

Article

Simultaneous retrieval of soil, leaf, and canopy parameters from Sentinel-3 OLCI and SLSTR multi-spectral top-of-canopy reflectances

Simon Blessing^{1*} , Ralf Giering¹ 

¹ FastOpt GmbH, Lerchenstraße 28a, D-22767 Hamburg, Germany

* Correspondence: Simon.Blessing@FastOpt.de

Abstract: Multi- and hyper-spectral, multi-angular top-of-canopy reflectance data call for an efficient retrieval system which can improve the retrieval of standard canopy parameters (as albedo, LAI, fAPAR), and exploit the information to retrieve additional parameters (e.g. leaf pigments). Furthermore consistency between the retrieved parameters and quantification of uncertainties are required for many applications. We present a retrieval system for canopy and sub-canopy parameters (OptiSAIL), which is based on a model comprising SAIL, PROSPECT-D (leaf properties), TARTES (snow properties), a soil model (BRDF, moisture), and a cloud contamination model. The inversion is gradient based and uses codes created by Automatic Differentiation. The full per pixel covariance-matrix of the retrieved parameters is computed. For this demonstration, single observation data from the Sentinel-3 SY_2_SYN (synergy) product is used. The results are compared with the MODIS 4-day LAI/fPAR product and PhenoCam site photography. OptiSAIL produces generally consistent and credible results, at least matching the quality of the technically quite different MODIS product. For most of the sites, the PhenoCam images support the OptiSAIL retrievals. The system is computationally efficient with a rate of 150 pixel s⁻¹ (7 ms per pixel) for a single thread on a current desktop CPU using observations on 26 bands. Not all of the model parameters are well determined in all situations. Significant correlations between the parameters are found, which can change sign and magnitude over time. OptiSAIL appears to meet the design goals, puts real-time processing with this kind of system into reach, seamlessly extends to hyper-spectral and multi-sensor retrievals, and promises to be a good platform for sensitivity studies. The incorporated cloud and snow detection adds to the robustness of the system.

Keywords: Sentinel-3;SAIL; PROSPECT; TARTES; PROSAIL; LAI; fAPAR; fPAR; leaf pigments

1. Introduction

Multi- and hyper-spectral multi-angular top-of-canopy reflectance data call for an efficient retrieval system which can improve the retrieval of standard canopy parameters (as albedo, LAI, fAPAR), and exploit the information to retrieve additional parameters (e.g. leaf pigments). Furthermore, consistency between the retrieved parameters and quantification of uncertainties are required for many applications. [1] give an overview of methods for the retrieval of canopy parameters from hyper-spectral data and make suggestions for future developments. Among the identified requirements are: per pixel uncertainties and computational speed.

In their review on LAI, [2] report from the literature on suggested ways forward for the retrieval algorithms. This comprises the inclusion of contributions from more complicated background elements, including water, snow and understory vegetation. For algorithms that do not rely on land cover information, it is suggested that multiple sensors, multiple spectral bands, and observational geometry are likely to improve the

retrieval accuracy.

The more components and parameters are included, the more demanding is the inversion of the model, which is necessary to retrieve the model parameters from given TOC reflectances (observations). Many techniques have been suggested and demonstrated with some success, including machine learning algorithms like Gaussian estimators ([3],[4]) and artificial neural networks ([5]).

While a physically based model promises to give consistent results over a wide range of possible scenarios, its inversion via look-up-tables or gradient-descent algorithms is among the more time-consuming methods. The reason lies in the dimensionality of the parameter space, which pushes look-up-tables to their practical limits and makes the computation of gradients with finite difference approximations computationally expensive.

The purpose of the development presented in the present paper, is to implement and demonstrate a mature algorithm for the joint retrieval of LAI, FAPAR, albedo, presence of snow, leaf physiological and soil parameters from top-of-canopy (TOC) reflectance data, which can be efficiently inverted by a gradient based algorithm.

Various models exist, which have to some degree a comparable set of parameter, e.g. SLC ([6]) or SCOPE ([7]). The motivation of the present study to create yet another retrieval system lies in the opportunities offered by Automatic Differentiation (AD). Automatic differentiation could be described as the symbolic differentiation of the implementation of a function as an algorithm. This can be done in forward (tangent-linear model) and reverse mode (adjoint model). With an adjoint model, gradient computation for many control variables becomes more computationally affordable and second derivatives enable the direct computation of posterior uncertainties and uncertainty propagation. Usually models are not written with differentiability or invertibility in mind (a trivial example resulting in a wrong zero derivative with respect to LAI is “if(LAI==0) skip canopy computations”). In a review study, [8] identify the combined PROSPECT and SAIL models as suitable candidates for the retrieval from hyper-spectral instruments. Therefore, in combination with SAIL (with hot spot), we chose to use PROSPECT-D ([9]), which can simulate green and senescent leaves, included TARTES ([10]) in order to be able to detect the presence of snow, and allowed for more complex bare soil spectra with moisture effect, and, optionally the soil reflectance anisotropy. Such a system simulates top-of-canopy (TOC) reflectances for any given geometry from its parameters. For given TOC reflectance observations, inversion yields these parameters, with which in turn the full spectrum can be simulated, yielding fAPAR, the BRDF, and albedo without the need of any band-to-band mappings.

2. Methods

2.1. Model Description

SAIL and its various combinations with other models, especially PROSAIL, has found wide use for decades to simulate and retrieve the optical properties of plant canopies ([11], [8]). The SAIL model requires spectral leaf reflection and transmittance of the leaves, and a spectral albedo of the surface at the lower boundary. Fig. 1 illustrates, how the sub-models described in the following provide these values, and thus reduce the spectral information of the surface-canopy system to a few parameters, as listed in Table 1.

2.1.1. PROSPECT-D

The optical leaf properties direct-hemispherical reflectance and direct-hemispherical transmission are simulated by PROSPECT-D ([9]) controlled by the seven parameters listed in Table 1. PROSPECT-D is a variant of the empirical leaf spectral properties model PROSPECT ([12]). It includes anthocyanins as additional leaf pigments in order to

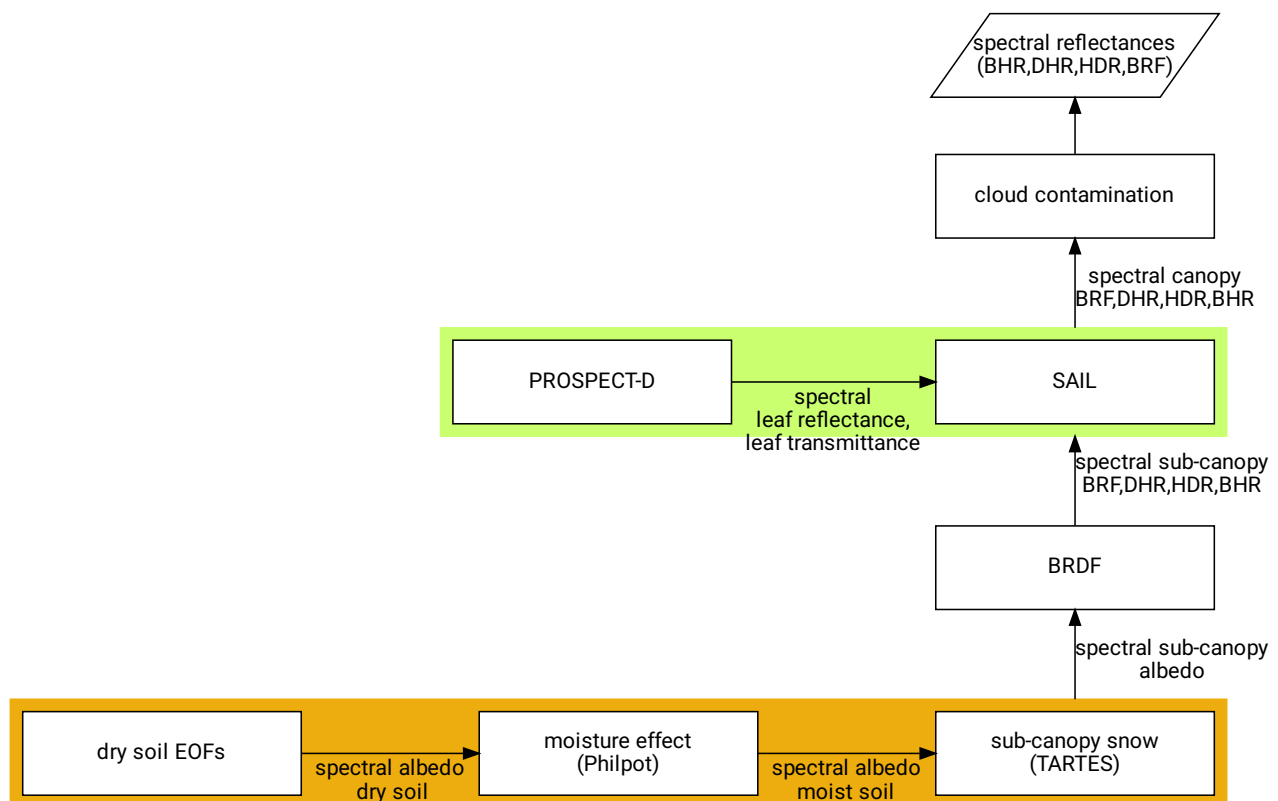


Figure 1. Overview of OptiSAIL model components. Inputs are wavelength-independent parameters and observation geometries.

improve the simulation of the leaf optical properties through a complete life cycle until senescence. Prospect-D's validation error is about 1.5 % of the simulated reflectances (cf. Fig. 7+10 of [9]).

2.1.2. Soil and Surface State Model

The full soil and surface state model computes the surface albedo below the canopy and has four independent parameters and three components: dry soil (Sec. 2.1.2.1), moisture effect (Sec. 2.1.2.2, and snow (Sec. 2.1.2.3. Together, they simulate the sub-canopy spectral albedo for given parameters. Structural effects are taken account of by an optional BRDF model (Sec. 2.1.3.1).

