vegetation parameters cci

CCI Vegetation

D2.1 Algorithm Theoretical Basis Document Cycle 1

Simon Blessing

October 2022

Distribution list

Author(s)	:	Simon Blessing
Reviewer(s)	:	Else Swinnen
Approver(s)	:	Clement Albergel
Issuing authority	:	VITO

Change record

Release	Date	Pages	Description of change	Editor(s)/Reviewer(s)
V1	15/09/2022	all	First version	Simon Blessing / Else Swinnen
V1.1	05/10/2022	all	Answer to RIDs	Simon Blessing / Else Swinnen

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 such as LAI and fAPAR from multi-platform and multi-mission satellite data and interact with the user community to match their requirements. The work plan includes three cycles, in which different data sources are combined and the algorithms' scientific and operational maturity is increased and user feedback is incorporated. This document describes the algorithms used in cycle 1, which use existing TOC reflectance datasets from Proba-V, SPOT-Vegetation, and Sentinel-3 OLCI. Two processing chains are described. One is the combination of OptiAlbedo and TIP (Two-stream Inversion Package), and the other one is OptiSAIL (based on the established PROSPECT, SAIL and other models).

Table of Contents

LIST OF ACRONYMS

BHR	Bi-Hemispherical Reflectance
BRDF	Bi-hemispheric Reflectance Distribution Function
CCI+	Climate Change Initiative Plus
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
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
TIP	Two-stream Inversion Package
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 describes the theoretical basis of the algorithms used cycle 1 of the CCI+ Vegetation Parameters project (ID1). For the retrieval from existing datasets of gridded TOC reflectance, two retrieval algorithms are used in this cycle. These are the OptiAlbedo albedo retrieval algorithm, which feeds into TIP (Two-stream-Inversion-Package) for the retrieval of fAPAR and effective LAI, and the innovative OptiSAIL which retrieves LAI and fAPAR together with other parameters directly from TOC reflectances. Both algorithmic chains are using the same interface for accessing the multi-sensor, multi-band, multi-angular data for the retrievals.

1.2 Related documents

Internal documents

Reference ID	Document
ID1	Climate Change Initiative Extension (CCI+) Phase 2 New ECVs: Vegetation Parameters – EXPRO+ (ITT)

External documents

Reference ID	Document
ED1	VGT and Proba-V ATBD, D1.3.4-v2.0_ATBD_CDR_SA_MULTI_SENSOR_v2.0_PRODUCTS_v1.1
ED2	Sentinel-3 surface albedo ATBD, CGLOPS1_ATBD_S3-AC-V1_I1.30
ED3	TIP ATBD, D1.4.5-v4.0_ATBD_CDR-ICDR_LAI_FAPAR_SENTINEL3_v4.0_PRODUCTS_v1.0
ED4	C3S TIP quality, C3S LAI/fAPAR v2-4 quality reports
ED5	C3S TIP product user guide C3S_D312b_Lot5.3.3.10-v4.1_PUGS_CDR-ICDR_LAI_FAPAR_SENTINEL3_v4.0_PRODUCTS_v1.1

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 [GCOS-200, 2016].

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 [GCOS-200, 2016].

2 Processing and input data

Figure 1 sketches out the structure for algorithm development and successively increasing data complexity in the three development cycles of this project. Cycle 1 uses existing, atmospherically corrected, and gridded TOC reflectances as input. The datasets are intermediate products of surface albedo production chains with validated output. The ATBDs for these chains are listed as external documents ED1 for VGT and Proba-V 1 km gridded TOC reflectances, and ED2 (using v2, documentation of which is not yet publicly available) for Sentinel-3 gridded TOC reflectances. Sentinel-3 data are re-gridded to 1 km as outlined in the Annex (p. 23). Figure 1 also gives an overview of the periods and satellite products used in cycles 1-3. 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. Both algorithmic chains, OptiAlbedo+TIP and OptiSAIL use the same data selection mechanism in order to use the same set of observations from multiple sources for a given temporal aggregation window. In order to improve the temporal resolution for situations with many usable observations, the TOC reflectance uncertainties are weighted with a time-dependent factor. It is exponential in time difference Δt between the window centre and the observation time, and 1 for Δt=0 and 2 or Δt =120h (5d).

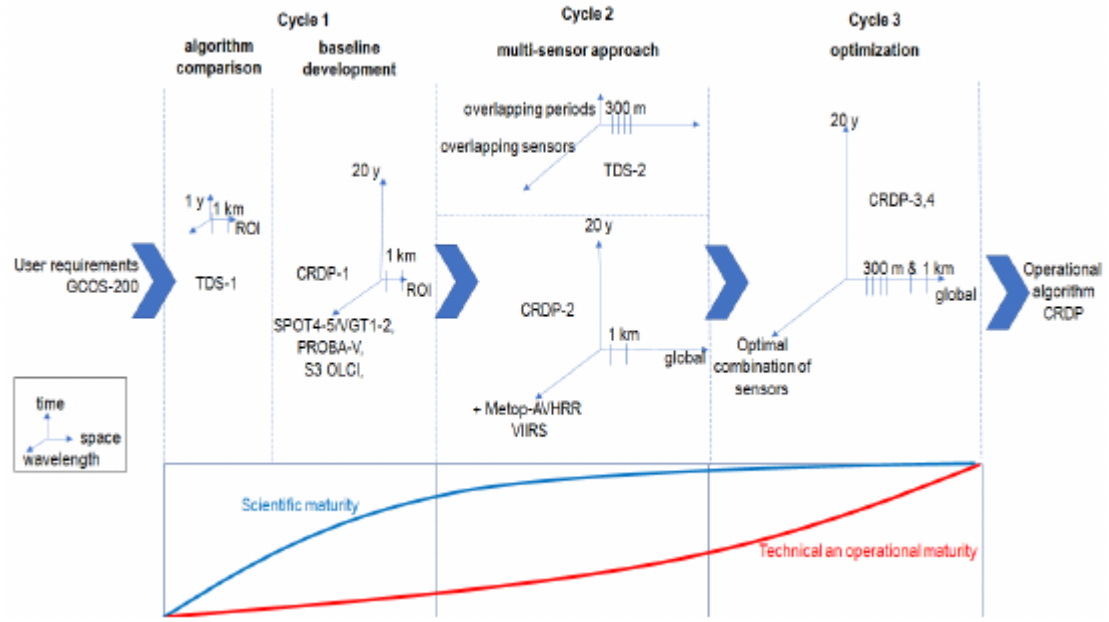


Figure 1: General concept of the three cycles, with progressive inclusion of sensors, spatial and temporal coverage and resolution, with the dimensions of the test datasets (TDS) and climate research data packages (CRDP)illustrated. The initially emphasis is on the implementation of an innovative approach,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

