PROBA-V instruments and data	21
	3.2.1	VEGETATION instrument and S1 data	21
	3.2.2	PROBA-V instrument and S1 data	23
	3.3	Rationale for the Algorithm Selection and Design	25
	3.4	Algorithm Outline	26
4	Alge	orithm Description	29
	4.1	Inputs	2 9
	4.1.1	L Top of canopy daily reflectances	29
	4.1.2	/ / /	
	4.1.3	GEOCLIM: climatology of Version 1 VGT LAI, FAPAR and FCover	29
	4.1.4	1 Algorithmic parameters	29
	4.2	Outputs	32
	4.2.1	L The LAI, FAPAR and FCover values	32
	4.2.2	2 Quality indicators	33
	4.3	Detailed Description	35
	4.3.1	Daily LAI, FAPAR and FCover estimates (Branch A)	35
	4.	3.1.1 Rejection of input data based upon their quality status (Step 1A)	36

Date Issued: 11.03.2019

Issue: I1.41



	4.	3.1.2	Spectral conversion of PROBA-V reflectances (Step 2A)	36
	4.	3.1.3	Input outlier rejection (Step 3A)	37
	4.	3.1.4	Instantaneous estimates of LAI, FAPAR and FCover using neural networks (Step 4A)	37
	4.	3.1.5	Rescaling the PROBA-V outputs of the networks (Step 5A)	38
	4.	3.1.6	Output outlier rejection (Step 6A)	38
	4.	3.1.7	Ensuring FAPAR-FCover consistency (Step 7A)	38
	4.	3.1.8	Computation of percentiles P5 and P90 of daily products (Step 8A)	39
	4.	3.1.9	Preparation of the LAI, FAPAR and FCover climatology (Step 9A)	39
	4.3.2	2 Prod	cessing the past series (Branch B)	41
	4.	3.2.1	Second outlier rejection (Step 1B)	42
	4.	3.2.2	Temporal smoothing and gap filling (TSGF) (Step 2B)	
	4.	3.2.3	Third outlier rejection (Step 3B)	44
	4.	3.2.4	Consistent Adjustment of the Climatology to Actual Observations (CACAO, Step 4B)	46
	4.	3.2.5	Computation of the dekadal Version 2 LAI, FAPAR, FCover products (Step 5B)	48
	4.3.3	B Real	time estimates (Branch C+)	48
	4.	3.3.1	Differences between branches B and C	49
	4.	3.3.2	Defining the consolidation period	50
	4.3.4	Prod	cessing the first dekads of the past series (Branch C-)	52
	4.3.5	5 Com	putation of the associated quality indicators	52
	4.4	Limita	tions	52
5	Eva	luatio	n of the algorithm	55
	5.1	Temp	oral profiles for selected sites	55
	5.2	Consis	stency between LAI, FAPAR and FCover	58
	5.3		arison with Version 1 VGT	
		•		
	5.4	Comp	arison with GEOV3/VGT	63
	5.5	Distrib	oution of values	65
	5.6	Temp	oral continuity	67
	5.7	Temp	oral smoothness	68
	5.8	Consis	stency between Version 2 PROBA-V and Version 2 VGT NRT estimates	69
	5.9	Summ	nary and Recommendations	71
_			•	
6	кеј	erence	S	/3
4	nnex 1	: GEO	CLIM, a climatology of GEOV1 LAI, FAPAR and FCover	77
4	nnex 2	: Neur	al Networks Calibration	81
4	nnex 3	: Resco	aling PROBA-V Estimates	89

Document-No. CGLOPS1_ATBD_LAI1km-V2

© C-GLOPS Lot1 consortium

6 of 93

Page:

Issue: **I1.41** Date: 11.03.2019

Date Issued: 11.03.2019

Issue: I1.41



LIST OF FIGURES

•	ng distance (GPS in m) as and SWIR bands (right)		•	
superimposed with a	oonse Functions of VGT1 spectrum of green grass (Blue, Red, NIR and SW	/IR bands, respect	tively).
reflectance products specific neural netw (Branch A). Second techniques are appliangle of the observathe EBF and Bare Sthe final dekadal V2	three processing branched are transformed into daily orks (NNTs) for Evergreed, dedicated filtering, smoothed. The inputs of this seconditions, the pixel's latitude, oil (BS) landcover classes -HIST estimates when protime (Branch C).	y estimates of LAI, FA en Broadleaf Forest (E pothing, gap filling and end step are the daily e a climatology of LAI, F derived from the climatocessing historical time	PAR and FCover BF) or non_EBF d compositing ter estimates, the sun APAR and FCove atology. The outpute e series (Branch E	using pixels mporal zenith er, and uts are
C-,C+) used to proce series at the time of	howing the several period ess the data. <i>d1</i> and <i>dx</i> co implementation of V2 algorithms.	rrespond to the first ar orithm. n is the numbe	nd last dekad in the r of dekads requir	e time red for
Figure 6: Flow chart desc	ribing the Branch A for dail	y product estimation fro	om VGT	35
•	scribing the Branch A for n A between VGT (Figure 6	• •		
sun zenith angle (Software)	cribing the processing of the ZA) of the observations a P5 from Step 9A. GEOV are ancillary information of	re coming from Step 7 1 corrected climatolog	7A, the parameter gy and the quality	s P90 / flags
Figure 8: Flow chart show	ving how the temporal smo	othing gap filling (TSGF	-) algorithm works	44
=	of the 3-iterations of Tailled circles). Empty circles	- · · · · · · · · · · · · · · · · · · ·		
and SZA are coming the Scale _{BS} and Sc	scribing the processing of the from Step 7A, the parame cale _{EBF} from Step 4B. The are ancillary information d	eters P90 from Step 8A e corrected climatolog	, P5 from Step 9A y and the quality	A, and r flags
(Branch B) estimates n of dekads after the	the differences between value over the BELMANIP2 site are date being processed. available only for the past	s for the year 2008 as Zero dekad $(n = 0)$	a function of the norresponds to the	umber e daily

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **7 of 93**

Date Issued: 11.03.2019

Issue: I1.41



percentile of residual outliers. The bold back solid line corresponds to the median value of the differences. The dotted line is the 0:0 line. Case of LAI
Figure 13: RMSE between NRT-n and HIST Version 2 VGT LAI estimates as a function of the noise in the data and the number of observations (Adapted from Verger et al. (2014a)) 50
Figure 14: Temporal profiles of near real time LAI estimates for n=0,3,6 dekads after the date being processed (NRT-n) as compared to Branch B estimates (HIST) over several BELMANIP2 sites for the year 2005. The original daily LAI data, filtered outliers, the climatology (GEOV1) product are also shown. The title of each plot indicates the BELMANIP2 site number, the GLOBCOVER biome class, the latitude and longitude in degrees
Figure 15: Temporal profile of Version 2 (black solid line) over a typical evergreen broadleaf forest site. Daily estimates derived from VGT-S1 products are indicated by the dots: black squares correspond to outliers. Empty circles to the valid LAI estimates used to compute the Version 2 product. The dashed green line corresponds to the GEOV1 corrected climatology. The solid green line to the CACAO estimates. The red line corresponds to Version 1 VGT product 56
Figure 16: Idem Figure 15 but for a typical Deciduous Broadleaf Forest site56
Figure 17: Idem Figure 15 but for a typical Needleleaf Forest site56
Figure 18: Idem Figure 15 but for two typical Crop-Grassland sites
Figure 19: Idem Figure 15 but for three typical Shrub-Savanna-Bare sites
Figure 20: Comparison between LAI-FAPAR (top), LAI-FCover (middle) and FAPAR-FCover (bottom) for Version 1 (GEOV1), Version 2 (GEOV2), GEOV3/VGT and MODIS products 59
Figure 21: Evaluation of the differences between GEOV2/VGT and GEOV1/VGT products over the BELMANIP2 sites for the years 2003-2007 as a function of the number n of valid daily estimates in the composition period (left), of the RMSE between the GEOV2/VGT product and the daily estimates (center), and of the GEOV2/VGT product value (right). The several gray values correspond to 75% (dark gray), 90% (medium gray) and 95% (light gray) of the population, and the dots to 5% percentile of residual outliers. The bold black solid line corresponds to the median value of the differences. The dotted line is the 0:0 line. The dashed line shows the distribution of values of the variable in the abscissa which frequencies are indicated in the vertical axis on the right. Case of LAI products
Figure 22: Idem Figure 21 but for FAPAR60
Figure 23: Idem Figure 21 but for FCover
Figure 24: Comparison between GEOV2/VGT and GEOV1/VGT LAI products per GLOBCOVER biome type over the 445 BELMANIP2 sites for the years from 2003 to 2007
Figure 25: Comparison between GEOV2/VGT and GEOV1/VGT FAPAR products per GLOBCOVER biome type over the 445 BELMANIP2 sites for the years from 2003 to 200762

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **8 of 93**

Date Issued: 11.03.2019

Issue: I1.41



Figure 26: Comparison between GEOV2/VGT and GEOV1/VGT FCover products pe GLOBCOVER biome type over the 445 BELMANIP2 sites for the years from 2003 to 200763
Figure 27: Comparison between GEOV2/VGT and GEOV3/VGT (PRO3_V1 algorithm) LA products over the 445 BELMANIP2 sites for the years from 2003 to 2007
Figure 28: Comparison between GEOV2/VGT and GEOV3/VGT (PRO3_V1 algorithm) FAPAF products over the 445 BELMANIP2 sites for the years from 2003 to 2007
Figure 29: Comparison between GEOV2/VGT and GEOV3/VGT (PRO3_V1 algorithm) FCove products over the 445 BELMANIP2 sites for the years from 2003 to 2007
Figure 30: Distribution of GEOV1/VGT, GEOV2/VGT, GEOV3/VGT and MODIS LAI products per biome type as sampled by the 445 BELMANIP2 sites over the period 2003-2007
Figure 31: Idem Figure 30 but for FAPAR.
Figure 32: Idem Figure 30 but for FCover.
Figure 33: Average fraction of valid GEOV2/VGT and GEOV1/VGT products per biome. The biomed classes are derived from the GLOBCOVER global landcover: Shrubs/Savana/Bare soil (SSB) Crops and Grassland (CG), Deciduous Broadleaf Forests (DBF), Needleleaf Forest (NF), and Evergreen Broadleaf Forest (EBF). For GEOV2/VGT, high quality products (grey) and products where the climatology was used to fill gaps (less than 12 valid daily estimates exist in the compositing period) (black) are distinguished. Evaluation over the BELMANIP2 sites for the 2003-2007 period.
Figure 34: Histogram of the δLAI absolute difference representing temporal smoothness fo GEOV2/VGT and GEOV1/VGT LAI products. Evaluation over the BELMANIP2 sites for the 2003-2007 period
Figure 35. Comparison between GEOV2/PROBA and GEOV2/VGT NRT-0, NRT-2 and NRT-6 LAI FAPAR and FCover estimates over the 445 BELMANIP2 sites for the period 2013-10-16 to 2014-05-31
Figure 36: Comparison between GEOV2/PROBA-V and GEOV2/VGT NRT-0 LAI estimates per biome type over the 445 BELMANIP2 sites for the period 2013-10-16 to 2014-05-31
Figure 37: Flow chart showing how the GEOV1/VGT climatology is corrected from residual artifacts
Figure 38: (a) Map of bare soil and evergreen broadleaf forest areas identified based or GEOV1/VGT climatology. (b) Simplified GLOBCOVER land-cover map after aggregating the 22 original classes into six main land-cover classes
Figure 39: Correction of GEOV1/VGT climatology. The blue line corresponds to the original GEOV1/VGT climatology LAI product. The red line corresponds to the corrected climatology based on prior knowledge. Green line is the final GEOCLIM climatology resulting from applying gap filling and temporal smoothing techniques to the first corrected climatology80

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **9 of 93**

Date Issued: 11.03.2019

Issue: I1.41



Figure 40: The weighing function used in GEOV2 for the fusion between CYCLOPES and MODIS LAI and FAPAR products. The dashed line corresponds to the weight used for generatin GEOV1 products. The dotted line corresponds to $w = 0.5$
Figure 41: Manual filtering of the outliers. The green line corresponds to the GEOV1 climatolog product, red circles to V0 (first neural net) valid estimates, black squares are outliers
Figure 42: Relationships between the values resulting from the fusion of MODIS and CYCLOPE products according to equation (1) as a function of composited MODIS and CYCLOPE products for LAI (top) and FAPAR (bottom) variables
Figure 43: The convex hull that corresponds to the definition domain using the manually filtere 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
Figure 44: Theoretical performances of the neural networks used for LAI, FAPAR and FCover products. Top: EBF, Bottom: Non EBF. Neural network predicted outputs and the observe fused MODIS and CYCLOPES products in the test dataset are displayed as a density plot: the more red, the denser the points are
Figure 45: Comparison between GEOV2 neural network LAI daily estimates (output of step 4A from PROBA-V and VGT S1 data before (a) and after (b) the rescaling applied to PROBA-V to mimic VGT estimates. Comparison per GLOBCOVER biome type over the 445 BELMANIP sites for the period 2013-10-16 to 2014-05-31.
Figure 46: Comparison between GEOV2 neural network FAPAR daily estimates (output of step 4A from PROBA-V and VGT S1 data before (a) and after (b) the rescaling applied to PROBA-V to mimic VGT estimates. Comparison per GLOBCOVER biome type over the 445 BELMANIP sites for the period 2013-10-16 to 2014-05-31.
Figure 47: Comparison between GEOV2 neural network FCover daily estimates (output of step 4A from PROBA-V and VGT S1 data before (a) and after (b) the rescaling applied to PROBA-V to mimic VGT estimates. Comparison per GLOBCOVER biome type over the 445 BELMANIP sites for the period 2013-10-16 to 2014-05-31.
Figure 48: Boxplots of the differences between GEOV2 neural network LAI (top), FAPAR (middle and FCover (bottom) daily estimates (output of step 4A) from PROBA-V (before rescaling) an VGT S1 data. Median value correspond to the central red mark, the edges of the box are th 25 th and 75 th percentiles, the whiskers extend to the most extreme data points not considere outliers, and outliers are plotted individually with red crosses. The green line corresponds to the 3 rd order polynomial fitted on the median of the residuals

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **10 of 93**

Date Issued: 11.03.2019

Issue: I1.41



LIST OF TABLES

2011)
Table 2: CGLOPS uncertainty levels for FAPAR and FCover products19
Table 3: WMOs requirements for global LAI and FAPARproducts (From http://www.wmosat.info/oscar/requirements); G=goal, B=breakthrough, T=threshold
Table 4: Spectral characteristics of the SPOT/VGT and PROBA-V sensor2
Table 5: VGT-S1 data descriptor22
Table 6: PROBA-V S1 data descriptor25
Table 7: The algorithmic parameters used in Branch A. * indicates whether a parameter is applied only in PROBA-V processing or also for VGT
Table 8: The algorithmic parameters used in Branch B and Branch C
Table 9: Minimum, maximum values and associated resolution for LAI, FAPAR and FCove products. Note that these values are also valid for the climatological products
Table 10: Minimum, maximum values and associated resolution for quantitative quality indicators of LAI, FAPAR and FCover
Table 11: Description of the quality flag provided for the LAI, FAPAR, FCover34
Table 12: Spectral conversion coefficients (α_{Bx} and β_{Bx}) between VEGETATION and PROBA-vectors [GIOGL1_ATBD_PROBA2VGT]
Table 13: Coefficients used to rescale the neural network PROBA-V outputs issued from step 4A Note that step 5A is not applied for VGT which is equivalent to set all the coefficients a ₀ , a ₁ , a ₂ a ₃ to zero.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **11 of 93**

Date Issued: 11.03.2019

Issue: I1.41



LIST OF ACRONYMS

ANN Artificial Neural Network

ATBD Algorithm theoretical based Document

BRDF Bidirectional Reflectance Distribution Function

BS Bare Soil

CACAO Consistent Adjustment of Climatology to Actual Observations

CEOS Committee for Earth Observation Satellite
CTIV VEGETATION image processing centre

EBF Evergreen Broadleaf Forest ECV Essential Climate Variable

ENVISAT Environment Satellite

FAPAR Fraction of Absorbed Photosynthetically Active Radiation

FCover Fraction of vegetation cover GAC Global Area Coverage GAI Green Area Index

GCOS Global Climate Observation System
GEOV1 GEOLAND2 Version 1 product
GEOV2 GEOLAND2 Version 2 product
GEOV3 GEOLAND2 Version 3 product

GMES Global Monitoring of Environment and Security

GTOS Global Terrestrial Observation System
HIST Offline processing of historical time series

JRC Joint Research Center

L2 Level 2 product
L3 Level 3 product
LAC Local Area Coverage
LAI Leaf Area Index

LTDR Long Time Data Record

LTS Long Time Series

MODIS Moderate Imaging Spectrometer

NDVI Normalized Difference Vegetation Index

NIR Near Infrared

NNT Neural Network Technique

NRT Near real Time

RMSE Root Mean Square Error RTM Radiative Transfer Model

SPOT Satellite Pour l'Observation de la Terre

SWIR Short Wave Infrared
TOA Top of Atmosphere
TOC Top of Canopy

TSGF Temporal Smoothing Gap Filling

VI Vegetation Index

Document-No. CGLOPS1_ATBD_LAl1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **12 of 93**

Date Issued: 11.03.2019

Issue: I1.41



EXECUTIVE SUMMARY

The Copernicus Global Land Service (CGLS) is earmarked as a component of the Land service to operate "a multi-purpose service component" that provides a series of bio-geophysical products on the status and evolution of land surface at global scale. Production and delivery of the parameters take place in a timely manner and are complemented by the constitution of long-term time series.

From 1st January 2013, the Copernicus Global Land Service is providing a set of biophysical parameters that describe the vegetation dynamics, such as the first version of Leaf Area Index (LAI), the fraction of absorbed photosynthetically active radiation (FAPAR) and the fraction of vegetation cover (FCover) products

The Copernicus Global Land Service provides a first version of Collection 1km LAI, FAPAR and FCover products also known as GEOV1 [GIOGL1_ATBD_LAI[FAPAR/FCOVER]1km-V1]. This version 1 product were derived from the SPOT/VEGETATION and updated to support PROBA-V reflectance data. Version 1 products are provided every 10 days, with a temporal basis for compositing of 30 days and delivered with a 12 days lag in Near Real Time (NRT) [CGLOPS1_PUM_LAI[FAPAR/FCOVER]1km-V1].

To comply with the Copernicus Global Land Service technical requirements [AD2], a version 2 of the Collection 1km LAI, FAPAR, FCover products, also known as GEOV2, was derived from SPOT/VEGETATION data since May 1998 until December 2013 and PROBA-V since January 2014 onwards. Similarly to Version 1, Version 2 capitalizes on the development and validation of already existing products: CYCLOPES version 3.1 and MODIS collection 5 and the use of neural networks (Baret et al. 2013; Verger et al. 2008). The Version 2 of the Collection 1km products are derived from top of canopy daily (S1-TOC) reflectances instead of normalized top of canopy 30-day composited reflectances as in the Version 1. As compared to Version 1, the compositing step is performed at the biophysical variable level instead of reflectance level. This allows reducing sensitivity to missing observations and avoiding the use of a BRDF model. The version 2 of products has a high consistency with the Version 1 but provides an improved continuity and smoothness and includes a near Real Time estimate.

This Algorithm Theoretical Based Document (ATBD) describes the proposed algorithm for the generation of Collection 1km LAI, FAPAR and FCover version 2 products derived from SPOT/VEGETATION and PROBA-V data. They are provided every 10 days, with a temporal basis for compositing between ±15 and ±60 days depending on the number of available valid observations. They are delivered with a maximum of 3 days lag in Near Real Time, followed by consolidations in the course of the next 6 dekads.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **13 of 93**

Date Issued: 11.03.2019

Issue: I1.41



1 BACKGROUND OF THE DOCUMENT

1.1 SCOPE AND OBJECTIVES

One of the main objectives of Copernicus program is to provide to the scientific community as well as other stakeholders including policy makers, the proper information required for several applications. The products are then operationally generated and delivered freely through the Copernicus Global Land Service portal (http://www.land.copernicus.eu/global) in near real time as well as in offline mode (time series from 1999 to present).

This document provides a detailed description and justification of the algorithm proposed for Version 2 of the algorithm used to assess the Collection 1km LAI, FAPAR and FCover from the daily S1 TOC SPOT/VEGETATION and PROBA-V reflectances.