Dry Soil: EOFs

The dry soil reflectance model is based on Empirical Orthogonal Functions (EOF,[13], [14]) of the spectra of lab-dried topsoils from the ICRAF-ISRIC Soil VNIR Spectral Library ([15]). The dry soil spectrum is simulated as the unweighted average spectrum of the libraries top soils plus a linear combination of their two leading EOFs. The EOFs are computed with a weighting function which emphasis the centre wavelengths of the observing sensor bands, in order to force near-orthogonality on this sub-set of the spectrum without losing the information between the bands. Fig. 2 shows the mean and the two EOFs, re-transformed into unweighted space. For the case of the Sentinel-3 bands (OLCI+SLSTR), they explain about 85 % and 10 %, respectively, of the variability of the dry soil spectra. It would have been a positive side-effect if the EOF directions would distinguish soil classes. However, as Fig. 3 shows for the example of FAO 74 soil

Table 1. parameters of OptiSAIL

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 zenith angle	°
or LIDF _{a,b}	leaf zenith angle distribution parameters	1
hspot	canopy hot-spot parameter	1
PROSPECT-D sub-model		
N	leaf structure parameter	1
C_{ab}	chlorophyll a+b content	$\mu\text{g}/\text{cm}^2$
C_{ar}	carotenoids content	$\mu\text{g}/\text{cm}^2$
C_{Anth}	Anthocyanin content	$\mu\text{g}/\text{cm}^2$
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_{iso})	isotropic kernel factor, diagnosed from snow albedo	1
f_{vol}	volumetric scattering kernel factor	1
f_{geo}	geometric scattering kernel factor	1
Snow sub-model (TARTES)		
snowheight	height of a single snow layer with fixed properties	m
Soil albedo model (EOF+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	1

classes, projection onto the EOFs is largely uncorrelated with soil class. One reason may be that FAO 74 soil classes (and many other soil classification systems) are focusing on the agricultural use of the soil rather than its immediate physical or chemical properties. Therefore it remains to be shown, how well the principle components align with Soil Organic Carbon (SOC) content or other physical soil properties (cf. [16], [17]).

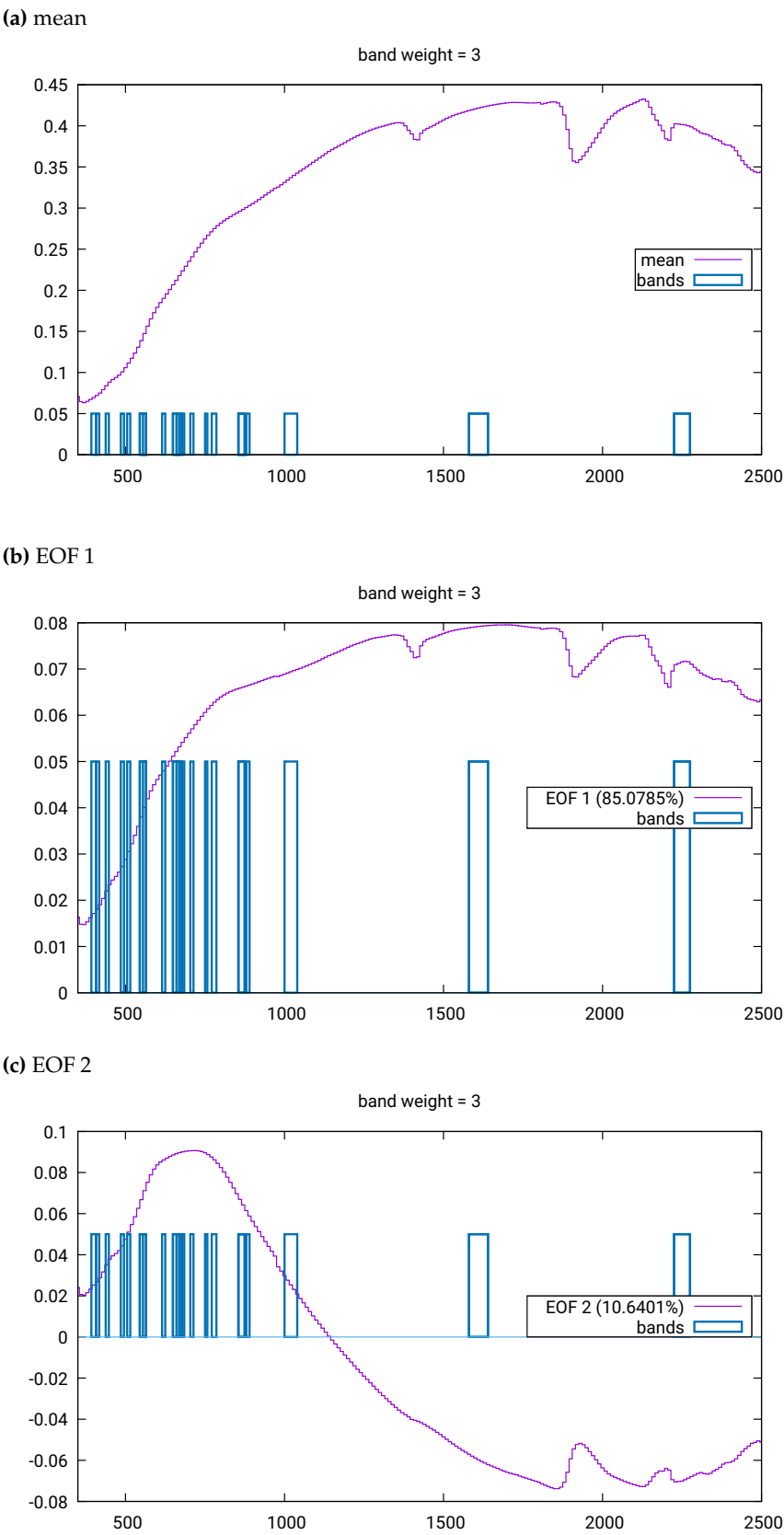


Figure 2. Dry soil spectrum components: (a) mean spectrum, (b) first, and (c) second EOF, computed to be orthogonal on SENTINEL-3 SYNERGY central wavelengths.

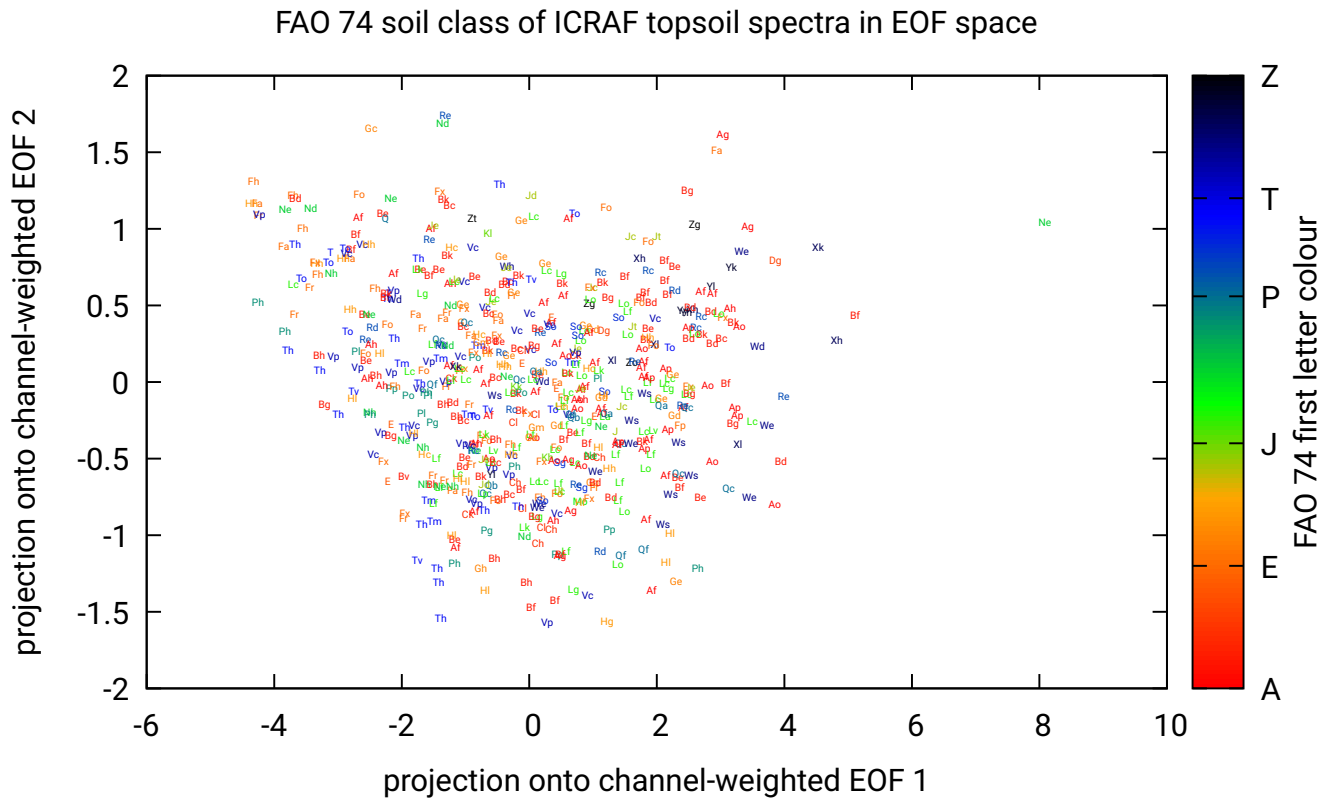


Figure 3. FAO 74 soil class of ICRAF topsoil spectra in EOF space.

The linear combination of the two EOFs introduces two parameters, which are limited in amplitude by their effect on the reflectance spectrum (no values above 1 and below 0 permitted). A bijective function is used to map between a quadrangle covering most of the legal area in the space of the principle components and two unbound control parameters.

Moisture Effect: Philpot

The dry soil spectrum is modified by a third parameter, namely the relative saturation of the soil with moisture, which drives a moist soil spectral model suggested by [18]. The advantage of this approach is that for a given dry spectrum the moist spectrum is simulated without a priori knowledge of the saturated spectrum. In addition to Philpot's formulation, the two parameters d (mean free path in pore water), and f_w (fraction of moisture on the surface), are parameterised by the soil moisture expressed as relative saturation. While this relationship obviously is influenced by the properties of the individual soils, a general tendency can be deduced from a curve fit (Fig. 4) to the data of the four soils presented in Fig. 8 of [18].

