vegetation parameters cci

CCI Vegetation

D2.1 Algorithm Theoretical Basis Document Retrieval Algorithm Cycle 2

Simon Blessing

November 2024

Distribution list

Author(s)	:	Simon Blessing
Reviewer(s)	:	Else Swinnen, Christiaan Van der Tol
Approver(s)	:	Clément Albergel
Issuing authority	:	VITO

Change record

Release	Date	Pages	Description of change	Editor(s)/Reviewer(s)
V1	03/09/2022	all	First version	Simon Blessing / Else Swinnen
V2	13/11/2024	multiple	Cycle 2/CRDP-2; added mixed prior approach	Simon Blessing / Else Swinnen
V2.1	16/12/2024			

Executive summary

CCI+ Vegetation Parameters is part of the ESA Climate Change Initiative. It aims at the identification, development and improvement of algorithms for the consistent retrieval of vegetation ECVs LAI and fAPAR from multi-platform and multi-mission satellite data, the generation of long-term climate data records, and the interaction with the user community to match their requirements. The work plan includes three cycles, in which different data sources are combined, the algorithms' scientific and operational maturity is increased, and user feedback is incorporated.

This document is an updated version of the D2.1 ATBD V1.2. It describes the OptiSAIL algorithm used in cycle 2 and its application to TOC reflectances from multiple sensors. During cycle-1, the OptiSAIL cloud contamination simulation was activated, and a pre-filtering algorithm was implemented. As a reaction to user preferences, the fAPAR absorbed by Chlorophyll A+B (fAPAR_Cab) is computed as additional output of OptiSAIL. The version for cycle-2 uses the posterior covariance matrix of the previous retrieval at the same location to modify the prior assumption, thus introducing a temporal correlation in the data, based on the assumption that the model variables have individual timescales on which they change.

Table of Contents

LIST OF ACRONYMS

BHR	Bi-Hemispherical Reflectance
BRDF	Bidirectional Reflectance Distribution Function
BRF	Bi-directional reflectance
CCI+	Climate Change Initiative Plus
DHR	Directional-Hemispherical reflectance
ECV	Essential Climate Variable
ED	External Document (as listed in section 1.2)
EOF	Empirical Orthogonal Function
fAPAR	*f*raction of Absorbed Photosynthetically Active Radiation
HDR	Hemispherical-Directional reflectance
ID	Internal Document (as listed in section 1.2)
LAI	Leaf Area Index
NIR	Near Infra-Red range of the electromagnetic spectrum, here 700--2500 nm
PCA	Principal Components Analysis
PROSPECT	PROperties of leaf SPECTtra
RT	Radiative Transfer
SAIL	Scattering of Arbitrarily Inclined Leaves
TAF	Transformation of Algorithms in Fortran
TARTES	Two-streAm Radiative TransfEr in Snow
TOA	Top-Of-Atmosphere
TOC	Top-Of-Canopy
VIS	VISible range of the electromagnetic spectrum, here 400–700 nm
VP	Vegetation Parameters

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Scope of this document

This document updates the theoretical basis of the algorithms (ATBD-V1.1) used in cycle 1 of the CCI+ Vegetation Parameters project (ID1) for the production of the CRDP-2 with processor version OptiSAIL-r37088M. The OptiSAIL algorithm retrieves LAI and fAPAR together with other parameters directly from TOC reflectances. Because OptiSAIL includes turbid-medium radiative transfer model and does not account for horizontal sub-pixel heterogeneity, the LAI is an effective LAI that disregards possible effects of vegetation clumping (partial vegetation coverage and shoot, branch and crown clumping).

1.2 Related documents

Internal documents

Reference ID	Document
ID1	Climate Change Initiative Extension (CCI+) Phase 2 New ECVs: Vegetation Parameters – EXPRO+ (ITT)
VP-CCI_D4.2_PUG_V2.0	Product User Guide (PUG) CRDP-2, ESA CCI+ Vegetation Parameters
VP-CCI_D2.4_PVASR_V1.0	Product Validation and Algorithm Selection Report, ESA CCI+ Vegetation Parameters
VP-CCI_D2.1_ATBD-pre-processing_V1.0	Algorithm Theoretical Basis Document of the pre-processing of the sensors used in the LAI and fAPAR retrieval.

External documents

Reference ID	Document
ED1	C3S ATBD of Surface Albedo, multi-sensor, D1.3.4-v2.0_ATBD_CDR_SA_MULTI_SENSOR_v2.0_PRODUCTS_v1.1
ED2	CGLS ATBD Sentinel-3 OLCI and SLSTR atmospheric correction , CGLOPS1_ATBD_S3-AC-V1_I1.30

1.3 General definitions

Leaf Area Index (LAI) is defined as the total one-sided area of all leaves in the canopy within a defined region, and is a non-dimensional quantity, although units of [m2/m2] are often quoted, as a reminder of its meaning (Zemp et al., 2022). The selected algorithm in the CCI-Vegetation Parameters project uses a 1-D radiative transfer model, and LAI is uncorrected for potential effects of crown clumping. Its value can be considered as an effective LAI, notably the LAI-parameter of a turbid-medium model of the canopy that would let the model have similar optical properties as the true 3-D structured canopy with true LAI (Pinty et al., 2006). Additional information about the geometrical structure may be required for this correction to obtain true LAI (Nilson, 1971), which involves the estimation of the clumping index, CI, defined as the ratio between the true and effective LAI [see Fang (2021) for a review of methods to estimate CI].

Fraction of Absorbed Photosynthetically Active Radiation (fAPAR) is defined as the fraction of Photosynthetically Active Radiation (PAR; solar radiation reaching the surface in the 400-700 nm spectral region) that is absorbed by a vegetation canopy (Zemp et al., 2022).

2 CRDP-2

For the production of the CRDP-2, the OptiSAIL algorithm was used, with some changes from the version used for CRDP-1. These are mainly the mixed prior approach described in section 4.2 and the use of data from more instruments. See Table 1 for the differences between the CRDP-1 inputs and the CRDP-2 input candidates. Section 5 presents the selected input data for CRDP-2. CRDP-2 contains a selection of data layers from OptiSAIL, applied to TOC reflectances retrieved with multiple combinations of sensors. This data is described in the PUG (VP-CCI_D4.2_PUG).

Table 1: Differences of Inputs datasets prepared for use in CRDP-1 and CRDP-2.