A theoretical validation is proposed along with a comparison with Version 1 of Collection 1km products, Version 1 of Collection 300m products, MODIS C5 products (Myneni et al. 2002); the consistency between VGT and PROBA-V V2 products is also assessed. Further validation of the products is completed by a full quality assessment analysis according to the Service Validation Plan [CGLOPS1_SVP]. The results are presented in the quality assessment reports [GIOGL1_QAR_LAI[FAPAR/FCOVER]1km-VGT-V2 and CGLOPS1_QAR_LAI[FAPAR/FCOVER]1km-PROBAV-V2]

1.2 CONTENT OF THE DOCUMENT

This document is structured as follows:

- Chapter 2 recalls the users requirements
- Chapter 3 contains the definition of the proposed products, a brief description of the used VGT and PROBA-V data and the outline of the algorithm
- Chapter 4 describes the algorithm in details
- Chapter 5 presents the algorithm performance

1.3 RELATED DOCUMENTS

1.3.1 Applicable documents

AD1: Annex I –Technical Specifications JRC/IPR/2015/H.5/0026/OC to Contract Notice 2015/S 151-277962 of 7thAugust 2015

AD2: Appendix 1 –Copernicus Global land Component Product and Service Detailed Technical requirements to Technical Annex to Contract Notice 2015/S 151-277962 of 7th August 2015

AD3: GIO Copernicus Global Land –Technical User Group –Service Specification and Product Requirements Proposal –SPB-GIO-3017-TUG-SS-004 –Issue I1.0 –26_{th}May 2015

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **14 of 93**

Date Issued: 11.03.2019

Issue: I1.41



1.3.2 Input

D	ocument ID	Descriptor
G	IOGL1_SSD	Service Specifications of the Copernicus Global Land Service.
С	GLOPS1_SVP	Service Validation Plan of the Copernicus Global Land Service
	IOGL1_ATBD_LAI[FAPAR/FCO ER]1km-V1	Algorithm Theoretical Basis Document of Version 1 of Collection 1km LAI, FAPAR, FCover products
G	IOGL1_ATBD_PROBA2VGT	Algorithm Theoretical Basis Document of the pre- processing module to convert PROBA-V data in VGT- like data
In	nagineS_RP2.1_ATBD-LAI300m	Algorithm Theoretical Basis Document of LAI, FAPAR, FCover products derived from PROBA-V data at 300m resolution
	GLOPS1_PUM_LAI[FAPAR/FC VER]1km-V1	Product User Manual of Collection 1km LAI, FAPAR, FCover Version 1 products
	IOGL1_VR_LAI[FAPAR/FCOVER km-V1	Validation Report describing the results of the scientific quality assessment of the SPOT/VGT LAI Collection 1km V1 products

1.3.3 Output

Document ID	Descriptor
CGLOPS1_PUM_LAI[FAPAR/FC OVER]1km-V2	Product User Manual summarizing all information about Collection 1km LAI, FAPAR, FCover Version 2 product
GIOGL1_QAR_LAI[FAPAR/FCO VER]1km-VGT-V2	Quality Assessment Report of the Collection 1km LAI, FAPAR, FCover Version 2 product derived from SPOT/VGT
CGLOPS1_QAR_LAI[FAPAR/FC OVER]1km-PROBAV-V2	Quality Assessment Report of the Collection 1km LAI, FAPAR, FCover Version 2 product derived from PROBA-V

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **15 of 93**

Date Issued: 11.03.2019

Issue: I1.41



1.3.4 External documents

Document ID Descriptor

PROBAV_PUM PROBA-V Products User Manual

Available at http://proba-v.vgt.vito.be/sites/proba-v.vgt.vito.be/files/Product_User_Manual.pdf

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **16 of 93**

Date Issued: 11.03.2019

Issue: I1.41



2 REVIEW OF USERS REQUIREMENTS

According to the applicable document [AD2] and [AD3], the user's requirements relevant for LAI, FAPAR, FCover are:

Definition:

- Leaf Area Index (LAI): One half of the total projected green leaf fractional area in the plant canopy within a given area; Representative of the total biomass and health of vegetation (CEOS)
- Fraction of absorbed PAR (FAPAR): Fraction of PAR absorbed by vegetation for photosynthesis processes (generally around the "red": PAR stands for Photosynthetically Active Radiation).
- Fractional cover (FCover): Fractional cover refers to the proportion of a ground surface that is covered by vegetation.

• Geometric properties:

- The baseline datasets pixel size shall be provided, depending on the final product, at resolutions of 100m and/or 300m and/or 1km.
- The target baseline location accuracy shall be 1/3 of the at-nadir daily field of view
- Pixel co-coordinates shall be given for centre of pixel

Geographical coverage:

o geographic projection: lat long

geodetical datum: WGS84

o pixel size: 1/112° - accuracy: min 10 digits

o coordinate position: pixel centre

o global window coordinates:

Upper Left: 180°W-75°NBottom Right: 180°E, 56°S

• Time definitions:

- As a baseline the biophysical parameters are computed by and representative of dekad, i.e. for ten-day periods ("dekad") defined as follows: days 1 to 10, days 11 to 20 and days 21 to end of month for each month of the year.
- As a trade-off between timeliness and removal of atmosphere-induced noise in data, the time integration period may be extended to up to two dekads for output data that will be asked in addition to or in replacement of the baseline based output data.
- The output data shall be delivered in a timely manner, I. e. within 3 days after the end of each dekad.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **17 of 93**

Date Issued: 11.03.2019

Issue: I1.41



• Ancillary information:

- o the number of measurements per pixel used to generate the synthesis product
- the per-pixel date of the individual measurements or the start-end dates of the period actually covered
- quality indicators, with explicit per-pixel identification of the cause of anomalous parameter result

Accuracy requirements:

- Baseline: wherever applicable the bio-geophysical parameters should meet the internationally agreed accuracy standards laid down in document "Systematic Observation Requirements for Satellite-Based Products for Climate". Supplemental details to the satellite based component of the "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC". GCOS-#154, 2011" (Table 1).
- Target: considering data usage by that part of the user community focused on operational monitoring at (sub-) national scale, accuracy standards may apply not on averages at global scale, but at a finer geographic resolution and in any event at least at biome level.

Table 1: GCOS requirements for LAI and FAPAR as Essential Climate Variables (GCOS-154, 2011)

Variable/ Parameter	Horizontal Resolution	Vertical Resolution	Temporal Resolution	Accuracy	Stability
LAI	250 m	N/A	2- weekly averages	Max(20%; 0.5)	Max(10%; 0.25)
FAPAR	250 m	N/A	2- weekly averages (based on daily sampling)	Max(10%; 0.05)	Max(3%; 0.02)

Note however that the uncertainty associated to LAI reference maps is expected to be around 1 LAI units for forest (Fernandes et al. 2003) or around 0.5 for croplands (Martínez et al. 2009). Therefore, with the available ground truth reference data we cannot achieve the GCOS target requirement for LAI satellite-based products. Further research on FAPAR should be conducted to evaluate the uncertainty attached to ground reference maps, which should be also slightly higher than the GCOS requirement for satellite-based products

Additionally, the Technical User Group of the Copernicus Global Land [AD3] has recommended new uncertainty levels for FAPAR and FCover (Table 2) while for LAI the users did not come to an agreement.

Document-No. CGLOPS1_ATBD_LAl1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **18 of 93**

Date Issued: 11.03.2019

Issue: I1.41



Table 2: CGLOPS uncertainty levels for FAPAR and FCover products

	Optimal	Target	Threshold	
FAPAR	5%	10%	20%	
FCover	370	1070	2070	

• Additional user requirements

The GCOS requirements are supplemented by application specific requirements identified by the WMO (Table 3). These specific requirements are defined at goal (ideal), breakthrough (optimum in terms of cost-benefit), and threshold (minimum acceptable). In most cases the GCOS requirements satisfy threshold levels (especially considering that GCOS requirements greatly exceed threshold spatial resolution requirements so random errors will cancel during spatial aggregation).

Table 3: WMOs requirements for global LAI and FAPARproducts (<u>From http://www.wmosat.info/oscar/requirements</u>); G=goal, B=breakthrough, T=threshold.

Application	Variable	Accuracy (%)		Spatial Resolution (km)		Temporal Resolution (days)				
		G	В	Т	G	В	Т	G	В	Т
Global Weather Prediction	LAI	5	10	20	2	10	50	1	5	10
Ciobai Weather Frediction	FAPAR		10	20	۷	10	30	•	3	10
Regional Weather	LAI	5	10	20	1	5	40	0.5	1	2
Prediction	FAPAR	FAPAR	5 10	10 20	'	3	20	0.5	'	
Hydrology	LAI	5	8	20	0.01	0.1	10	7	11	24
Agricultural Meteorology	LAI	5	7	10	0.01	0.1	10	5	6	7
Agricultural Meteorology	FAPAR		8	20	5	13.6	100	1 h	0.25	7
Seasonal and Inter-annual Forecasts	FAPAR	5	7	10	50	100	500	7	12	30
Climate-Carbon Modelling	LAI	5	7	10	0.25	0.85	35 10	1	3	30
Offiniate Oarborr Modelling	FAPAR	J	,	10	0.20	0.5	2	•		50

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **19 of 93**

Date Issued: 11.03.2019

Issue: I1.41



3 OVERVIEW

3.1 THE CONSIDERED PRODUCTS

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

3.1.1 FAPAR

FAPAR corresponds to the fraction of photosynthetically active radiation absorbed by the canopy. The FAPAR value results directly from the radiative transfer model in the canopy which is computed instantaneously. 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). 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 daily 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 during the FP5/CYCLOPES project. Results show that the daily 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, and MODIS products, the daily 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 will be little sensitive to scaling issues (Hilker et al. 2010; Weiss et al. 2000). Note also that the FAPAR refers only to the green parts of the canopy.

3.1.2 Cover fraction (FCover)

FCover is defined as the fraction of ground surface covered by green vegetation as seen from the 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 will be only marginally scale dependent (Weiss et al. 2000). Note that similarly to LAI and FAPAR, only the green elements will be considered.

3.1.3 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

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **20 of 93**

Date Issued: 11.03.2019

Issue: I1.41



(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 will be scale dependent (Garrigues et al. 2006; Weiss et al. 2000). Note that vegetation LAI as estimated from remote sensing will include 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 element: the correct term to be used would be GAI (Green Area Index) although we propose to still use LAI for the sake of simplicity.

3.2 VGT AND PROBA-V INSTRUMENTS AND DATA

The input data of Version 2 algorithm are the daily synthesis (S1) Top of Canopy (TOC) reflectances in 3 bands (B2, B3, MIR) generated and provided by the SPOT/VEGETATION and PROBA-V programmes through VITO (http://www.vito-eodata.be).

3.2.1 VEGETATION instrument and S1 data

From April 1998 to May 2014, the VEGETATION sensor has been operational on board the SPOT 4 and 5 Earth observation satellite systems. It provided a global observation of the world on a daily basis. The instrumental concept relied on a linear array of 1728 CCD detectors with a large field of view (101°) in four optical spectral bands described in Table 4 and Figure 2. Although very similar, some differences between VEGETATION 1 and VEGETATION 2 instruments have to be noticed, particularly regarding the spectral sensitivity (Figure 2).

Table 4: Spectral characteristics of the SPOT/VGT and PROBA-V sensor

Acronym	Cen	itre (nm)	Width (nm)		Potential Applications
	VGT	PROBA-V	VGT	PROBA-V	
B0	450	463	40	46	Continental Ecosystem -Atmosphere
B2	645	655	70	79	Continental Ecosystem
B3	835	845	110	144	Continental Ecosystem
SWIR	1665	1600	170	73	Continental Ecosystem

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **21 of 93**

Issue: I1.41



Table 5: VGT-S1 data descriptor

VGT-S1TOC	Description
planes	
PHYS_VOL	Information about the delivery
LOG	Information about the product.
Logical	- map projection information (general information, geodetic system parameters,
volume	projection parameters)
descriptor	- cartographic location
	- geographic location
	- image coordinates (corresponding to carto and geographic location)
	- geometric correction
	- radiometric correction
	- orbit parameters
	- date and time
	- algorithms references
	- production
RIG	Copyright descriptor
B0	B0 spectral band, Radiometry data
B2	B2 spectral band, Radiometry data
B3	B3 spectral band, Radiometry data
SWIR	SWIR spectral band, Radiometry data
NDV	NDVI
SM	Status Map
	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	view zenith angles
VAA	view azimuth angles
SZA	sun zenith angles
SAA	sun azimuth angles
TG	Time Grid
QL	Quick Look

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **22 of 93**

Date Issued: 11.03.2019

Issue: I1.41



The spatial resolution is 1.15 km at nadir and presents minimum variations for off-nadir observations. The 2200 km swath width implies a maximum off nadir observation angle of 50.5°. About 90% of the equatorial areas are imaged each day, the remaining 10% being imaged the next day. For latitudes higher than 35° (North and South), all regions are acquired at least once a day. The multi-temporal registration is about 300 meters.

The VGT Collection 3 daily S1 TOC reflectance products are used in Version 2 of Collection 1km algorithm. VGT-S1 products are daily VEGETATION images, corrected from system errors (error registration of the different channels, calibration of all the detectors along the line-array detectors for each spectral band, dark current, geometry) 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 VEGETATION bands. The TOA resulting reflectances are then corrected from atmospheric effects using the SMAC algorithm. SMAC inputs are derived from the Numerical Weather Prediction (NWP) model for water vapour, a climatology for ozone, the Global Land Surface Digital Elevation model for pressure and using the optimization algorithm from Maisongrande et al (2004). The data are then flagged thanks to dedicated cloud/ice/snow/shadow detection algorithm. Finally, the 1-day synthesis is composed of the 'best' ground reflectance measurements of images received during one day. The VGT-S1 TOC product planes are described in Table 5.

3.2.2 PROBA-V instrument and S1 data

The PROBA-V Collection 0 S1 TOC reflectance products were used to set-up Version 2 of algorithm.

The PROBA-V sensor has been launched on 6th May 2013 onboard the PROBA platform. It was designed to bridge the gap in space-borne vegetation measurements between SPOT-VGT (March 2018 – May 2014) and the Sentinel-3 satellites launched in 2016. The mission objective is to ensure the continuity with the heritage of the SPOT-VGT mission.

PROBA-V operates at an altitude of 820 km altitude in a sun-synchronous orbit with a local overpass time at launch of 10:45 h. Because the satellite has no onboard propellant, the overpass time is expected to gradually differ from the at-launch value. After launch, the local overpass time first increased to 10:50h in October 2014, followed by a decrease to 10:45h in June 2016. By end-of-mission in March 2020, the Local Time of Descending Node will be at ~09:30h.

The instrument has a Field Of View of 102.6°, resulting in a swath width of 2295 km. This swath width ensures a daily near-global coverage (90%) and full global coverage is achieved every 2 days. 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 1, 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 1, right). This obviously poses a problem regarding the consistency of the radiometric information between the VIS-NIR and SWIR domains.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **23 of 93**

Issue: I1.41



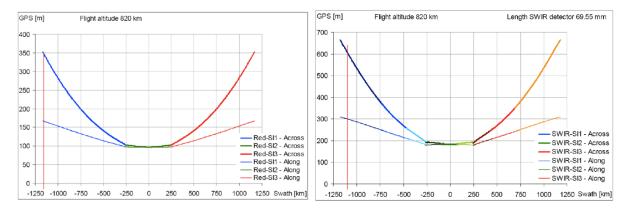


Figure 1: Ground sampling distance (GPS in m) as a function of the position on the swath (in km) for the VIS-NIR (left) and SWIR bands (right).

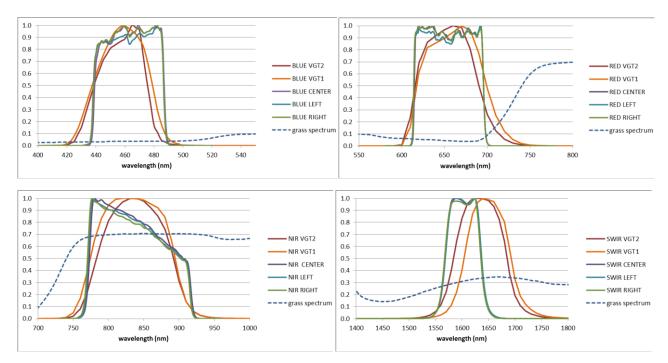


Figure 2: Spectral Response Functions of VGT1, VGT2, and the three PROBA-V cameras, superimposed with a spectrum of green grass (Blue, Red, NIR and SWIR bands, respectively).

The optical design of PROBA-V consists of three cameras. Each camera has two focal planes, one for the short wave infrared (SWIR) and one for the visible and near-infrared (VNIR) bands. The VNIR detector consists of four lines of 5200 pixels. Three spectral bands were implemented, comparable with SPOT-VGT: BLUE, RED, and NIR (Table 4). The SWIR detector is a linear array composed of three staggered detectors of 1024 pixels. The normalized spectral response functions of the four spectral bands of PROBA-V are shown in Figure 2.

Document-No. CGLOPS1_ATBD_LAI1km-V2

© C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **24 of 93**

Issue: I1.41



The PROBA-V processing is described in Sterckx et al. (2014) and Dierckx et al. (2014). The description of the PROBA-V S1 TOC products is summarized in Table 6 and detailed in the Product User Manual [PROBAV_PUM].

Table 6: PROBA-V S1 data descriptor

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					

3.3 RATIONALE FOR THE ALGORITHM SELECTION AND DESIGN

The objective is to develop an algorithm dedicated to the estimation of Version 2 of Collection 1km LAI, FAPAR and FCover from the VEGETATION and PROBA-V 1km series of observations. The algorithm should provide improved estimates as compared to Version 1 products although derived from the same series of observations. These Version 2 of LAI, FAPAR and FCover products should have the same temporal sampling frequency of 10 days. Similarly to Version 1, the Version

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **25 of 93**

Date Issued: 11.03.2019

Issue: I1.41



2 products capitalize on the development and validation of already existing products: CYCLOPES Version 3.1 and MODIS Collection 5 and the use of neural networks calibrated with VGT reflectances. The basic underlying assumption is that a strong link exists between VGT reflectances and the fused product resulting from CYCLOPES and MODIS products. Products should also be associated with quality assessment flags as well as quantified uncertainties. The algorithm runs at the pixel level. The main improvements targeted for Version 2 as compared to Version 1 are:

- Version 2 offers near real time (RT) estimation, in addition of historical (H) data series.
- Version 2 improves the smoothness as compared to Version 1. Despite being one of the smoothest available products, Version 1 still include some problems, particularly over cloudy areas [GIOGL1_VR_LAI[FAPAR/FCOVER]1km-V1GIOGL1_VR_LAI[FAPAR/FCOVER]1km-V1] (Camacho et al. 2013).
- Version 2 has no missing data due to the use of the climatological gap filling while Version 1 has 20% of missing data as evaluated over the BELMANIP2.1 sites (http://calvalportal.ceos.org/web/olive/site-description) from 1999 to 2010 (Verger et al. 2014b).

3.4 ALGORITHM OUTLINE

The algorithm starts from the daily S1 top of the canopy reflectances. Daily LAI, FAPAR and FCover variables are estimated first using neural networks (Branch A in Figure 3). The output is the daily first guess of the LAI, FAPAR and FCover. Then, dekadal composites of LAI, FAPAR and FCover are produced from the daily estimates issued from Branch A using specific processing branches depending on the considered time series:

- The past-time series (Branch B in Figure 3) corresponds to the historical period where, for a given dekad 'd' to be processed, the 'n' dekads before 'd' and after 'd' in the time series are available. 'n' is the number of dekads required for convergence of LAI, FAPAR and FCover values. 'n' is fixed to 6 (cf. section 4.3.2). In Branch B, a climatology coming from Version 1 VGT products is used and the estimation of the final LAI, FAPAR and FCover is achieved through the integration of CACAO module in TSGF processing.
- The real time products (Branch C in Figure 3) are derived for the most recent limited season (around 2 months) using similar principles as those for the past-time series. Note that, each time a new dekad is processed (real time estimates), the recent past values of the variables are updated. This results in successive updates of the LAI, FAPAR, FCover that converge towards the past time series values after the 'convergence period' (up to about 2 months).

Note that for the first past dekads, no past data is available (from d1 to d1+n). Here, Branch C is run in reverse mode (it is called C- as opposed to the forward mode for real time estimation called C+).

Document-No. CGLOPS1_ATBD_LAl1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **26 of 93**

Date Issued: 11.03.2019

Issue: I1.41



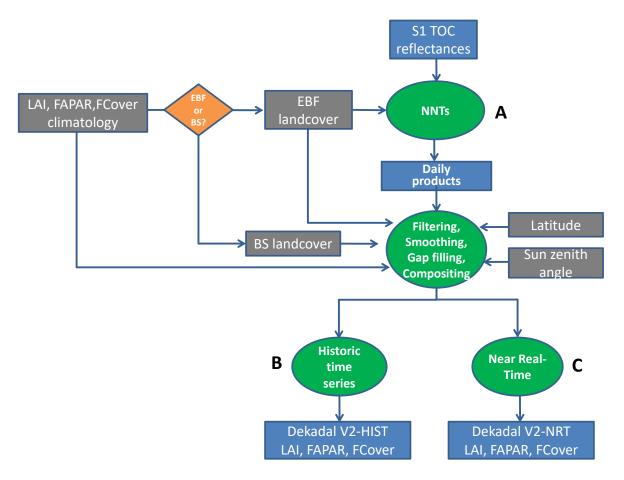


Figure 3: Flow chart the three processing branches (A, B and C). First, daily S1 top of canopy reflectance products are transformed into daily estimates of LAI, FAPAR and FCover using specific neural networks (NNTs) for Evergreen Broadleaf Forest (EBF) or non_EBF pixels (Branch A). Second, dedicated filtering, smoothing, gap filling and compositing temporal techniques are applied. The inputs of this second step are the daily estimates, the sun zenith angle of the observations, the pixel's latitude, a climatology of LAI, FAPAR and FCover, and the EBF and Bare Soil (BS) landcover classes derived from the climatology. The outputs are the final dekadal V2-HIST estimates when processing historical time series (Branch B) and V2-NRT in near real time (Branch C).

Figure 4 shows how the various branches are applied over the several time periods:

- **Branch B** is first applied over the d1+n to the dx-n dekads, where d1 is the first dekad of the time series, dx the dekad corresponding to the time when the Branch C processing starts and 'n' is the number of dekads required for convergence.
- Branch C+ is applied over the dx-n to dx period for the real time estimation at dx. Then, for each new dekad, Branch C is applied over the convergence period 'n' to account for the lack of observations in the future. The product is 'consolidated' over a dekad dc at the dekad dc+n. The convergence period n was fixed to 6 dekads, when the estimates by Branch C converge towards that of Branch B.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **27 of 93**

Issue: I1.41



• **Branch C-** is applied over the d1 to d1+n period in reverse mode **(C-)** to account for the lack of past data for consolidation of the algorithm.