Hence this parameterisation is chosen as:

$$f_w = 0.15 \times \log(1 + 27.85 \times f_{\text{moist}}) \quad (1)$$

$$d = d_0 f_{\text{moist}}, \quad (2)$$

with d_0 being the maximum mean free path in pore water of saturated soil (set to $305 \mu\text{m} = 0.0305 \text{ cm}$). A similar attempt is made by [19], empirically fitting gravimetric water content θ_m to the model parameters f_w and d as $f_w = 5.7 \times \theta_m$ and $d = 2.5 \text{ cm} \times \theta_m$. However, the values they determine for d in their Fig. 5 of a few millimetres are about a factor of 10 larger than the ones found by Philpot for different soil types. Also, their f_w does not seem to saturate with water content. For these reasons, the above Eqs. (1) and (2) appear to be of more general validity and are adopted here. Moreover, the systematic

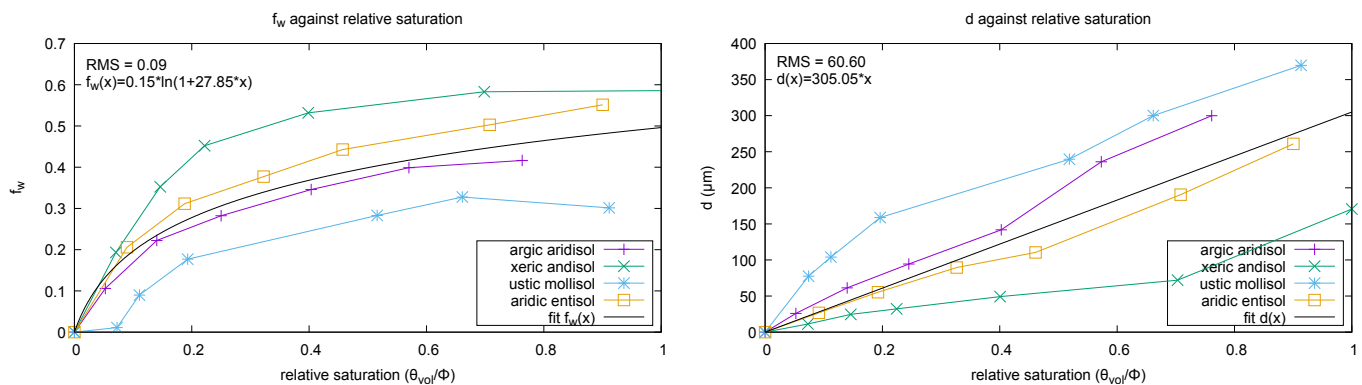


Figure 4. Parameterisation of parameter f_w and d of Philpot's model with relative saturation, based on the data of for four soil samples presented in his Fig. 8. Volumetric water content θ_{vol} was converted to relative saturation by dividing with porosity ϕ .

misfits for a given soil type seem to balance to a certain extent, for instance, for a given f_w , the relative saturation of the ustic mollisol sample is systematically underestimated, while for a given d it is overestimated. The reliability of the estimated relative saturation will have to be shown in the verification.

Snow: TARTES

The TARTES model (Two-streAm Radiative TransfER in Snow,[10]) model is included in order to compute the albedo of snow lying on the ground below the canopy. It is run with a single layer, the height of which is determined by the fourth parameter of the soil and surface state sub-model. In order to facilitate the optimisation of this parameter, it never goes to zero fully, but can take very low values, which enables the simulation of virtually snow-unaffected spectra.

2.1.3. SAIL canopy reflectance model

SAIL simulates the reflectance of a horizontal, homogeneous and infinitely extended canopy layer, made up of small, flat leaves. The azimuth of the leaf normals is evenly distributed, while their nadir angle follows a parameterised distribution (one or two parameters) to simulate either leaves of planophile or erectophile plants, at the extremes. SAIL uses spectral leaf optical properties for the leaves, here provided by PROSPECT-D (forming PROSAIL-D). Further parameters of SAIL are the Leaf Area Index (LAI) and a hot-spot parameter, to account for the reduction of visible shadows when the observer nearly eclipses the sun. SAIL has been extended to optionally handle reflectance anisotropy at the lower boundary in analogy to the SLC-model ([6]).

BRDF

The surface reflectance anisotropy below the canopy is simulated by the Ross-Thick-Li-Sparse-Reciprocal model as documented in [20], including the approximate integral formulas for white-sky and black-sky albedo. Just two global parameters are used, corresponding to the factors of the geometric and the volume scattering kernels. They are scaled for each band to match the spectral albedo prescribed by the surface model. The factor for the isotropic kernel is thus fully determined and has a spectral dependency. This approach is motivated by the observation that normalised BRDF kernel factors for bare soils show substantial correlation across bands (e.g [21], [22]). We eliminate the spectral dependency of the kernel factors by normalising them with the spectral BHR. This leads to a frequency-independent anisotropy factor (ANIF, see [23]) and preserves the Lambertian equivalent albedo computed by the soil model. Note that this is not the same as normalising with f_{iso} , since the other kernels also contribute to the BHR.

The RossLi BRDF model makes reference to vegetated surfaces, but is intended for general surfaces, such that its parameters have similar interpretation for bare soils. If the leaf canopy is negligibly thin or sparse (LAI near zero), the surface BRDF model parameters are well determined, and the surface BRDF dominates the TOC reflectance anisotropy. In the opposite case of a high LAI, its parameters are not well determined and tend to the default values given by the optimisation prior.

2.1.4. Cloud contamination model

Atmospheric correction and cloud flagging both depend on precise and up-to-date surface reflectance spectra, which are often substituted by climatological values. Therefore TOC reflectances can contain undetected cloud contamination. In order to detect such situations and possibly neutralise their effect, a cloud model is included in OptiSAIL, the single parameter of which is determined together with those of the other sub-models. Fig. 5 gives an overview of the cloud model. The model uses analytical solutions for weakly absorbing optically thick, plane-parallel, homogeneous water clouds from [24] for the computation of cloud reflectance and transmission. The semi-infinite cloud reflectance is modelled as in Appendix B of [25]. These results are extended for the uncollided light path through the cloud. Collided radiation is treated as ideally diffuse, with the exception of BRF_{cloud} , for which the escape function from [24] is used. Also, it is assumed that all sensors look through or at the same cloud, which is not generally true in practise. The resulting formula for the semi-infinite cloud is based on the computation for asymmetry parameter g , absorption coefficient σ_{abs} , and extinction coefficient σ_{ext} as shown in [25], eq. B1–4:

$$1 - g = 0.12 + 0.5(ka_{eff})^{-2/3} - 0.15\kappa a_{eff} \quad (3)$$

$$\sigma_{ext} = C_v \frac{1.5}{a_{eff}} \left(1 + \frac{1.1}{(ka_{eff})^{2/3}} \right) \quad (4)$$

$$\sigma_{abs} = C_v \frac{5\pi\chi(1 - \kappa_{eff})}{\lambda} \left(1 + 0.34 \left(1 - \exp\left(\frac{-8\lambda}{a_{eff}}\right) \right) \right), \quad (5)$$

$$(6)$$

where C_v is the volumetric droplet concentration, a_{eff} the effective droplet size, λ is the wavelength, $k = 2\pi/\lambda$, χ is the imaginary part of the refractive index of water, $\kappa = 4\pi\chi/\lambda$. Using eqs. 4.42 and 4.48 of [24], we have for the bi-hemispherical reflectance of a weakly absorbing semi-infinite cloud, r_∞ ,

$$y = \sqrt{\frac{16\sigma_{abs}}{3\sigma_{ext}(1 - g)}} \\ r_\infty = \exp(-y)$$

x and y are called generalised radiative transfer parameters. From Section 2.4 of [25] it follows that

$$x = \sqrt{3\sigma_{abs}\sigma_{ext}(1 - g)}L, \quad (7)$$

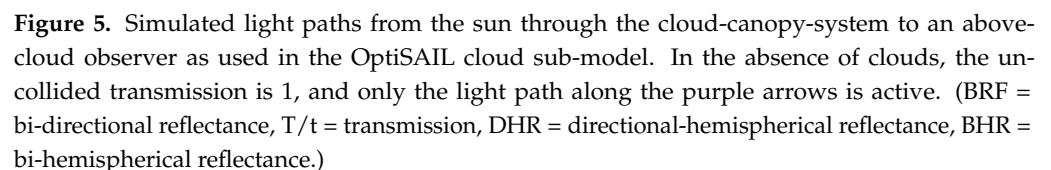
thus introducing the cloud thickness L , which is the single parameter of this sub-model (effective droplet size a_{eff} is fixed to 12 μm , and the volumetric droplet concentration C_v to 1.0×10^{-7}). To compute the reflectance of a finite cloud with black background, we extend eq. 4.30 from [24] with the uncollided radiation,

$$BRF_{cloud,bb} = r_\infty K_0(\mu_0)(1 - T_{uncollided, down})K_0(\mu) - t_{coll,down} \exp(-x - y)K_0(\mu). \quad (8)$$

$$T_{\text{uncoll, down}} = \exp(-\tau/\mu_0), \quad (9)$$

where $\tau = \sigma_{\text{ext}} L$ is the optical thickness. Collided directed radiation is treated as diffuse after its passage through the cloud, with factor

The single cloud thickness parameter L is expected to show a certain degree of variation



2.1.5. Diagnostics: fAPAR, DHR, BHR

In addition to the simulation of TOC reflectances, the model diagnoses a set of quantities for given parameter sets. These include the fraction of Absorbed Photosynthetically Active Radiation (fAPAR), direct-to-hemispheric reflectance (DHR) and bi-hemispheric reflectance (BHR). These diagnostics (quantities not actively involved in the simulation, but derived from the model state) are computed at TOC level, i.e. without the cloud effects.

In order to convert spectral information to broadband quantities, e.g. compute the fAPAR, a solar reference spectrum is used. This data is identified as: ASTM G173-03 Reference Spectra Derived from SMARTS v. 2.9.2¹. Fig. 6 gives an example of the simulated spectra from different sub-models and the broadband integrated quantities for a given parameter set.

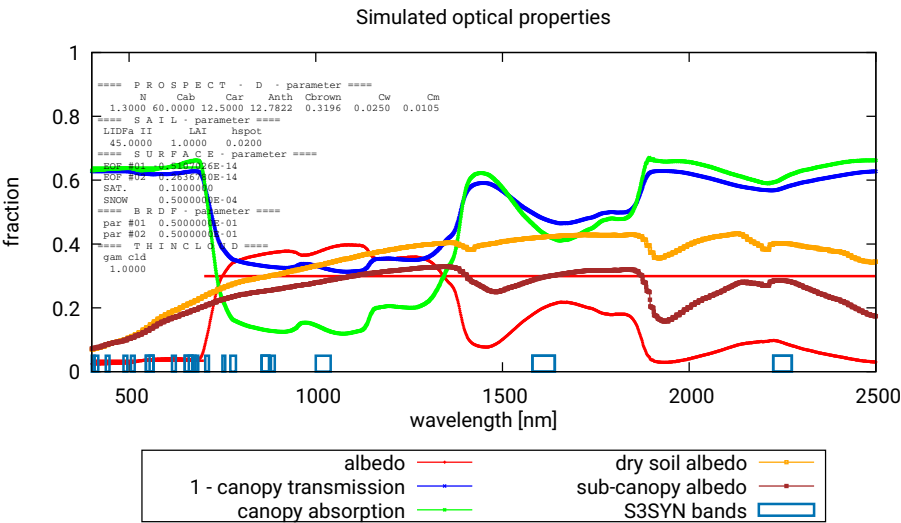


Figure 6. Simulated spectra using prior parameter values as defined in the forthcoming section. The horizontal green bar gives the value and wavelength range of the simulated fAPAR, while the horizontal red bars are BHR-VIS and BHR-NIR.

216

217 2.2. Data

218 2.2.1. Sentinel-3 Synergy product (SY_2_SYN)

219 The Sentinel-3 OLCI and SLSTR channels present in the SY_2_SYN product are
220 shown in table 2.

¹ available from <https://www.nrel.gov/grid/solar-resource/spectra-am1.5.html>

Table 2. Sentinel-3 OLCI and SLSTR channels present in SY_2_SYN.