Platform	Sensor	Band name In data	Central Wavelength of SRF (nm)	CRDP-1	CRDP-2 Candidate	band flags	Snow Identification	Cloud Identification	ac_flag (CRDP-2)	SZA/VZA filter (deg; CRDP-2)	remarks
METOP-A	AVHRR/3	TOC_1	633		x	NA	status_map is „ice_snow“	status_map is „cloudy“	Drop „Climato“	65/65	
		TOC_2	864		x	NA					
		TOC_3a	1606.5		x	NA					
METOP-B	AVHRR/3	TOC_1	633		x	NA	status_map is „ice_snow“	status_map is „cloudy“	Drop „Climato“	65/65	
		TOC_2	862		x	NA					
		TOC_3a	1607.5		x	NA					
METOP-C	AVHRR/3	TOC_1	628.96		(x)	NA	status_map is „ice_snow“	status_map is „cloudy“	Drop „Climato“	65/65	unreliable geolocation
		TOC_2	837.81		(x)	NA					unreliable geolocation
		TOC_3a	1607.22		(x)	NA					unreliable geolocation
S-NPP	VIIRS	TOC_M01	410.57		x	TOC_Mnn_quality_flags is not any of „Substitute_Cal“ or „Out_of_Range“ or „Saturation“ or „Temp_not_Nominal“ or „DG_Anomaly“ or „Some_Saturation“ or „Missing_EV“ or „Cal_Fail“ or „Dead_Detector“ or „Noisy_Detector“	NA	Integer_Cloud_Mask is not any of „probably clear“ or „confident clear“	Drop „Climato“	65/65	
		TOC_M02	443.47		x						
		TOC_M03	486.19		x						
		TOC_M04	550.47		x						
		TOC_M05	671.25		x						
		TOC_M06	745.27		x						high uncertainty
		TOC_M07	861.61		x						
		TOC_M08	1238.26		x						
		TOC_M10	1601.16		x						
		TOC_M11	2256.99		x						

NOAA-20	VIIRS	TOC_M01	410.76		x	TOC_Mnn_quality_flags is not any of „Substitute_Cal“ or „Out_of_Range“ or „Saturation“ or „Temp_not_Nominal“ or „DG_Anomaly“ or „Some_Saturation“ or „Missing_EV“ or „Cal_Fail“ or „Dead_Detector“ or „Noisy_Detector“	NA	Integer_Cloud_Mask is not any of „probably clear“ or „confident clear“	Drop „Climato“	65/65	
		TOC_M02	444.56		x						
		TOC_M03	488.23		x						
		TOC_M04	558.48		x						
		TOC_M05	668.14		x						
		TOC_M06	745.88		x						high uncertainty
		TOC_M07	867.56		x						
		TOC_M08	1238.51		x						
		TOC_M10	1604.38		x						
		TOC_M11	2258.8		x						
PROBA-V	VNIR	LEVEL2B/band1/TOC	463.5	x	x	Good_Blue	SM_probav_v2 is „IceSnow“	SM_probav_v2 is „Shadow“ or „Undefined“ or „Cloud“	NA	65/65	
		LEVEL2B/band2/TOC	655	x	x	Good_Red					
		LEVEL2B/band3/TOC	839	x	x	Good_Nir					
	SWIR	LEVEL2B/band4/TOC	1602.5	x	x	none					Also use interpolated pixels
SPOT 4/VGT1	VEGETATION 1	LEVEL2B/band1/TOC	457.5	x	x	Good_Blue	SM is „IceSnow“	SM is „Shadow“ or „Undefined“ or „Cloud“	NA	65/65	
		LEVEL2B/band2/TOC	658.75	x	x	Good_Red					
		LEVEL2B/band3/TOC	833.75	x	x	Good_Nir					
		LEVEL2B/band4/TOC	1648.75	x	x	none					Also use interpolated pixels
SPOT 5/VGT2	VEGETATION 2	LEVEL2B/band1/TOC	457.5	x	x	Good_Blue	SM is „IceSnow“	SM is „Shadow“ or „Undefined“ or „Cloud“	NA	65/65	
		LEVEL2B/band2/TOC	653.75	x	x	Good_Red					
		LEVEL2B/band3/TOC	837.5	x	x	Good_Nir					
		LEVEL2B/band4/TOC	1635	x	x	none					Also use interpolated pixels
Sentinel-3A	OLCI	Oa02_toc	411.82		x		Quality_flag is not any of			65/65	

		Oa02_toc	442.95		x	*Already Quality Screened*	„SNOW_ICE“ or „MIXED_CLEAR_SNOW_ICE“	*already Cloud Screened*	Drop „Climato“		
		Oa04_toc	490.45		x						
		Oa05_toc	510.42		x						
		Oa06_toc	560.43		x						
		Oa07_toc	620.38		x						
		Oa08_toc	665.24		x						
		Oa09_toc	674.01		x						
		Oa10_toc	681.56		x						
		Oa11_toc	709.12		x						
		Oa12_toc	754.18		x						
		Oa16_toc	779.22		x						
		Oa17_toc	865.59		x						
		Oa18_toc	884.36		x						
		Oa21_toc	1012.93		x						
Sentinel -3B	OLCI	Oa02_toc	411.94		x	Already Quality Screened	Quality_flag is not any of „SNOW_ICE“ or „MIXED_CLEAR_SNOW_ICE“	*already Cloud Screened*	Drop „Climato“	65/65	
		Oa02_toc	443.02		x						
		Oa04_toc	490.37		x						
		Oa05_toc	510.36		x						
		Oa06_toc	560.35		x						
		Oa07_toc	620.25		x						
		Oa08_toc	665.1		x						
		Oa09_toc	673.86		x						
		Oa10_toc	681.38		x						
		Oa11_toc	708.97		x						
		Oa12_toc	754.03		x						
		Oa16_toc	779.09		x						
		Oa17_toc	865.43		x						
		Oa18_toc	884.19		x						
		Oa21_toc	1012.76		x						

3 Processing and input data

Figure 1 sketches the structure for algorithm development and successively increasing data complexity in the three development cycles of this project. Atmospherically corrected, gridded TOC reflectance data have been used as input. The processing chain resulting in these input data is described in the atmospheric correction and pre-processing ATBD [VP-CCI_D2.1_ATBD-pre-processing_V1.0]. Figure 2 In each cycle of the project, different satellite products have been used. Figure 1 gives an overview of the data use per period and per cycle. The TOC reflectance datasets which will be produced for this project in cycles 2 and 3 will use the same approach for atmospheric correction as the products used in cycle 1.

Figure 2 gives an overview of the processing structure. OptiSAIL uses a mechanism to select reflectance data from multiple sources for a given temporal aggregation window. To improve the temporal resolution for situations with many usable observations, the TOC reflectance uncertainties are inflated with a time-dependent factor ($f(\Delta t)$). It is exponential in time difference Δt between the window centre and the observation time, and 1 for Δt=0 and 2 for Δt =120h (5d), computed as $f(\Delta t) = 2^{\left(\frac{\Delta t}{120\,h}\right)}$, thus a doubling of the uncertainty per 5 days.

Cloud and quality flags of the data are used to exclude pixels from processing whenever they are raised. Snow flags are used to select a higher value of the snow prior in the inversion.