Although the proposed scheme includes many branches, they share some commonalities: in Branch B and Branch C, a common algorithm deals with possible dissymmetry and lack of observations (existing gaps or projection). Branch C+ is similar to Branch B except that the product needs to be updated until reaching convergence. Further, Branch C+ uses some inputs from Branch B. Branch C- is the same as Branch C+ except that it is run in the reverse mode to account for the lack of observations before the dekadal date where estimates have to be computed. Obviously, no updates are proposed for Branch C-.

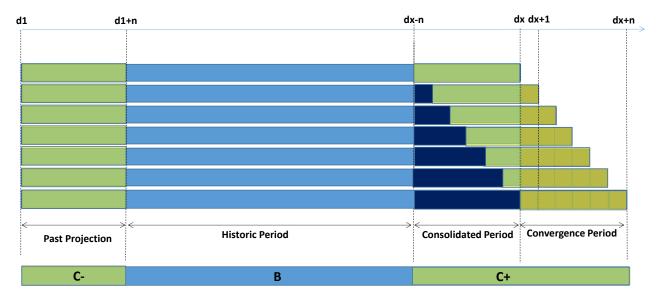


Figure 4: Chronograph showing the several periods considered and the associated branches (B, C-,C+) used to process the data. *d1* and *dx* correspond to the first and last dekad in the time series at the time of implementation of V2 algorithm. *n* is the number of dekads required for convergence.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **28 of 93**

Date Issued: 11.03.2019

Issue: I1.41



4 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.

4.1 INPUTS

All these inputs are required for each considered pixel.

4.1.1 Top of canopy daily reflectances

The daily synthesis (S1) of Top of canopy reflectances (TOC) in the 3 bands (B2, B3, MIR) of VGT and PROBA-V are required as inputs. The blue band is not considered to minimize the impact of residual atmospheric effects. Reflectances should be expressed in terms of reflectance factor, mainly varying between 0.0 and 0.7 for most land surfaces outside hot-spot or specular directions and snow or ice cover.

4.1.2 Geometry of acquisition

Since S1 products are not normalized, geometry information is required as input to the neural network for the three variables. It includes:

- 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))

4.1.3 GEOCLIM: climatology of Version 1 VGT LAI, FAPAR and FCover

GEOCLIM (Verger et al. 2015), the climatology of LAI, FAPAR and FCover defined as the average inter-annual value from Version 1 of Collection 1km products from VGT time series, is used as background information for the processing of the past time series (Branch B) and projections for real time estimates (Branch C). Based on the LAI climatology, pixels corresponding to evergreen broadleaf forest (EBF) and permanent bare soils (BS) were identified by a particular quality flag (QF). This work is described in detail in [GIOGL1_ATBD_LAI[FAPAR/FCOVER]1km-V1]. It is also described in Annex 1.

For the pixels identified as EBF, the corresponding flag QC(11) in Table 11 is set to 1. For the pixels identified as BS, the flag QC(12) in Table 11 is set to 1.

4.1.4 Algorithmic parameters

The algorithm uses a series of parameters listed in Table 7 for Branch A and

Table 8 for Branch B and Branch C. Their roles and usages are further described in the various paragraphs of Section 4.3.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **29 of 93**

Date Issued: 11.03.2019

Issue: I1.41



Table 7: The algorithmic parameters used in Branch A. * indicates whether a parameter is applied only in PROBA-V processing or also for VGT.

Parameter	Description	Value	Unit	Step*	Section
$lpha_{B_2/B_3/SWIR}$	Slope of the spectral conversion for B ₂ , B ₃ and SWIR input reflectance			2A- PV	4.3.1.2
$eta_{B_2/B_3/SWIR}$	Offset of the spectral conversion for B ₂ , B ₃ and SWIR input reflectance	See Table 12		2A- PV	4.3.1.2
$a_{0/1/2/3}^{LAI/FAPAR/FCOVER}$	Rescaling a_0 , a_1 , a_2 a_3 coefficient for LAI, FAPAR and FCover	See Table 13		5A- PV	4.3.1.5
tol^{min}_{LAI}	Tolerance minimum on LAI used to reject estimated values outside the expected range of variation	-0.20	-	6A	4.3.1.6
tol_{FAPAR}^{min}	Tolerance minimum on FAPAR used to reject estimated values outside the expected range of variation	-0.05	1	6A	4.3.1.6
tol_{FCOVER}^{min}	Tolerance minimum on FCover used to reject estimated values outside the expected range of variation	reject estimated values outside the -0.05		6A	4.3.1.6
tol_{LAI}^{max}	Tolerance maximum on LAI used to reject estimated values outside the expected range of variation	7.2	-	6A	4.3.1.6
tol_{FAPAR}^{max}	Tolerance maximum on FAPAR used to reject estimated values outside the expected range of variation	0.99	-	6A	4.3.1.6
tol_{FCOVER}^{max}	Tolerance maximum on FCover used to reject estimated values outside the expected range of variation		-	6A	4.3.1.6
Lat_{EBF}^{max}	Latitude max at which EBF can be found	28.5	0	9A	4.3.1.9
Lat ^{min}	Latitude min for each dekad at which the GEOCLIM LAI/FAPAR/FCover is adjusted to the daily Product_1	[42.5 43.5 45.5 48.5 51.5 55.5 59 64 68.5 73.5 78 82 85.5 88.5 90 90 90 90 90 90 97 83.5 80 75.5 71.5 67 63 59 55 51.5 48.5 46 44 42.5 42]	0	9A	4.3.1.9

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **30 of 93**

Issue: I1.41



Table 8: The algorithmic parameters used in Branch B and Branch C

Parameter	Description	Value	Unit	Step	Section
Lat_{winter}^{min}	Latitude min at which daily estimates are filtered in winter time	55	0	1B	4.3.2.1
SZA ^{HL}	Sun Zenith Angle (SZA) for norther high latitudes at which daily estimates are filtered	70	0	1B	4.3.2.1
LAI_{winter}^{min}	LAI min in winter time at which daily estimates are filtered	0.5	ı	1B	4.3.2.1
LAI_{EBF}^{min}	LAI min in EBFs at which daily estimates are filtered	5.5	1	1B	4.3.2.1
n_{TSGF}^{min}	Minimum number of daily estimates within each side of the dekadal date being smoothed for the application of TSGF	6	-	2B	4.3.2.2
$length_{TSGF}^{max}$	Maximum length of the half compositing window in TSGF	60	days	2B	4.3.2.2
$length_{TSGF}^{min}$	Minimum length of the half compositing window in TSGF	15	days	2B	4.3.2.2
$Scale_W^{Climato}$	Scale factor applied to the weights of the climatology values in TSGF	0.5	-	2B	4.3.2.2
$W_{iter=1}^{Product_1}$	Weights of Product_1 estimates in the first iteration of TSGF	1	-	2B	4.3.2.2
W ^{Climato} iter=1	Weights of the climatology values in the first iteration of TSGF	0.5	-	2B	4.3.2.24.3.2.2
$niter_{gap}$	Number of iteration of the gap filling procedure in TSGF	2	-	2B	4.3.2.2
niter	Number of iterations TSGF is applied	3		3B	4.3.2.3
$length_{Outlier}$	Length of the half compositing window for the third outlier rejection	15	days	3B	4.3.2.3
$tol_{Outlier}^{abs}$	Value of outlier threshold (absolute value) used to detect outliers	0.10	-	3B	4.3.2.3
$tol^{rel}_{Outlier}$	Value of outlier threshold (relative value) used to detect outliers	15	%	3B	4.3.2.3
$tol_{Outlier}^{base}$	Threshold distance to the base level used to detect outliers	0.5	-	3B	4.3.2.3
$tol_{Outlier}^{TSGF}$	Threshold distance to TSGF estimates used to detect outliers	0.5	-	3B	4.3.2.3
LAI ^{base} outlier	Threshold LAI value at the base level used to detect oultiers	0.5	-	3B	4.3.2.3
$P90_{Outlier}^{min}$	Minimum Percentile 90 used to detect outliers	0.5	-	3B	4.3.2.3

Document-No. CGLOPS1_ATBD_LAI1km-V2

© C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019

Page: 31 of 93

Date Issued: 11.03.2019

Issue: I1.41



Threshold LAI value (absolute value) used for CACAO computation	0.10	-	4B	4.3.2.4
Threshold FAPAR and FCover value (absolute value) used for CACAO computation	0.025	-	4B	4.3.2.4
Threshold (relative value) used for CACAO computation	15	%	4B	4.3.2.4
Relative value of amplitude used to define the sub-seasons in CACAO	30	%	4B	4.3.2.4
Relative value of length period used to define the sub-seasons in CACAO	30	%	4B	4.3.2.4
Shift max of CACAO	60	days	4B	4.3.2.4
Temporal step of application of CACAO	5	days	4B	4.3.2.4
Threshold (relative value) of available daily estimates in a sub-season for the application of CACAO		%	4B	4.3.2.4
Threshold (relative value) of the amplitude of estimates in a sub-season for the application of CACAO	30	%	4B	4.3.2.4
Minimum number of daily estimates required to fit CACAO over BS and EBF pixels		-	4B	4.3.2.4
Extended period (in months) to fit CACAO	6	months	4B	4.3.2.4
Minimum number of daily estimates required to compute the uncertainty, RMSE, of the product value		-	QA	4.3.5
	Threshold FAPAR and FCover value (absolute value) used for CACAO computation Threshold (relative value) used for CACAO computation Threshold (relative value) used for CACAO computation Relative value of amplitude used to define the sub-seasons in CACAO Relative value of length period used to define the sub-seasons in CACAO Shift max of CACAO Temporal step of application of CACAO Threshold (relative value) of available daily estimates in a sub-season for the application of CACAO Threshold (relative value) of the amplitude of estimates in a sub-season for the application of CACAO Minimum number of daily estimates required to fit CACAO over BS and EBF pixels Extended period (in months) to fit CACAO Minimum number of daily estimates required to compute the uncertainty,	used for CACAO computation Threshold FAPAR and FCover value (absolute value) used for CACAO computation Threshold (relative value) used for CACAO computation Threshold (relative value) used for CACAO computation Relative value of amplitude used to define the sub-seasons in CACAO Relative value of length period used to define the sub-seasons in CACAO Shift max of CACAO Temporal step of application of CACAO Threshold (relative value) of available daily estimates in a sub-season for the application of CACAO Threshold (relative value) of the amplitude of estimates in a sub-season for the application of CACAO Minimum number of daily estimates required to fit CACAO over BS and EBF pixels Extended period (in months) to fit CACAO Minimum number of daily estimates required to compute the uncertainty, 2	used for CACAO computation Threshold FAPAR and FCover value (absolute value) used for CACAO computation Threshold (relative value) used for CACAO computation Threshold (relative value) used for CACAO computation Relative value of amplitude used to define the sub-seasons in CACAO Relative value of length period used to define the sub-seasons in CACAO Shift max of CACAO Shift max of CACAO Threshold (relative value) of available daily estimates in a sub-season for the application of CACAO Threshold (relative value) of the amplitude of estimates in a sub-season for the application of CACAO Minimum number of daily estimates required to fit CACAO over BS and EBF pixels Extended period (in months) to fit CACAO Minimum number of daily estimates required to compute the uncertainty, Augusta 15	used for CACAO computation Threshold FAPAR and FCover value (absolute value) used for CACAO computation Threshold (relative value) used for CACAO computation Threshold (relative value) used for CACAO computation Relative value of amplitude used to define the sub-seasons in CACAO Relative value of length period used to define the sub-seasons in CACAO Shift max of CACAO Threshold (relative value) of available daily estimates in a sub-season for the application of CACAO Threshold (relative value) of the amplitude of estimates in a sub-season for the application of CACAO Minimum number of daily estimates required to fit CACAO over BS and EBF pixels Extended period (in months) to fit CACAO Minimum number of daily estimates required to compute the uncertainty, AB U.10 4B 4B 4B 4B 4B 4B 4B 4B 4B 4

4.2 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)

They are computed by application of the algorithm over each pixel at each dekadal date.

4.2.1 The LAI, FAPAR and FCover values

The range of variation and resolution of LAI, FAPAR and FCover are presented in Table 9. The physical values are retrieved from the Digital Number (DN) by:

PhyVal = DN * Scaling_factor + Offset

where the scaling factor and the offset are given in the table below.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **32 of 93**

Issue: I1.41



Table 9: Minimum, maximum values and associated resolution for LAI, FAPAR and FCover products.

Note that these values are also valid for the climatological products.

Variable	Physical Minimum	Physical Maximum	Max DN value	Missing value	Scaling factor	Offset
LAI	0.0	7.0	210	255	1/30	0
FAPAR	0.0	0.94	235	255	1/250	0
FCover	0.0	1.0	250	255	1/250	0

4.2.2 Quality indicators

In addition to the product values, other quantitative quality indicators (Table 10) and quality flags are also provided.

The quantitative metrics NOBS, RMSE and LENGTH_BEFORE, LENGTH_AFTER are ancillary layers describing the quality of the product (Table 10). See Section 4.3.5 for more details.

Table 10: Minimum, maximum values and associated resolution for quantitative quality indicators of LAI, FAPAR and FCover.

	Description	Physical Minimum	Physical Maximum	Max DN value	Missing value	Scaling factor	Offset
NOBS	Number of valid daily estimates in the compositing window	0 (*)	120	120	0	1	0
LENGTH_ BEFORE	Length in days of the semi- period of compositing before the dekad d	5	60	60	255	1	0
LENGTH_ AFTER	Length in days of the semi- period of compositing after the dekad d	5	60	60	255	1	0
RMSE-LAI	RMSE between the final dekadal LAI value and the daily estimates in the compositing period	0.0	7.0	210	255	1/30	0
RMSE- FAPAR	RMSE between the final dekadal FAPAR value and the daily estimates in the compositing period	0.0	0.94	235	255	1/250	0
RMSE- FCover	RMSE between the final dekadal FCover value and the daily estimates in the compositing period	0.0	1.0	250	255	1/250	0

Document-No. CGLOPS1_ATBD_LAI1km-V2

© C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019

Page: **33 of 93**

Issue: I1.41



(*) Note that for NOBS=0 (no observations in the ± 60 -day period), LAI, FAPAR and FCover values may still be provided through the gap-filling, indicated by the highest bits of QFLAG (value >= 4096).

The qualitative quality flag (QC) indicator is coded as a 16-bit (2 byte) pattern specified in Table 11. Bit number 1 is the least significant bit (right-most). The QC value 65535 is used for missing (non-processed) pixels.

- Bit 1 refers to the land-sea mask based upon the GLC-2000 land cover map (Bartholomé and Belward 2005).
- Bit 3 is activated when gap-filling is applied.
- Bit 6 is activated when no available observations exist in the compositing period (NOBS=0).
- Bits 7-9 are activated when LAI, FAPAR and FCover values, respectively, are not available (out of range or invalid).
- Bit 10 is activated when a specific correction is applied for northern high latitudes at extreme illumination conditions.
- Bit 11 is activated when the pixel is recognized as Evergreen Broadleaf Forest based on GEOCLIM climatology.
- Bit 12 is activated when the pixel is recognized as Bare Soil based on GEOCLIM climatology. Note that some pixels correspond indeed to inland-water bodies not identified in the land-sea mask (Bit 1).
- Bit 13 is activated when gap-filling is based on the climatology.
- Bit 14 is activated when gap-filling is based on linear interpolation.

Table 11: Description of the quality flag provided for the LAI, FAPAR, FCover.

	Bit = 0	Bit = 1
Bit 1: Land/Sea	Land	Sea
Bit 2: Not used		
Bit 3: Filled	No filled	The number of valid observations at (at least)
		one side (the left side in the NRT0 case) of the
		±60-day period is lower than 6 and a gap filling
		procedure (Bit 13-14) is applied
Bit 4 : Not used		
Bit 5: Not used		
Bit 6: Input status	OK	No valid observations within ±60 days (-60 days
		in the NRT0 case)
Bit 7: LAI status	OK, in expected range	Out of range or invalid
	including tolerance	
Bit 8: fAPAR status	OK, in expected range	Out of range or invalid
	including tolerance	
Bit 9: FCover status	OK, in expected range	Out of range or invalid
	including tolerance	
Bit 10: HLAT status	No specific correction for	A specific correction for high latitudes ($lat >$
	high latitudes is applied	55º) and SZA>70º is applied

Document-No. CGLOPS1_ATBD_LAI1km-V2

© C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **34 of 93**

Date Issued: 11.03.2019

Issue: I1.41



Bit 11: EBF status	Pixel is not recognized as	Pixel is recognized as
	Evergreen Broadleaf Forest	Evergreen Broadleaf Forest
Bit 12: BS status	Pixel is not recognized as	Pixel recognized as Bare Soil
	Bare Soil	
Bit 13: Climatology	Not filled	Filled with climatology
Bit 14: Gap filling	Not filled	Filled with interpolation

4.3 DETAILED DESCRIPTION

The several steps of the algorithm are presented here.

4.3.1 Daily LAI, FAPAR and FCover estimates (Branch A)

The Branch A corresponds to a first estimate of daily LAI, FAPAR and FCover (called here Inst. Product_1) and to the preparation of the climatology and the percentiles 5 and 90% of LAI, FAPAR, FCover which are used as background information for the temporal composition in Branch B and Branch C. Some differences exist in Branch A for VGT (Figure 5 and PROBA-V (Figure 6). In particular, the steps 8A and 9A required, respectively, for the computation of the percentiles and the climatology data are specific for VGT (Figure 5). Note that only VGT provides an archive of time series long enough for the computation of this background information. Conversely, steps 2A and 5A are only applied for processing PROBA-V. These steps are required to minimize the differences of the inputs and outputs of the neural networks when run with PROBA-V but trained with VGT data.

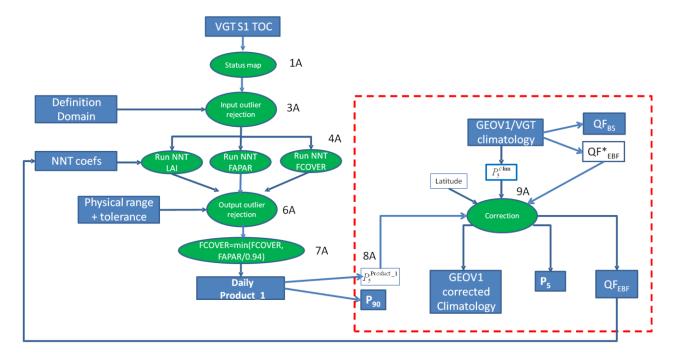


Figure 5: Flow chart describing the Branch A for daily product estimation from VGT.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **35 of 93**

Issue: I1.41



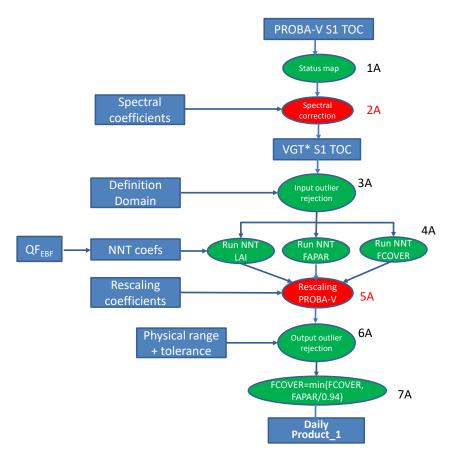


Figure 6: Flow chart describing the Branch A for daily product estimation from PROBA-V. The differences in Branch A between VGT (Figure 5) and PROBA-V are highlighted in red.

4.3.1.1 Rejection of input data based upon their quality status (Step 1A)

The status map plane of VGT and PROBA-V S1-TOC reflectances is first used to keep only the best quality pixels. The neural network is only applied to pixels that are considered valid i.e. if their value in the status map (Table 5 and Table 6) is equal to 248 = 11111000: clear pixels with all the 3 bands (B2, B3 and SWIR) having good radiometric quality, located over land, not covered by ice or by snow. The neural networks are only applied to these pixels.

4.3.1.2 Spectral conversion of PROBA-V reflectances (Step 2A)

This step 2A is only applied when the Version 2 algorithm is run using PROBA-V S1 reflectances. It is not applied when using VGT data as input.

Since the training of neural networks was based on VGT S1 TOC data (Annex 2), VGT-like reflectances are required for the application of neural networks. Because the spectral

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **36 of 93**

Issue: I1.41



characteristics of PROBA-V sensor are slightly different from those of VEGETATION (Table 4), a spectral conversion was applied on the actual PROBA-V TOC reflectances as:

$$\widehat{\rho_{VGT(Bx)}} = \alpha_{Bx} \cdot \rho_{PROBA}(Bx) + \beta_{Bx}$$

where $\rho_{VGT(Bx)}$ is the converted PROBA-V TOC reflectance (called VGT* S1 TOC in (Figure 6) and $\rho_{PROBA}(Bx)$ is the TOC PROBA-V reflectance and α_{Bx} is the conversion coefficient for band Bx (Table 12).

Table 12: Spectral conversion coefficients (α_{Bx} and β_{Bx}) between VEGETATION and PROBA-V sensors [GIOGL1_ATBD_PROBA2VGT]

	B2	В3	SWIR
α_{Bx}	1.001869321	0.998005748	0.986722946
β_{Bx}	0.002362609	0.000112021	0.002070232

4.3.1.3 Input outlier rejection (Step 3A)

To determine if a set of inputs (reflectance in the B2, B3 and SWIR bands, cosine of the view zenith, sun zenith and relative azimuth angle) is valid for the neural networks, the following criteria need to be met (Annex 2):

- Air mass test: $\frac{1}{\cos SZA} + \frac{1}{\cos VZA} \le 5$
- Soil line: Reflectances above the soil line in the B2, B3 and SWIR bands:

$$\rho_{B3} \ge 0.54 \frac{\rho_{B2} - 0.04}{0.5 - 0.04} \quad \text{or} \quad \rho_{SWIR} \ge 0.70 \frac{\rho_{B2} - 0.08}{0.5 - 0.08}$$

where ρ_{B2} , ρ_{B3} and ρ_{SWIR} represent respectively the S1 TOC reflectance for bands B2, B3 and SWIR. The data points lying below the soil line were flagged as outliers due to their high probability of being contaminated by significant fraction of water-bodies or clouds.