SYN level 2 channel	Associated channel in L1b product band	centre (nm)	width (nm)
OLCI			
1	Oa01	400.00	15
2	Oa02	412.50	10
3	Oa03	442.50	10
4	Oa04	442.00	10
5	Oa05	510.00	10
6	Oa06	560.00	10
7	Oa07	620.00	10
8	Oa08	665.00	10
9	Oa09	673.75	7.5
10	Oa10	681.25	7.5
11	Oa11	708.75	10
12	Oa12	753.75	7.5
13	Oa16	778.75	15
14	Oa17	865.00	20
15	Oa18	885.00	10
16	Oa21	1020.00	40
SLSTR-nadir, -oblique			
17, 22	S1N, S1O	555	20
18, 23	S2N, S2O	659	20
19, 24	S3N, S3O	865	20
20, 25	S5N, S5O	1610	60
21, 26	S6N, S6O	2250	50

The SY_2_SYN product combines the information from the Ocean and Land Colour Instrument (OLCI), Sea and Land Surface Temperature Radiometer (SLSTR) on-board ESA's Sentinel-3A and Sentinel-3B satellites. From the Sentinel-3 orbit (814.5 km), OLCI's 68.5° field of view covers a swath width of 1270 km. For the nominal orbit, at sub-satellite point of the OLCI instrument, the resolution is approximately 300 m. The nadir swath of the SLSTR instrument fully overlaps with this, while the SLSTR oblique view (tilted 55° against nadir) only covers a swath 740 km wide. In total up to 26 (= 16 + 2 × 5) observations with three different viewing geometries per pixel are used where OLCI, SLSTR-nadir and SLSTR-oblique are available simultaneously. This reduces to 21 observations with two different geometries outside the field of view of SLSTR-oblique, or even below that, when individual bands are missing in the data. For the two satellite constellation, the revisit frequency of an individual point is less than one day, exceeding twice daily towards the poles. Note that in this study each overpass is processed individually, using no time aggregation and no combination of the two satellites.

Each Product Dissemination Unit (PDU) of the SY_2_SYN product covers 19902715 (4865×4091) pixels, corresponding to three minutes of satellite swath. Presently no uncertainties are available for the TOC reflectance synergy product. Therefore we use a heuristics, which is purely based on an educated guess. The absolute uncertainty of a reflectance is computed as $\sigma_{\rho} = \max(0.0025, \sigma_{rel} \times \rho)$ with σ_{rel} being either *base uncertainty* or *low quality uncertainty*, depending on the flags as detailed in Tab. 3.

In addition, cloud and partly cloudy flags are used to increase the prior value of the residual cloud thickness of the cloud model (2.1.4).

Table 3. Heuristics used for relative albedo uncertainty.

Sensor	minimal absolute uncertainty	base uncertainty [%]	low quality uncertainty [%]
OLCI	0.0025	5	^a 20
SLSTR-N	0.0025	5	^b 20
SLSTR-O	0.0025	7	^c 20

(a) flags OLC_dubious, OLC_saturated, SYN_border, SYN_SDR_OOR
(b) flags SLN_twilight, SLN_summary_invalid_radiance, SLN_summary_no_signal, SLN_summary_saturation, SYN_border, SYN_SDR_OOR
(c) flags SLO_twilight, SLO_summary_invalid_radiance, SLO_summary_no_signal, SLO_summary_saturation, SYN_border, SYN_partly_cloudy, SYN_SDR_OOR

Note that during the period studied here, repeated adjustments of the processing baseline have been carried out, with the intention to improve various aspects of the product. For instance, on 2020-01-15, a new processing baseline was implemented which reduces the geolocation offset in the oblique view for SLSTR-A and B. This can lead to discontinuities in the retrieved time series.

2.2.2. MODIS 4-day 500 m LAI/fAPAR

The 4-day LAI/FPAR product based on NASA’s MODIS/Terra+Aqua sensors (MCD15A3H, [26]) is used for verification of our algorithm. The product is distributed on a 500m sinusoidal grid. The product was selected for its established status, the relatively short aggregation window, and the convenient access to pre-processed time-series for an extensive list of observational sites through the MODIS and VIIRS Land Products Fixed Sites Subsetting and Visualization Tool ([27]). Note, that in contrast to OptiSAIL, the MODIS LAI/fAPAR algorithm uses a biome map as auxiliary input.

2.2.3. PhenoCam network

To verify individual events, a visual inspection of the conditions at the site is a valuable means to distinguish the plausible from the outlier. The PhenoCam Network ([28]) provides such permanent monitoring of many sites, which are included in the list of the above-mentioned MODIS and VIIRS Land Products Fixed Sites Subsetting and Visualization Tool. Table 4 gives an overview of the sites, which are used in this study, and their characteristics. The sites were selected according to the following criteria: Be part of the Fixed Sites Subsetting and Visualization Tool, be in continuous operation at least since 2019, cover as many vegetation classes as possible, give a homogeneous impression in a radius of at least a couple of hundred meters around the field of view of the PhenoCam, cover as many continents and latitudes as possible.

Table 4. Overview of the PhenoCam sites used in this study. The position is the position of the camera.

name	lat (°N)	lon (°E)	elev. (m)	veg. class	reference (https://phenocam.sr.unh.edu/webcam/sites/...)
canadaOBS	53.9872	-105.1178	628	EN,DN	canadaOBS/
eslm1	39.9427	-5.7787	265	GR,EB	eslm1/
eucflux	-22.9679	-48.7280	750	EB,-	eucflux/
gcesapelo	31.4440	-81.2835	0	WT,-	gcesapelo/
howland1	45.2041	-68.7403	80	EN,DB	howland1/
innsbruck	47.1162	11.3203	972	GR,-	innsbruck/
jurong	31.8068	119.2173	15	AG,-	jurong/
torgnon-ld	45.8238	7.5609	2091	DN,EN	torgnon-ld/
mead1	41.1651	-96.4766	361	AG,-	mead1/
montebondonegrass	46.0147	11.0458	1550	GR,-	montebondonegrass/
NEON.D02.SERC	38.8901	-76.5600	30	DB,DB	NEON.D02.SERC.DP1.00033/
NEON.D04.GUAN	17.9696	-66.8687	136	EB,-	NEON.D04.GUAN.DP1.00033/
NEON.D10.STER	40.4619	-103.0293	1366	AG,-	NEON.D10.STER.DP1.00033/
NEON.D14.SRER	31.9107	-110.8355	999	SH,-	NEON.D14.SRER.DP1.00033/
NEON.D18.TOOL	68.6611	-149.3705	827	TN,-	NEON.D18.TOOL.DP1.00033/
turkeypointenf39	42.7102	-80.3573	232	EN,-	turkeypointenf39/

Table 5. Vegetation Type Codes used in Table 4.

Abbreviation	Description
AG	agriculture
DB	deciduous broadleaf
DN	deciduous needleleaf
EB	evergreen broadleaf
EN	evergreen needleleaf
GR	grassland
MX	mixed vegetation
SH	shrubs
TN	tundra
WT	wetland

Table 6. Altered positions at some stations used for the pixel extraction in the OptiSAIL processing chain, in order to cover a more homogeneous area.

name	lat (°N)	lon (°E)
innsbruck	47.1165	11.3209
torgnon-ld	45.8250	7.5615
mead1	41.1652	-96.4778
NEON.D10.STER	40.4632	-103.0319
NEON.D02.SERC	38.8903	-76.5618
turkeypointenf39	42.7099	-80.3583

2.3. Model inversion: OptiSAIL

2.3.1. Control variables

All model parameters p are mapped to a vector of control variables x . Prior assumptions about the distribution of the model parameters are used to find mappings which make the expected distribution of the control parameters standard normal. This

has numerical advantages for the minimisation and facilitates the computation of the posterior uncertainties. Also, it maps bounded model parameters to unbounded control variables, thus making constraints implicit. Fig 7 shows the consequences of the assumptions made, computed as $\partial\Phi_{\|x(p)}/\partial p$, that is the probability density of x plotted over p . The crosses in the figure mark the region which is mapped to 1 standard deviation of the control variables and therefore get a relatively low penalty from the prior term of the cost function (see 2.3.2). The centre cross corresponds to the unpenalised value which is assumed in the optimisation when not enough or no other information is contained in the TOC reflectance data. The ranges are mostly chosen according to the limitations imposed by the models, while the distributions are chosen by educated guess. Especially for LAI, care was taken that the prior is effectively only a very weak constraint, by emphasising the tails of the distribution.

2.3.2. Objective function

The objective function to be minimised in order to retrieve the parameters, is a combination of a data term and a prior term.

$$J = J_{\text{data}} + J_{\text{prior}} \quad (11)$$

The former measures the misfit between model and observation, weighted with the uncertainties, while the latter measures the deviation from prior assumptions as described in section 2.3.1. This is the standard Bayesian approach to find the a posterior probability of variables which are approximately normally distributed. The degrees of freedom for the final χ^2 -test are updated for every pixel, depending on the number of available observations. Failed inversions are marked as untrusted, badly failed inversions are discarded. J_{data} could use inter-band correlation information, if it was available. Presently it is formulated as

$$J_{\text{data}}(x) = \sum_{i \in N} \frac{(\rho_{i,\text{data}} - \rho_{i,\text{OptiSAIL}}(x))^2}{\sigma_{\rho_i}^2}, \quad (12)$$

which corresponds to using a diagonal variance-covariance matrix. N is an index set of valid observations. Since it is assumed that the distribution of the x is standard normal, the prior term is simply

$$J_{\text{prior}}(x) = \sum_{i \in N} x_i. \quad (13)$$

2.3.3. Minimisation

The optimal parameters for a single pixel are determined by minimising the cost function. Here we use a quasi Newton, reduced memory, BFGS algorithm ([29]). Efficient minimisation algorithm like BFGS use the gradient of the cost function. The exact gradient is computed by code generated in reverse mode Automatic Differentiation (AD) by FastOpt's tool Transformation of Algorithms in Fortran (TAF, [30,31]). The so called adjoint code generated by source-to-source transformation is very efficient in terms of computer time and memory resources ([32]). Here, the time to compute a single cost function and its gradient is about four times the time to compute the function only. This factor is independent of the number of parameters chosen. In this application the increase in memory resources is minor and without implications. The minimisation starts from the first guess value and stops when J does not decrease significantly any more between iterations, and the gradient approaches zero.