To avoid outliers in the observations, a pre-filtering criterion based on a maximum anisotropy threshold was implemented. Bright outliers are filtered out as follows: for all sensors with bands at wavelengths shorter than 650 nm (CRDP-1: 500 nm) the band with the shortest wavelength is selected for comparison. All reflectances at this band higher than two times the lowest reflectance at this band are excluded from the retrieval. For this comparison no angular effects are considered, that is, the reflectances are treated as if they were Lambertian. This is expected to avoid the use of observations that suffer from undetected cloud, high aerosol, or similar effects in the retrieval.

A second filtering step selects out of the remaining observations only those at the three bands closest to the centre date of the observation window. This helps to increase temporal resolution and processing speed, and somewhat limits the effect of unknown reflectance error correlations, which are currently not considered. In order to have a better constraint of the cloud thickness parameter, observations are grouped together when they have an observation time that lies within the same period of 5 minutes. This can have the effect that more than three observations on the same band of one sensor can be used for the retrieval.

A common land sea mask is used for all TOC input datasets to define if a pixel is processed or not. The land-sea mask is derived from the ESA Sentinel-3 OLCI surface classification but resampled to 1 km resolution. All pixels with a fraction of 0.3 land cover are classified as land.

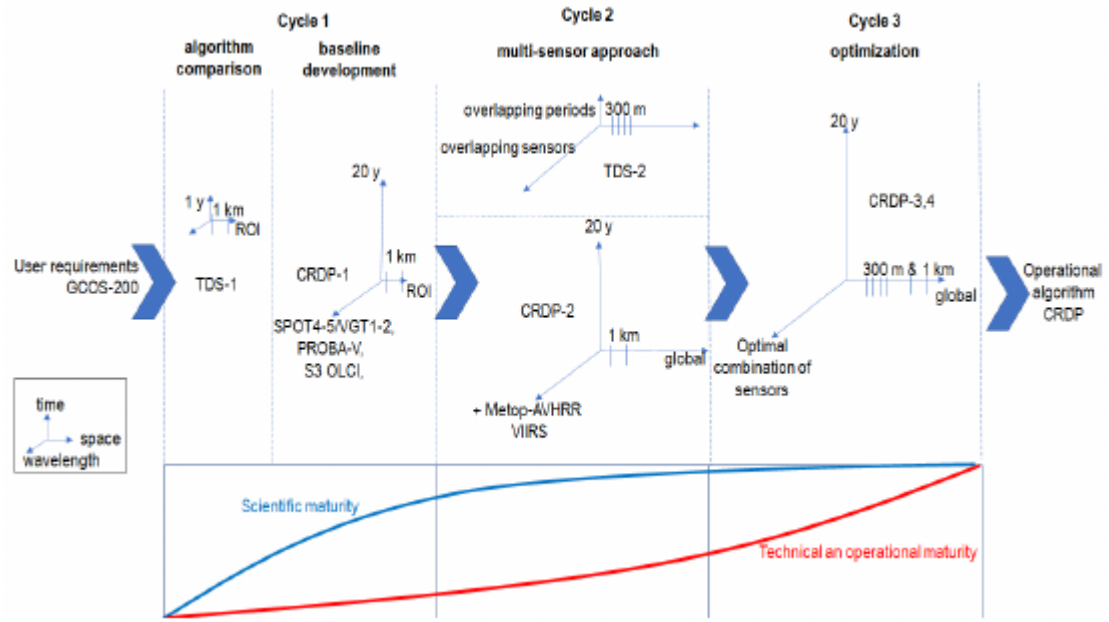

Figure 1: General concept of the three cycles, with progressive inclusion of sensors, spatial and temporal coverage and resolution, with the dimensions of the test datasets (TDS) and climate research data packages (CRDP) illustrated. The initially emphasis is on the implementation of an innovative approach, gradually shifting towards selection and optimization for an operational context.

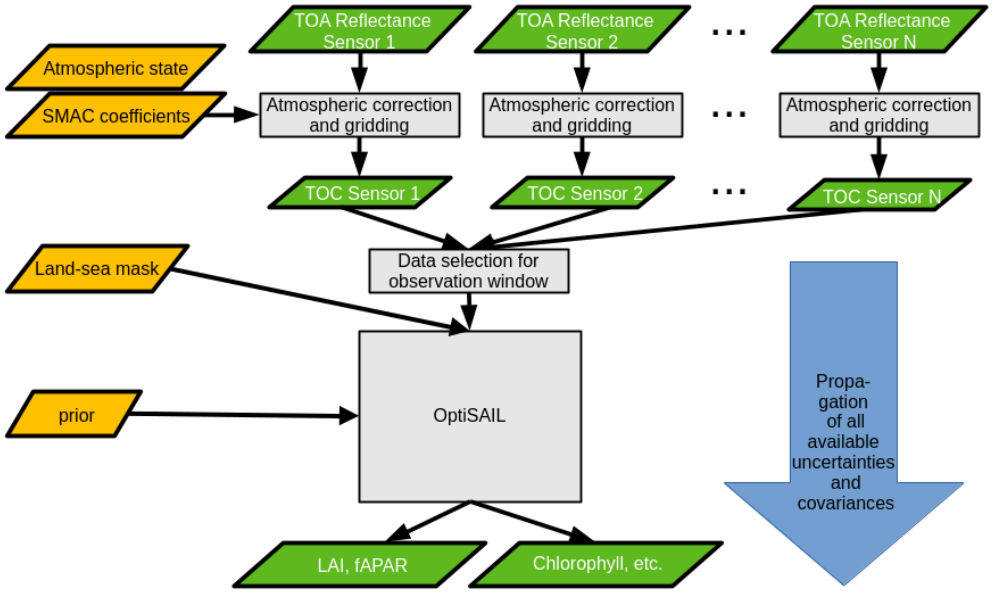

Figure 2: Processing diagram for CCI+ VP CRDP-2.

4 OptiSAIL

4.1 Algorithm Summary

OptiSAIL is a retrieval and error propagation framework. It uses automatic differentiation for gradient, Jacobian and Hessian computations. It is built around the established components 4SAILH (Scattering of Arbitrarily Inclined Leaves, with 4-stream extension and hot-spot), PROSPECT-D (simulation of leaf spectra, version D including senescence, (Féret et al., 2017)), TARTES (Two-streAm Radiative Transfer in Snow (Libois et al., 2013), with the addition of an empirical soil reflectance model, a semi-empirical soil moisture model (Philpot, 2010), the Ross-Thick-Li-Sparse BRDF model, and a cloud contamination simulation. Table 2 shows the parameters of all sub-models that are retrieved simultaneously, and Figure 3 their prior distribution. In addition to the data pre-filtering for bright outliers in the band with the shortest wavelength less than 500 nm mentioned in section 2, the cloud contamination detection built into OptiSAIL (Blessing et al., 2024) was extended for the multi-sensor approach. It identifies and to some degree corrects for observations with residual cloud contamination. Test runs during cycle-1 showed quality issues when this option was not used.
Figure 4 gives an overview of the reflectance simulation and Figure 5 of the retrieval framework. The model is described with further references and demonstrated in (Blessing et al., (2024).