The top of canopy reflectances are within the definition domain

If any of these criteria is not met, the corresponding input values are considered as outliers. This should allow rejecting cloud/snow/water/directional contaminated values.

4.3.1.4 Daily estimates of LAI, FAPAR and FCover using neural networks (Step 4A)

Two specific neural networks for EBF and nonEBF were previously calibrated for each of the 3 variables considered (LAI, FAPAR, and FCover) (see Annex 2). The EBF NNTs are then used for pixels identified as EBF (QF_{EBF} =1) based on GEOCLIM (See Section 4.3.1.9) and the nonEBF NNTs are used for the nonEBF pixels (QF_{EBF} =0). They are then applied to each individual

Document-No. CGLOPS1_ATBD_LAl1km-V2 © C-GLOPS Lot1 consortium Issue: I1.41 Date: 11.03.2019 Page: 37 of 93

Date Issued: 11.03.2019

Issue: I1.41



observation (one pixel at a given date) using the S1 top-of-canopy reflectances in the red, NIR and SWIR bands and the cosine of the three angles characterizing the sun and view directions to estimate the corresponding daily LAI, FAPAR and FCover values.

4.3.1.5 Rescaling the PROBA-V outputs of the networks (Step 5A)

This step 5A is only applied when the Version 2 algorithm is run using PROBA-V S1 reflectances. It is not applied when using VGT data as input.

Since the neural networks were trained with VGT S1 TOC reflectances, when applied to PROBA-V S1 data, the PROBA-V V2 outputs are scaled with respect to VGT V2 outputs to reduce marginal discrepancies. For each product, we fitted a third order polynomial function of each product over the median of the residuals, using the learning dataset (Annex 3). The neural network outputs *P* have thus to be rescaled using the following equation with the coefficients of Table 13.

$$P_{step5A} = P_{step4A} - (a_3 P_{step4A}^3 + a_2 P_{step4A}^2 + a_1 P_{step4A} + a_0)$$

Table 13: Coefficients used to rescale the neural network PROBA-V outputs issued from step 4A.

Note that step 5A is not applied for VGT which is equivalent to set all the coefficients a₀, a₁, a₂, a₃ to zero.

	LAI	FAPAR	FCover
a_3	-0.0137	-0.6921	-0.6148
a_2	0.0774	0.9837	0.7901
a_1	0.0947	-0.3290	-0.1548
a_0	-0.0640	0.0064	0.0022

4.3.1.6 Output outlier rejection (Step 6A)

Daily values that were outside the physical range of variation of the variables extended by the tolerance limits (Table 7) were rejected. Values that were within the tolerance limits but higher (lower) than the physical maximum (minimum) (Table 9) were fixed to the physical maximum (minimum).

4.3.1.7 Ensuring FAPAR-FCover consistency (Step 7A)

The FAPAR value divided its maximum theoretical value (0.94) must remain higher than FCover in any situations since it can be approximated by the FIPAR (Fraction of Intercepted Photosynthetically Active Radiation). Indeed, FIPAR is the complementary of the gap fraction at a higher angle than nadir (that corresponds to FCover). Therefore, to avoid possible physical

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **38 of 93**

Date Issued: 11.03.2019

Issue: I1.41



inconsistencies between the estimated FAPAR and FCover values, a final condition must be set to the outputs of neural networks:

$$FCOVER = \min\left(FCOVER_{NNET}, \frac{FAPAR_{NNET}}{0.94}\right)$$

This way a corrected FCover is generated ensuring the consistency with FAPAR.

The resulting filtered time series are the daily LAI, FAPAR and FCover Product_1 estimates.

4.3.1.8 Computation of percentiles P5 and P90 of daily products (Step 8A)

This step 8A is only applied when the Version 2 algorithm is run in offline (HIST) mode for processing VGT time series. Its outputs are used for PROBA-V processing.

For each of the LAI/FAPAR/FCover variables, the percentile 5% of daily Product_1 estimates is computed for each pixel: $P5^{Product_1}$

P90 is defined as the percentile 90 of daily LAI Product_1, $P90 = P90^{Product_1}$.

P5^{Product_1} is an input of step 9A for the computation of P5. P5 and P90 are subsequently used in Steps 1B and 3B of Branch B and Steps 1C and 3C of Branch C.

4.3.1.9 Preparation of the LAI, FAPAR and FCover climatology (Step 9A)

This step 9A is only applied when the Version 2 algorithm is run in offline (HIST) mode for time series of at least 5 years (in practice, when using VGT S1 reflectances but not PROBA-V).

The GEOCLIM (Verger et al. 2015) product, a global climatology of LAI, FAPAR, and FCover from Version 1 (GEOV1) VGT products for 1999–2010 and the quality flags identifying EBF and bare soil areas (Annex 1: GEOCLIM, a climatology of GEOV1 LAI, FAPAR and FCover) are used as an input of the Version 2 algorithm as described in section 4.1.3. Some preparatory steps are required to ingest the GEOCLIM climatology in the processing chain as described in

Figure 5. First, the quality flag QF_{EBF} needs to be checked and corrected when necessary. Second, the P5 parameter needs to be computed for each pixel based on the percentiles 5% of the daily Product_1 estimates (step 8A) and the GEOCLIM values. To correct possible bias between the climatology and the daily estimates, the GEOCLIM product is then adapted to the available daily Version 2 VGT estimates for northern latitudes during the winter period. A daily climatology is finally computed from the corrected GEOCLIM product based on linear interpolation. These steps are described hereafter more in detail:

CHECKING THE LANDCOVER QUALITY FLAG

The quality flags QF_{EBF} and QF_{BS} resulting from the Version 1 VGT climatology are used in subsequent steps of Branches B and C. A high agreement with the GLOBCOVER landcover class was found for EBF and BS classes (compare Figure 37a and Figure 37b). However, very few

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **39 of 93**

Date Issued: 11.03.2019

Issue: I1.41



pixels were misclassified as EBF in the West coast of USA due to possible artefacts in Version 1 VGT climatology (Figure 37a). To prevent from misclassification problems observed at northern latitudes, the QF was forced to 0 for pixels at northern latitudes higher than $Lat_{EBF}^{max} = 28.5^{\circ}$ since EBFs are not expected in these locations:

$$QF_{FRF} (Lat > Lat_{EBF}^{max}) = 0$$

COMPUTATION OF P5

For each of the LAI/FAPAR/FCover variables, the P5 values are computed as the minimum values between P5^{clim} and P5^{Product_1}:

where P5^{clim} is the percentile 5 of the GEOCLIM products and P5^{Product_1} is the percentile 5 of daily Product_1 estimates (step 8A).

P5 and P90 (step 8A) are subsequently used in Steps 1B and 3B of Branch B and Steps 1C and 3C of Branch C.

ADAPTATION OF GEOCLIM TO DAILY VERSION 2 VGT ESTIMATES FOR NORTHERN HIGH LATITUDES

The specific correction applied to GEOCLIM (Annex 1) removed the instabilities in the solution but cannot correct possible biases in the magnitude of original Version 1 LAI, FAPAR, FCover products used as input data for GEOCLIM. For example, previous studies have also shown that Version 1 products produce slightly higher values than MODIS for needleleaf forest in winter (Fang et al. 2013). The specific correction applied in GEOCLIM at northern high latitudes (Annex 1) reduced these differences but may still result in some overestimation of the seasonal amplitude in winter time. Accurate estimation of LAI/FAPAR/FCover in needleleaf forests in winter is challenging because contamination by clouds and snow limits the reliability of the reflectances used as inputs in the algorithms (Camacho et al. 2013). Further, the strong bidirectional effects of surface-reflectance at very high latitudes are not well simulated by the radiative transfer models currently used for product generation (Yang et al. 2006). In addition, the understory and foliage clumping are not well accounted for (Jiao et al. 2014; Pisek et al. 2010).

The possible bias between GEOCLIM and Version 2 VGT daily estimates may introduce some artefacts in the final solution of Version 2 products. To correct the overestimation of GEOCLIM as compared to Version 2 VGT, the GEOCLIM LAI/FAPAR/FCover values are adapted to the daily Version 2 VGT estimates for high northern latitudes (latitude> 40°) during the winter time (defined here as the period for which the sun zenith angle $SZA > SZA^{HL}$ with $SZA^{HL} = 70^{\circ}$). The GEOCLIM LAI/FAPAR/FCover products higher than the percentile P5 are fixed to P5 for pixels at latitudes higher than a threshold corresponding to the latitude for which the $SZA > SZA^{HL}$ at the time of VGT overpass:

 $GEOCLIM(Lat > Lat_{SZA}^{min} \& GEOCLIM > P5) = P5;$

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **40 of 93**

Issue: I1.41



where Lat_{SZA}^{min} is the threshold of latitude which depends on the dekad (from 1 to 36) of the year:

INTERPOLATION AT THE DAILY TIME STEP OVER THE ENTIRE PERIOD

The 36 dekadal climatology is repeated for all the years of the time series extended by 1 year on each side to prevent border effects. Finally, this dekadal climatology is linearly interpolated at the daily time step for further use as background information for gap filling.

4.3.2 Processing the past series (Branch B)

The flow chart of the processing of past series is depicted in Figure 7.

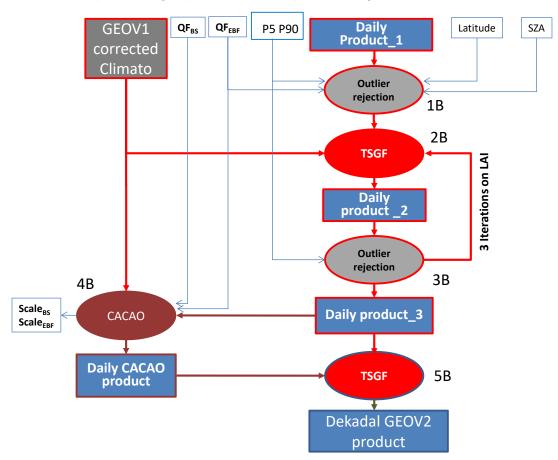


Figure 7: Flow chart describing the processing of the past series (Branch B). *Daily Product_1*, the sun zenith angle (*SZA*) of the observations are coming from Step 7A, the parameters P90 from Step 8A and P5 from Step 9A. *GEOV1 corrected climatology* and the quality flags QF_{EBF}/QF_{BS} / Latitude are ancillary information described Annex 1: GEOCLIM, a climatology of GEOV1 LAI, FAPAR and FCover.

Document-No. CGLOPS1_ATBD_LAl1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **41 of 93**

Date Issued: 11.03.2019

 $else\ P_{sten1B} = P_{sten3A}$

Issue: I1.41



4.3.2.1 Second outlier rejection (Step 1B)

For this second outlier rejection, emphasis was put on LAI products that show the highest sensitivity to possible problems in reflectance values. Therefore, when an outlier is detected on LAI data, it is also considered as an outlier for FAPAR and FCover to keep a high level of consistency between the three variables.

We filter the noisy data based on expert knowledge of the expected seasonality:

- For the high northern latitude ($Lat > Lat_{winter}^{min}$, with $Lat_{winter}^{min} = 55^{\circ}$), the LAI (FAPAR, FCover) values in winter time ($SZA > SZA^{HL}$ where SZA is the sun zenith angle associated to daily VGT observations and SZA^{HL} , with $SZA^{HL} = 70^{\circ}$, being the minimum value in winter time) are expected to be relatively stable and low due to the low temperatures, short days, and low illumination. However, observations are affected by snow cover or very poor illumination conditions that introduce a positive bias in the LAI estimates. For the pixels at $Lat > Lat_{winter}^{min}$ in winter time, $SZA > SZA^{HL}$, a specific correction is applied and the corresponding flag QC(10) in Table 11 is set to 1. In this case, the LAI values >P5 and $>LAI_{winter}^{min}$, with $LAI_{winter}^{min} = 0.5$ being the default minimum value of LAI in winter time, are considered as outliers and rejected.
- For pixels identified as EBF (QC(11)=1 in Table 11), a minimum seasonality and high values of LAI are assumed. The observed artifacts in EBFs are mostly associated to the high cloud cover observed in the equatorial and tropical latitudes which introduce a negative bias in LAI. The LAI values <P90 and <LAI_{EBF}^{min}, with LAI_{EBF}^{min} = 5.5 being the minimum default value of LAI in EBFs are rejected.

```
 \begin{cases} if Lat > Lat_{winter}^{min} \ and \ SZA > SZA^{HL} \ and \ LAI > P5 \ and \ LAI > LAI_{winter}^{min}, \ P_{step1B} = invalid \ value \\ if \ QF_{EBF} = 1 \ and \ LAI < P90 \ and \ LAI < LAI_{EBF}^{min}, \ P_{step1B} = invalid \ value \end{cases}
```

4.3.2.2 Temporal smoothing and gap filling (TSGF) (Step 2B)

The daily estimates ("daily Products_1" in Figure 7) after the first outlier rejection are smoothed and gap filled using the same techniques (Verger et al. 2011) as the ones considered for processing the climatology (Annex 1: GEOCLIM, a climatology of GEOV1 LAI, FAPAR and FCover). This is achieved similarly at the dekadal time step with however some particularities that account for the important noise associated to the S1 daily estimates (Figure 8):

Adjusting the length of the compositing window to get 6 observations on each side
of the dekadal date. After evaluating several widely used temporal filters, a simple but
robust method based on the adaptive Savitzky-Golay (SG) filter (Savitzky and Golay, 1964;
Chen et al., 2004) was selected to smooth the climatology. The fixed and symmetric
smoothing compositing window of the standard SG polynomial fitting method was replaced

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **42 of 93**

Date Issued: 11.03.2019

Issue: I1.41



by an adaptive process with variable length and asymmetric (in time) compositing window. The minimum number of daily estimates within each side of the dekadal date being smoothed was fixed to n_{TSGF}^{min} with $n_{TSGF}^{min}=6$. The length of the window is therefore variable, depending on the available observations in the vicinity of the dekadal date considered. However, a maximum length of the half compositing window, $length_{TSGF}^{max}=60$ days, was used to allow adaptation to the local variations in the data. Further, a minimum length of the half compositing window was imposed to $length_{TSGF}^{min}=15$ days even if more than n_{TSGF}^{min} observations exist within each semi-period since it increases the robustness of the fitting.

- Use climatology to fill values. If the number of available daily estimates in the $length_{TSGF}^{max}$ semi-compositing window is lower than n_{TSGF}^{min} , the corresponding flag QC(3) in Table 11 is set to 1 and a gap filling procedure is activated. If the climatology is available, it is used to fill values and the corresponding flag QC(13) in Table 11 is set to 1. In this case, the available estimates are systematically completed with n_{TSGF}^{min} climatology values located every $\frac{length_{TSGF}^{max}}{n_{TSGF}^{min}} = 10 \ days$ evenly distributed over the $length_{TSGF}^{max}$ day period. The considered climatology for filling gaps is a daily climatology that results from the linear interpolation of the original dekadal climatology as described in step 9A (Section 4.3.1.9). If there are no climatology values, this results in a missing value at the considered dekadal date.
- *Fit polynomial model.* A weighted polynomial fitting is applied with weighting factors, *W*, computed according to the distance of the daily estimates to the TSGF outputs derived from the previous iteration. A sigmoidal function was considered for computing the weights *W* with less weight associated to the values smaller than the previous TSGF estimates since the low values have more chances to be contaminated by residual clouds or snow.

$$W = 2/(1 + \exp(-2 * \delta))$$

Where δ is the difference between the daily estimates and the TSGF outputs, $\delta = Product_1 - TSGF$. To put less emphasis on the climatology values used to fill gaps, weights of the climatology fill values were multiplied by a scale factor of $Scale_W^{Climato} = 0.5$, i.e. $W^{Climato} = Scale_W^{Climato} * W^{Climato}$. For the first iteration of TSGF, weights of daily estimates were fixed to $W^{Product_1}_{iter=1} = 1$ and to $W^{Climato}_{iter=1} = 0.5$ for the climatology.

• Use interpolation to fill values. Finally, we use interpolation to fill the residual gaps. The corresponding flag QC(14) in Table 11 is set to 1. A simple linear interpolation based on a local moving window of $\pm length_{TSGF}^{max}$ days was applied to fill gaps in the time series. To improve the efficiency of gap filling, an iterative process ($niter_{gap} = 2$ iterations) was considered: the gain of available data achieved with the first iteration allows improving robustness and continuity in the gap-filled data through a second iteration. Gaps longer than $2 * length_{TSGF}^{max} = 120$ days are not filled and will result in missing data.

Finally, to avoid possible artifacts introduced by divergences of the temporal filters being applied, the TSGF values were forced to the physical range of variation of the LAI, FAPAR and FCover (Table 9).

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **43 of 93**

Date Issued: 11.03.2019

Issue: I1.41



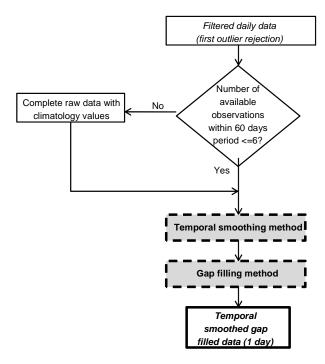


Figure 8: Flow chart showing how the temporal smoothing gap filling (TSGF) algorithm works.

4.3.2.3 Third outlier rejection (Step 3B)

Similarly to the second outlier rejection, emphasis was put on LAI product for identifying outliers. The daily LAI values (daily Products_2 in Figure 7) are filtered by comparison between LAI values and the temporal smoothed and gap filled (TSGF) LAI series. A point under (respectively over) the TSGF estimate (interpolated at daily step from the dekadal values) is considered as an outlier if its minimum absolute distance to all the TSGF LAI values within a $\pm length_{Outlier} = \pm 15 \ day$ window is greater than $tol_{Outlier}^{abs} = 0.10$ and the $tol_{Outlier}^{rel} = 15\%$ of the TSGF value. Considering a ± 15 -day window prevents from eliminating too many observations during periods of high rate of variation of LAI.

Similarly, to prevent rejecting too many low values in the base level due to a possible limitation of TSGF to fit the data in these regions, the estimates that are within $\pm tol_{outlier}^{base} = \pm 0.5$ of the base level and having a distance lower than $tol_{outlier}^{TSGF} = 0.5$ to TSGF are considered valid. The base level is defined as the maximum of percentile-5 (P5) and a threshold fixed to $LAI_{outlier}^{base} = 0.5$. This last condition to avoid removing valid data in the base level is only applied when the percentile-90 of the data is higher than $P90_{outlier}^{min} = 0.5$.

The process is repeated *niter* times, *niter* being the number of iterations used for TSGF (*niter* was fixed to 3). To avoid rejecting false outliers the upper values are only filtered in the last iteration. The more restrictive criteria imposed here to the values under TSGF upper envelope accounts for the effect of residual clouds that yield systematically lower values than the realistic LAI.

To apply the third outlier rejection, the following steps must be completed:

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **44 of 93**

Issue: I1.41



• For the LAI values lower than the TSGF upper envelope, the following conditions are applied to filter outliers in the *niter* iterations:

$$\text{If } \left\{ \underbrace{\text{Min} \left(\left| LAI_{t} - \left(\overrightarrow{TSGF}_{t-15,t-4\dots t+15} \right) \right| \right)}_{LAI_{t}} < TSGF_{t} \right) > \max \left(tol_{outlier}^{abs} =, tol_{outlier}^{rel} = *TSGF_{t} \right) \right\}, P_{step3B} = invalid \ value = tol_{outlier}^{rel} =$$

$$else \ if \left\{ \begin{aligned} & \underbrace{LAI_{t} < TSGF_{t}} \\ \min(\left| LAI_{t} - \left(\overrightarrow{TSGF_{t-15,t-4\dots t+15}} \right) \right|) > \max(tol_{outlier}^{abs} =, tol_{outlier}^{rel} = *TSGF_{t}) \\ & P90 > P90_{outlier}^{min} \\ & \left| LAI_{t} - \max(P5, LAI_{outlier}^{base}) \right| < tol_{outlier}^{base} \\ & \left| LAI_{t} - TSGF_{t} \right| < tol_{outlier}^{TSGF} \end{aligned} \right\}, P_{step3B} = P_{step2B}$$

else
$$P_{step3B} = P_{step2B}$$

• For the LAI values higher than TSGF, the following conditions are applied to filter outliers in the last iteration:

$$\label{eq:lambda} \begin{split} \text{If} \left\{ & \underbrace{LAI_t > TSGF_t}_{\min\left(\left|LAI_t - \left(\overrightarrow{TSGF_{t-15,t-4\dots t+15}}\right)\right|\right) > \max\left(tol_{Outlier}^{abs}, tol_{Outlier}^{rel} * TSGF_t\right)} \right\}, P_{step3B} = invalid \ value \\ & else, \\ & P_{step3B} = P_{step2B} \end{split}$$

For FAPAR and FCover variables, the Step 3B is not applied but an outlier detected on LAI is also rejected on FAPAR and FCover. For pixels identified as EBF, the Step 3B is not applied and a specific procedure is applied in Step 1B (4.3.2.1) for removing estimates contaminated by residual clouds. The performance of this automatic method is illustrated in Figure 9 which shows the improvements in TSGF outputs through the proposed iterative process.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **45 of 93**

Date Issued: 11.03.2019

Issue: I1.41



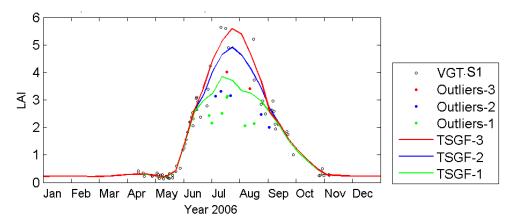


Figure 9: Illustration of the 3-iterations of TSGF filtering (continuous line) to eliminate contaminated data (filled circles). Empty circles correspond to valid data.