2.3.4. Posterior uncertainties

Having minimised the cost function we want to specify the posterior uncertainties of the optimal model parameters, i.e. the covariance matrix Σ^x . It is the inverse of the Hessian matrix at the minimum. The Hessian is the second order derivative of the cost function and computed here by vector forward over scalar reverse mode of AD ([33]). In

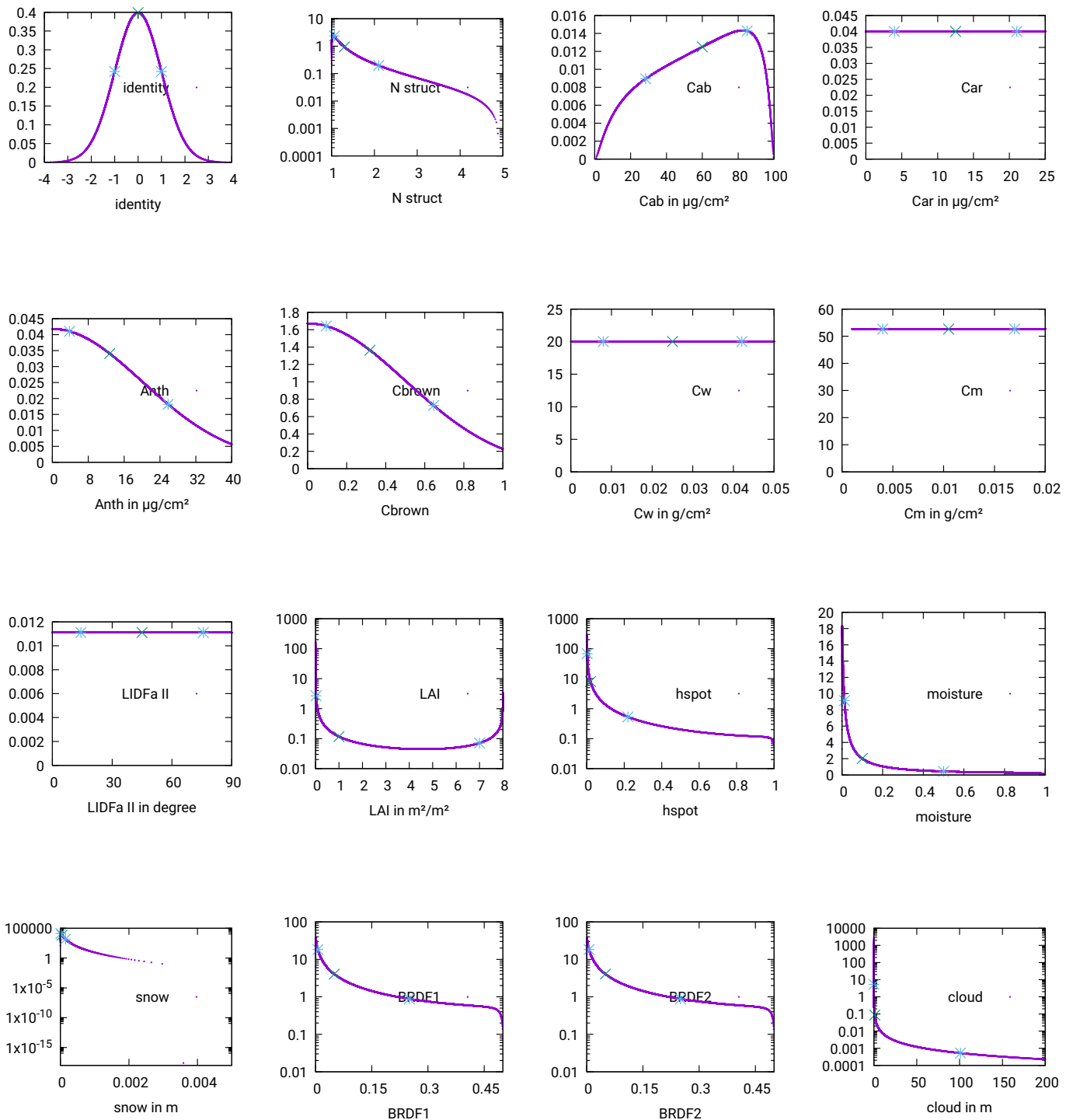


Figure 7. Assumptions on model parameter distribution. x-axes are in the unit of the model parameter, y-axes are probability densities. Blue stars mark the points corresponding to -1 and +1 standard deviations of the control variable, a green cross marks its mean. Probability densities are zero outside the plotted region, except for snow height, which is unbounded for positive values. The top left sub-figure shows the distribution of identity mapping from control to model variable as a reference.

essence, that means that the generated gradient code is differentiated again, this time in forward mode, to get the code for the computation of the second order derivatives. In order to compute the full Hessian the vector forward mode is used. It propagates several perturbations simultaneously, allowing the extraction of the full Hessian in a single sweep, and such gaining a speed up for increased temporal and spatial locality ([34]). Since the number of parameters is small we can directly invert the Hessian matrix. The posterior covariance matrix Σ^y of the non-parameter values, such as the diagnosed quantities (sec. 2.1.5), is achieved by applying the Jacobian $\partial_x y$ of a result vector y to the prior covariance:

$$\Sigma^y = \partial_x y \Sigma^x (\partial_x y)^T. \quad (14)$$

Code for the computation of $\partial_x y$ is generated in vector forward mode AD, which is then used to extract $\partial_x y$ in a single sweep. Fig. 8 gives an overview of the full OptiSAIL system.

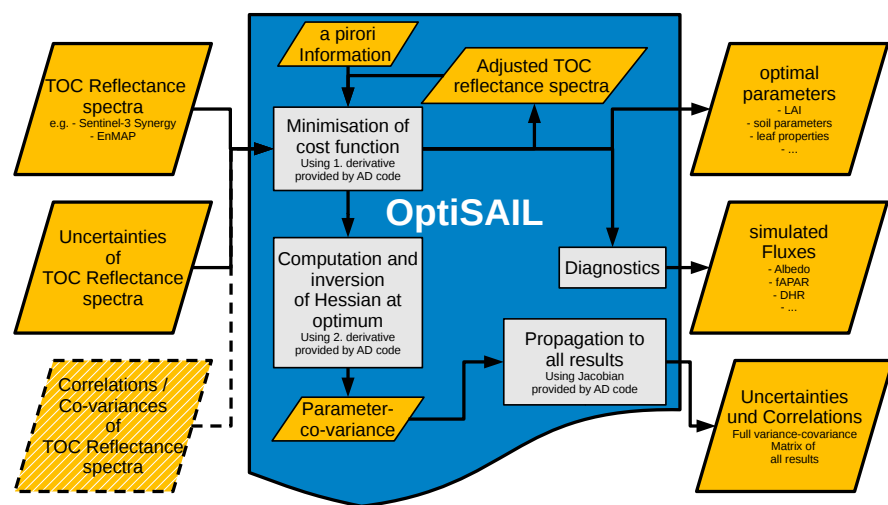


Figure 8. OptiSAIL flow chart depicting data (rhombi), processes (rectangles) and data flow (arrows).

3. Results and Discussion

3.1. Model performance

Using a state-of-the-art Fortran-compiler with machine-specific optimisation switches, the performance of the full procedure typically is above 150 pixel/s and thread (7 ms/pixel) on a modern desktop computer. This also includes computation of the posterior covariances and compressed output of the results to disk.

3.2. LAI and fAPAR

Figures 9 and 10 show LAI and fAPAR at a deciduous broadleaf forest site in the North-American East (NEON.D02.SERC, Mid-Atlantic, Smithsonian Environmental Research Center). The results from OptiSAIL were smoothed with an inverse variance weighted 4-day running mean to make them better comparable to the MODIS 4-day product. Also note that the product footprint is not fully identical, while being in the range of a few decametres. The overall agreement is quite favourable, with the OptiSAIL fAPAR going to more extreme values (both high and low) with more confidence, than the MODIS 4-day fPAR in this example. OptiSAIL shows some outliers, but usually with large uncertainties, especially outside the vegetation period, which may be due to the fact, that during periods with few credible observations, the running mean does not reduce their impact.

Figure 11 shows the comparisons of OptiSAIL and MODIS products at the remaining selected PhenoCam stations. The overall impression is that the products match

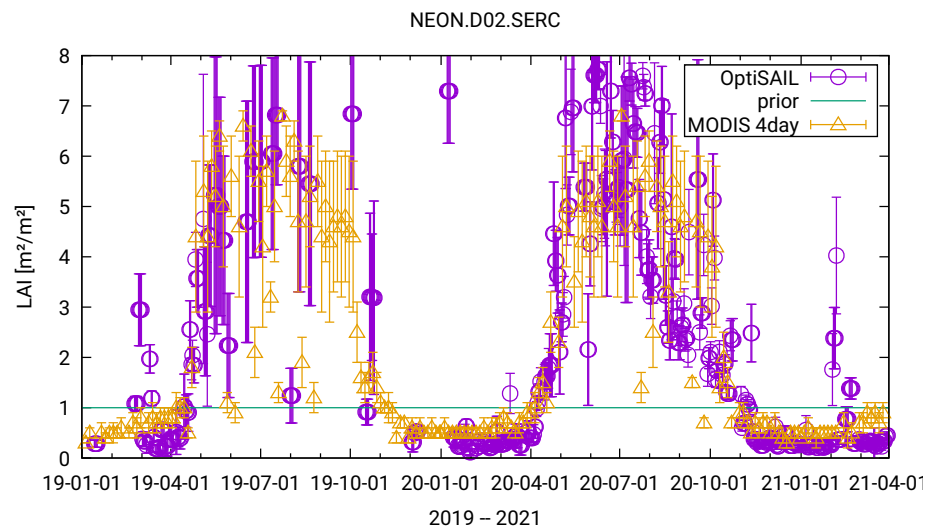


Figure 9. Comparison of LAI from OptiSAIL and MODIS. Here, OptiSAIL results are smoothed with a four-day running mean with inverse variance weights.

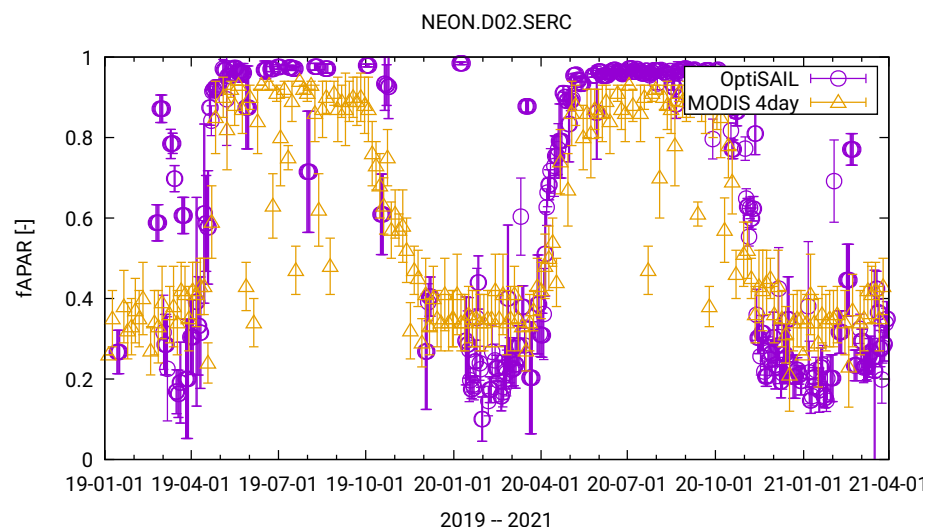


Figure 10. As Fig. 9, but for fAPAR.