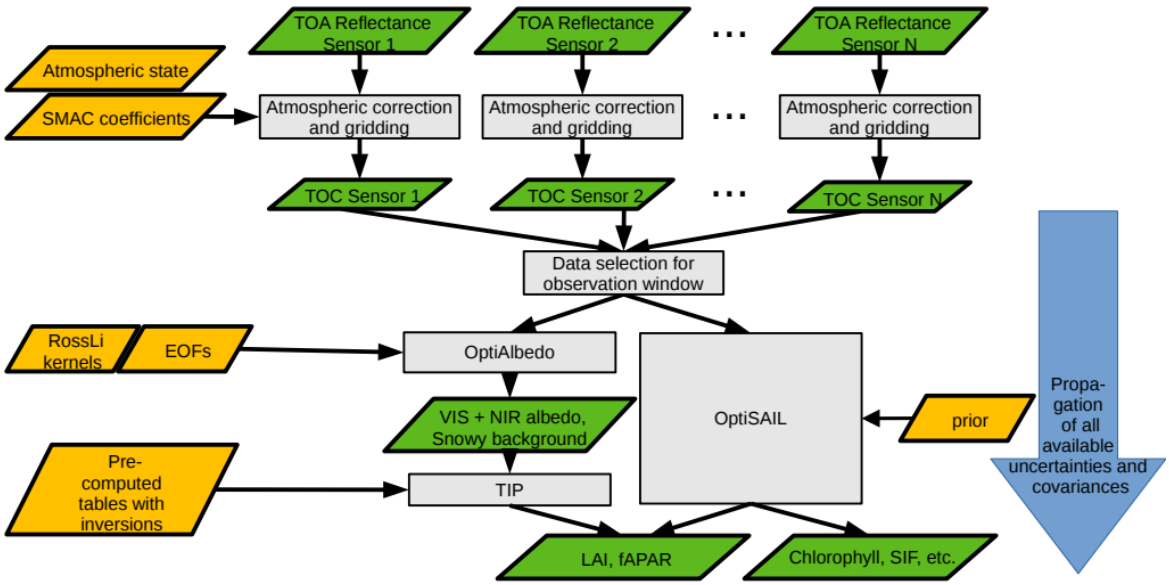


Figure 2: Processing diagram for CCI+ VP.

3 OptiAlbedo

3.1 Algorithm Summary

OptiAlbedo is a two-step algorithm for the estimation of surface albedo from a given arbitrary constellation of TOC reflectances. It is based on a simulated dataset of Ross-thick-Li-sparse model kernel factor spectra in order to derive a model formulation with parameters which are independent from viewing and illumination geometries.

In the first step, OptiAlbedo uses the results of a PCA (Principal Components Analysis) of the synthetic Ross-thick-Li-sparse kernel model parameter spectra projected onto the band wavelengths of the given sensor combination. By using not more than the first N empirical vectors, a reduced order model of TOC reflectance is formulated, which has parameters that are independent of wavelength and of geometry. The parameters of this model are determined for the observed TOC reflectance data by minimising the simulated misfit.

In the second step, a regression model is used to convert the kernel factor estimated in the first step into albedo. A similar regression model is used to retrieve the snow height below the canopy, which is a relevant input for TIP (see section 4) in order to identify bright backgrounds.

The algorithm makes use of the reported TOC reflectance uncertainties, and, if available, covariances, and provides the covariance matrix for all retrieved quantities.

3.2 Kernel factor dataset

Wanner et al. (1995) give the following formulae for the combination of the Ross-Thick with the Li-Sparse BRDF kernel model which are a direct consequence of the assumptions made in the derivation of these models (their Eqs. 67–69):

$$f_{iso} = \alpha C + (1-\alpha)\left(\frac{s}{3} + exp(-LAIB)\left(\rho_0 - \frac{s}{3}\right)\right) \quad (1)$$

$$f_{geo} = \alpha C \lambda \pi r^2 \quad (2)$$

$$f_{vol} = (1-\alpha)\frac{4s}{3\pi}\left(1 - exp(-LAIB)\right) \quad (3)$$

The meaning of the symbols is given in Table 1. For the term $\lambda\pi r^2$, the interpretation as a "crown area index" is illustrative, namely the fraction of the area that is covered with crowns. For the coupling with OptiSAIL or its components, ρ_0 ,LAI, and *s* have direct correspondence, while *C* is less obvious. In the current implementation it is set to reflectance simulated by OptiSAIL for an observer facing nadir with illumination from zenith.

Table 1: Symbols used in the Ross-Thick Li-Sparse BRDF model.

Symbol	Meaning
f_{iso}	Factor for isotropic reflectance kernel
f_{geo}	Factor for geometric reflectance kernel (Ross-Thick)
f_{vol}	Factor for volumetric reflectance kernel (Li-Sparse)
a	Areal proportion of α or $(1-\alpha)$ of land cover representing one or the other type of scattering (note that other interpretations of the parameter α are possible)
C	Reflectance of sunlit crown and sunlit ground alike, as seen by the sensor; from derivation of Li-Sparse kernel
LAI	Leaf Area Index of volume scattering area proportion
B	1.5, approximation of average 1/2(sec ϑ_i + sec ϑ_v)
ρ_0	Lambertian reflectance of flat horizontal surface under canopy, from Ross-thick kernel
λ	n/A, density of objects on a plane (e.g. crowns)
πr^2	Vertically projected area of single crown

3.3 Parameter space and sampling

The full parameter set for the simulation of the kernel factor spectra is a combination of the OptiSAIL parameters (see description of OptiSAIL in section 5) and the additional a and $\lambda\pi r^2$. Of the original OptiSAIL parameters, only those related to isotropic soil reflection, LAI and leaf single scattering albedo remain. Latin Hypercube sampling is used to generate sample parameter sets. The Latin Hypercube sampling uses ten disjunct intervals from the prior probability distribution as it is used for OptiSAIL (see section 5) with equal probability between the limits corresponding to the +/-2σ interval of a Gaussian distribution. Table 2 lists the parameters and the sampled range. The soil EOF parameter range is not fully representable of the actually used range, because the control vector is a non-linear mapping in order to avoid spectra with physically meaningless values. The maximum value of $\lambda\pi r^2$ is limited to 1/1.377622 in order to make sure that the computation of the albedo (by combining eqs. 1-3 with eq. 12 in section 3.6 below) always yields values greater or equal zero. As an alternative, such samples could just be discarded. From the drawn intervals, a pseudo-random value is drawn, which gives the actual parameter value. If, for one of the parameters, the Latin Hypercube is exhausted, that is, all intervals have been used once, then it is reset. It is ensured that no duplicate sets are drawn. For the database currently used, 2 million of such parameter sets are drawn and used for the simulation of the corresponding kernel factor spectra. An example of 100 such spectra is shown in Figure 3.

Table 2: Overview of parameter intervals used for the kernel factor samples.

Symbol	Meaning	Minimum	Maximum
LAI	Leaf Area Index	0.1744E-02	7.915
N	Leaf structure parameter	1.025	3.059
C_{ab}	Leaf chlorophyll a+b content	14.07	93.21
C_{ar}	Leaf carotenoids content	1.196	23.80
C_{Anth}	Leaf anthocyanin content	1.145	33.79
C_{brown}	Leaf brown pigments content	0.2863E-01	0.8447
C_w	Leaf equivalent water thickness	0.2439E-02	0.4761E-01
C_m	Leaf dry matter content	0.1909E-02	0.1909E-01
EOF1	Factor for empirical soil spectrum 1	-0.63747	0.9580
EOF2	Factor for empirical soil spectrum 2	-1.0674	1.8799
moist	Soil relative moisture saturation	0.2848E-02	0.8121
a	Areal proportion as in Table 1	0	1
$\lambda\pi r^2$	"Crown area Index" as in Table 1	0	0.6533