Table 2: OptiSAIL retrieved parameters by sub-model.

Parameter	Description	unit
	Cloud contamination sub-model	
L	Cloud thickness parameter vector (one entry per retrieval time)	m
	SAIL sub-model	
LAI	Leaf Area Index	m^2/m^2
ALIA	Average Leaf Inclination Angle (normal against zenith)	°
hspot	canopy hot-spot parameter	1
	PROSPECT-D sub-model	
N	leaf structure parameter	1
C_{ab}	chlorophyll a+b content	$\mu g/cm^2$
C_{Car}	carotenoids content	$\mu g/cm^2$
C_{Anth}	Anthocyanin content	$\mu g/cm^2$
C_{brown}	brown pigments content	1
C_w	equivalent water thickness	cm
C_m	dry matter content	g/cm^2
	Soil BRDF sub-model (Ross-Li-R)	
f_{vol}	volumetric scattering kernel factor	1
f_{geo}	geometric scattering kernel factor	1
	Snow sub-model (TARTES)	
h_{snow}	height of a single snow layer with fixed properties	1
	Soil albedo model (empirical+Philpot)	
EOF1	factor for empirical soil spectrum variation 1	1
EOF2	factor for empirical soil spectrum variation 2	1
moist	relative moisture saturation of soil (to field capacity)	1

4.2 Mixed prior extension

In the algorithm for CRDP-1, the retrieval was carried out for all pixels independently of each other, allowing for high flexibility in the parallelisation. It can, however, be assumed that the retrieved parameters follow temporal dynamics which has typical time scales for changes, with dry soil spectra changing slowest, and cloud cover fastest. In the algorithm for CRDP-2, we use this information by

taking the previous retrieval as prior information. Based on the difference in retrieval time, each parameter is relaxed towards the standard prior with an individual time scale (Table 3) with the following formula:

$$x(\Delta t) = \exp\left(-\frac{\Delta t}{\tau}\right) x_{previous} + \left(1 - \exp\left(-\frac{\Delta t}{\tau}\right) x_{default\ prior}\right)$$

For the elements of the prior covariance matrix, we then have:

$$K_{ij}^{x}(\Delta t) = \exp\left(-\frac{\Delta t}{\tau_j}\right) K_{ij}^{x_{previous}} \exp\left(-\frac{\Delta t}{\tau_i}\right) + \left(1 - \exp\left(-\frac{\Delta t}{\tau_j}\right)\right) K_{ij}^{x_{default\ prior}} \left(1 - \exp\left(-\frac{\Delta t}{\tau_i}\right)\right)$$

The mixed-prior approach can be configured to be used with or without the covariance matrix of the previous retrieval. For the relaxation times τ we have used literature values of (Yang et al., 2021), which are based on expert knowledge, with few modifications. For the leaf chlorophyll content a lower τ was used, considering that rapid changes may occur at senescence. For the thickness of residual clouds a τ of zero was used and for snow a low τ of 1 day.
In some cases, this could trap the algorithm in extreme and unlikely parts of the parameter space. Therefore, some filtering is applied to the previous retrieval before it is used for the computation of the new prior. First, it is checked whether any flags are raised which indicate sub-optimal quality. In a second step, the values are clipped in the dimensionless control parameter space, not to exceed 1.5 standard deviations of the default prior. The control parameters for Cab, Car, and Cm are in addition clipped at –0.5 standard deviations, enforcing the prior assumption that leaves on plants contain a certain level of pigments.

*Table 3: Time scales in **days** used for τ in the equations for the mixed prior*

Cloud contamination sub-model	
L	0
SAIL sub-model	
LAI	30
ALIA	30
Hspot	30
PROSPECT-D sub-model	
N	60
C_{ab}	7.5
C_{Car}	30
C_{Anth}	30
C_{brown}	30
C_{w}	30
C_{m}	30
Soil BRDF sub-model (Ross-Li-R)	
f_{vol}	60
f_{geo}	60
Snow sub-model (TARTES)	
h_{snow}	1
Soil albedo model (empirical+Philpot)	
EOF1	60
EOF2	60
Moist	2