4.3.2.4 Consistent Adjustment of the Climatology to Actual Observations (CACAO, Step 4B)

Adjusting climatological patterns to actual observations was shown to efficiently capture the product dynamics and the inter-annual anomalies while filling gaps and smoothing biophysical products in a robust way (Baret et al. 2011; Verger et al. 2013). CACAO consists in fitting the climatology to the daily estimates ("daily Product_3" in Figure 7) by shifting and scaling the climatology values over portions of the seasonal cycle (sub-seasons). The CACAO process is applied to each sub-season and will result in estimates at the daily time step. A sub-season is defined by the period between two consecutive extrema in the climatology. The following steps are thus followed:

- **Decompositing the climatology into sub-seasons.** The global extrema (minima and maxima) points from the climatology time series should be identified first. However, to exclude possible false extrema in the time series due to the effect of residual noise in the signal, extrema values that differ by less than $tol_{CACAO-LAI}^{abs}=0.10$ for LAI $(tol_{CACAO-FAPAR-FCOVER}^{abs}=0.025$ for FAPAR and FCover) or the $tol_{CACAO}^{rel}=15\%$ of the median value of the climatology are excluded. The sub-seasons are slightly extended by considering the minimum extra time window before and after the period containing either $\Delta_{CACAO}^{Amplitude}=30\%$ of the season amplitude or $\Delta_{CACAO}^{Period}=30\%$ of the period length (in days) of the adjacency sub-seasons. This allows more robust fit by providing clearer temporal features on which the adjustment could grasp.
- **Shifting and scaling the climatology for each sub-season.** The climatology is fitted to the actual daily estimates ("daily products_3" in Figure 7) for each year and each sub-season. The daily climatology P^{clim} was fitted to the daily P data by considering a scale factor, scale, and a temporal shift, shift:

$$\widehat{P}(t) = scale \cdot P^{clim}(t + shift)$$

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **46 of 93**

Date Issued: 11.03.2019

Issue: I1.41



If \widehat{P} follows the usual climatological pattern, P^{clim} , scale = 1 and shift = 0. The two parameters (scale, shift) are found by minimizing the cost function defined by the RMSE between daily estimates and the calculated values after fitting the climatology.

The shift parameter is allowed to vary between $-shift^{max} < shift < shift^{max}$ with $shift^{max} = 60$ by steps of $Step_{CACAO} = 5$ days which results in 25 adjustments between the shifted climatology and the P data.

Two conditions are required for applying CACAO: (i) a minimum number of available daily Product_3 estimates representing $Thr_n^{sub-season}=10\%$ of the potential estimates in the subseason given by the length in days of the sub-season, $length^{sub-season}$, and (ii) the amplitude of available Product_3 estimates in the sub-season, $Amplitude_{Product_3}^{sub-season}$, represent a threshold of $Thr_{Amplitude}^{sub-season}=30\%$ of the amplitude of the climatology in the subseason, $Amplitude_{Climato}^{sub-season}$. Otherwise, the original climatology is considered as a backup solution.

$$| f \begin{cases} n_{Product_3}^{sub-season} \geq Thr_n^{sub-season} * length^{sub-season} \\ Amplitude_{Product_3}^{sub-season} \geq Thr_{Amplitude}^{sub-season} * Amplitude_{Climato}^{sub-season} \end{cases}, P_{step_4B}^{sub-season} = CACAO^{sub-season}$$

$$e | SeP_{step_4B}^{sub-season} = Climato^{sub-season}$$

For pixels identified as bare soil or EBF with almost no seasonality, CACAO is applied if the number of available Product_3 estimates over the whole time time series, $n_{Product_3}$, is higher than a minimum of $n_{CACAO}^{BS-EBF}=10$ daily estimates. In that case the climatology is adjusted over the whole time series and the resulting scale factor is kept for further use in Branch C. Otherwise, the original climatology is considered as a backup solution.

$$if \ n_{Product_3} \ge n_{CACAO}^{BS-EBF}, P_{step4B}^{BS-EBF} = CACAO$$

$$else \ P_{step4B}^{BS-EBF} = Climato$$

In the transition between sub-seasons, a weighted average between \hat{P} estimates from the two sub-seasons is considered as the final solution. A linear weight contribution varying between 1 and 0 (between 0 and 1) was assigned to the first (second) sub-season estimates in the overlapping period.

• To avoid problems at the beginning and the end of the time series, an extended period of $\Delta_{month} = 6$ months before and $\Delta_{month} = 6$ months after the date being processed is required for consistently adjusting CACAO.

Finally to avoid possible artifacts introduced by divergences of the temporal filters being applied, the CACAO values were forced to the physical range of variation of LAI, FAPAR and FCover (Table 9).

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **47 of 93**

Issue: I1.41



4.3.2.5 Computation of the dekadal Version 2 LAI, FAPAR, FCover products (Step 5B)

TSGF is finally applied over the filtered daily estimates ("daily products_3" in Figure 7) at each dekadal step for generating the final Version 2 products. $n_{TSGF}^{min} = 6$ observations within a maximum period of $length_{TSGF}^{max} = 60$ days and a minimum period of $length_{TSGF}^{min} = 15$ days on each side of the dekadal dates are necessary. The CACAO product is used to fill gaps in the time series in place of the standard climatology as it was used previously in the Step 2B. Since CACAO is expected to be closer to the data as compared to the original climatology, the use of CACAO helps in the estimation of the final product over period with missing observations.

The outputs of Step 5B are the final temporal smoothed and gap filled 10-day V2 LAI, FAPAR and FCover historical products. When the LAI, FAPAR and FCover values are out of range or invalid, the corresponding flag QC(7), QC(8) and QC(9) in Table 11 are set to 1.

4.3.3 Real time estimates (Branch C+)

Branch C (Figure 10) is identical to the Branch B (Figure 7) described previously for the processing of historical series. Only some operational considerations need to be specified.

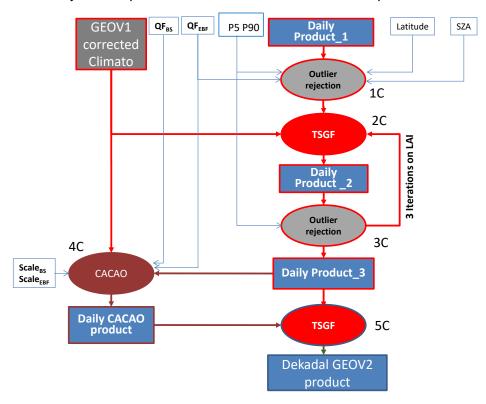


Figure 10: Flow chart describing the processing of the real-time series (Branch C). *Daily Product_1* and SZA are coming from Step 7A, the parameters P90 from Step 8A, P5 from Step 9A, and the *Scale_{BS}* and *Scale_{EBF}* from Step 4B. The corrected climatology and the quality flags QF_{EBF}/QF_{BS} /Latitude are ancillary information described in Annex 1: GEOCLIM, a climatology of GEOV1 LAI, FAPAR and FCover.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **48 of 93**

Date Issued: 11.03.2019

Issue: I1.41



4.3.3.1 Differences between branches B and C

Some of the parameters used for several processing steps are computed within data preparation phase (Branch A) or the processing of historical data (Branch B). This concerns:

- the QF indicating that a pixel is identified as EBF (BS) and used in Step 1C and 4C results from the GEOV1 climatology correction (Annex 1: GEOCLIM, a climatology of GEOV1 LAI, FAPAR and FCover).
- P5, i.e. the 5% percentiles of LAI/FAPAR/FCover, and P90, i.e. 90% percentile of LAI for filtering the outliers (Steps 1C and 3C) are also pixel specific and were computed previously in Steps 8A and 9A of branch A (see sections 4.3.1.9 and 4.3.1.8).
- the Scale_{BS} and Scale_{EBF} parameters used in CACAO (Step 4C) are pixel specific and were computed in Step 4B of Branch B (see 4.3.2.4).
- the corrected GEOV1 climatology at daily step was generated as described in Step 9A of Branch A (see 4.3.1.9).

The dissymmetry in the observations with regards to the date for which the product has to be computed is solved using the climatology values similarly as in the Branch B for gap filling. The corrected GEOV1 climatology is used here in the *niter* iterations of Step 2C and CACAO in the last iteration at Step 5C. As CACAO is expected to be closer to the observations, it will help the temporal projection.

The product value for the date dx is estimated in Step 5C as the output of TSGF applied to the filtered daily estimates using CACAO output for filling gaps.

An extended period of $\Delta_{month} = 6$ months before and $\Delta_{month} = 6$ months after the date being processed is required for consistently adjusting CACAO (Step 4C).

For pixels identified as EBF or BS, the climatology is not fitted to the data in Step 4C but CACAO is estimated by multiplying the climatology value by the scale factors $Scale_{BS}$ and $Scale_{EBF}$ computed in Step 4B when processing historical VGT time series.

For the TSGF application in Step 2C and Step 5C,

- the past period is defined as the minimum $length_{TSGF}^{min}=15$ day and maximum $length_{TSGF}^{max}=60$ day period with enough $n_{TSGF}^{min}=6$ valid daily estimates for TSGF processing.
- the future period for the daily estimates (no data in the future period) is defined as $length_{TSGF}^{max}$ period filled with the n_{TSGF}^{min} climatology values located every $\frac{length_{TSGF}^{max}}{n_{TSGF}^{min}} = 10$ days evenly distributed over this period.

The outputs of Step 5C are the final temporal smoothed and gap filled 10-day V2 LAI/FAPAR/FCover near real time products. When the LAI, FAPAR and FCover values are out of range or invalid, the corresponding flag QC(7), QC(8) and QC(9) in Table 11 are set to 1.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **49 of 93**

Issue: I1.41



4.3.3.2 Defining the consolidation period

The C+ branch is applied first over the dx-n to dx period, n being the number of past dekads. Then, the time window of n dekads is shifted by one dekad each time real time estimates have to be provided for a new dekad. This process results in n updates of the products value. The last update is considered as the consolidated value. The difference between product estimates from Branch B and Branch C+ (Figure 11) shows that after 6 dekads the real time estimates (Branch C+) converges closely towards the historical processing (Branch B). As a consequence, it was decided to consider the product as consolidated after 6 dekads. This means that there will be 6 consecutive updates of the product value.

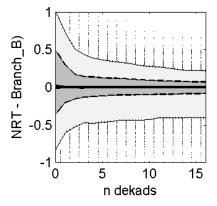


Figure 11: Evaluation of the differences between Version 2 real time (Branch C+) and historical (Branch B) estimates over the BELMANIP2 sites for the year 2008 as a function of the number n of dekads after the date being processed. Zero dekad (n=0) corresponds to the daily estimates with data available only for the past. The several gray values correspond to 75% (dark gray), 90% (medium gray) and 95% (light gray) of the population, and the dots to 5% percentile of residual outliers. The bold back solid line corresponds to the median value of the differences. The dotted line is the 0:0 line. Case of LAI.

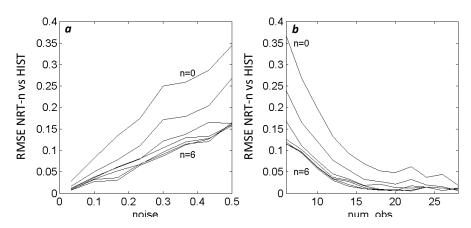


Figure 12: RMSE between NRT-n and HIST Version 2 VGT LAI estimates as a function of the noise in the data and the number of observations (Adapted from Verger et al. (2014a)).

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **50 of 93**

Issue: I1.41



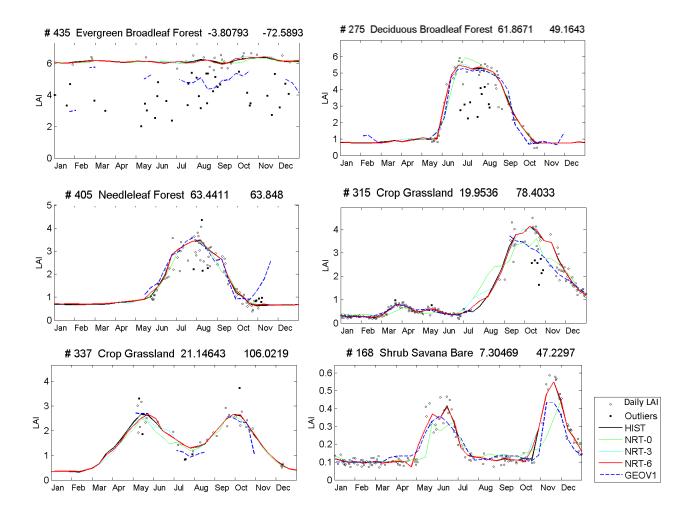


Figure 13: Temporal profiles of near real time LAI estimates for n=0,3,6 dekads after the date being processed (NRT-n) as compared to Branch B estimates (HIST) over several BELMANIP2 sites for the year 2005. The original daily LAI data, filtered outliers, the V1 of Collection 1km products (GEOV1) product are also shown. The title of each plot indicates the BELMANIP2 site number, the GLOBCOVER biome class, the latitude and longitude in degrees.

Detailed investigation of the temporal profiles corresponding to the several updates (Figure 13) confirms that after 6 dekads the differences between real time estimates and historical series are marginal except in particular situations with few observations and scattering in the daily estimates. In such conditions, the first real time estimates (NRT-0) with data only in the past shows some discrepancies and higher instability in the solution compared with the processing of the historical series. An exhaustive sensitivity analysis of NRT-n (from the initial solution NRT-0 to the consolidated product NRT-6) to the number of observations and data noise is provided in Verger et al. (2014a). Results confirm a rapid convergence for the intermediate solutions towards HIST which show a similar pattern after two dekads (NRT-2) (Figure 12). Since the product value after the second consolidation NRT-2 remains stable, the third, fourth and fifth consolidations are not distributed; only NRT-0, NRT-1, NRT-2 and NRT-6 are distributed.

Document-No. CGLOPS1_ATBD_LAI1km-V2

© C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **51 of 93**

Date Issued: 11.03.2019

Issue: I1.41



4.3.4 Processing the first dekads of the past series (Branch C-)

Branch C is run in reverse mode (C-) over the d1 to d1+n period. This allows accounting for the missing past data before d1 while keeping consistent with the processing of present observations.

4.3.5 Computation of the associated quality indicators

The following quality indicators associated to each Version 2 product value, $P^{GEOV2}(d)$, are computed:

- NOBS: the number of valid daily estimates in the composition period that are used to compute the product value, P^{GEOV2}(d) (Step 5B or 5C). Note that NOBS accounts only for the valid daily estimates from actual observations and not for the climatological-filled values. The more observations, the more reliable are the products. If NOBS = 0, the corresponding QC(6) in Table 11 is set to 1.
- LENGTH_BEFORE and LENGTH_AFTER: the length in days of the semi-periods of composition between the dekad being processed and the outermost daily estimate or climatological-filled value in the compositing window, which have been required to find $n_{TSGF}^{min} = 6$ valid estimates (see section 4.3.2.5). The shorter the semi-periods of composition, the more reliable are the products.
- *RMSE*: the uncertainty of the product value. Two cases are considered:
 - o If $NOBS \ge n_{RMSE}^{min}$ (the default value for the minimum number of required available daily estimates, n_{RMSE}^{min} , is set to 2), the uncertainties are computed as the RMSE between the final dekadal value, $P^{GEOV2}(d)$, and the daily estimates, P(i), in the compositing period:

$$RMSE(d) = \sqrt{\frac{\sum_{i=1}^{NOBS} (P^{GEOV2}(d) - P(i))^{2}}{NOBS}}$$

Note that only valid daily estimates from actual observations are used for the computation of the RMSE. Climatological values are not used.

o If $NOBS < n_{RMSE}^{min}$, the RMSE is not computed and it results in a missing value (RMSE(d)) is set to 255) for the uncertainty at the dekad d.

The lower the RMSE, the more reliable are the products.

4.4 LIMITATIONS

Version 2 algorithm capitalizes on the development and validation of already existing products: CYCLOPES version 3.1 and MODIS collection 5 and the use of neural networks. The CYCLOPES and MODIS products used in the training dataset, the efficacy of the training process and the criteria used to define the input outliers will determine, respectively, the magnitude and range of

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **52 of 93**

Date Issued: 11.03.2019

Issue: I1.41



variation of the final products, their reliability and the definition domain used to remove contaminated values.

The final product is also dependent on the criteria used to filter the output outliers, particularly for the tropical forests and high latitudes. Outlier rejection constitutes a critical step in the algorithm. The efficiency of the temporal methods used in the composition of products from daily estimates depends on the level of noise and gaps in the time series and on the reliability of the auxiliary data (climatology) used as a background information to fill gaps.

The main identified limitations are associated to the input data:

- The consistency across missions of long time series of Version 2 products depends mainly on the temporal consistency of input TOC reflectances between sensors and across the different processing chains (e.g. cloud-aerosol screening). Since the Version 2 neural networks were trained over the VGT-S1 TOC inputs, the final estimates resulting from the application of networks to PROBA-V data depend on the coefficients of the spectral correction applied to PROBA-V reflectances to mimic VGT-S1. An adaptation of the rescaling of Version 2 PROBA-V with respect to Version 2 VGT estimates will be necessary after the planned reprocessing of PROBA-V that includes a recalibration of sensors and a correction of cloud screening.
- When the climatology derived from Version 1 VGT products is missing, no ancillary information is available to fill gaps in Version 2 time series or to make NRT projections.
- Despite the specific corrections applied to the original Version 1 VGT climatology, remaining artifacts and possible biases in the magnitude and seasonality of LAI/FAPAR/FCover in winter time may limit the reliability of the final Version 2 products at very high northern latitudes.
- The approach used to process pixels flagged as evergreen broadleaf forest (EBF, QC(11)=1 in Table 11) constitutes an oversimplification of the reality because of the possible seasonality of EBFs. The high uncertainty associated with the data due to poor atmospheric correction and very high cloud occurrence in equatorial and tropical latitudes prevented the extraction of meaningful phenology at the resolution of the individual pixels of 1 km. The high spatial and temporal resolution of Sentinel2 sensors should improve the monitoring of vegetation in these problematic areas.
- In cases of a wrong identification of a pixel as an EBF, Version 2 products only reproduce the high values but not the actual seasonality of the pixel.
- The algorithm uses a static mask for EBF based on the climatology for the period 1999-2010. Consequently, for pixels flagged as EBF, the Version 2 product may not capture deforestation processes. This mask may require to be updated in the future.
- V2 LAI, FAPAR and FCover variables are retrieved over inland waters not discriminated
 as water in the land/sea mask (QC(1)=0 in Table 11). Inland water bodies are mostly
 identified as Bare Soil (BS, QC(12)=1 in Table 11) based on GEOCLIM climatology.
 Although it would bring an improvement to mask the inland waters, it does not exist

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **53 of 93**

Date Issued: 11.03.2019

Issue: I1.41



inland water mask reliable enough to be used without taking the risk to create side effects in the products.

• The values of LAI, FAPAR and FCover over pixels identified as BS (QC(12)=1 in Table 11) are close to zero but not strictly zero. Some users may prefer forcing the values of biophysical variables to zero for pixels flagged as BS.

The user should use the product with due attention to the quality flags values as well as the associated uncertainties (§4.3.5), in particular for areas with long periods of cloudiness.

Document-No. CGLOPS1_ATBD_LAl1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **54 of 93**

Date Issued: 11.03.2019

Issue: I1.41



5 EVALUATION OF THE ALGORITHM

An evaluation of the algorithm has been achieved, mostly based on the temporal consistency as well as on the comparison with Version 1 VGT and MODIS (Myneni et al. 2002) products. This was completed over the BELMANIP2 sites during 2003-2007 period. Finally the agreement between NRT GEOV2 estimates from PROBA-V and VGT data was assessed over the overlapping 2013-2014 period. For further details and evaluation of the Version 2 algorithm, we refer to (Verger et al. 2014a) and (Verger et al. 2014b).

A full quality assessment was performed according to the procedure described in the Service Validation Plan [GIOGL1-SVP]. The results are presented into the validation reports [GIOGL1_QAR_LAI[FAPAR/FCOVER]1km-VGT-V2 and CGLOPS1_QAR_LAI[FAPAR/FCOVER]1km-PROBAV-V2].

In addition, the algorithm developed in the FP7/ImagineS project to be applied on 300m PROBA-V data [ImagineS_RP2.1_ATBD-LAI300m] was applied to SPOT/VGT data to create the so-called "GEOV3/VGT" products used hereafter for the comparison with Version 2 VGT products. The main differences between both algorithms are:

- Different reflectances used as inputs of the neural networks: Version 2 uses Top Of Canopy reflectances in the red, NIR and SWIR spectral bands of VEGETATION (blue band is not used) while GEOV3 uses Top of Aerosols reflectances in the blue, red and NIR VEGETATION bands (SWIR is not used).
- 2. The training dataset of neural networks and the definition domain are different.
- 3. Different neural networks for EBFs and non EBFs are applied in GEOV3 while only one set of networks across biomes are applied in Version 2.
- 4. Different criteria were considered for the identification of EBFs.
- 5. No climatology is used in GEOV3 methodology.