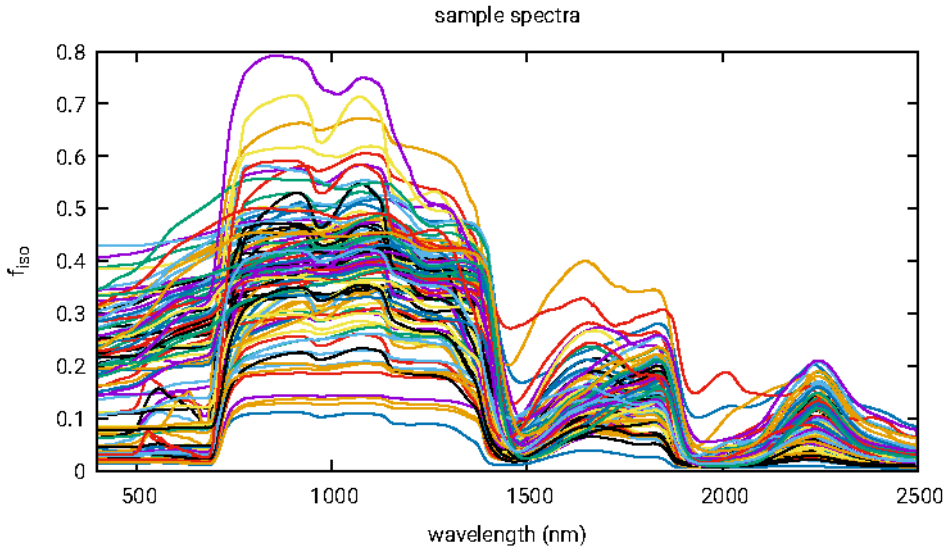


Figure 3: 100 examples of the 2 million generated kernel factor spectra (isotropic part).

3.4 Low-order spectral BRDF kernel factor model

In order to derive a low-order empirical model of the kernel factor spectra for a given sensor combination, the full spectra $B_{n\_full\,x\ n\_samles}$ (denoted B in the following) are projected onto the effective band wavelengths, yielding $B_{n\_0\,x\ n\_samples}$ (denoted B_0 in the following) by applying the sensors' spectral response functions, denoted by P:

$$P: B \rightarrow B_0 \tag{4}$$

In the vectors of the B-matrices, the spectra for the three kernel factors are concatenated. Thus, by computing the Empirical Orthogonal Functions (EOF) of the Principal Components Analysis (PCA), a joint model is achieved by currently using the 15 vectors explaining most of the variance of the data in B_0. The reflectance ρ for a given wavelength and viewing and illumination geometry is then computed as

$$\rho(\lambda,\theta_i,\theta_o,\phi_{rel}) = f_{iso}(\lambda)K_{iso}(\theta_i,\theta_o,\phi_{rel}) + f_{geo}(\lambda)K_{geo}(\theta_i,\theta_o,\phi_{rel}) + f_{vol}(\lambda)K_{vol}(\theta_i,\theta_o,\phi_{rel}) \quad (5)$$

with

$$f_{xxx}(\lambda) = \sum_{i=1}^{15} p_i \, EOF_{i,xxx}(\lambda), \quad (6)$$

where " xxx" denotes one of *iso, geo, vol*, and p_i are the wavelength-independent parameters of the model which have to be determined in the retrieval. The kernels K_{xxx} of the Ross-Thick-Li-Sparse-reciprocal BRDF model describe the angular dependence of the reflectance.

3.5 Kernel factor model parameter retrieval

The p_i from equation 6 are determined by minimising a cost function with a gradient-based algorithm. Gradient information is obtained from automatic differentiation of the forward model implementation. The cost function measures the difference between simulated and observed reflectances for the given scene geometries, thus allowing for an arbitrary combination of wavelengths and geometries. It is regularised with a prior term based on the explained variance from the PCA:

$$J(p,\rho_{obs}) = 1/2\left(\overrightarrow{\rho_{obs}} - \vec{\rho}(p)\right)^T \cdot C^{-1} \cdot \left(\overrightarrow{\rho_{obs}} - \vec{\rho}(p)\right) + \sum_{i=1}^{15} p_i \Big/ \sigma^2 \quad (7)$$

Observed and simulated reflectance are written as vectors here, containing the reflectances at the different band wavelengths as entries. The inverse covariance matrix C^{-1} of the TOC reflectance observations is diagonal, if no covariance terms are available in the data.

3.6 Bands-to-albedo regression model

In order to estimate the albedo, a regression model R is derived from the kernel factor spectral data B, which maps the B_0 back to the full spectrum:

$$R: B_0 \rightarrow \vec{B} \quad (8)$$

The model is chosen to minimise the following function:

$$L = (RB_0 - B)^T C^{-1}(RB_0 - B). \quad (9)$$

The weight matrix C^{-1} is included here, without loss of generality. Setting the derivative with respect to R to zero, we have to solve:

$$(C^{-1} + (C^{-1})^T)(RB_0 - B)B_0^T = 0. \quad (10)$$

It follows that

$$R = BB_0^T(B_0B_0^T)^{-1}, \quad (11)$$

Independent from any weights C^{-1}. The expression is equivalent to matrix $A_{nch\ x\ nch0}$ from eq. 19 of Yang et al. 2020 when their matrices $U_{nch\ x\ npc}U_{nch\ x\ npc}^T$ reduce to the unit matrix for nch = npc.
The estimated full kernel factor spectra are converted to albedo using the approximate integration formula of the Ross-Thick Li-Sparse model as given in Strahler et al. (1999):

$$BHR(\lambda) = f_{iso}(\lambda) + 0.189184 f_{vol}(\lambda) - 1.377622 f_{geo}(\lambda). \quad (12)$$

The spectral bi-hemishperic reflectance BHR(λ) is then weighted with a reference solar spectrum over the VIS (400–700 nm) and NIR (700-2500 nm) ranges in order to yield the visible and near-infrared

albedo. The solar reference spectrum used is the same as in OptiSAIL (*ASTM G173-03 Reference Spectra Derived from SMARTS v. 2.9.2.).*

The regression matrix R and the conversion to BHR VIS/NIR can be combined, yielding a set of regression coefficients for the given sensor combination. This is very similar to Liang's coefficients (Liang 2001) but mapping directly from the Ross-Li-kernel factors to albedo.

3.7 Identification of snowy backgrounds

The sub-canopy snow height parameter of OptiSAIL is included in the simulation of the kernel factor dataset B. Equivalent to the computation of R for the mapping from the band wavelengths to the full spectrum, a mapping R_{snow} is computed, to get an estimate of the snow height from the retrieved kernel factor spectra:

$$h_{snow} = R_{snow}\tilde{f_0} \tag{13}$$

The only difference to eqs. 8-11 is, that B is substituted with the know snow height control parameter used for the generation of the spectra. Snow status flags from the TOC reflectance data are currently ignored by the implementation.

3.8 Propagation of Covariance and Error Budget

The uncertainty of the TOC retrievals enters the computation through the inverse covariance matrix C^{-1} in the minimisation of J from eq. 7. Using automatic differentiation of the implemented retrieval, the posterior covariance matrix of the retrieved kernel factors is computed. This is a two-step process. First, the Hessian matrix at the minimum of J from eq. 7 is computed. Its inverse gives Σ^p, the posterior covariance of the retrieved model parameters p. Then this matrix is propagated through the algorithm using the Jacobians (again, from automatic differentiation), to the posterior covariance matrix of the retrieved kernel factors:

$$\Sigma^{\tilde{f}} = dJ/dp\,\Sigma^{\tilde{p}}(dJ/dp)^T \tag{14}$$

Since only the first 15 EOFs are used in the model, there are some weak but unresolved directions, which can be quantified by computing their covariance matrix:

$$\Sigma^{funresolved} = dJ/dp\,EOF_{16-n}EOF_{16-n}^T(dJ/dp)^T \tag{15}$$

Thus, the total posterior covariance of the kernel factor retrieval is

$$\Sigma^{ftotal} = \Sigma^{\tilde{f}} + \Sigma^{funresolved}. \tag{16}$$

For the regression step, we denote the actions of eqs. 12 and 13 as matrix A, and compute

$$\Sigma^{out} = AR\Sigma^{ftotal}(AR)^T, \tag{17}$$

in order to obtain the covariance of all output quantities of OptiAlbedo, namely BHR VIS, BHR NIR, and sub-canopy snow height h_{snow}.