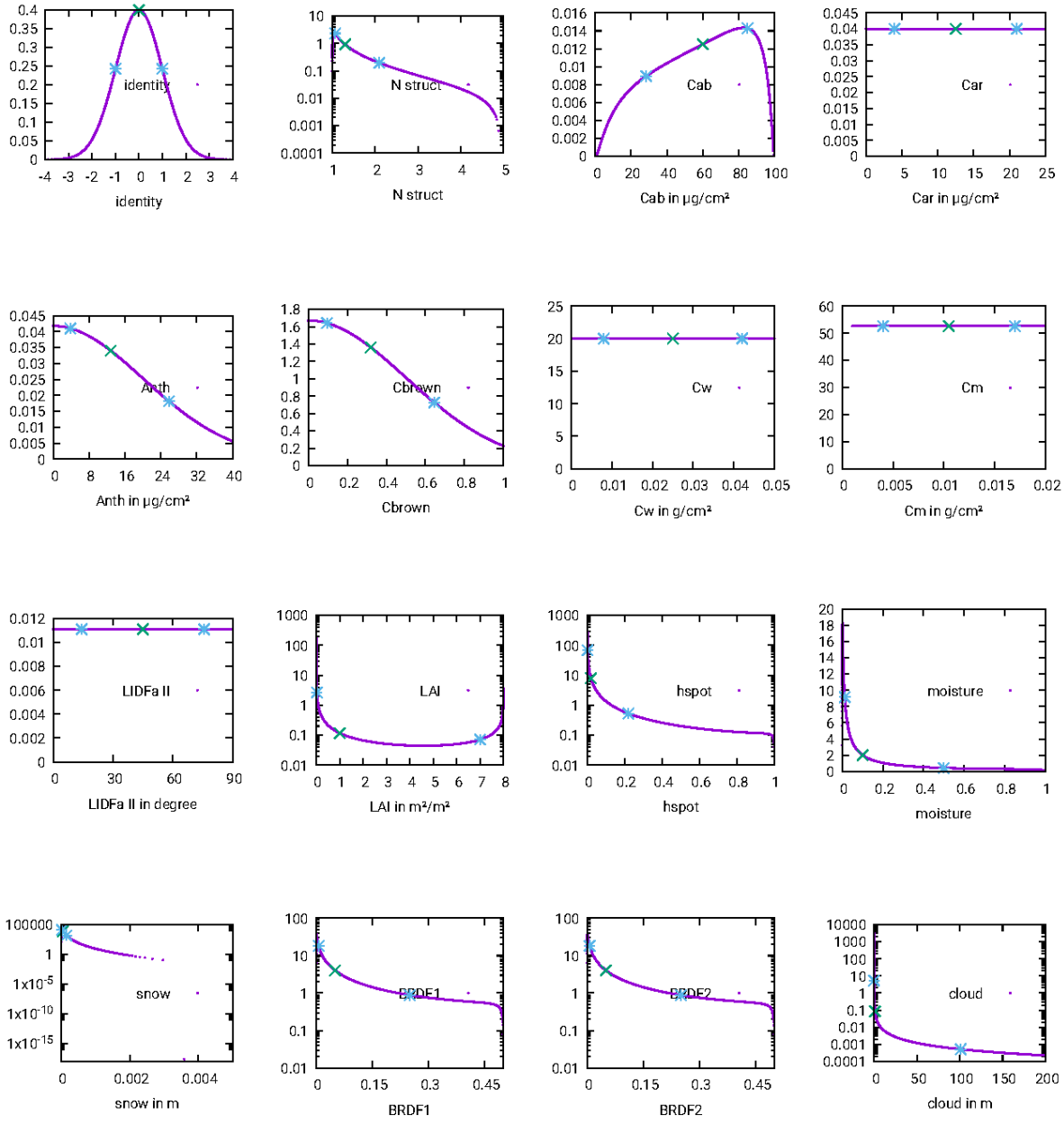

Figure 3: Prior distributions used in OptiSAIL. All model parameters are mapped to Gaussian control parameters for the minimisation, using these distributions.

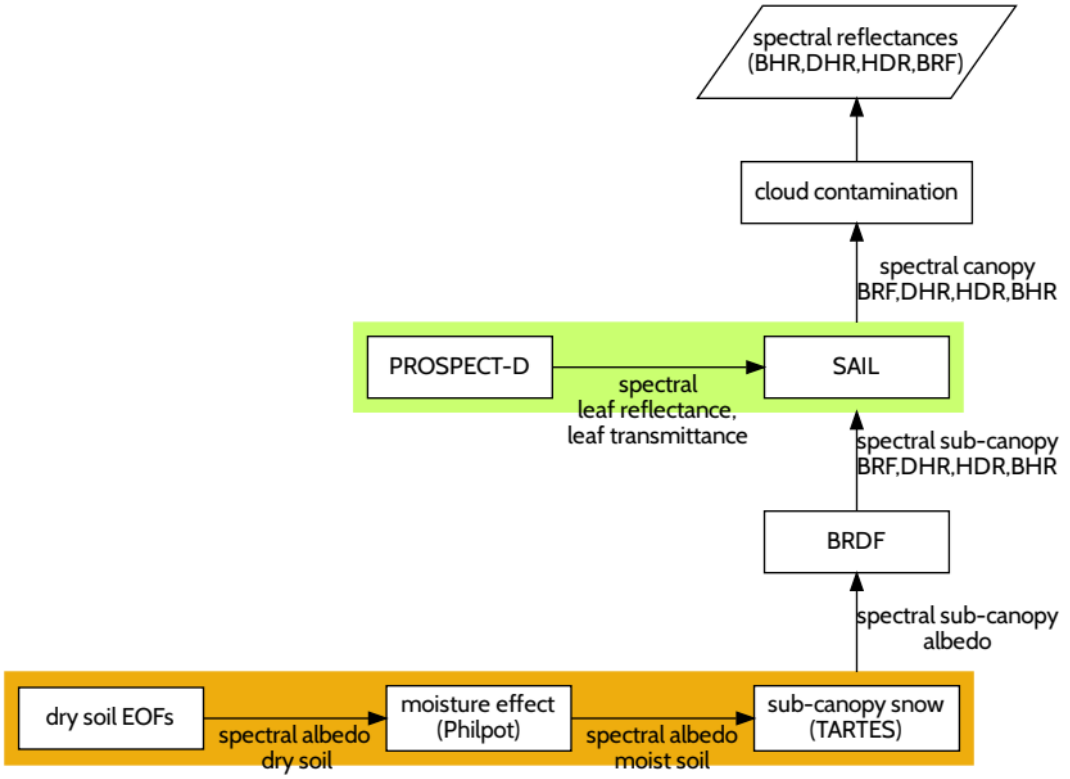

Figure 4: OptiSAIL reflectance simulation

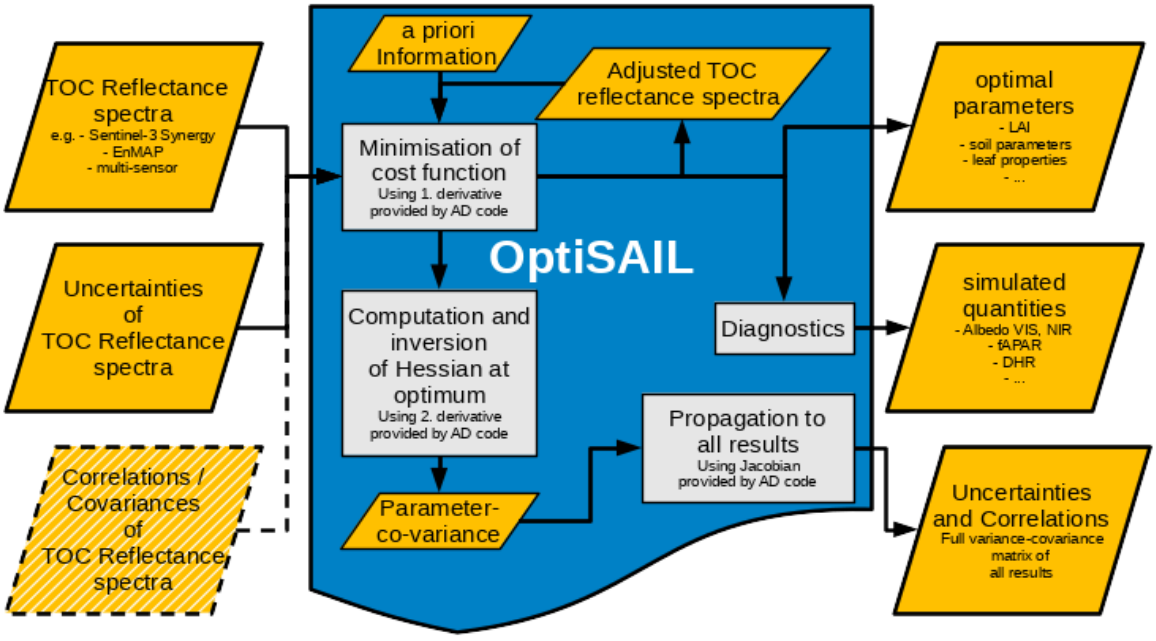

Figure 5: OptiSAIL retrieval framework with covariance propagation.

4.3 OptiSAIL output

All outputs for the CRDP-2 are on the same 1 km regular lat-lon grid as the TOC reflectance data used for input (ED1–3). The format is netCDF. For all retrieved and diagnosed quantities, the uncertainty corresponding to one standard deviation of a Gaussian distribution and the correlation of the uncertainty with all other retrieved and diagnosed quantities is given. In production, the correlation information can optionally be directed to a second output file. Table 4 gives an overview of the potential data layers. See the PUG (VP-CCI_D4.2_PUG) for the layers actually available to the user.