5.1 TEMPORAL PROFILES FOR SELECTED SITES.

The analysis is organized per large biome type, selecting few sites showing typical temporal profiles. For the evergreen broad leaf forest (Figure 14), the effect of residual clouds is very pronounced, creating strongly negative biased estimates of daily products. These are efficiently filtered thanks to the frequency criterion used. This results in a very flat temporal profile with a high level of LAI as expected. As compared to Version 2, Version 1 shows generally lower LAI values, discontinuous and shaky temporal profiles.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **55 of 93**

Issue: I1.41



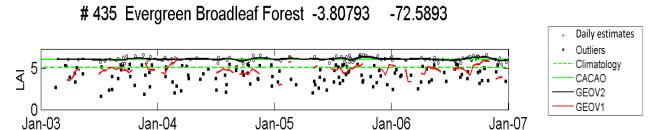


Figure 14: Temporal profile of Version 2 (black solid line) over a typical evergreen broadleaf forest site. Daily estimates derived from VGT-S1 products are indicated by the dots: black squares correspond to outliers. Empty circles to the valid LAI estimates used to compute the Version 2 product. The dashed green line corresponds to the GEOV1 corrected climatology. The solid green line to the CACAO estimates. The red line corresponds to Version 1 VGT product.

For deciduous broadleaf forest (Figure 15), similarly, the negative bias due to cloud contamination is efficiently filtered by the algorithm. For this high latitude site, the winter period is continuously and consistently gap filled although very few observations are available.

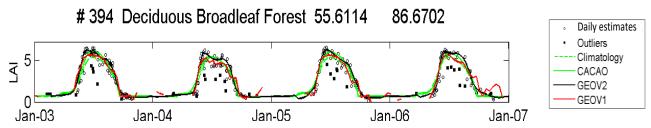


Figure 15: Idem Figure 14 but for a typical Deciduous Broadleaf Forest site.

For this high latitude needle leaf forest (Figure 16), the Version 2 temporal profile is very consistent with that of Version 1 during the vegetation season, with however no interruption during the winter period conversely to Version 1.

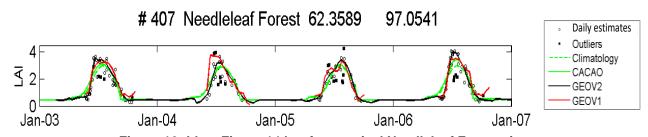


Figure 16: Idem Figure 14 but for a typical Needleleaf Forest site.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **56 of 93**

Issue: I1.41



The two crop-grassland sites (Figure 17) show also a good consistency between Version 1 and Version 2. However, the discontinuities noticed for Version 1 have disappeared in Version 2, thanks to the use of the corrected climatology background information to provide dekadal values in absence of daily estimates. These cases of double seasonality are well described.

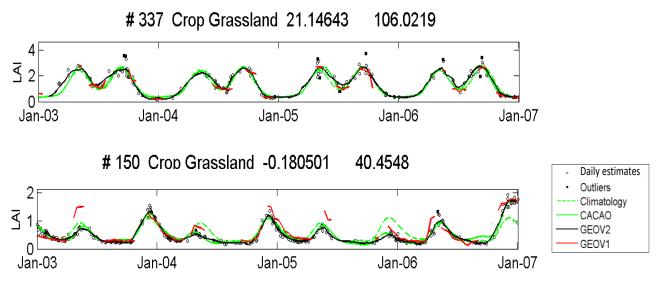
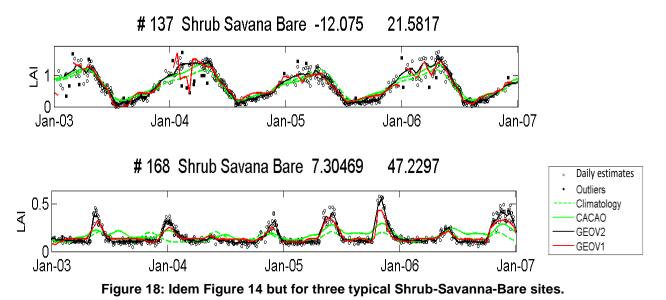


Figure 17: Idem Figure 14 but for two typical Crop-Grassland sites.

The temporal profiles of Shrub-savana-bare soils (Figure 18) are also very well captured by Version 2, with large improvements as compared to Version 1 regarding continuity and smoothness of estimates (e.g. site #137).



Document-No. CGLOPS1_ATBD_LAl1km-V2 © C-GLOPS Lot1 consortium

Issue: I1.41 Date: 11.03.2019 Page: 57 of 93

Date Issued: 11.03.2019

Issue: I1.41



5.2 CONSISTENCY BETWEEN LAI, FAPAR AND FCOVER

Version 2 VGT improves Version 1 VGT, GEOV3/VGT and MODIS products in terms of the internal consistency between LAI, FAPAR and FCover variables. The LAI-FAPAR for the different products show an exponential relationship as expected (Myneni and Williams 1994). For MODIS products (Figure 19), higher scattering is observed and different patterns in the LAI-FAPAR relationship that may be introduced by the biome dependency of the retrieval approach (Myneni et al. 2002). The comparison of LAI-FAPAR for Version 2 VGT shows higher consistency for all the range of values with non-artifacts introduced by the use of two NNTs (one for EBFs and one for non EBFs classes).

The LAI-FCover for Version 2 VGT shows an exponential relationship as expected (Roujean and Lacaze 2002). Conversely no clear relationship between LAI and FCover is found for Version 1 VGT and GEOV3/VGT due to the high scattering for FCover>0.4.

The FAPAR-FCover relationship is again more robust for Version 2 VGT than for Version 1 VGT and GEOV3/VGT. Note that the theoretical constrain of FCover<FAPAR/0.94 is only meet for Version 2 VGT.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **58 of 93**

Issue: I1.41



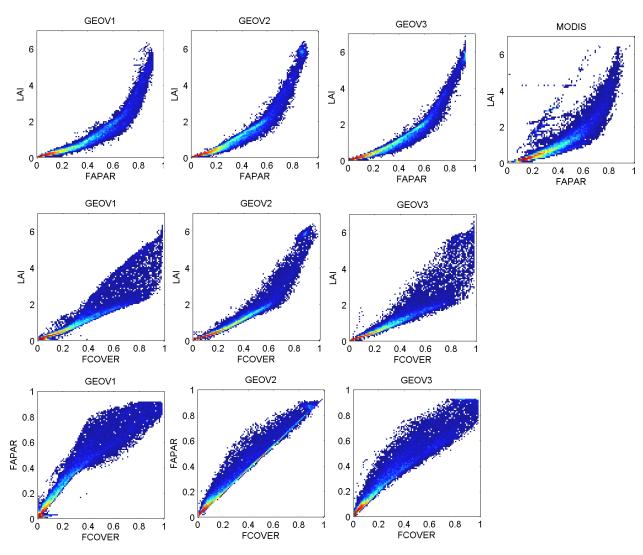


Figure 19: Comparison between LAI-FAPAR (top), LAI-FCover (middle) and FAPAR-FCover (bottom) for Version 1 (GEOV1), Version 2 (GEOV2), GEOV3/VGT and MODIS products

5.3 COMPARISON WITH VERSION 1 VGT

The comparison of Version 2 and Version 1 shows an overall good agreement. However, the agreement degrades when the number of available observations in the composition period, n, is lower than approximately n=10 (Figure 20, left). The distribution of n (Figure 20, left, dashed bold line) shows that, in most situations, n is in between 10 to 25. For the lower values of n, a small positive bias is observed, due to the effect of cloud contamination not detected in Version 1. Note that the maximum number of observations is 30 since the minimum semi-period of composition was set to ± 15 days for the most favorable conditions (when >6 daily estimates exist at each side of the date being processed).

The RMSE computed between actual valid daily LAI estimates and Version 2 (Figure 20, center) is closely linked with the difference between Version 2 and Version 1. This results the increasing

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **59 of 93**

Issue: I1.41



number of available observations, *n*, as well as from the 'scaling' effect: the RMSE GEOV2-GEOV1 (Figure 20, right) tend to increase with the Version 2 value. Note that the distribution of RMSE values (Figure 20, center, dashed bold line) shows that most values are lower than 0.1.

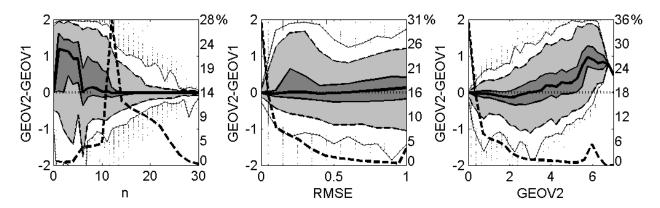


Figure 20: Evaluation of the differences between GEOV2/VGT and GEOV1/VGT products over the BELMANIP2 sites for the years 2003-2007 as a function of the number n of valid daily estimates in the composition period (left), of the RMSE between the GEOV2/VGT product and the daily estimates (center), and of the GEOV2/VGT product value (right). The several gray values correspond to 75% (dark gray), 90% (medium gray) and 95% (light gray) of the population, and the dots to 5% percentile of residual outliers. The bold black solid line corresponds to the median value of the differences. The dotted line is the 0:0 line. The dashed line shows the distribution of values of the variable in the abscissa which frequencies are indicated in the vertical axis on the right. Case of LAI products.

Similar behavior is observed for FAPAR (Figure 21) and FCover (Figure 22), with however, a more limited range of variation of the difference GEOV2-GEOV1 and the associated RMSE values.

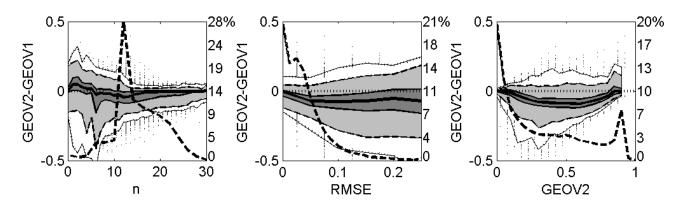


Figure 21: Idem Figure 20 but for FAPAR

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **60 of 93**

Date Issued: 11.03.2019

Issue: I1.41



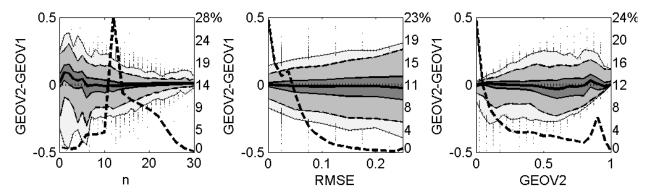


Figure 22: Idem Figure 20 but for FCover.

The scatterplots between Version 1 and Version 2 LAI products show a very good consistency for LAI values lower than about 4 (Figure 23). For the larger values of LAI, there is higher scattering between Version 2 and Version 1 partially due to the noise in the data. LAI Version 2 is significantly larger than Version 1 for EBF, mainly because of the more efficient filtering of cloud contamination for Version 2. The overall RMSE across biomes is 0.29 LAI with very high correlation between Version 2 and Version 1 (R=0.98).

For FAPAR (Figure 24) and FCover (Figure 25) variables, less affected by noise in the data, very strong consistency is observed between Version 2 and Version 1 for all the biomes and values. Note however that the FAPAR Version 2 values are systematically lower than the Version 1 for FAPAR between 0.25 and 0.85. This slight negative bias may be due to the differences in the training process of neural networks (Step 4A, Figure 6).

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **61 of 93**

Issue: I1.41



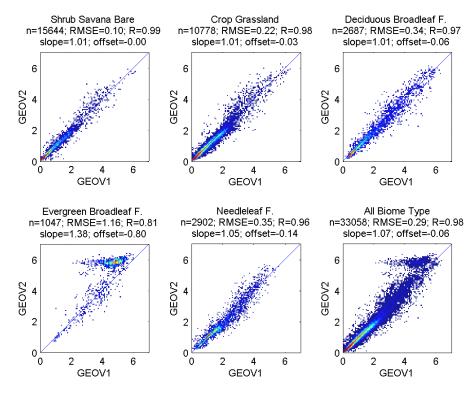


Figure 23: Comparison between GEOV2/VGT and GEOV1/VGT LAI products per GLOBCOVER biome type over the 445 BELMANIP2 sites for the years from 2003 to 2007.

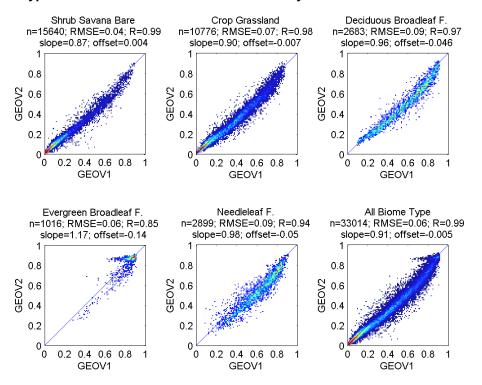


Figure 24: Comparison between GEOV2/VGT and GEOV1/VGT FAPAR products per GLOBCOVER biome type over the 445 BELMANIP2 sites for the years from 2003 to 2007.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **62 of 93**

Issue: I1.41



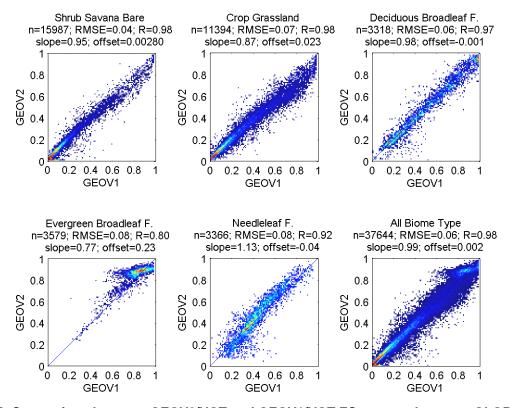


Figure 25: Comparison between GEOV2/VGT and GEOV1/VGT FCover products per GLOBCOVER biome type over the 445 BELMANIP2 sites for the years from 2003 to 2007.

5.4 COMPARISON WITH GEOV3/VGT

The scatter plots of the comparison of Version 2 VGT with GEOV3/VGT for LAI products (Figure 26) shows a similar pattern as when compared with Version 1 (compare Figure 26 and Figure 23) with an improvement in the performance for EBF (overall RMSE of 0.80 as compared to 1.16) due to the reduction of scattering (improved cloud filtering in GEOV2 and GEOV3 compared to GEOV1) and bias (similar algorithms between GEOV2 and GEOV3). Some differences are observed in the identification of EBF between GEOV2 and GEOV3 which partially explain the remaining differences (note that the EBF identified in Version 2 processing are not necessarily the same as in GLOBCOVER although a good agreement was found).

Very good agreement between GEOV2 and GEOV3 is also observed for FAPAR (overall RMSE of 0.05, Figure 27) although GEOV2 values are systematically lower than GEOV3 ones for FAPAR>0.2.

Similarly, GEOV2 and GEOV3 highly agree (RMSE of 0.06) for FCover products although a slight negative bias of GEOV2 compared to GEOV3 is observed for very high FCover values (Figure 28).

Document-No. CGLOPS1_ATBD_LAl1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **63 of 93**

Issue: I1.41



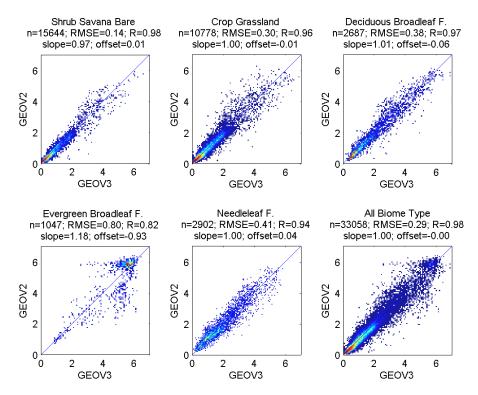


Figure 26: Comparison between GEOV2/VGT and GEOV3/VGT (PRO3_V1 algorithm) LAI products over the 445 BELMANIP2 sites for the years from 2003 to 2007.

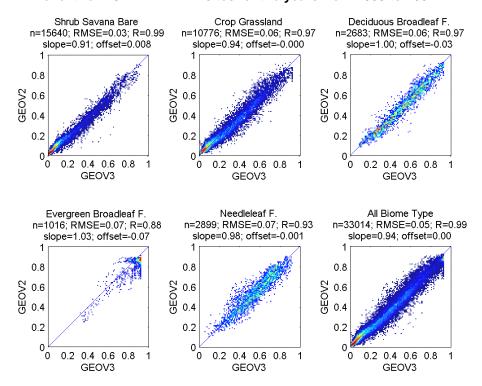


Figure 27: Comparison between GEOV2/VGT and GEOV3/VGT (PRO3_V1 algorithm) FAPAR products over the 445 BELMANIP2 sites for the years from 2003 to 2007.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **64 of 93**

Issue: I1.41



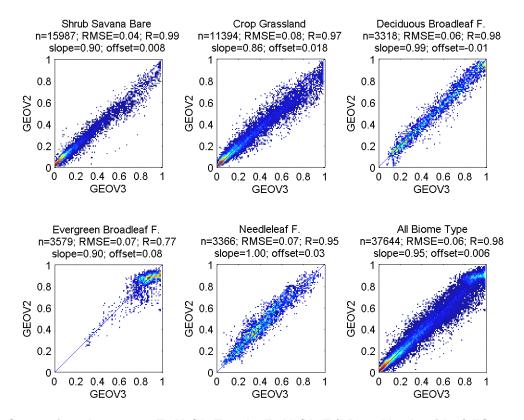


Figure 28: Comparison between GEOV2/VGT and GEOV3/VGT (PRO3_V1 algorithm) FCover products over the 445 BELMANIP2 sites for the years from 2003 to 2007.

5.5 DISTRIBUTION OF VALUES

The distributions of LAI (Figure 29) are very consistent between GEOV1, GEOV2, GEOV3 and MODIS, except, again, for EBF because of the higher uncertainties in the data and improved cloud filtering associated to GEOV2 and GEOV3. The distributions of FAPAR (Figure 30) show higher discrepancies between products: MODIS (GEOV1) FAPAR shows a positive bias as compared to the other products for low (high) FAPAR values, while GEOV2 and GEOV3 show a high agreement and constitute an intermediate solution. The distributions of FCover (Figure 31) are very consistent between GEOV1, GEOV2 and GEOV3 with high agreement in the location of peaks although GEOV2 shows a narrower distribution for the EBF class.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **65 of 93**

Issue: I1.41



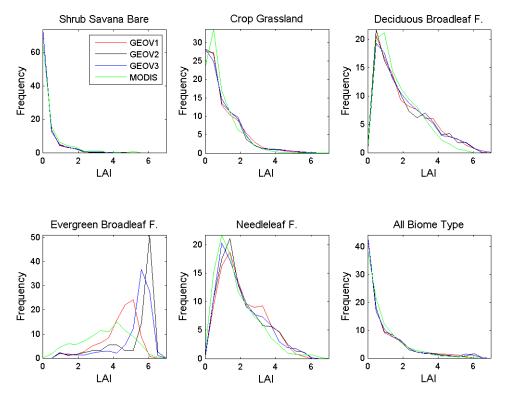


Figure 29: Distribution of GEOV1/VGT, GEOV2/VGT, GEOV3/VGT and MODIS LAI products per biome type as sampled by the 445 BELMANIP2 sites over the period 2003-2007.

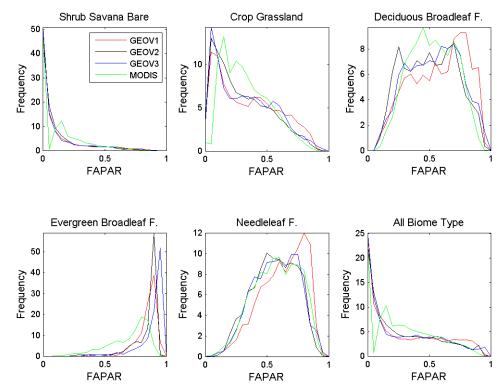


Figure 30: Idem Figure 29 but for FAPAR.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **66 of 93**

Issue: I1.41



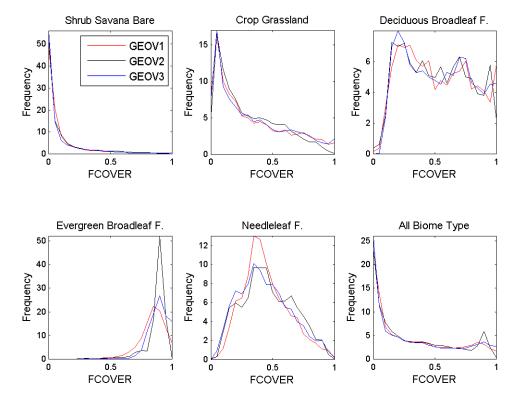


Figure 31: Idem Figure 29 but for FCover.

5.6 TEMPORAL CONTINUITY

The continuity of Version 2 VGT time series is evaluated by the fraction of missing data based on the 445 BELMANIP2 sites during the 2003-2007 period. Only 1% of the potential dekads are missing globally, i.e. much less than for Version 1 VGT showing about 20% of missing data (Figure 32). Version 1 VGT products are missing if less than 2 observations exist in the 30-day compositing window. Version 2 VGT benefits from the use of the TSGF and the CACAO climatology based techniques for filling, respectively, the small gaps within 120 days or the larger gaps.

Version 2 VGT data are missing only if the GEOCLIM climatology is not available due to too large discontinuities in the data. The advantages of this gap filling climatological approach as compared to classical temporal methods for the cases with high discontinuities and noise in the data was analyzed in Verger et al. (2013) and Kandasamy et al. (2013).

The climatological values are used to fill large gaps in 15% to 80% of the cases depending on the biome type as displayed in Figure 32. The fraction of dekads that were not filled but with 12 available observations are slightly lower than those observed for Version 1 VGT due to the more restrictive condition imposed in the number of available observations (12 observations in a ±60-day window for Version 2 compared to 2 observations in a ±10-day window for Version 1) and the outlier rejection process used in Version 2. Identical patterns per biome are observed with lower fraction of high quality observations for evergreen broadleaf forests which are mostly located in

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **67 of 93**

Issue: I1.41



areas with continuous cloud cover around the Equator. In this particular case, a high fraction of data is filtered out in the outlier rejection process applied in Version 2.