3.9 OptiAlbedo output

All outputs are on the same 1 km regular lat-lon grid as the TOC reflectance data used for input (ED1–3). The output of OptiAlbedo is an intermediate product, as it provides the albedo and snow information required by TIP. The format is netCDF. Table 3 gives an overview of the layers used by TIP. Further layers exist, which may be removed as the algorithms evolves.

Table 3: Main OptiAlbedo output data layers.

Name	Long or standard name	Unit
time	time (dImension)	days since 1970-01-01 00:00
lon	Longitude (dimension)	degrees_east
lat	Latitude (dimension)	degrees_north
AL_BH_VI	bi-hemispherical reflectance (white-sky albedo) in the visible range	1
AL_BH_NI	bi-hemispherical reflectance (white-sky albedo) in the near infra-red range	1
AL_BH_VI_ERR	uncertainty (one standard deviation) of AL_BH_VI	1
AL_BH_NI_ERR	uncertainty (one standard deviation) of AL_BH_NI	1
AL_BH_VI_AL_BH_NI_correl	uncertainty correlation of AL_BH_VI and AL_BH_NI	1
invcode	Various diagnostics; bit corresponding to 0x32 is snow flag	(-)

4 Two-stream Inversion Package (TIP)

4.1 Algorithm Summary

TIP-LAI and TIP-fAPAR are effective Leaf Area Index and Fraction of Absorbed Photosynthetically Active Radiation, respectively. They are retrieved by applying the TIP to visible (VIS) and near-infrared (NIR) broadband albedos. TIP is based on the Two-stream Model developed by Pinty et al. 2006, which implements the two-stream approximation of radiative transfer for a homogeneous canopy ("1D-canopy"). For efficient processing, the retrievals are taken from tables of pre-computed inversions, generated with the TIP. It allows for the retrieval of effective LAI and fAPAR and their covariance from Top-Of-Canopy (TOC) VIS and NIR Bi-Hemispheric Reflectances (BHRs) and their joint variance-covariance matrix. By using the full variance-covariance matrix of the BHRs, it is a consistent enhancement beyond previous implementations of the TIP (e.g., Disney et al. 2013; Lavergne et al. 2006; Muller et al. 2013).

The retrievals of LAI and fAPAR using TIP are tied to the assumptions used in the Two-stream Model. This is the assumption of horizontal homogeneity over the whole pixel and vertical homogeneity over the canopy (1D approach), which yields effective values of the model's state variables, including the LAI. By means of a clumping factor, as suggested by Pinty et al. (2006), and a domain average, this can be related to the properties of realistic canopies, which are clumped at multiple scales. Direct measurements or retrievals of LAI-based 3D Radiative Transfer (RT) simulations will typically yield higher values (Jonckheere et al. 2004; Weiss et al. 2004). However, the assumptions used in the Two-stream Model are motivated by the need for consistency with those made in many surface flux retrievals, as well as those used in large-scale Land Surface Models (e.g., Widlowski et al. 2011). Verification studies of the retrieval algorithm have been done by Pinty et al. 2011a, b, c, d; Widlowski et al. 2011, and in the framework of the Copernicus Climate Change Service (C3S, see ED4).

The implementation of TIP used here is in essence identical to the one used in the C3S documented in ED3. An important difference is the availability of correlation information for the input BHRs coming from OptiAlbedo. Whenever this correlation exceeds +/-0.5, different inversion tables are used, which are pre-computed for a correlation of +/-0.5.

4.2 TIP LAI/fAPAR output

All outputs are on the same 1 km regular lat-lon grid as the TOC reflectance data used for input (ED1–3). The format is netCDF and the file contents are similar to those of the C3S LAI/fAPAR products v2–4 which are documented in the corresponding Product User Guide (ED4). Table 4 gives an overview of the data layers.

Table 4: Data layers in TIP output.

Name	Long/Standard name	Unit
time	Time (dimension)	days since 1970-01-01 00:00
lon	Longitude (dimension)	degrees_east
lat	Latitude (dimension)	degrees_north
LAI	Effective Leaf Area Index	1
LAI_ERR	Standard deviation of Effective Leaf Area Index	1
fAPAR	Fraction of Absorbed Photosynthetically Active Radiation	1
fAPAR_ER	Standard deviation of fAPAR	1
unc_correl	uncertainty correlation (LAI/fAPAR)	1
retrieval_flag	TIP retrieval procedure flags (diagnostic and traceability information)	(-)

5 OptiSAIL

5.1 Algorithm Summary

OptiSAIL is a retrieval and error propagation framework and 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 5 shows the parameters of all sub-models that are retrieved simultaneously, and Figure 4 their prior distribution.
Figure 3 gives an overview of the reflectance simulation and Figure 2 of the retrieval framework. The model is described with further references and demonstrated in Blessing et al. (2021).