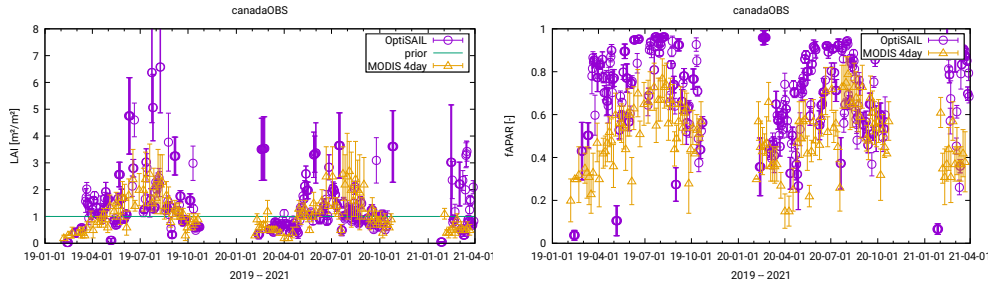
within their uncertainty bounds for the majority of cases. In some cases there seems to be a permanent (eucflux, NEOD.D04.GUAN) and sometimes a temporal divergence (NEON.D14.SRER).

The eucflux site is an Eucalyptus plantation (commercial clone of *Eucalyptus urophylla*). PhenoCam images (not shown, follow link in Tab. 4) of January 2019 show sparsely set, very young trees, more in keeping with an effective leaf area index below 1 for the observed area. We have to speculate that the MODIS product for this site suffers either from inappropriate prior information or a geolocation issue, while the OptiSAIL results are in good keeping with a developing Eucalyptus plantation, even though the differently coloured upper and lower sides of eucalyptus leaves violate a model assumption. The differences at the evergreen broadleaf site NEOD.D04.GUAN may again be a result of inaccurate prior information and a regime switch in the MODIS product, since the LAI abruptly goes from quite high to medium values (most notably in March 2020), which is not supported by the corresponding PhenoCam images. The OptiSAIL LAI evolution is, while noisy, more credible for an evergreen forest.

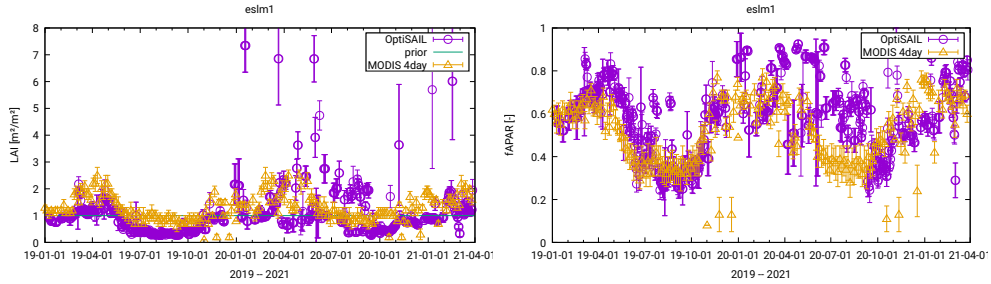
In the case of NEON.D14.SRER, an inspection of the PhenoCam images and rain gauge

data of different years did not seem sufficient to explain the fundamental difference observed in fAPAR after September 2020 (Fig. 11l right) and the persistent abrupt loss of Canopy Chlorophyll Content in January 2020 (12). Since the OptiSAIL processing is time-independent, a qualitative difference in the data must have occurred. It is interesting to note, that also the MODIS product shows less variability of fAPAR for the period after September 2020.