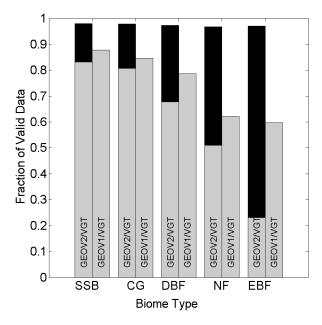


Figure 32: Average fraction of valid GEOV2/VGT and GEOV1/VGT products per biome. The biome classes are derived from the GLOBCOVER global landcover: Shrubs/Savana/Bare soil (SSB), Crops and Grassland (CG), Deciduous Broadleaf Forests (DBF), Needleleaf Forest (NF), and Evergreen Broadleaf Forest (EBF). For GEOV2/VGT, high quality products (grey) and products where the climatology was used to fill gaps (less than 12 valid daily estimates exist in the compositing period) (black) are distinguished. Evaluation over the BELMANIP2 sites for the 2003-2007 period.

5.7 TEMPORAL SMOOTHNESS

LAI variable results from incremental bio-physical processes. It is therefore expected to show relatively smooth temporal variations except in extreme situations such as flooding, fire or changes in the land-use. High variability in the temporal profiles would indicate a lack of reliability of the derived products. The smoothness of the LAI temporal series was evaluated based on the absolute value of the difference, δLAI , between LAI(t) product value at date t and the mean value between the two closest bracketing dates in a maximum Δt period of 60 days: $\delta LAI = |1/2(LAI(t+\Delta t)+LAI(t-\Delta t))-LAI(t)|$ (Verger et al. 2011). The smoother the temporal evolution, the smaller the δ difference should be. The histogram of δ over the whole dataset of BELMANIP2 sites in the 2003-2007 period (Figure 33) shows that both Version 1 VGT and Version 2 VGT products are very smooth with differences lower than 0.25 for most of cases. However, GEOV2/VGT shows generally smoother temporal profiles as attested in Figure 33. The differences in terms of smoothness between GEOV2/VGT and GEOV1/VGT mostly rely on the different composition approaches applied in the space of output biophysical variables or input reflectances, respectively.

Document-No. CGLOPS1_ATBD_LAI1km-V2

© C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **68 of 93**

Issue: I1.41



GEOV2/VGT benefits from the use of temporal smoothing techniques applied to daily neural networks estimates while GEOV1/VGT results from the application of neural network to 30-day composited reflectances.

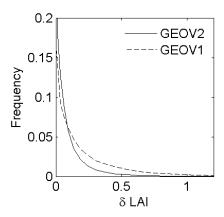


Figure 33: Histogram of the δLAI absolute difference representing temporal smoothness for GEOV2/VGT and GEOV1/VGT LAI products. Evaluation over the BELMANIP2 sites for the 2003-2007 period.

5.8 CONSISTENCY BETWEEN VERSION 2 PROBA-V AND VERSION 2 VGT NRT ESTIMATES

Version 2 PROBA-V and Version 2 VGT NRT LAI, FAPAR and FCover estimates show a high agreement (RMSE <0.3 for LAI and <0.04 for FAPAR and FCover) with unbiased estimates as evaluated over the 445 BELMANIP2 sites for the overlapping period 2013-10-16 to 2014-05-31 (Figure 34) and for all the biome types (Figure 35).

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **69 of 93**

Issue: I1.41



© C-GLOPS Lot1 consortium

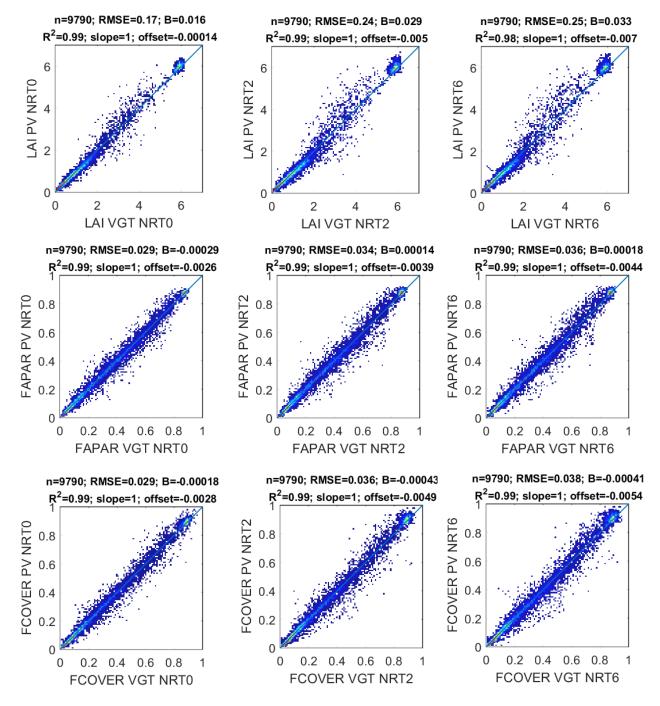


Figure 34. Comparison between GEOV2/PROBA and GEOV2/VGT NRT-0, NRT-2 and NRT-6 LAI, FAPAR and FCover estimates over the 445 BELMANIP2 sites for the period 2013-10-16 to 2014-05-31.

Document-No. CGLOPS1_ATBD_LAI1km-V2

Issue: **I1.41** Date: 11.03.2019 Page: **70 of 93**

Issue: I1.41



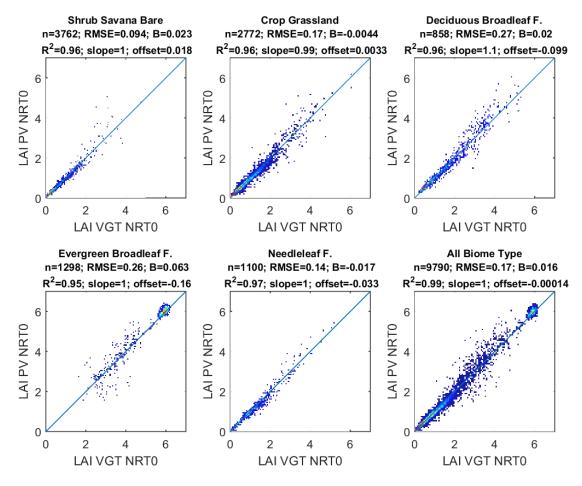


Figure 35: Comparison between GEOV2/PROBA-V and GEOV2/VGT NRT-0 LAI estimates per biome type over the 445 BELMANIP2 sites for the period 2013-10-16 to 2014-05-31.

5.9 SUMMARY AND RECOMMENDATIONS

This Version 2 of Collection 1km algorithm shows improvement as compared to the Version 1 method at least for three aspects:

- It allows real time estimates
- It is much more continuous with very few missing data
- It is very smooth as expected for the dynamics of such variables at this moderate resolution.

It shows very good consistency with Version 1 in situations where few observations are missing. For evergreen broadleaf forests, the LAI values are enhanced, mainly because of the improved cloud filtering. This would suggest that the principles used within this algorithm, i.e. daily estimates of products followed by a careful gap filling and smoothing based on the climatology as prior information is very efficient. This is confirmed by the results of the exhaustive quality assessment

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **71 of 93**

Date Issued: 11.03.2019

Issue: I1.41



exercise performed on SPOT/VGT [GIOGL1_QAR_LAI[FAPAR/FCOVER]1km-VGT-V2] and PROBA-V [CGLOPS1_QAR_LAI[FAPAR/FCOVER]1km-PROBAV-V2] Version 2 products.

The user should use the product with due attention to the QF values as well as proxy of uncertainties coming mainly from the RMSE values and the number of actual observations used. In case of very little number of actual observations, the product will mainly derive from the climatology values, although some seasonal variability or land cover change may happen.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **72 of 93**

Date Issued: 11.03.2019

Issue: I1.41



6 REFERENCES

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., Weiss, M., & Kandasamy, S. (2010). Climatology of LAI, FAPAR and FCOVER products (V1). In (p. 30). Avignon (France): INRA-EMMAH

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., & Smets, B. (2013). GEOV1: LAI, FAPAR Essential Climate Variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production. *Remote Sensing of Environment*, 137, 299-309

Baret, F., Weiss, M., Verger, A., & Kandasamy, S. (2011). BioPar Methods Compendium - LAI, FAPAR and FCOVER from LTDR AVHRR series. In, *Report for geoland2 EC contract FP-7-218795. Available at http://www.geoland2.eu/portal/documents/CA80C881.html* (p. 46). Avignon: INRA-EMMAH

Bartholomé, E., & Belward, A.S. (2005). GLC2000: a new approach to global land cover mapping from Earth Observation data. *International Journal of Remote Sensing*, 26, 1959 - 1977

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., & Weiss, M. (2013). GEOV1: LAI, FAPAR Essential Climate Variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products. *Remote Sensing of Environment,* 137, 310-329

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high quality NDVI time series data set based on the Savitzky-Golay filter. Remote Sensing of Environment, 91, 332-344

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., Bicheron, P., Brockmann, C., Bontemps, S., Van Bogaert, E., Vancutsem, C., Pekel, J.F., Huc, M., Henry, C., Ranera, F., Achard, F., di Gregorio, A., Herold, M., Leroy, M., & Arino, O. (2009). The first 300 m global land cover map for 2005 using ENVISAT MERIS time series: a product of the GlobCover system,. In, *Proceedings of the 33rd International Symposium on Remote Sensing of Environment*. Stresa (Italy)

Dierckx, W., Sterckx, S., Benhadj, I., Livens, S., Duhoux, G., Van Achteren, T., Francois, M. Mellab, K. and G. Saint, 2014. PROBA-V mission for global vegetation monitoring: standard products and image quality. Int. J. Remote Sens., vol. 35, issue 7, p. 2589-2614

Fang, H., Jiang, C., Li, W., Wei, S., Baret, F., Chen, J.M., Garcia-Haro, J., Liang, S., Liu, R., Myneni, R.B., Pinty, B., Xiao, Z., & Zhu, Z. (2013). Characterization and intercomparison of global

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **73 of 93**

Date Issued: 11.03.2019

Issue: I1.41



moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties. *Journal of Geophysical Research: Biogeosciences*.

Fernandes, R., Burton, C., Leblanc, S., & Latifovic, R. (2003). Landsat-5 TM and Landsat-7 ETM+ based accuracy assessment of leaf area index products for Canada derived from SPOT-4 VEGETATION data. Canadian Journal of Remote Sensing., 20, 241-258

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 Environment, 105,* 286-298

Garrigues, S., Lacaze, R., Baret, F., Morisette, J., Weiss, M., Nickeson, J., Fernandes, R., Plummer, S., Shabanov, N.V., Myneni, R., & Yang, W. (2008). Validation and Intercomparison of Global Leaf Area Index Products Derived From Remote Sensing Data. *Journal of Geophysical Research*, 113

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

Jiao, T., Liu, R., Liu, Y., Pisek, J., & Chen, J.M. (2014). Mapping global seasonal forest background reflectivity with Multi-angle Imaging Spectroradiometer data. *Journal of Geophysical Research: Biogeosciences, 119*, 2013JG002493

Kandasamy, S., Baret, F., Verger, A., Neveux, P., & Weiss, M. (2013). A comparison of methods for smoothing and gap filling time series of remote sensing observations: application to MODIS LAI products. *Biogeosciences*, *10*, 4055-4071

Maisongrande, P., Duchemin, B., & Dedieu, G. (2004). VEGETATION/SPOT: an operational mission for the Earth monitoring; presentation of new standard products. *International Journal of Remote Sensing*, *25*, 9-14

Martínez, B., García-Haro, F.J., & Camacho-de Coca, F. (2009). Derivation of high-resolution leaf area index maps in support of validation activities: Application to the cropland Barrax site. Agricultural and Forest Meteorology, 149, 130-145

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, R.R., & Running, S.W. (2002). Global products of vegetation leaf area and absorbed PAR from year one of MODIS data. *Remote Sensing of Environment, 83*, 214-231

Myneni, R.B., & Williams, D.L. (1994). On the relationship between FAPAR and NDVI. *Remote sensing of the environment, 49*, 200-211

Pisek, J., Chen, J.M., Alikas, K., & Deng, F. (2010). Impacts of including forest understory brightness and foliage clumping information from multiangular measurements on leaf area index mapping over North America. *Journal of Geophysical Research: Biogeosciences, 115*, G03023

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **74 of 93**

Date Issued: 11.03.2019

Issue: I1.41



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

Roujean, J.L., & Lacaze, R. (2002). Global mapping of vegetation parameters from POLDER multiangular measurements for studies of surface-atmosphere interactions: a pragmatic method and validtion. *Journal of Geophysical Research*, 107, ACL 6 1-14

Roujean, J.L., Leroy, M., & Deschamps, P.Y. (1992). A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*, 97, 20455-20468

Sterckx, S., Benhadj, I., Duhoux, G., Livens, S., Dierckx, W., Goor, E., Adriaensen, S., Heyns, W., Van Hoof, K., Strackx, G., Nackaerts, K., Reusen, I., Van Achteren, T., Dries, J., Van Roey, T., Mellab, K., Duca, R. and Zender, J. (2014). The PROBA-V mission: image processing and calibration. Int. J. Remote Sens., 35(7), 2565 – 2588.

Verger, A. (2008). Analysis of operational algorithms for the estimation of biophysical variables from satellite data (Analisi comparativa d'algorismes operacionals d'estimacio de parametres biofisics de la coberta vegetal amb teledeteccio). In, *Departament of Earth Physics and Thermodynamics* (p. 277). Valencia (Spain): University of Valencia

Verger, A., Baret, F., & Weiss, M. (2008). Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products. *Remote Sensing of Environment, 112*, 2789-2803

Verger, A., Baret, F., & Weiss, M. (2011). A multisensor fusion approach to improve LAI time series. *Remote Sensing of Enviroment, 115*, 2460-2470

Verger, A., Baret, F., & Weiss, M. (2014a). Near real time vegetation monitoring at global scale. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*, 3473-3481

Verger, A., Baret, F., Weiss, M., Filella, I., & Peñuelas, J. (2015). GEOCLIM: A global climatology of LAI, FAPAR, and FCOVER from VEGETATION observations for 1999–2010. *Remote Sensing of Environment, 166*, 126-137

Verger, A., Baret, F., Weiss, M., Kandasamy, S., & Vermote, E. (2013). The CACAO method for smoothing, gap filling and characterizing seasonal anomalies in satellite time series. *IEEE transactions on Geoscience and Remote Sensing*, *51*, 1963-1972

Verger, A., Baret, F., Weiss, M., Smets, B., Lacaze, R., & Camacho, F. (2014b). Near real time estimation of biophysical variables within Copernicus global land service. In, *Global vegetation monitoring and modeling (available at https://colloque.inra.fr/gv2m/Poster-Sessions/Poster-S7)*. Avignon (France)

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **75 of 93**

Date Issued: 11.03.2019

Issue: I1.41



Weiss, M., Baret, F., Block, T., Koetz, B., Burini, A., Scholze, B., Lecharpentier, P., Brockmann, C., Fernandes, R., Plummer, S., Myneni, R., Gobron, N., Nightingale, J., Schaepman-Strub, G., Camacho, F., & Sanchez-Azofeifa, A. (2014). On Line Validation Exercise (OLIVE): A Web Based Service for the Validation of Medium Resolution Land Products. Application to FAPAR Products. *Remote Sensing*, *6*, 4190-4216

Weiss, M., Baret, F., Garrigues, S., Lacaze, R., & Bicheron, P. (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. part 2: Validation and comparison with MODIS Collection 4 products. *Remote Sensing of Environment, 110*, 317-331

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

Yang, W., Shabanov, N.V., Huang, D., Wang, W., Dickinson, R.E., Nemani, R.R., Knyazikhin, Y., & Myneni, R.B. (2006). Analysis of leaf area index products from combination of MODIS Terra and Aqua data. *Remote Sensing of Environment, 104*, 297-312

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **76 of 93**

Date Issued: 11.03.2019

Issue: I1.41



ANNEX 1: GEOCLIM, A CLIMATOLOGY OF GEOV1 LAI, FAPAR AND FCOVER

The principles of the generation of the GEOCLIM dataset are here described (Figure 36). Further information is provided in Verger et al. (2015).

Inter-annual average values from the Copernicus Global Land time series of biophysical LAI, FAPAR and FCover Version 1 (GEOV1) at 1-km resolution and 10-day frequency were first computed for 1999 to 2010 (Baret et al. 2010). Since the climatology plays an important role for gap filling and temporal smoothing of the data, possible artifacts in the climatology were corrected through a two-step process:

- Correction of specific artifacts
- · Gap filling and temporal smoothing

Correction of specific artifacts

The climatology was first corrected over specific problematic conditions (Figure 36) based on the prior knowledge:

- o Some artifacts are observed at *northern high latitudes during the winter time when the sun zenith angle, SZA>70°*. Anomalous seasonality and unexpected increases in LAI (FAPAR, FCover) with an artificial maximum peak in winter and high inter-annual variability. These artefacts mainly due to snow cover or very poor illumination conditions that limited the number of valid observations and the reliability of the bidirectional reflectance model applied for the correction of VEGETATION data (Roujean et al. 1992). The LAI (FAPAR, FCover) values are expected to be relatively stable and low due to the low temperatures, short days, and low illumination during winter at these high latitudes. To correct these artefacts at northern high latitudes (latitude>40°) for winter (defined here as the period for which the *sun zenith angle, SZA>70°)*, the climatology values higher than the 20-percentile (P20^{clim}) were fixed at minima by preferentially selecting the values computed from at least three valid observations. If none of the dekads meets this condition, the minimum value computed over all the dekads was used.
- o Significant artifacts were also detected in *Equatorial and tropical latitudes* due to the permanent presence of clouds which results in high instabilities in the temporal profiles of GEOV1/VGT climatology. Since most of these cases correspond to evergreen broadleaf forests (EBF), a minimum seasonality and high values were assumed. A pixel was identified as being an EBF if the 90-percentile (P90^{clim}) of LAI climatology is > 4.5 and the 20-percentile P20^{clim} is >P90^{clim} -1.5. This method for the detection of EBF is based only on GEOV1 products (Figure 37a): it agrees well with the GLOBCOVER land-cover map (Defourny et al. 2009) (Figure 37b). For EBFs, the climatology values were fixed to the 90-percentile. A quality flag indicating that the pixel was identified as EBF is activated (QF_{EBF}=1). It is subsequently used in Steps 1-4 of Branches B and C.
- Some artifacts were also detected in GEOV1/VGT climatology for *bare soil (BS) areas*. A pixel was identified as being a BS if the 90-percentile of LAI climatology P90^{clim} is <0.05. For those cases the climatology values were fixed to the median value (50-percentile)

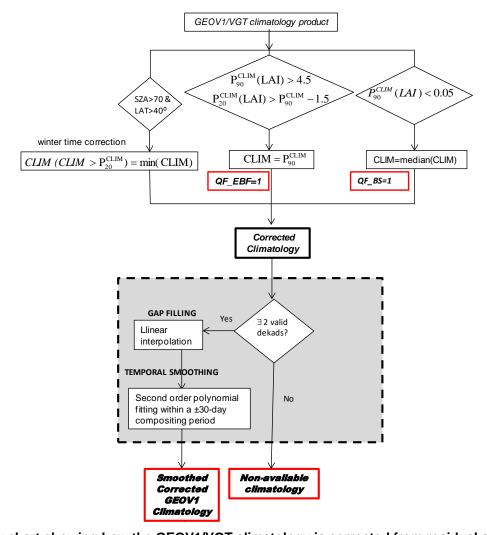
Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **77 of 93**

Issue: I1.41



computed over the entire period. In addition, a quality flag indicating that the pixel was identified as BS is activated ($QF_{BS}=1$). It is subsequently used in Step 4 of Branches B and C.



 $\label{thm:continuous} \textbf{Figure 36: Flow chart showing how the GEOV1/VGT climatology is corrected from residual artifacts.}$

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **78 of 93**

Issue: I1.41



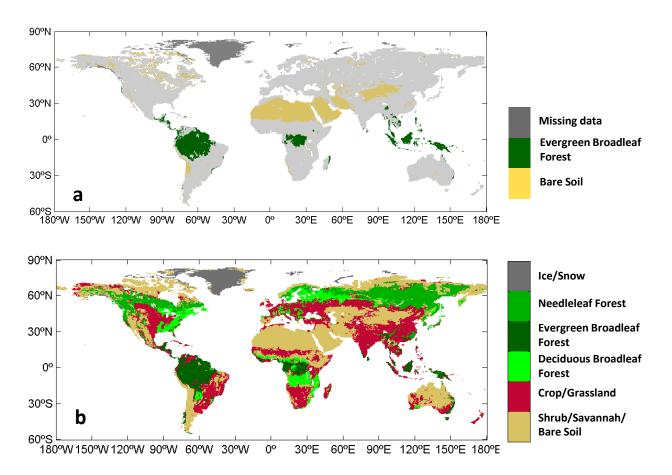


Figure 37: (a) Map of bare soil and evergreen broadleaf forest areas identified based on GEOV1/VGT climatology. (b) Simplified GLOBCOVER land-cover map after aggregating the 22 original classes into six main land-cover classes.

Gap filling and smoothing

The climatology was then gap filled and smoothed to eliminate possible high temporal frequency residual artifacts. The gap filling (GF) and temporal smoothing (TS) techniques proposed by (Verger et al. 2011) were applied here.

- *Gap filling.* A simple linear interpolation was applied if two valid dekads are available along the 36 potential dekads of the climatology
- Temporal smoothing. A second order polynomial order was fit to the data within a ±30-day compositing period centered on the date being smoothed. This polynomial fitting applied at a dekadal time step.