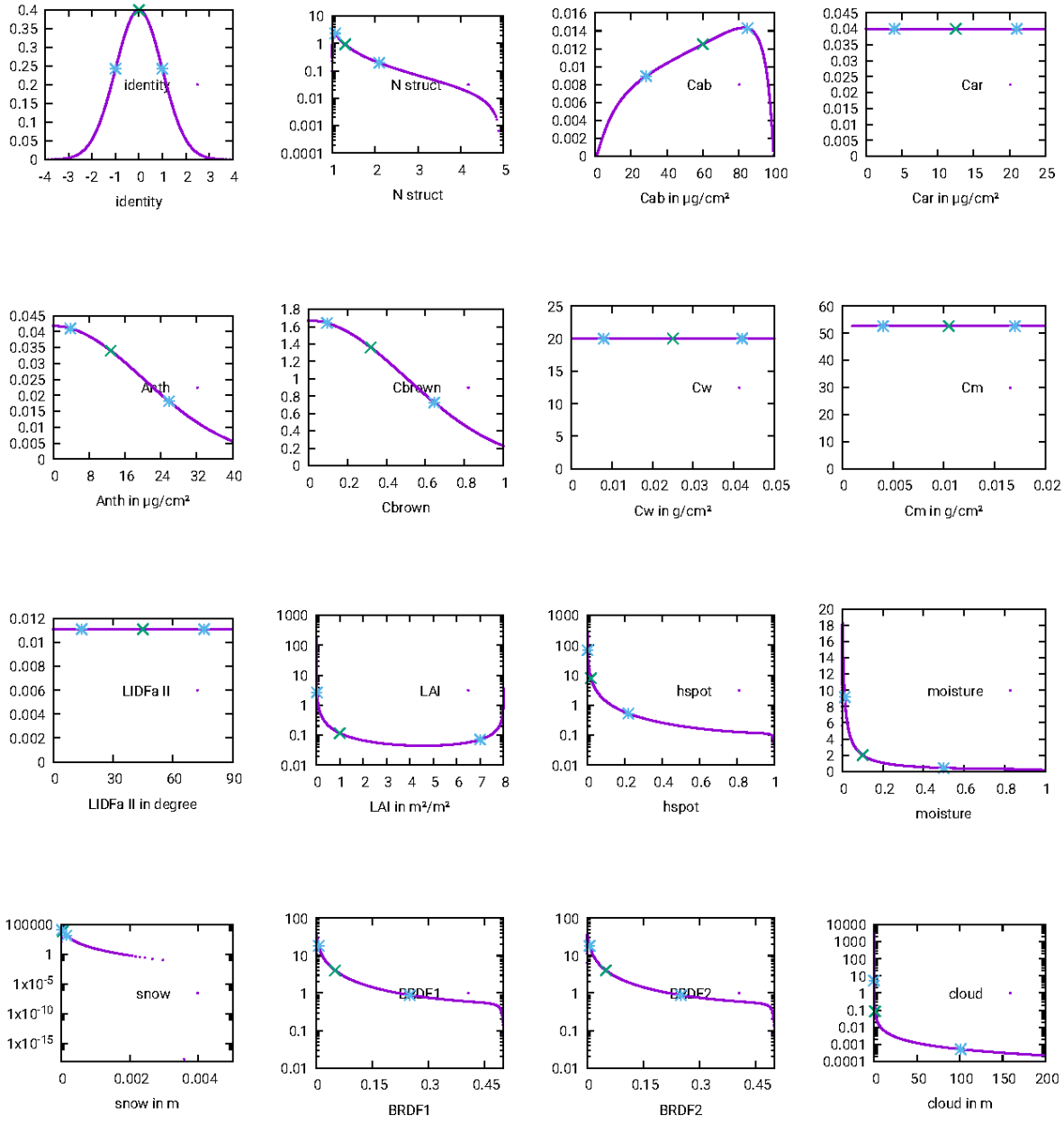


Figure 4: Prior distributions used in OptiSAIL. All model parameters are mapped to Gaussian control parameters for the minimisation, using these distributions.

Table 5: OptiSAIL retrieved parameters by sub-model.

Parameter	Description	unit
	Cloud contamination sub-model	
L	Cloud thickness parameter	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	µg/cm^2
C_{Car}	carotenoids content	µg/cm^2
C_{Anth}	Anthocyanin content	µg/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

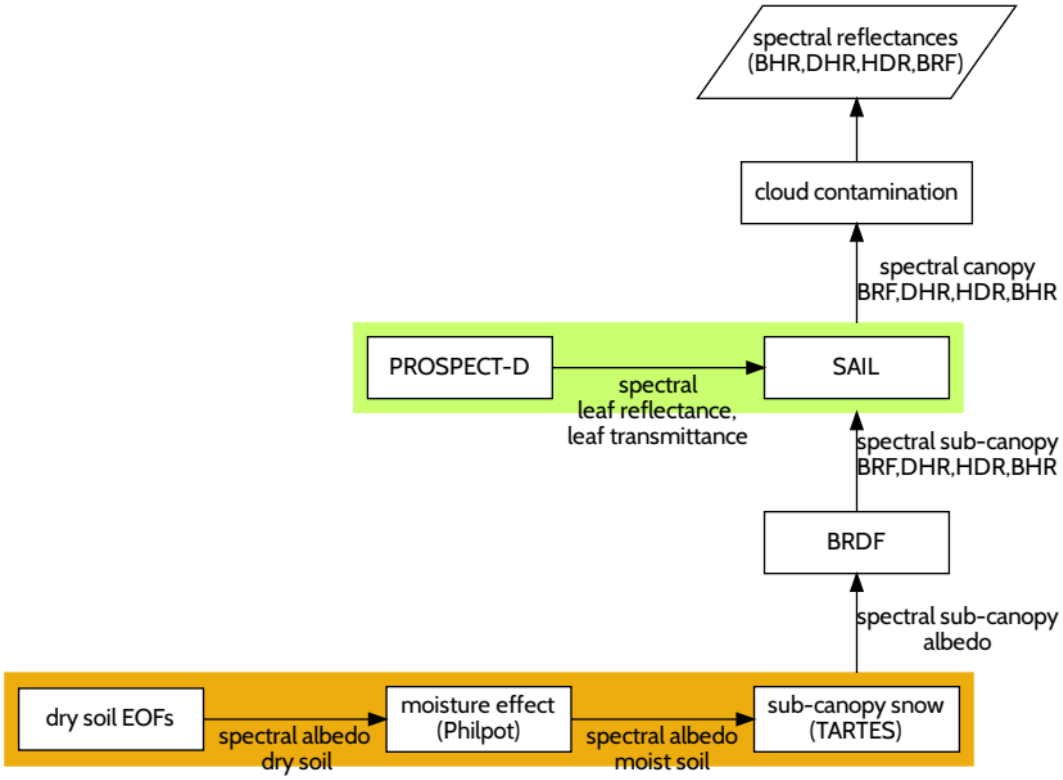


Figure 5: OptiSAIL reflectance simulation

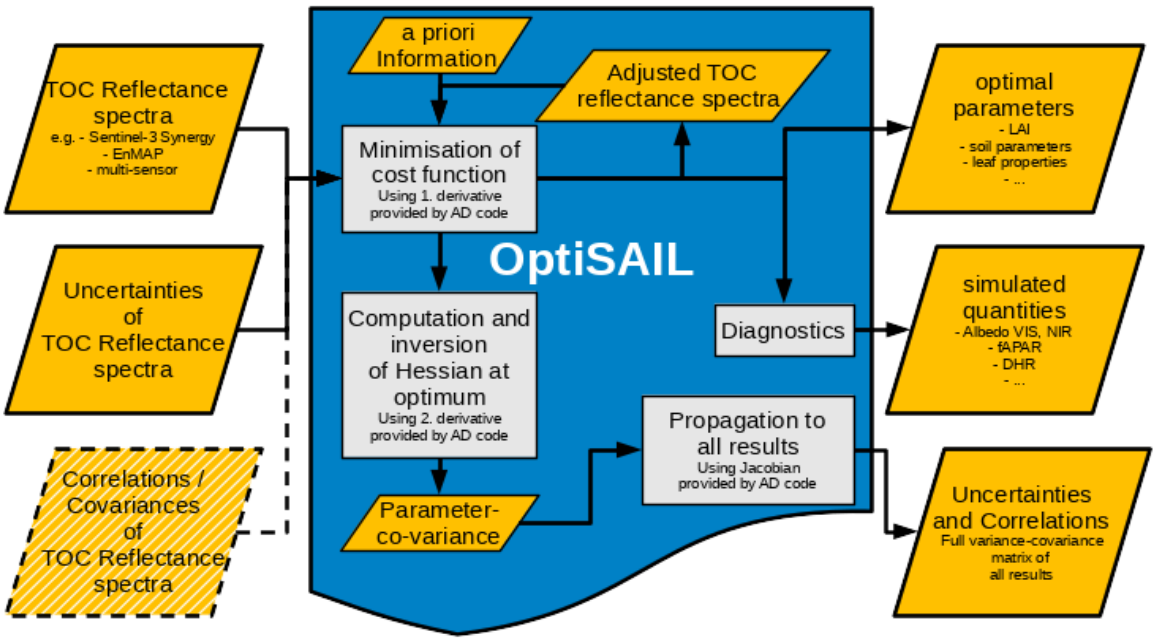


Figure 6: OptiSAIL retrieval framework with covariance propagation.