During cycle-1, the capability to compute fAPAR specifically absorbed by the leaf pigments Chlorophyll-A+B (*f*APAR_Cab) and Carotenoids (*f*APAR_Car) was added to OptiSAIL, by adopting the approach from SCOPE. In order to obtain the absorption by the pigment, the full leaf absorption spectrum (a_λ) is multiplied with the relative contribution to the total absorption of the respective pigment (Cab, Car):

$$a_{\lambda,pigment} = a_\lambda \frac{C_{pigment} \cdot k_{\lambda,pigment}}{\Sigma_{i=1}^{n} C_i k_{\lambda,i}} \quad (18)$$

The coefficient k_λ is the specific absorption coefficient as defined in PROSPECT-D, and *C* the pigment content. The summation in the denominator is over all pigments (van der Tol et al., 2019). Although the total absorption a_λ depends on many factors including multiple scattering between the soil and the vegetation, Eq 18 remains valid because the relative contribution to the absorption is conservative.
Using ASTMG173 solar spectrum for the irradiance $E_{e,\lambda}$, *f*APAR_pigment is then computed as

$$fAPAR_{pigment} = \frac{\sum_{\lambda=400\,nm}^{700\,nm} a_\lambda E_{e,\lambda}}{\sum_{\lambda=400\,nm}^{700\,nm} E_{e,\lambda}} \quad (19)$$

In cycle-1 and 2, this sum is computed with 10 nm steps and an irradiance spectrum averaged over 10 nm-intervals, for computational speed.

Some additional layers exist, containing further information about the retrieval.

- "**n_bands_used**" gives the number of observations on individual bands, which were used in the inversion. Currently a cut-off of three observations per band and sensor is used to limit the influence of potential error correlations of data retrieved with the same sensor and platform or using the same ancillary data in the atmospheric correction, and to improve computational speed. For a sensor with four bands, for example, "n_bands_used" has a maximum value of 12 (=3*4).
- "**p_chisquare**" gives the probability of a χ^2-distribution with the same number of degrees of freedom as the retrieval, to have a cost function value greater or equal than the one reached in the inversion of the pixel ($p_{\chi^2}(J_{min,n}) = p(x \geq J_{min,n} | X \sim \chi_n^2)$). Low values of "p_chisquare" are an indicator that model and data are inconsistent, and hence the retrieval quality is low. In CRDP-1 and 2, retrievals with p_chisquare < 0.001 are discarded (invcode is set to "RETR_UNTRUSTED" and data to the missing value). Retrievals with 0.001<p_chisquare<0.01 are marked as "RETR_UNTRUSTED" in "invcode".

Quality flags are collected in "invcode" (Table 5). Those beginning with "OPTIERR" and "XHESSERR" are mainly of technical interest and are kept to identify eventual numerical issues. "RETR_UNTRUSTED" and "RETR_LOW_QUALITY" are intended as guidance for the user. "RETR_UNTRUSTED" can be raised with or without missing data values. If there is data, it should only be used with great caution (if at all).
"RETR_LOW_QUALITY" combines a number of criteria to identify unreliable or bad data. These criteria are:

- "RETR_UNTRUSTED" is raised
- cloud_thk > 1 m (even "best" observation has cloud contamination above 1m of effective thickness)
- LAI > 3 and Chloropyll-A+B < 5 µg.cm-2 (dense canopy with extremely low Chlorophyll content, typically a bad solution triggered by complicated conditions, as varying snow cover during time window)
- LAI > 5 and Chloropyll-A+B < 15 µg.cm-2 (as above, but for even denser canopy)

This combination of criteria has proven quite effective excluding retrievals with quality issues. However, especially if the detection of senescent leaves is of interest to the user, then the LAI/Cab criteria may turn out to be problematic because in these cases the low-quality flag may not be justified. For the mixed prior extension, CRDP-2 – products contain three additional flags (bits 10, 11, and 12 listed in Table 4). Bit 12 (PRIOR_LAST_RETRIEVAL) indicates that the mixed prior approach was used and contains information from the previous retrieval. If the previous retrieval is not usable due to quality concerns, the default prior is used and bit 11 (PRIOR_UNTRUSTED) is set. If the current retrieval is invalid, the mixed prior is written to the file and bit 10 (RETR_GAP_FILLED) is set. This ensures that the information from previous retrievals is retained through periods without successful retrievals. As these values are relaxed towards the default prior, they should not be seen as an actual gap-filling of the data. Especially in extreme locations (e.g. desert, rain forest), the RETR_GAP_FILLED pixels should not be used or included in an analysis as they will not contain good results. Consequently, all non-parameter quantities are set to "missing" for such pixels.

Table 4: Potential data layers in OptiSAIL output. For all quantities, the standard error and the correlation with all other main layers is given. For layers included in CRDP-2, please see the PUG (VP-CCI_D4.2_PUG).

Name	Standard/long name	Unit
Time	time (dimension)	days since 1970-01-01 00:00
Lon	Longitude (dimension)	degrees_east
lat	Latitude (dimension)	degrees_north
N_struct	PROSPECT-D leaf structure parameter	1
Cab	PROSPECT-D leaf chlorophyll a+b content	ug.cm-2
Car	PROSPECT-D leaf carotenoids content	ug.cm-2
Anth	PROSPECT-D leaf Anthocyanin content	ug.cm-2
Cbrown	PROSPECT-D leaf brown pigments content in arbitrary units	1
Cw	PROSPECT-D leaf equivalent water thickness	g.cm-2
Cm	PROSPECT-D leaf dry matter content	g.cm-2
LIDFa_II	SAIL average leaf angle (degrees) for type II	degree
LAI	SAIL Leaf Area Index	m2.m-2
hspot	SAIL hot spot parameter (av. leaf size / canopy height)	1

soilEOF1	SURF soil reflectance model parameter 1	1
soilEOF2	SURF soil reflectance model parameter 2	1
moisture	SURF relative volumetric moisture saturation of soil (theta/theta_sat)	1
snowheight	SURF snow height below canopy	m
k_vol	RossLi-reciprocal RossThick kernel parameter k_vol	1
k_geo	RossLi-reciprocal LiSparse kernel parameter k_geo	1
cloud_thk	Optical thickness from cloud contamination detection, only reporting cloud thickness of observation with smallest cloud thickness.	m
*f*APAR	fraction of Absorbed Photosynthetically Active Radiation using diffuse ASTMG173	1
*f*APAR_Cab	*f*APAR absorbed by Chlorophyll-A+B	1
*f*APAR_Car	*f*APAR absorbed by Carotenoids	1
BHR_VIS	bi-hemispherical reflectance (albedo) in the visible range	1
BHR_NIR	bi-hemispherical reflectance (albedo) in the near infra-red range	1
BHR_SW	bi-hemispherical reflectance (albedo) in the shortwave range	1
DHR_VIS	directional-hemispherical reflectance (black-sky albedo), VIS, at local solar noon	1
DHR_NIR	directional-hemispherical reflectance (black-sky albedo), NIR, at local solar noon"	1
DHR_SW	directional-hemispherical reflectance (black-sky albedo), SW, at local solar noon	1
*name*_ERR	*name* standard_error	Unit of *name*
*name1_name2*_correl	*name1 name2* standard_error_correlation	1