Figure 38 shows the original and corrected climatology at four problematic sites.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **79 of 93**

Issue: I1.41



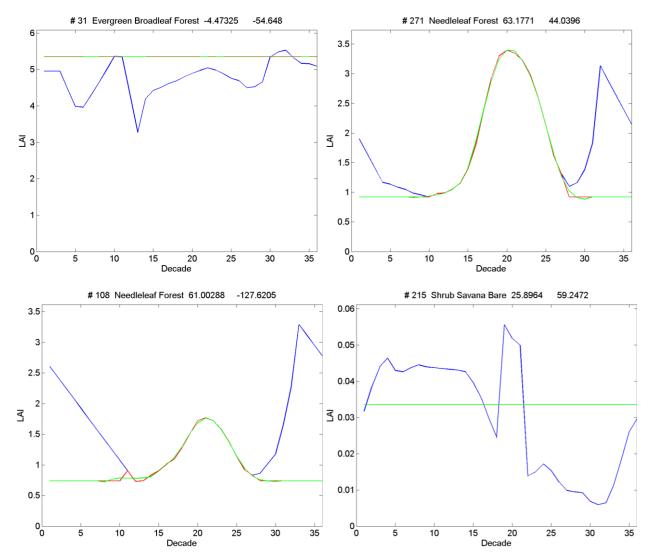


Figure 38: Correction of GEOV1/VGT climatology. The blue line corresponds to the original GEOV1/VGT climatology LAI product. The red line corresponds to the corrected climatology based on prior knowledge. Green line is the final GEOCLIM climatology resulting from applying gap filling and temporal smoothing techniques to the first corrected climatology.

Document-No. CGLOPS1_ATBD_LAI1km-V2

Date: 11.03.2019

Issue:

I1.41

Page: **80 of 93**

Date Issued: 11.03.2019

Issue: I1.41



Annex 2: Neural Networks Calibration

PREPARATION OF THE INPUTS AND OUTPUTS OF THE LEARNING DATA SET

The Version 2 algorithm was calibrated on the reprocessed version of the VGT-S1 TOC data produced by CTIV during summer 2015. The neural networks were calibrated using the BELMANIP2 sites extracts over the 2003-2007 period. The BELMANIP2 set of sites is an update of the original BELMANIP sites (Baret et al. 2006). The 445 BELMANIP2 sites present the same distribution of vegetation types and conditions as the Earth's surface while showing little topography and good level of homogeneity. The land cover homogeneity of each site was double checked using the GLOBCOVER map, and the Google Earth engine (Weiss et al. 2014). For the 3 variables, the same inputs are used:

- Daily top of canopy reflectance in 3 VEGETATION bands (B2, B3, SWIR),
- 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)),

The output is the corresponding V0 (first neural net) daily values of the biophysical variables. 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 weighing, w, is designed to enhance the specific advantage of each product while limiting their deficiencies.

Garrigues et al. (2008) and (Weiss et al. 2007) reported that CYCLOPES LAI was showing some saturation for LAI values around 4. Conversely, MODIS LAI and FAPAR values were generally higher than expected for the very low vegetation amounts. Further, the MODIS algorithm assigns zero values for LAI and FAPAR over pixels classified as bare soil, which may pose problems in case of misclassification. It was thus proposed to fuse the products by reducing the contribution of MODIS products for low LAI and FAPAR values and enhancing the MODIS contribution for the large LAI and FAPAR values as sketched in Figure 39. The weight, w, is driven by LAI_{CYCV31} since LAI_{CYCV31} appears more stable as compared to MODIS LAI_{MODC5} (Verger et al. 2011). The threshold of $LAI_{CYCV31} = 4$ corresponds to the value when LAI_{CYCV31} starts to saturate. The parallel processing applied to both LAI and FAPAR (Eq. 1) is expected to keep a good consistency between these two variables.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **81 of 93**

Date Issued: 11.03.2019

Issue: I1.41



As compared to GEOV1 (Figure 31), this function is smoother, limiting the brutal change in the weight observed for $LAI_{CYCV31} = 4$. Similarly to GEOV1, w = 0.5 when $LAI_{CYCV31} = 2$.

Note that for FCover, no fusion was completed since CYCLOPES was the only existing product at the global scale. The only other existing product is the LSA SAF FCover but it is not global since derived from geostationary sensors. However, as Verger (2008) demonstrated, the CYCLOPES FCover product was significantly underestimating both the FCover measured at ground and the LSA SAF product. To mitigate the underestimation of CYCLOPES product, the FCover values are scaled afterwards in the training of neural networks.

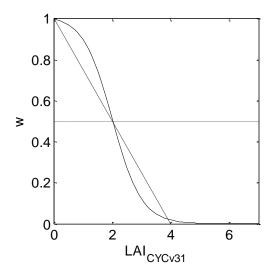


Figure 39: The weighing function used in GEOV2 for the fusion between CYCLOPES and MODIS LAI and FAPAR products. The dashed line corresponds to the weight used for generating GEOV1 products. The dotted line corresponds to w=0.5.

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 visually removing data that appear to be outliers (Figure 40). Several operators were completing this first filtering process to get more robust results. Despite the inherent subjectivity associated to manual filtering, at this algorithm level, a manual method was preferred to automatic methods to avoid possible systematic biases which would affect neural networks estimates. Note however that manual filtering was only applied to the training database while completely automatic methods of outlier detection were applied in subsequent steps (steps 1A, 3A, 6A, 1B, 3B, 1C and 3C) of the algorithm. The performance of this manual method is illustrated in Figure 40.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **82 of 93**

Date Issued: 11.03.2019

Issue: I1.41



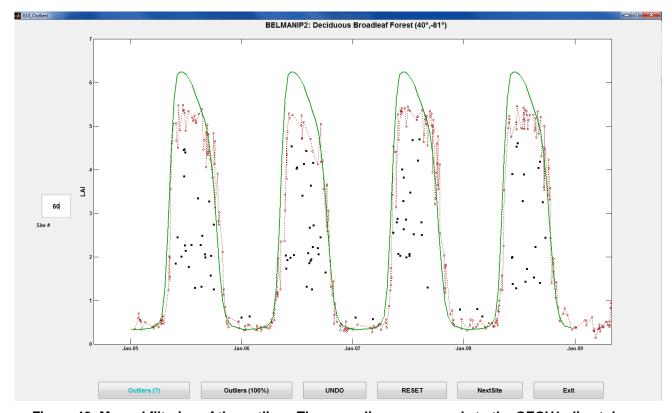


Figure 40: Manual filtering of the outliers. The green line corresponds to the GEOV1 climatology product, red circles to V0 (first neural net) valid estimates, black squares are outliers.

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 and 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 which is mainly the case by construction of the ensemble of BELMANIP2 sites. It is achieved for each date where a daily VGT-S1 product is available through three steps:

- Selection of the MODIS and CYCLOPES products: all the MODIS and CYCLOPES products available within ±10 days around the VGT-S1 daily product date are gathered over the 3x3 pixels. For MODIS, only the main and main + saturation LAI and FAPAR products are considered. This results in nMOD (0<nMOD≤18) 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 Δ MOD and Δ CYC between the 70% and 90% percentiles within, respectively, the nMOD and nCYC LAI values available is computed.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **83 of 93**

Issue: I1.41



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 LAI 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 Δ MOD and Δ CYC 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 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 41). 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.

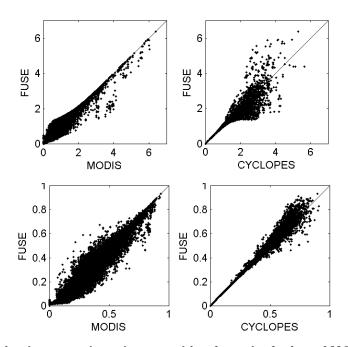


Figure 41: Relationships between the values resulting from the fusion of MODIS and CYCLOPES products according to equation (1) as a function of composited MODIS and CYCLOPES products for LAI (top) and FAPAR (bottom) variables

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **84 of 93**

Date Issued: 11.03.2019

Issue: I1.41



To eliminate outliers in the daily input VGT-S1 TOC products, reflectances were tentatively transformed into LAI by training a specific neural network that relates the VGT-S1 TOC reflectances with the corresponding fused MODIS and CYCLOPES LAI values. 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 variables. 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. Inputs and output were normalized.

The NNT was applied over the 2003-2007 period and then the outliers in the daily Product_0 (first neural net) estimates were manually filtered out considering the GEOV1/VGT climatology as background information and using a graphical user interface (GUI) 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.

The data were further filtered using three additional criterions:

- **Air mass.** The total air mass, m, computed as $m = \frac{1}{\cos \theta_s} + \frac{1}{\cos \theta_o}$ where θ_s and θ_o are respectively the sun and view zenith angles. The values for which m > 5 were considered as unreliable because of too large atmospheric BRDF effects. They were thus flagged as outliers.
- **Soil line.** Points lying below the soil line in the B2, B3 and SWIR bands were considered with a high probability of being contaminated by significant fraction of water-bodies or clouds. They correspond to the following condition:

$$\rho_{B3} < 0.54 \frac{\rho_{B2} - 0.04}{0.5 - 0.04} \quad \text{or} \quad \rho_{SWIR} < 0.70 \frac{\rho_{B2} - 0.08}{0.5 - 0.08}$$

where ρ_{B2} , ρ_{B3} and ρ_{SWIR} represent respectively the TOC reflectance for bands B2, B3 and SWIR. These data points were flagged as outliers.

NDVI: the NDVI value must be higher than 0 (NDVI>0)

The available data in the training dataset were randomly split into a learning dataset made of 70% of the data available (n=189285), and a test dataset (15% of the data) used for testing the hyperspecialization of the training process and the remaining 15% for evaluating the theoretical performances.

GENERATION OF THE DEFINITION DOMAIN

The convex hull made by the reflectances retained after this manual selection process and additional filtering was formed in the B2, B3 and MIR feature space to generate the definition domain that will be used to check if the inputs are valid or not when applying the neural network. The definition domain was gridded using 30⁴ cells by dividing each dimension (band) into 30 equally spaced steps between the minimum and maximum values. The definition domain was finally defined by the cells containing no data flagged as outliers. Figure 42shows the final definition domain for all the band combinations. The characteristics of the definition domain are reported in the Excel file under sheet 'Definition Domain'.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **85 of 93**

Issue: I1.41



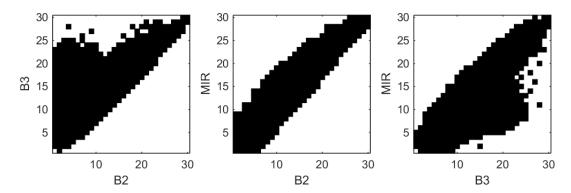


Figure 42: The convex hull that corresponds to the definition domain using the manually filtered outliers. Pixel will be declared as valid if they are within the area defined by the black areas. The 30 cells are distributed equally over the different reflectance ranges.

TRAINING THE NETWORKS FOR OPERATIONAL APPLICATIONS

Evergreen Broadleaf Forests are characterized by two specific behaviors as compared to other biomes: (i) although they are characterized by high variables' values, the near infrared reflectance level is lower than for the other biomes for a given value of the variables. This is mainly due to thicker leaves. And, (ii) the temporal profiles of LAI, FAPAR and FCover are characterized by a high level of noise while the values should remain high and stable during the whole year. To avoid the interference of these factors on the training for the other biomes, it was therefore decided to consider two different processings whether the biome is an EBF or not. EBF identification is achieved using the GEOV1/VGT climatology EBF flag (Annex 1: GEOCLIM, a climatology of GEOV1 LAI, FAPAR and FCover).

A second set of neural networks with the same architecture as the one used to filter the VGT-S1 TOC reflectances was trained over the VGT-S1 TOC inputs. These networks are those that are implemented in the processing chain. The FCover CYCLOPES product scaling was incorporated in the 'denormalization' step of the network outputs.

The training data sets were different for EBF and non EBF: for non EBF, we discarded all the BELMANIP2 sites that were identified as EBF by the GEOV1 climatology to prevent the training to be disturbed by relatively low near-infrared reflectance of EBF, and perform better estimates on high reflectances for non EBF biomes. Conversely, for EBF, we did not use EBF BELMANIP2 sites solely but the whole set of BELMANIP2 sites to be able to get valuable results in case the EBF climatology flag is erroneous. The neural net architecture is the same for all the variables and the two biomes (one hidden layer with 5 neurons with a tangent sigmoid function, and a linear output layer with a single neuron).

The theoretical performances of NNT (both for EBF and non EBF) were evaluated over the test dataset. The training process was quite successful (Figure 43): high correlation coefficient close to the line 1:1.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **86 of 93**

Issue: I1.41



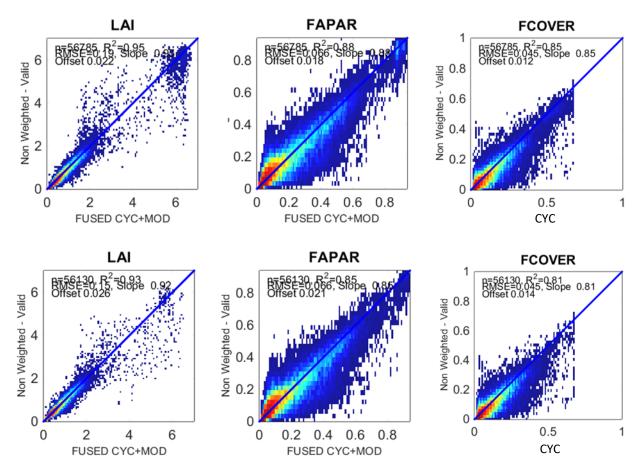


Figure 43: Theoretical performances of the neural networks used for LAI, FAPAR and FCover products. Top: EBF, Bottom: Non EBF. Neural network predicted outputs and the observed fused MODIS and CYCLOPES products in the test dataset are displayed as a density plot: the more red, the denser the points are.

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).

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **87 of 93**

Date Issued: 11.03.2019

Issue: I1.41



• **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 FCover**. To address the CYCLOPES FCover underestimating issue, the neural network estimated output is scaled by dividing the maximum theoretical FCover (1.00) by the 99% percentile value computed over the cumulated distribution frequency of the output *Y* values of the neural network.

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

Ymax = Scaling.Ymax* Ymin = Scaling.Ymin*

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **88 of 93**

Date Issued: 11.03.2019

Issue: I1.41



ANNEX 3: RESCALING PROBA-V ESTIMATES

Neural networks were trained over the VGT-S1 TOC inputs as described in Annex 2. Although neural networks provide unbiased estimates when applied to PROBA-V S1 reflectances (slopes close to 1 with low offset and little bias) (Figure 44, Figure 45, Figure 46), marginal discrepancies with GEOV2/VGT may be observed locally as illustrated by the boxplots in Figure 47: an overestimation of the neural network PROBA-V output appears mainly for the intermediate-high LAI/FAPAR/FCover values. Indeed, these values are (i) less represented in the training data set, which is in agreement with the global distribution of the products and (ii) they correspond to situations closed to saturation of the reflectance with LAI (FAPAR, FCover).

To better match the expected range of variation of the outputs with the theoretical one, the GEOV2/PROBA-V outputs were scaled with respect to GEOV2/VGT over the 445 BELMANIP2 sites for the period 2013-10-16 to 2014-05-31. For each product, we fitted a third order polynomial function of each product over the median of the residuals (Figure 47, Table 13).

The rescaled daily PROBA-V estimates showed a better agreement with VGT estimates for all biomes and variables with lower systematic bias and RMSE closer to 0 (Figure 44, Figure 45, Figure 46). Note however that some residual scattering is observed due mainly to unmasked clouds and atmospheric effects. This residual noise in the daily estimates is highly reduced in the final dekadal GEOV2 estimates (Figure 34, Figure 35) through the application of outlier filtering and temporal smoothing in Branches B and C of the algorithm.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **89 of 93**

Issue: I1.41



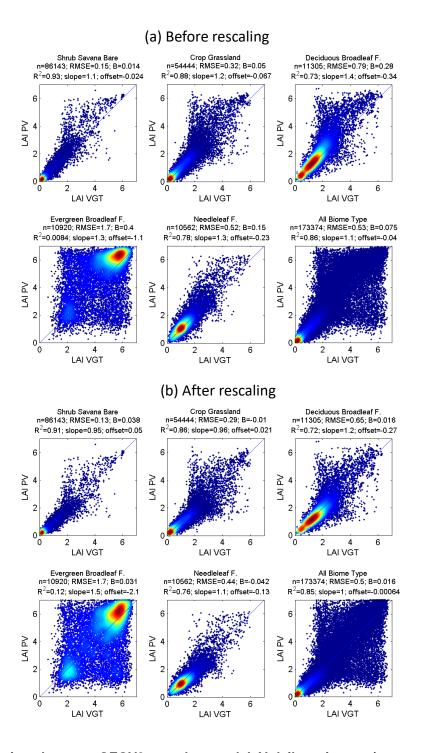


Figure 44: Comparison between GEOV2 neural network LAI daily estimates (output of step 4A) from PROBA-V and VGT S1 data before (a) and after (b) the rescaling applied to PROBA-V to mimic VGT estimates. Comparison per GLOBCOVER biome type over the 445 BELMANIP2 sites for the period 2013-10-16 to 2014-05-31.

Document-No. CGLOPS1_ATBD_LAI1km-V2

© C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **90 of 93**

Issue: I1.41



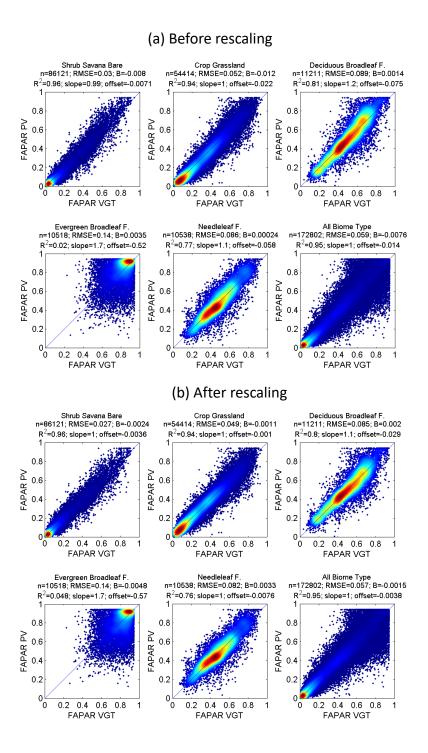


Figure 45: Comparison between GEOV2 neural network FAPAR daily estimates (output of step 4A) from PROBA-V and VGT S1 data before (a) and after (b) the rescaling applied to PROBA-V to mimic VGT estimates. Comparison per GLOBCOVER biome type over the 445 BELMANIP2 sites for the period 2013-10-16 to 2014-05-31.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **91 of 93**

Issue: I1.41



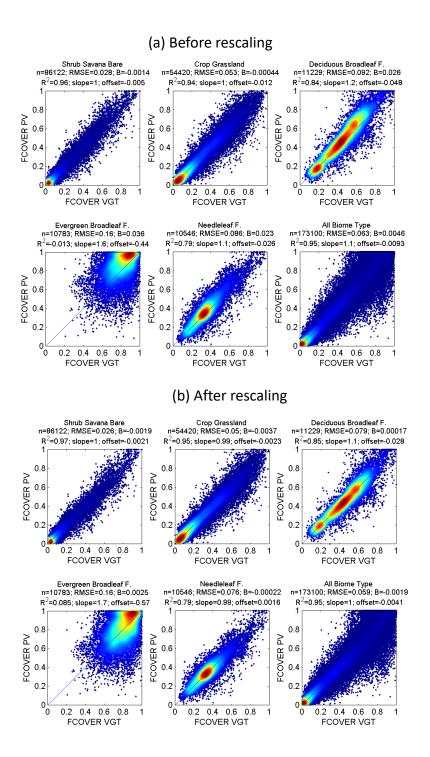


Figure 46: Comparison between GEOV2 neural network FCover daily estimates (output of step 4A) from PROBA-V and VGT S1 data before (a) and after (b) the rescaling applied to PROBA-V to mimic VGT estimates. Comparison per GLOBCOVER biome type over the 445 BELMANIP2 sites for the period 2013-10-16 to 2014-05-31.

Document-No. CGLOPS1_ATBD_LAI1km-V2 © C-GLOPS Lot1 consortium

Issue: **I1.41** Date: 11.03.2019 Page: **92 of 93**

Issue: I1.41



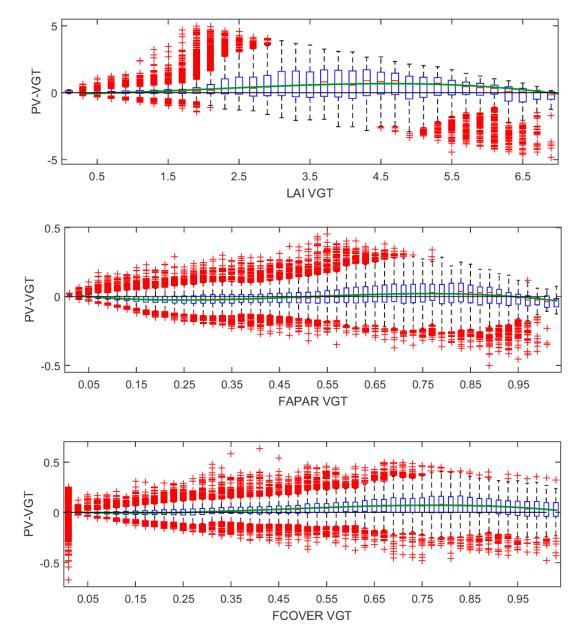


Figure 47: Boxplots of the differences between GEOV2 neural network LAI (top), FAPAR (middle) and FCover (bottom) daily estimates (output of step 4A) from PROBA-V (before rescaling) and VGT S1 data. Median value correspond to the central red mark, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually with red crosses. The green line corresponds to the 3rd order polynomial fitted on the median of the residuals.

Document-No. CGLOPS1_ATBD_LAI1km-V2

Dogg, 02 of 02

Issue: **I1.41** Date: 11.03.2019

Page: **93 of 93**

© C-GLOPS Lot1 consortium