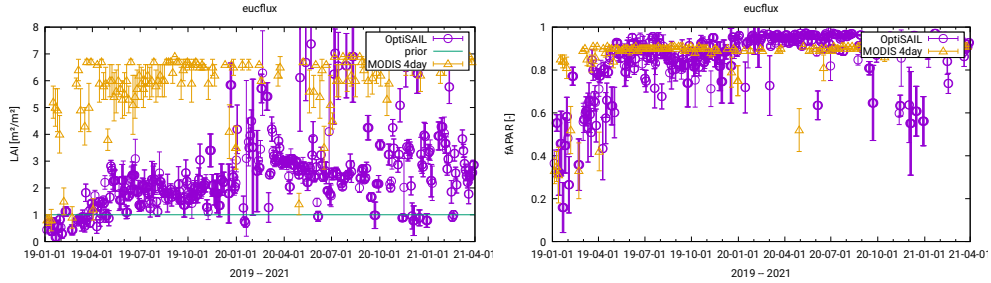
(a) canadaOBS



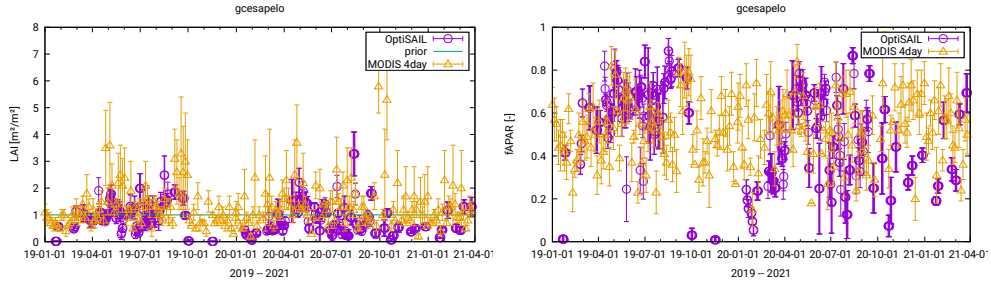
(b) eslm1



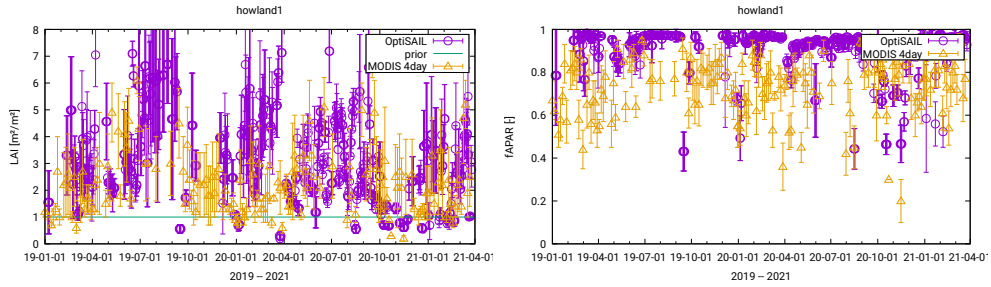
(c) eucflux



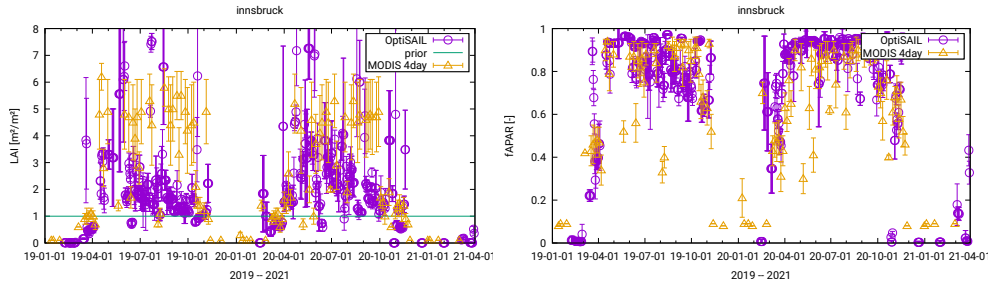
(d) gcesapelo



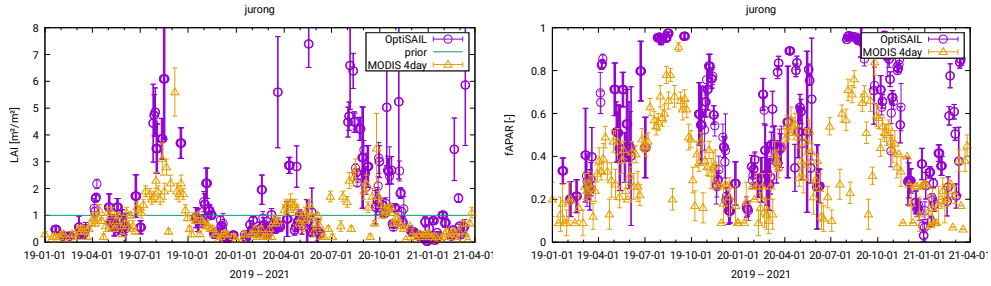
(e) howland1



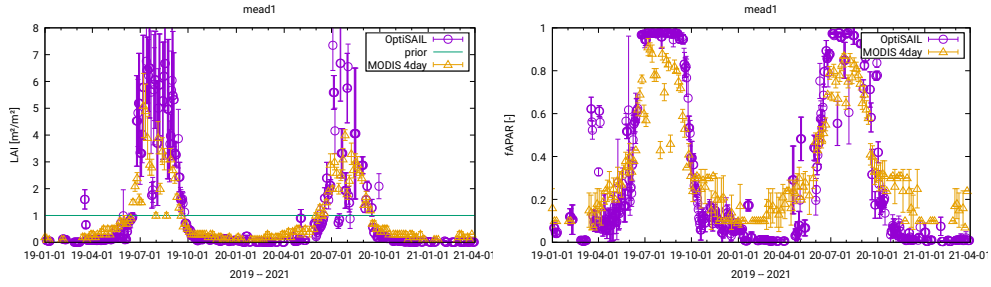
(f) innsbruck



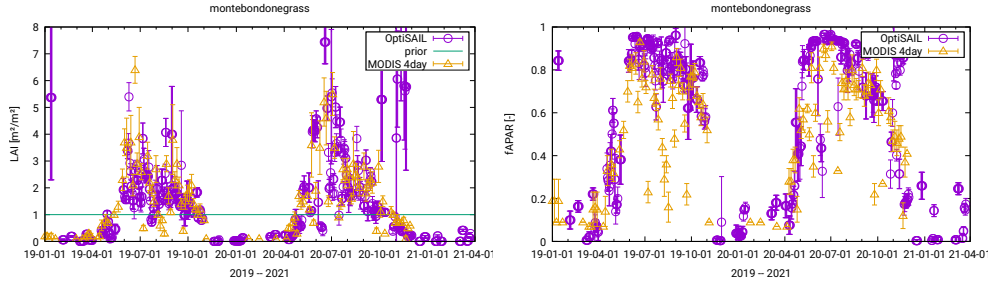
(g) jurong



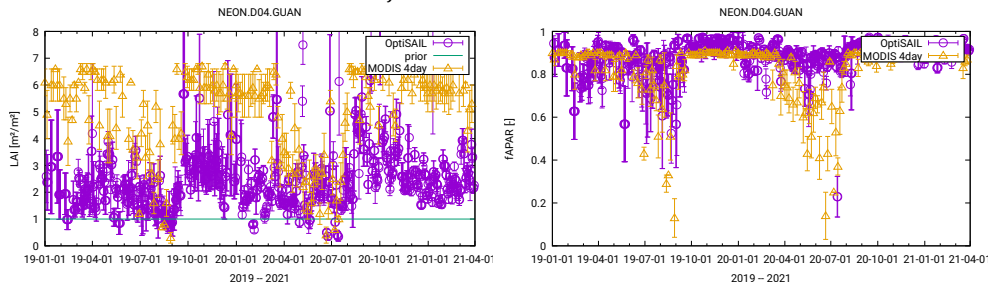
(h) mead1



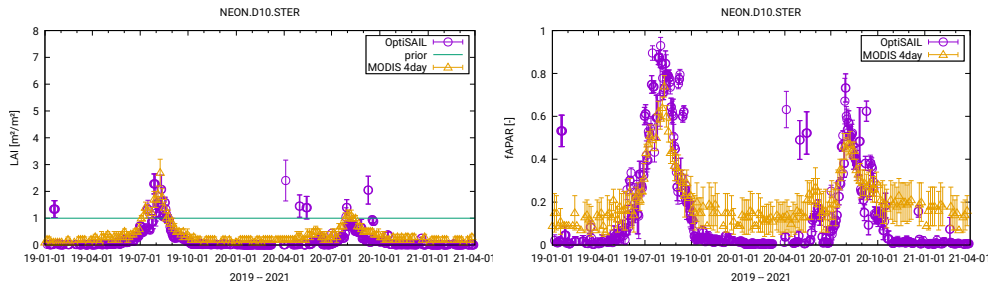
(i) montebandonegrass



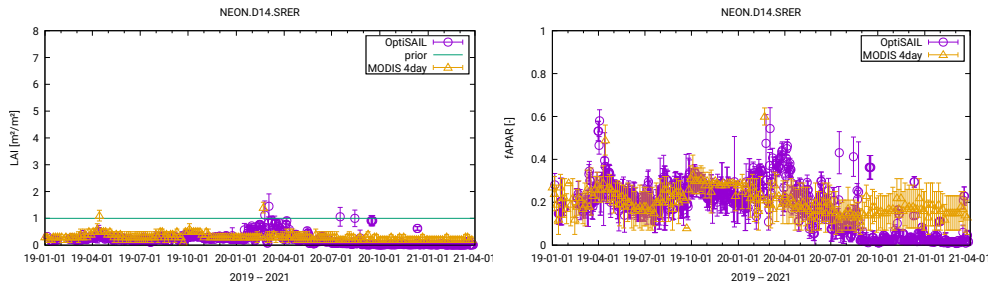
(j) NEON.D04.GUAN



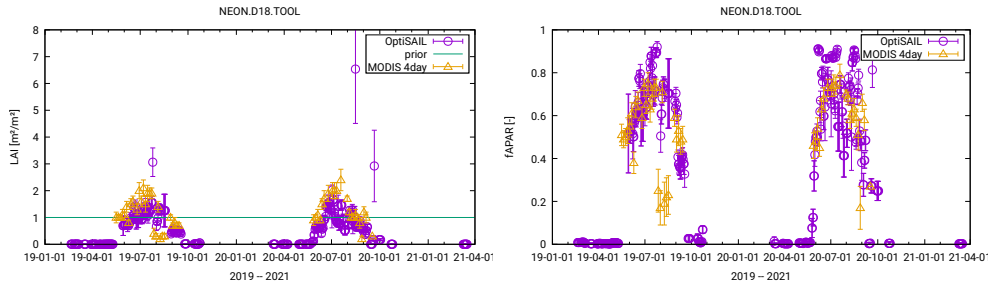
(k) NEON.D10.STER



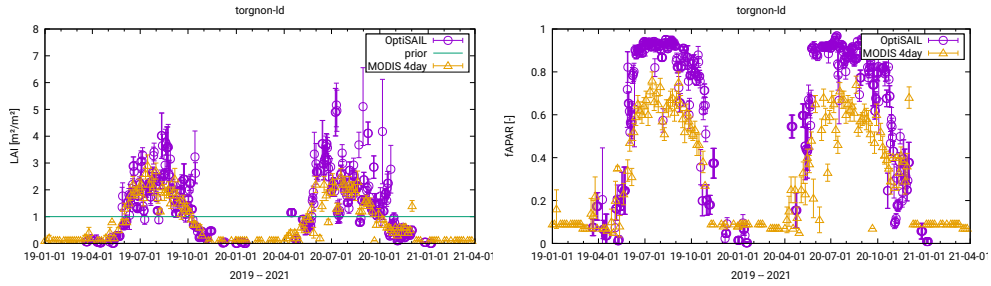
(l) NEON.D14.SRER



(m) NEON.D18.TOOL



(n) torgnon-ld



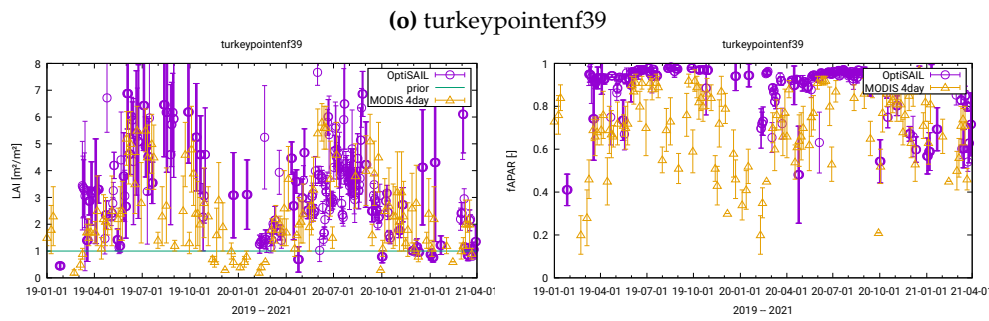


Figure 11. As Fig. 9 and 10, but for the remaining stations, showing LAI (left) and fAPAR (right).

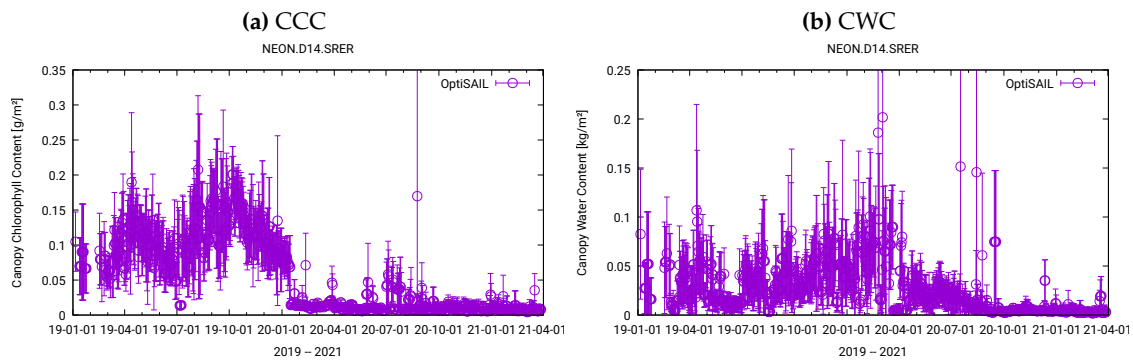


Figure 12. CCC and CWC for shrubland site NEON.D14.SRER.

3.3. Brown matter, leaf pigments, CCC, and CWC

Fig. 13 shows further retrievals at the deciduous broadleaf forest site NEON-D02-SERC. A fairly clear signal is retrieved for the leaf brown matter fraction (Fig. 13a), and for the leaf and canopy chlorophyll content (Fig. 13d). Note that for deciduous sites outside the growing season, leaf pigments are only constrained by the weak prior assumptions and can therefore show quite noisy behaviour. By looking at the pigment content of the canopy rather than the individual leaf, the corresponding temporal evolution is more regular, as for instance for Canopy Chlorophyll Content (CCC). We compute the propagation of the variance-covariance matrix to CCC as

$$\sigma_{CCC} = \sqrt{Cab^2 \sigma_{LAI}^2 + LAI^2 \sigma_{Cab}^2 + 2 Cab LAI \sigma_{LAI,Cab}}, \quad (15)$$

with $\sigma_{LAI,Cab} = r_{LAI,Cab} \sigma_{LAI} \sigma_{Cab}$ being the covariance of LAI and Cab, computed from the correlation $r_{LAI,Cab}$, as provided by OptiSAIL. To demonstrate the importance of the correlation information, the CCC has once been computed neglecting the correlation of LAI and Cab in the uncertainty propagation (Fig. 13c), and once using the full Eq. 15 (Fig. 13d). The consequence is a systematically reduced uncertainty, which in turn influences the four-day running mean through the inverse variance weights, thus resulting even in different values of CCC (see for instance end-of-season 2020). Both effects are present to a comparable degree at all stations, most easily visually discernible at the tundra site NEON.D18.TOOL (Fig. 14). Fig. 15 shows the temporal evolution of the correlation of LAI with the other retrieved quantities at the above-mentioned deciduous broadleaf site NEON.D02.SERC. Note the correlation with Car, which is often negative, and thus leads to the reduction of uncertainty of CCC when taken into account in Eq. (15).