5.2 OptiSAIL output

All outputs 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 6 gives an overview of the data layers.

Table 6: Data layers in OptiSAIL output. For all quantities, the standard error and the correlation with all other main layers is given.

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	m
fAPAR	DIAG fraction of Absorbed Photosynthetically Active Radiation using diffuse ASTMG173	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

6 References

Blessing, S. and R. Giering (2021). Simultaneous Retrieval of Soil, Leaf, and Canopy Parameters from Sentinel-3 OLCI and SLSTR Multi-spectral Top-of-Canopy Reflectances, preprints.org, doi:10.20944/preprints202109.0147.v1.

Disney, M., J.-P. Muller, S. Kharbouche, T. Kaminski, and M. Vossbeck (2013). fAPAR/LAI Product Validation Report. Tech. rep. WP2230. ESA, 50 p.,
URL: http://aramis.obspm.fr/~jimenez/Docs/WACMOSET/WACMOSET_WP2230_approved.pdf.

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

Jonckheere, I., S. Fleck, K. Nackaerts, B. Muys, P. Coppin, M. Weiss, and F. Baret (2004). "Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography". In: Agricultural and Forest Meteorology 121.1-2, pp. 19–35. DOI: 10.1016/j.agrformet.2003.08.027.

Lavergne, T., M. Voßbeck, B. Pinty, T. Kaminski, and R. Giering (2006). Evaluation of the Two-Stream Model Inversion Package. Tech. rep. EUR – Scientific and Technical Research series. Luxembourg, Office for Official Publications of the European Communities: JRC-ies, 82 p.

Liang, S. (2001). Narrowband to broadband conversions of land surface albedo I: Algorithms, Remote Sensing of Environment, 76(2), 213-238, doi:10.1016/S0034-4257(00)00205-4

Libois, Q., G. Picard, J. France, L. Arnaud, M. Dumont, C. Carmagnola, and M. D. King (2013). Influence of grain shape on light penetration in snow, The Cryosphere, 1803–1818, doi:10.5194/tc-7-1803-2013.

Muller, J.-P., P. Lewis, and M. Disney (2013). Design of the Albedo/fAPAR/LAI Products. Tech. Rep. WP2210, WP2220. ESA, 24 p.
URL:
http://aramis.obspm.fr/~jimenez/Docs/WACMOSET/WACMOSET_WP2210_WP2220_approved.pdf.

Philpot, W. (2010), Spectral Reflectance of Wetted Soils, Congress paper, doi:10.13140/2.1.2306.0169.

Pinty, B., I. Andredakis, M. Clerici, T. Kaminski, M. Taberner, M. M. Verstraete, N. Gobron, S. Plummer, and J. L. Widlowski (2011a). "Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 1. Effective leaf area index, vegetation, and soil properties". In: J. Geophys. Res. 116, 20 p. DOI: 10.1029/2010JD015372.

Pinty, B., M. Clerici, I. Andredakis, T. Kaminski, M. Taberner, M. M. Verstraete, N. Gobron, S. Plummer, and J. L. Widlowski (2011b). "Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 2. Fractions of transmitted and absorbed fluxes in the vegetation and soil layers". In: J. Geophys. Res. 116, 15 p. DOI: 10.1029/2010JD015373.

Pinty, B., M. Jung, T. Kaminski, T. Lavergne, M. Mund, S. Plummer, E. Thomas, and J.-L. Widlowski (2011c). "Evaluation of the JRC-TIP 0.01 degree products over a mid-latitude deciduous forest site". In: Remote Sensing of Environment 115.12, pp. 3567–3581. DOI: 10.1016/j.rse.2011.08.018.

Pinty, B., T. Lavergne, R. E. Dickinson, J. -L. Widlowski, N. Gobron, and M. M. Verstraete (2006). "Simplifying the interaction of land surfaces with radiation for relating remote sensing products to climate models," J. Geophys. Res., vol. 111, no. D2, p. D02116, doi:10.1029/2005JD005952.

Pinty, B., J. L. Widlowski, M. Verstraete, I. Andredakis, O. Arino, M. Clerici, T. Kaminski, and M. Taberner (2011d). "Snowy backgrounds enhance the absorption of visible light in forest canopies". In: Geophys. Res. L. 38. DOI : 10.1029/2010GL046417.

Strahler, A. H., J.-P. Muller, and MODIS Science Team Members (1999). MODIS BRDF/Albedo Product: Algorithm Theoretical Basis Document, Version 5.0, NASA, https://modis.gsfc.nasa.gov/data/atbd/atbd_mod09.pdf.

Wanner W., X. Li, and A. H. Strahler (1995). On the derivation of kernels for kernel-driven models of bidirectional reflectance, J. Geophys. Res., 100 (D10), 21,077—21089, doi:10.1029/95JD02371.

Weiss, M., F. Baret, G. J. Smith, I. Jonckheere, and P. Coppin (2004). "Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling". In: Agricultural and Forest Meteorology 121.1-2, pp. 37–53. DOI: 10.1016/j.agrformet.2003.08.001.

Widlowski, J.-L., B. Pinty, M. Clerici, Y. Dai, M. De Kauwe, K. de Ridder, A. Kallel, H. Kobayashi, T. Lavergne, W. Ni-Meister, A. Olchev, T. Quaife, S. Wang, W. Yang, Y. Yang, and H. Yuan (2011). "RAMI4PILPS: An intercomparison of formulations for the partitioning of solar radiation in land surface models". In: J. Geophys. Res.: Biogeosciences 116.G2. G02019, 25 p. ISSN: 2156-2202. DOI: 10.1029/2010JG001511.

7 Annex: Re-gridding of the Sentinel-3 TOC reflectance data to 1 km

7.1 Introduction

The TOC reflectance of Sentinel-3 generated in the Copernicus Land Monitoring Service are gridded at 333 m. To be able to use this data set for the retrieval of LAI and fAPAR at 1 km, these data are first re-gridded to 1 km, considering the associated per-pixel quality information. The procedure of the re-gridding is described in this annex.

7.1.1 Related documents

ID	Document and link
ARD-1	CGLOPS1 - ATBD Atmospheric Correction Sentinel-3 OLCI & SLSTR V1 https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_ATBD_S3-AC-V1_I1.20.pdf
ARD-2	Sentinel-3 OLCI Product Data Format Specification – OLCI Level 1 products https://sentinel.esa.int/documents/247904/1872756/Sentinel-3-OLCI-Product-Data-Format-Specification-OLCI-Level-1
ARD-3	CGLOPS1 – Preliminary Product User Manual – S3 Top of Canopy (TOC) reflectance 333 m https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_ATBD_S3-AC-V1_I1.30.pdf

7.2 Input data description

The input TOC reflectances used are those for the OLCI sensor coming from the Copernicus Global Land Service – Lot1 (CGLOPS1). The data are NetCDF files, and the layers are summarized in Table 7. A more comprehensive explanation of the input data is given in the PUM [ARD-3], and the relevant information is copied below.