Table 5: Quality flags as collected in "invcode" data layers in OptiSAIL output. Bits 3,7,10-31 are currently not used.

bit	value	Flag_meaning	Comment
0	1	NOT_PROCESSED	Pixel not processed (sea point or missing data)
1	2	OPTIERR_TOO_MANY_ITER	Inversion stopped at iteration limit
2	4	OPTIERR_LNSRCH	Inversion stopped for numerical reasons
4	16	XHESSERR_NOTSYM	The computed Hessian matrix is not symmetric, and uncertainties and correlations cannot be computed.
5	32	XHESSERR_INVERSION	The computed Hessian matrix cannot be inverted, and uncertainties and correlations cannot be computed.
6	64	XHESSERR_NOTPOSDEF	The computed Hessian matrix is not positive definite (e.g. if no cost function minimum was reached), and uncertainties and correlations cannot be computed.
8	256	RETR_UNTRUSTED	The retrieval is not trusted, because any of the

			previous bits with “ERR” in their name are raised, or the chi-square-criterion is violated.
9	512	RETR_LOW_QUALITY	The retrieval matches one or more criteria defined for low quality (see text for explanation).
10	1024	RETR_GAP_FILLED	Retrieval contains previous retrieval, time-relaxed towards the standard prior, in order to bridge a missing data/or unsuccessful retrieval gap.
11	2048	PRIOR_UNTRUSTED	Previous retrieval cannot be used for “"mixed prior” approach, falling back to standard prior.
12	4096	PRIOR_LAST_RETR	Mixed-prior retrieval using previous retrieval together with standard prior

5 Selection of the input data

The optimal combination of input data is defined by analysing the retrievals for 2 years (2012 and 2019) with each different sensor availability. For each of the years, various scenarios were evaluated. Table 6 and Table 7 define the scenarios for the year 2019 and 2012 respectively.

Table 6: Scenarios to evaluate input data (2019).

No	Sensor	all	1/family	No AVHRR
1	SuomiNPP-VIIRS	X	X	X
2	NOAA20-VIIRS	X		
3	MetopA-AVHRR	X		
4	MetopB-AVHRR			
5	MetopC-AVHRR	X	X	
6	PROBA-V	X	X	X
7	Sentinel3-OLCI	X	X	X

Table 7: Scenarios to evaluate input data (2012).

No	Sensor	all	No AVHRR
1	SuomiNPP-VIIRS	X	X
2	MetopA-AVHRR	X	
3	SPOT5-VGT2	X	X

Both the product accuracy and processing performance are evaluated and are used to select the optimal selection of the input data.

5.1 Evaluation method

5.1.1 Product quality

The quality of the different multi-sensor scenarios is assessed following the procedures described in the product validation plan [VP-CCI_D1.3_PVP_V1.1], which is compliant with the CEOS LPV recommendations (Fernandes et al., 2014). The analysis is focused on key criteria including spatial consistency, -temporal consistency, accuracy and precision of the test products. The proposed methodology relies on validation with ground-based reference (so-called direct validation) and satellite product intercomparison approaches.

The validation with ground reference is computed using three datasets:

(DIRECT 2.1) up-scaled maps (Camacho et al., 2024) with the CEOS LPV recommendations (Fernandes et al., 2014; Morisette et al., 2006) available for the 2012 year (5 sites over rice crops).

Copercnius GBOV V2 dataset, available for 2019 is also used, providing multi-temporal valuable information over 15 sites.

AMMA sites (Redelsperger et al., 2006), where variables were derived from the acquisition and the processing of hemispherical photographs taken along 1 km linear sampling transects

The product intercomparisons are computed over the LAND VALidation (LANDVAL) network (Fuster et al., 2020; Sánchez-Zapero et al., 2023, 2020) of sites, composed of 720 sites, is used for sampling global conditions. The different multi-sensor scenarios are mainly compared with the baseline mono-sensor CRDP-1 dataset based on SPOT/VGT (2012) and PROBA-V (2019). Two additional satellite-based products are also included for benchmarking:

Copernicus Land Monitoring Service (CLMS) collection 1 km V2 (Verger et al., 2023). The algorithm starts from the daily SPOT/VGT or PROBA-V and instantaneous first guess of the LAI/FAPAR variables. Then, a temporal smoothing and gap filling method is applied, using several techniques including the Savitzky-Golay filter, a climatology (Verger et al., 2013) or interpolation methods to smooth the time profile and fill the gaps.

NASA MODIS C6.1 (Knyazikhin et al., 1998). The operational LAI/fAPAR algorithm consists of a main algorithm that is based on 3D radiative transfer equation and a backup algorithm. Given atmosphere corrected Bidirectional Reflectance Factors (BRF) and their uncertainties, the main algorithm finds candidates of LAI and fAPAR by comparing observed and modeled BRFs that are stored in biome type specific Look-Up-Tables. The main algorithm may fail to localize a solution if uncertainties of input BRFs are larger than threshold values or due to deficiencies of the RT model that result in incorrect simulated BRFs. In such cases, a backup empirical method based on relations between NDVI and LAI/fAPAR (Knyazikhin et al., 1998; Myneni and Williams, 1994) is utilized to output LAI/fAPAR with relatively poor quality (called the backup algorithm).

The following criteria are analysed:

Product completeness: Completeness corresponds to the absence of spatial and temporal gaps in the data. Percentage of missing values are analysed over the LANDVAL network of sites.

Spatial consistency: refers the realism and repeatability of the spatial distribution of retrievals. Maps over the whole transect of different scenarios are analysed in order to check the reliability of values and to identify spatial inconsistencies for further analysis.

Temporal consistency: the realism of temporal variations is assessed over LANDVAL plus additional sites with availability of ground measurements (DIRECT 2.1, AMMA, GBOV).

Error evaluation: accuracy, precision and uncertainty (APU) of different scenarios are evaluated. Scatter-plots and validation metrics are generated for the different scenarios versus ground-based and satellite-based references for the available matchups. Additionally, the intra-annual precision (also known as smoothness) (Weiss et al., 2007), which correspond to temporal noise assumed to have no serial correlation within a season, is evaluated for the different scenarios and per biome type.

5.1.2 Processing performance

The runtime of the processing is calculated for the tile and site processing. Statistics will be derived from these per sensor combination.