Table 7: Sentinel-3 input data

Parameter	Description
Oa*xx*_toc (for *xx* in 2-12, 16-18, 21)	TOC reflectances for OLCI bands
Oa*xx*_toc_error (for *xx* in 2-12, 16-18, 21)	Error on TOC reflectances for OLCI bands
SAA_OLCI	Solar azimuth angle for OLCI
SZA_OLCI	Solar zenith angle for OLCI
VAA_OLCI	Viewing azimuth angle for OLCI
VZA_OLCI	Viewing zenith angle for OLCI
Quality_flags	Quality flags from OLCI level 1b input data (see Table 8)
Pixel_classif_flags	Output from OLCI-IdePix input data (see Table 9)
AC_process_flag	Quality flag of the atmospheric correction processing (Table 10)
Latitude	Latitude
Longitude	Longitude

Table 8: Quality_flags encoding

Bit	Name	Value
0	saturated_Oa21	1
1	saturated_Oa20	2
2	saturated_Oa19	4
3	saturated_Oa18	8
4	saturated_Oa17	16
5	saturated_Oa16	32
6	saturated_Oa15	64
7	saturated_Oa14	128
8	saturated_Oa13	256
9	saturated_Oa12	512
10	saturated_Oa11	1024
11	saturated_Oa10	2048
12	saturated_Oa09	4096
13	saturated_Oa08	8192
14	saturated_Oa07	16384
15	saturated_Oa06	32768
16	saturated_Oa05	65536
17	saturated_Oa04	131072
18	saturated_Oa03	262144
19	saturated_Oa02	524288
20	saturated_Oa01	1048576
21	dubious	2097152
22	sun_glint_risk	4194304
23	duplicated	8388608
24	cosmetic	16777216
25	invalid	33554432
26	straylight_risk	67108864
27	bright	134217728
28	tidal_region	268435456
29	fresh_inland_water	536870912
30	coastline	1073741824
31	land	2147483648

Table 9: pixel_classif_flags encoding

Bit	Name	Value
0	IDEPIX_INVALID	1
1	IDEPIX_CLOUD	2
2	IDEPIX_CLOUD_AMBIGUOUS	4
3	IDEPIX_CLOUD_SURE	8
4	IDEPIX_CLOUD_BUFFER	16
5	IDEPIX_CLOUD_SHADOW	32
6	IDEPIX_SNOW_ICE	64
7	IDEPIX_BRIGHT	128
8	IDEPIX_WHITE	256
9	IDEPIX_COASTLINE	512
10	IDEPIX_LAND	1024
11	IDEPIX_MOUNTAIN_SHADOW	2048
DN no data value		**-1**

The A/C processing flags are defined as in Table 10.

Table 10: A/C processing flags

Bit	Name	Flag value
0	Reserved – not implemented	0 fixed
2,1	Bit 2: 0, Bit 1: 0 Nominal AOT value (AOT ≤0.5): High confidence level	0
	Bit 2: 0, Bit 1: 1 High AOT value (0.5 < AOT ≤1.0): Moderate confidence level	2
	Bit 2:1, Bit 1: 0 Very high AOT value (1.0 < AOT ≤1.5): Low confidence level	
	Bit 2:1, Bit 1: 1 Critical AOT value (1.5 < AOT): Very low confidence level	6
3	High solar zenith angle (SZA>65°): Low confidence level	8
4	High viewing zenith angle (VZA>65°): Low confidence level	16
5	Use of a climatology (This flag value indicates the use, as backup, of the MERRA-2 climatology in case of missing CAMS NRT ancillary data)	32
6	Not used	64
7	Not used	128

7.3 Re-gridding to 1 km

7.3.1 Grid

The S3 333 m data should be re-gridded to exactly the same grid as PROBA-V 1 km data (used as reference):

- The grid uses the World Geodetic System 1984 (WSG 84) projection as defined by the following Well-Known Text (WKT):

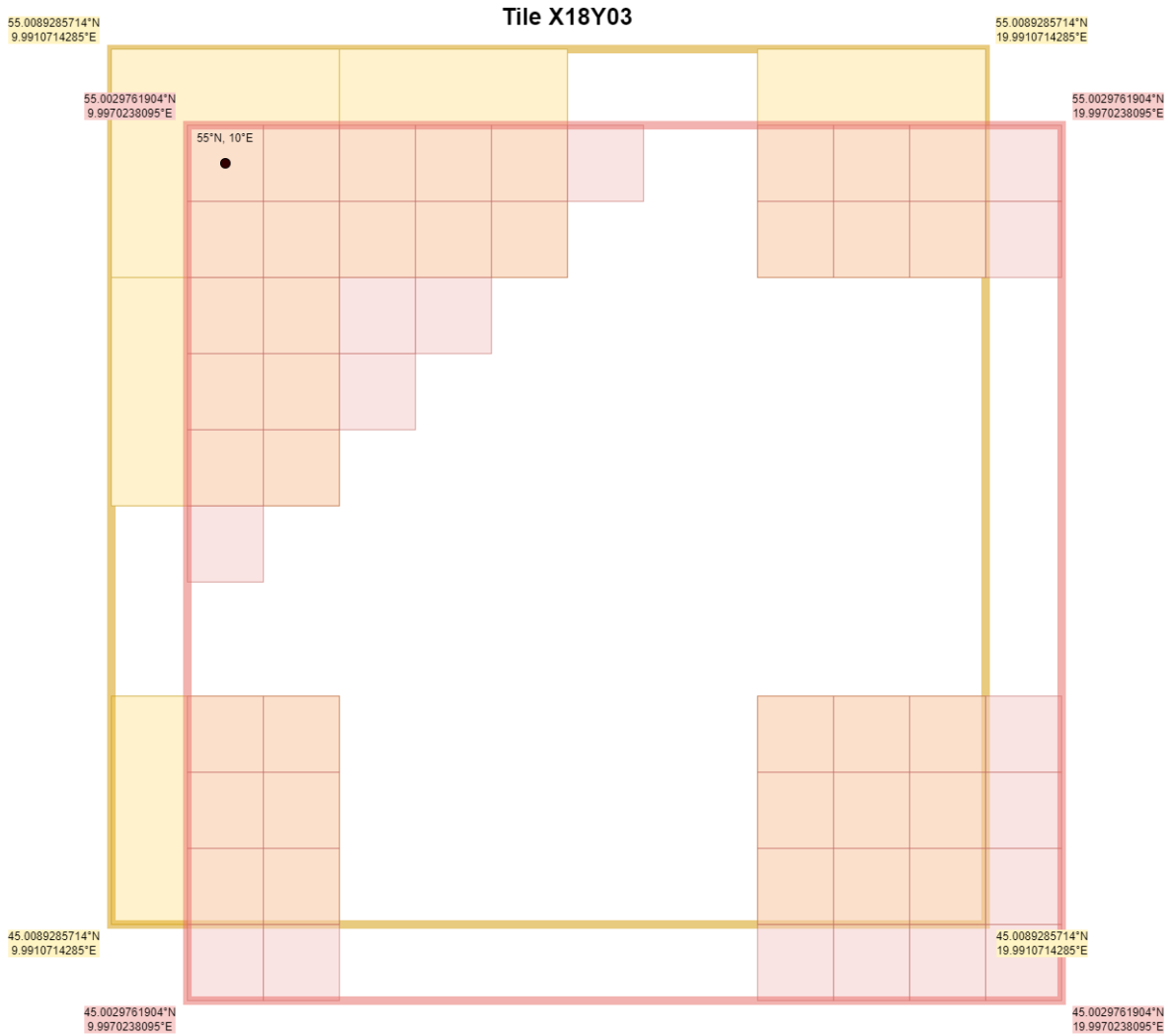
```
GEOGCS["WGS                                                    84",
        DATUM["WGS_1984",
              PHEROID["WGS                                     84",
                      6378137,
                      298.257223563,
                      'AUTHORITY["EPSG","7030"]],
              TOWGS84[0,0,0,0,0,0,0],
              AUTHORITY["EPSG","6326"]],
        PRIMEM["Greenwich",
               0,
               AUTHORITY["EPSG","8901"]],
        UNIT["degree",
             0.0174532925199433,
             AUTHORITY["EPSG","9108"]],
        AUTHORITY["EPSG","4326"]]
```