5.2 Results

5.2.1 Product quality

5.2.2 Processing performance

5.3 Conclusion

7 References

Blessing, S., Giering, R., & van der Tol, C. (2024). OptiSAIL: A system for the simultaneous retrieval of soil, leaf, and canopy parameters and its application to Sentinel-3 Synergy (OLCI+SLSTR) top-of-canopy reflectances. *Science of Remote Sensing*, *10*, 100148. https://doi.org/10.1016/J.SRS.2024.100148

Camacho, F., Sánchez-Zapero, J., Fang, H., Weiss, M., Brown, L.A., 2024. CEOS LPV DIRECT V2.1: A database of upscaled LAI, FAPAR and Fcover values for satellite biophysical product validation. [Data set]. Zenodo. https://doi.org/10.5281/ZENODO.11235157

Fang, H. (2021). Canopy clumping index (CI): A review of methods, characteristics, and applications. *Agricultural and Forest Meteorology*, *303*, 108374.

Féret, J.-B., Gitelson, A. A., Noble, S. D., & Jacquemoud, S. (2017). PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sensing of Environment*, *193*, 204–215. https://doi.org/10.1016/j.rse.2017.03.004

Fernandes, R., Plummer, S.E., Nightingale, J., Baret, F., Camacho, F., Fang, H., Garrigues, S., Gobron, N., Lang, M., Lacaze, R., Leblanc, S.G., Meroni, M., Martinez, B., Nilson, T., Pinty, B., Pisek, J., Sonnentag, O., Verger, A., Welles, J.M., Weiss, M., Widlowski, J.-L., Schaepman-Strub, G., Román, M.O., Nicheson, J., 2014. Global Leaf Area Index Product Validation Good Practices. Version 2.0. In G. Schaepman-Strub, M. Román, & J. Nickeson (Eds.), Best Practice for Satellite-Derived Land Product Validation (p. 76): Land Product Validation Subgroup (WGCV/CEOS), doi:10.5067/do [WWW Document]. https://doi.org/10.5067/doc/ceoswgcv/lpv/lai.002

Fuster, B., Sánchez-Zapero, J., Camacho, F., García-Santos, V., Verger, A., Lacaze, R., Weiss, M., Baret, F., Smets, B., 2020. Quality Assessment of PROBA-V LAI, fAPAR and fCOVER Collection 300 m Products of Copernicus Global Land Service. Remote Sens. 12, 1017. https://doi.org/10.3390/rs12061017

Knyazikhin, Y., Martonchik, J. V., Myneni, R.B., Diner, D.J., Running, S.W., 1998. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. J. Geophys. Res. 103, 32257. https://doi.org/10.1029/98JD02462

Libois, Q., Picard, G., France, J. L., Arnaud, L., Dumont, M., Carmagnola, C. M., & King, M. D. (2013). Influence of grain shape on light penetration in snow. *The Cryosphere*, *7*(6), 1803–1818.

Morisette, J.T., Baret, F., Privette, J.L., Myneni, R.B., Nickeson, J.E., Garrigues, S., Shabanov, N. V., Weiss, M., Fernandes, R.A., Leblanc, S.G., Kalacska, M., Sánchez-Azofeifa, G.A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A.N., Lewis, T.E., Iiames, J.S., Colombo, R., Meroni, M., Busetto, L., Cohen, W.B., Turner, D.P., Warner, E.D., Petersen, G.W., Seufert, G., Cook, R., 2006. Validation of global moderate-resolution LAI products: A framework proposed within the CEOS land product validation subgroup. IEEE Trans. Geosci. Remote Sens. 44, 1804–1814. https://doi.org/10.1109/TGRS.2006.872529

Myneni, R.B., Williams, D.L., 1994. On the relationship between FAPAR and NDVI. Remote Sens. Environ. 49, 200–211. https://doi.org/10.1016/0034-4257(94)90016-7

Nilson, T. (1971). A theoretical analysis of the frequency of gaps in plant stands. *Agricultural Meteorology*, *8*, 25–38.

Philpot, W. (2010). Spectral reflectance of wetted soils. *Proceedings of ASD and IEEE GRS*, *2*, 1–12.

Pinty, B., Lavergne, T., Dickinson, R. E., Widlowski, J.-L., Gobron, N., & Verstraete, M. M. (2006). Simplifying the interaction of land surfaces with radiation for relating remote sensing products to climate models. *Journal of Geophysical Research: Atmospheres*, *111*(D2).

Redelsperger, J.L., Thorncroft, C.D., Diedhiou, A., Lebel, T., Parker, D.J., Polcher, J., 2006. African Monsoon Multidisciplinary Analysis: An International Research Project and Field Campaign. Bull. Am. Meteorol. Soc. 87, 1739–1746. https://doi.org/10.1175/BAMS-87-12-1739

Sánchez-Zapero, J., Camacho, F., Martínez-Sánchez, E., Lacaze, R., Carrer, D., Pinault, F., Benhadj, I., Muñoz-Sabater, J., 2020. Quality Assessment of PROBA-V Surface Albedo V1 for the Continuity

of the Copernicus Climate Change Service. Remote Sens. 2020, Vol. 12, Page 2596 12, 2596. https://doi.org/10.3390/rs12162596

Sánchez-Zapero, J., Martínez-Sánchez, E., Camacho, F., Wang, Z., Carrer, D., Schaaf, C., García-Haro, F.J., Nickeson, J., Cosh, M., 2023. Surface ALbedo VALidation (SALVAL) Platform: Towards CEOS LPV Validation Stage — Application to Three Global Albedo Climate Data Records. Remote Sens. 2023, Vol. 15, Page 1081 15, 1081. https://doi.org/10.3390/RS15041081

van der Tol, C., Vilfan, N., Dauwe, D., Cendrero-Mateo, M. P., & Yang, P. (2019). The scattering and re-absorption of red and near-infrared chlorophyll fluorescence in the models Fluspect and SCOPE. *Remote Sensing of Environment*, *232*. https://doi.org/10.1016/j.rse.2019.111292

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 Trans. Geosci. Remote Sens. 51, 1963–1972. https://doi.org/10.1109/TGRS.2012.2228653

Verger, A., Sánchez-Zapero, J., Weiss, M., Descals, A., Camacho, F., Lacaze, R., Baret, F., 2023. GEOV2: Improved smoothed and gap filled time series of LAI, FAPAR and FCover 1 km Copernicus Global Land products. Int. J. Appl. Earth Obs. Geoinf. 123, 103479. https://doi.org/10.1016/J.JAG.2023.103479

Weiss, M., Baret, F., Garrigues, S., Lacaze, R., 2007. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. Remote Sens. Environ. 110, 317–331. https://doi.org/10.1016/j.rse.2007.03.001

Yang, P., Verhoef, W., Prikaziuk, E., & van der Tol, C. (2021). Improved retrieval of land surface biophysical variables from time series of Sentinel-3 OLCI TOA spectral observations by considering the temporal autocorrelation of surface and atmospheric properties. *Remote Sensing of Environment*, *256*, 112328. https://doi.org/10.1016/J.RSE.2021.112328

Zemp, M., Chao, Q., Han Dolman, A. J., Herold, M., Krug, T., Speich, S., Suda, K., Thorne, P., & Yu, W. (2022). *GCOS 2022 implementation plan*.