- All coordinates are centre pixel coordinates. The bounding box coordinates (e.g. to be used by GDAL) can be calculated as:

- $Y_{upper} = Y_{centerpixel} + \frac{1}{112}$and $Y_{lower} = Y_{centerpixel} - \frac{1}{112}$
- $X_{X_{centerpixel} - \frac{1}{112}}$ and $X = X_{centerpixel} + \frac{1}{112}$

- The upper left tie point of the grid is 75°N,180°W and covers the area up to 65°S,180°E.
- Individual tiles in the Grid have a tile index in their names (e.g., X10Y04), to represent the tiles relative horizontal (X) and vertical (Y) position on the globe. Note that the original PROBA-V grid also uses a XY index in their tile names. However, the S3 is extended 10° North and as such, the Y indexes have an offset of 1, e.g., PROBA-V tile X10Y04 corresponds with a Sentinel-3 tile X10Y05.

This means that tiles with the same upper left coordinate, but different resolutions, don't have the same bounding box coordinates.
Illustrated below are the coordinates for Sentinel3-grid 10°x10° tile X18Y03, both in 333 m and 1 km pixel resolutions.



7.3.2 Selection of pixels

The aggregation of 3x3 pixels of the different layers will be performed only on pixels that meet certain conditions. We follow the recommendations on using the annotation flags as defined in CGLOPS1 (see [ARD-3]).

For OLCI reflectance bands, pixels with the following flags should be omitted:

- quality_flags layer (See Table 9 in [ARD-3]):
  - 'saturated_Oa*' depending on which band is used, exclude when raised.
  - 'sea
- pixel_classif_flags (See Table 10 in [ARD-3]):
  - IDEPIX_LAND: include when raised
  - IDEPIX_MOUNTAIN_SHADOW: include when raised
  - IDEPIX_INVALID: exclude when raised
  - IDEPIX_CLOUD: exclude when raised
  - IDEPIX_CLOUD_AMBIGUOUS: exclude when raised
  - IDEPIX_CLOUD_BUFFER: exclude when raised
  - IDEPIX_CLOUD_SHADOW: exclude when raised
- AC_process_flag (See Table 11 in [ARD-3]):
  - Exclude when AOT > 1 (Bit 2 set)
  - Exclude when SZA > 65° (Bit 3 set)

For all layers, only if 5 or more pixels are not flagged, then the values are resampled to 1 km. If only 4 or less pixels are not flagged, then the output value will be flagged (see section 7.3.3.4).

7.3.3 Resampling method per layer

7.3.3.1 OLCI Reflectances

Reflectances are averaged per band. Only pixels that are not flagged (see 7.3.2) are averaged.
Further distinction should be made between clear land pixels and snow/ice land pixels according to the pixel_classif_flag:

- IDEPIX_SNOW_ICE

These two classes should never be mixed in the average. Only the pixels where the majority is (not snow/ice) or (snow/ice) should be averaged. The following rule should be applied:
After elimination of other flagged values, take the average of IDEPIX_LAND pixels if

- At least 5 pixels are retained after elimination of other flags, AND
- The majority of these pixels is NOT labelled as IDEPIX_SNOW_ICE, AND
- At least 4 pixels are retained.
- The output flag is then set to LAND.

A similar rule should be applied for snow/ice, take the average of IDEPIX_SNOW_ICE pixels if:

- At least 5 pixels are retained after elimination of other flags, AND
- The majority of these pixels is labelled as IDEPIX_SNOW_ICE, AND
- At least 4 pixels are IDEPIX_SNOW_ICE.
- The output flag is then set to LAND and SNOW_ICE.

If there are less than 4 pixels either IDEPIX_SNOW_ICE, then

- the average of all these non-flagged pixels will be taken
- the output flag is set to MIXED SNOW/ICE/LAND

$$Oaxx_{toc\,g,h,1000} = \frac{1}{N}\sum_{n=1}^{N} Oaxx_{toc\,i,j,333}$$

with

N: ranging between 4 and 9.

xx: band
1000, 333: spatial resolution
i,j: pixel position in 333 m
g,h: pixel position in 1000 m

7.3.3.2 OLCI angles

The value of the middle pixel will be retained.

7.3.3.3 Uncertainty of TOC reflectance

Uncertainties are propagated only for the pixels selected for the TOC reflectance averaging (see sections 7.3.2 and 7.3.3.1).
The method to calculate the output uncertainty is:

$$Oaxx_{g,h,1000} = \frac{1}{N}\sqrt{\sum_{n=1}^{N} Oaxx^{2}_{i,j,333}}$$

with

N: ranging between 4 and 9.
xx: band
1000, 333: spatial resolution
i,j: pixel position in 333 m
g,h: pixel position in 1000 m

7.3.3.4 Output Flags

Output flags will be summarized in one quality layer as defined in Table 11: Output flag definitions..
If TOC reflectances are not calculated, then there can be a number of reasons why pixels within the 3x3 window were flagged. Although it would be possible to propagate this information in the middle of large clouds (e.g.), this will inevitably result in multiple reasons why TOC reflectance was not calculated at 1000 m and these are not easy to label. In any case, it would not improve the retrievals.

Table 11: Output flag definitions.

Bit	Name	Description	Value
0	LAND	from IDEPIX_LAND	1
1	SNOW/ICE (1)	(TOC reflectance is calculated) AND (Majority of good pixels are IDEPIX_SNOW) AND (at least 4 non-flagged pixels with IDEPIX_SNOW)	2
2	MIXED CLEAR/SNOW/ICE (1)	(TOC reflectance is calculated) AND (Majority of good pixels are IDEPIX_SNOW_ICE or not IDEPIX_SNOW_ICE) AND (less than 4 non-flagged pixels with IDEPIX_SNOW_ICE or not IDEPIX_SNOW_ICE)	4
3	BRIGHT	(TOC reflectance is calculated) AND (at least 1 non-flagged pixel with IDEPIX_BRIGHT is used)	8
4	WHITE	(TOC reflectance is calculated) AND (at least 1 non-flagged pixel with IDEPIX_WHITE is used)	16
5		Not used	
6		Not used	

7	MISSING (1)	TOC reflectance is not calculated	126

7.3.3.5 Latitude and longitude

Lat, lon will be determined by the grid specification, the 1km pixel will have the same lat/lon as the 3x3 center pixel.

7.4 Output layers

The output layers are summarized in Table 12.
Data type and scaling are identical to the input data.

Table 12: Output layers definition

Parameter	Description
Oa*xx*_toc (for *xx* in 2-12, 16-18, 21)	TOC reflectances for OLCI bands
Oa*xx*_toc_error (for *xx* in 2-12, 16-18, 21)	Error on TOC reflectances for OLCI bands
SAA_OLCI	Solar azimuth angle for OLCI
SZA_OLCI	Solar zenith angle for OLCI
VAA_OLCI	Viewing azimuth angle for OLCI
VZA_OLCI	Viewing zenith angle for OLCI
Quality flag	Defined in Table 11
Latitude	Latitude
Longitude	Longitude