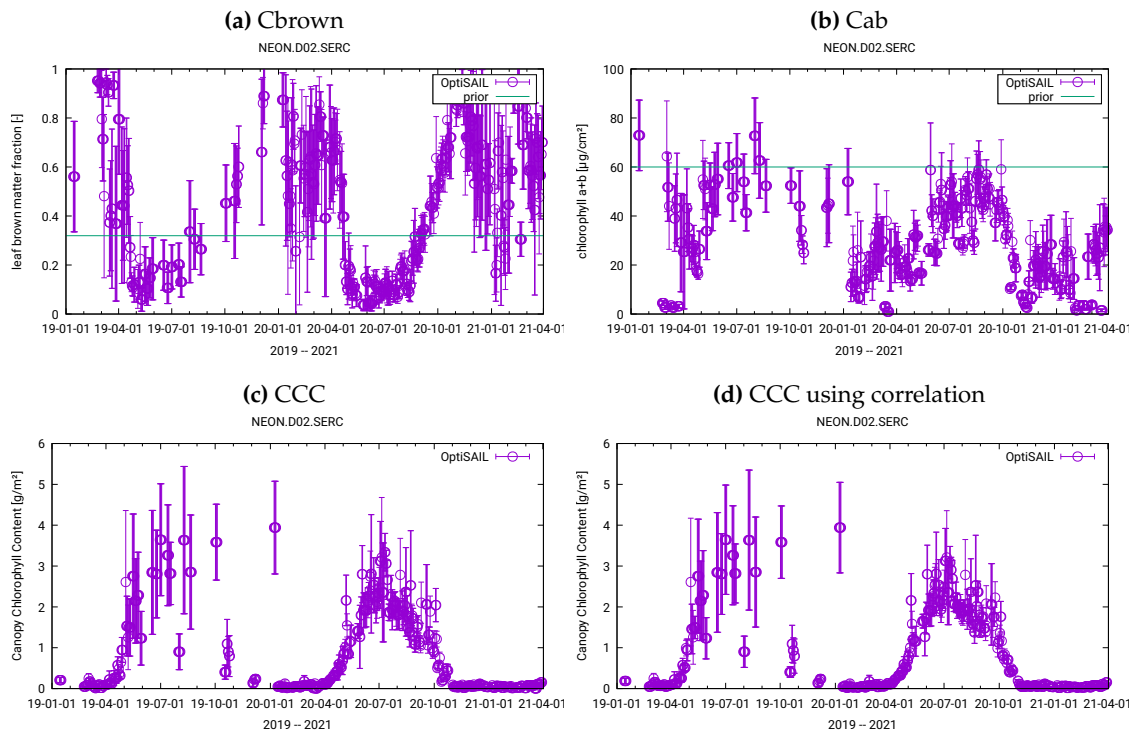


Figure 13. OptiSAIL results at the deciduous broadleaf site NEON.D02.SERC: **(a)** leaf brown matter fraction, **(b)** leaf chlorophyll content. Canopy Chlorophyll Content (CCC) is once computed **(c)** without taking the correlation of LAI and Cab into account, and once **(d)** with. All results are smoothed with a four-day running mean with inverse variance weights.

3.4. Other leaf parameters

For trivial reasons, leaf parameters are not well determined by single observations from the Synergy product in situations of low LAI, for instance outside the growing season at deciduous sites. We see a good signal in the leaf dry matter content (C_m), anthocyanin concentration, carotenoid concentration. For brevity, we omit the corresponding figures.

3.5. Other canopy structural parameters

The leaf structure parameter (N) shows a good signal at some sites, while the hotspot (hspot) and leaf angular distribution parameter (ALIA) tend to be somewhat noisy with large uncertainties. For brevity, we omit the corresponding figures.

3.6. Soil parameters

For brevity, we omit the figures corresponding to this section. At the deciduous and agricultural sites, during the vegetation phase, the soil parameters are not well determined. Here, they mainly serve to reflect the uncertainty at the lower boundary of the canopy. This is different where more of the bare soil is directly exposed to the observer.

The surface moisture saturation parameter (“moist” in Tab. 1) can exhibit a very high natural variability due to rain and dew. Even a superficial assessment of the quality is beyond the scope of this study, even though, especially at some of the NEON sites, soil moisture measurements very near the surface are available.

The sub-canopy surface BRDF parameters (f_{vol} , f_{geo}) show no consistent signal and large uncertainties.

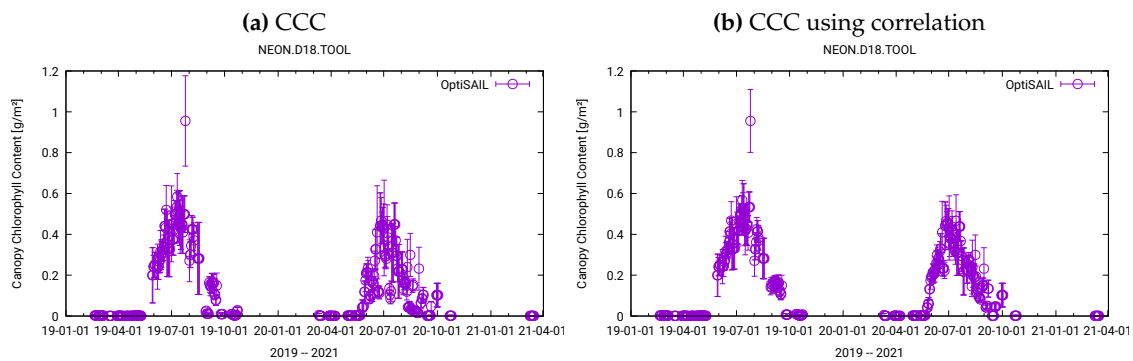


Figure 14. As bottom row of Fig. 13, but for tundra site NEON.D18.TOOL.

3.7. Spurious cloud and snow detection

Fig. 16 shows a false colour image with France in the centre, retrieved from data of a single overpass from 2021-02-28. The cloud_margin flags were used to avoid processing near clouds. The blue colour in the image starts at about 1 m of spurious cloud thickness and the red colour at about 1 mm of snow height. The patterns appear credible and consistent. For a more detailed presentation, Fig. 17 shows the retrieved values at the mixed evergreen and deciduous needle leaf site canadaOBS, chosen for its high probability of occurrence for both phenomena. Both presentations exhibit credible patterns, which seem to indicate a good separation of cloud and snow detection. This is supported by visual inspection of the corresponding PhenoCam images (not shown).

4. Conclusions

With OptiSAIL, we have documented and demonstrated an inversion system for the retrieval of canopy and soil parameters from multi-angular, multi- to hyper-spectral data. The application to the synergy (SY_2_SYN) top-of-canopy reflectances based on the OLCI and SLSTR sensors of ESA's Sentinel-3A and -3B satellites shows good performance when directly compared to the 4-day LAI and fPAR product based on the MODIS sensor. Existing discrepancies tend to be more in favour of OptiSAIL, even though or because OptiSAIL is not using the biome type as ancillary data. Note that for OptiSAIL, the individual satellites and visits were process independently.

OptiSAIL has integrated cloud contamination detection in order to complement the cloud flags of the satellite product. The simple assumptions of the underlying cloud model (spectrum of a semi-infinite water cloud with fixed effective droplet radius and droplet concentration) appear to be effective for the detection of spurious clouds.

In order to extend the retrieval into snowy regions, a sub-canopy snow layer is simulated, a situation not generally recognised by snow flags in satellite retrievals. Also here, the results are encouraging.

The inclusion of a sub-canopy BRDF model does not yield usable retrievals of the corresponding parameters for the low angular sampling used in this study. It does however, capture the resulting uncertainty in the other retrievals, and enables the system to do instantaneous retrievals directly from TOC reflectances.

Some of the other parameters are also less well determined by the observations, also depending on the situation (e.g. dense or sparse canopy). However, by the same argument as for the BRDF parameters, by retrieving them all together, credit is given to their impact on the overall uncertainty budget.

OptiSAIL provides the full posterior covariance matrix of the retrieved quantities, which makes it possible to propagate the uncertainty to quantities derived from these, as demonstrated for the Canopy Chlorophyll Content (CCC) and the Canopy Water Content (CWC). It is also demonstrated that the correlations between the retrieved parameters have a time dependency in magnitude and sign. This underlines the importance of the

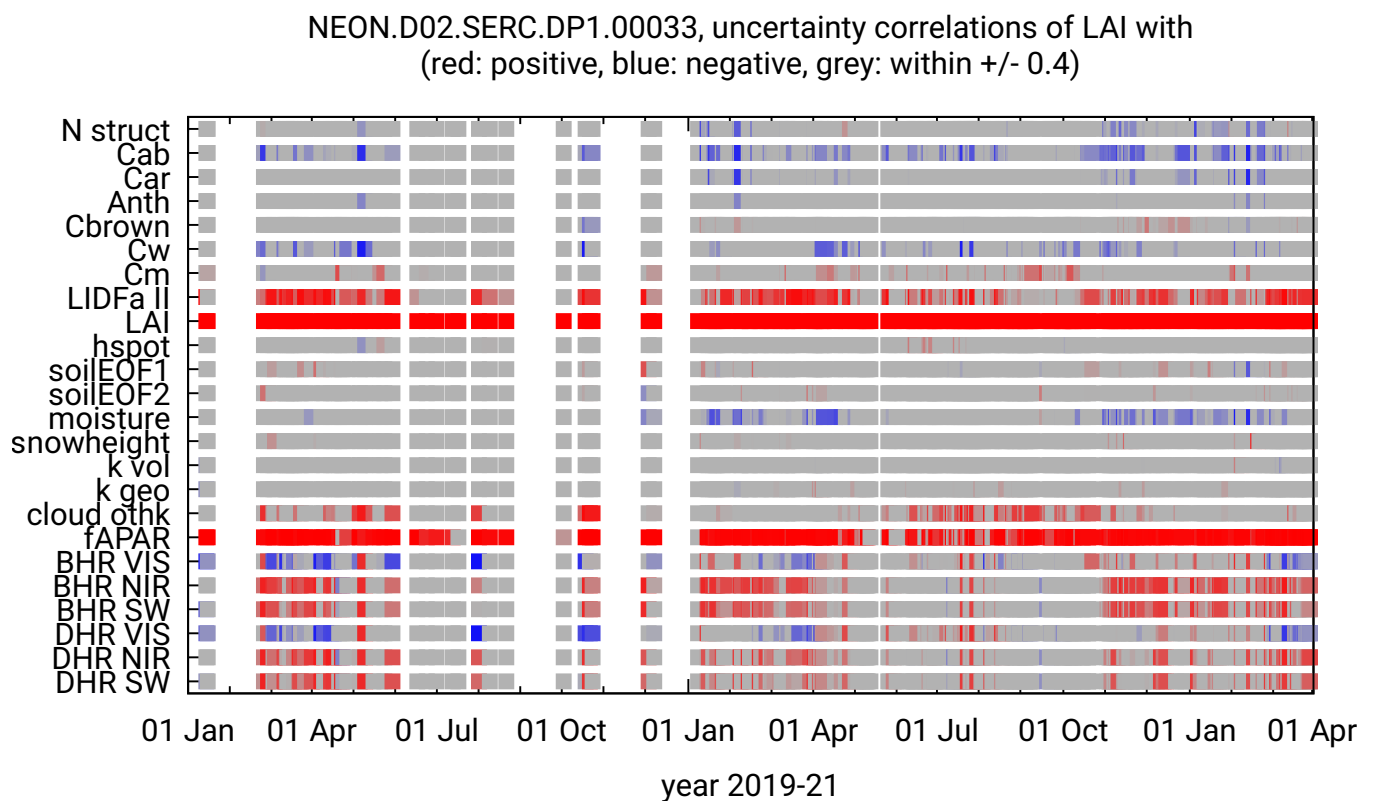


Figure 15. Temporal evolution of the correlation of LAI with the other retrieved quantities for deciduous broadleaf site NEON.D02.SERC. Correlations below 0.4 are shown in grey.

posterior covariance matrix as part of the retrieval product.

A Product Dissemination Unit (PDU) of SY_2_SYN comprises 19902715 (4865×4091) pixels from three minutes of the satellite swath. Of these pixels, on average roughly one third are land pixels and a substantial number is flagged as clouds. Therefore, a real-time processor would need to process at least 37 kpixels^{-1} ($27 \mu\text{s}$ per pixel). A single thread of OptiSAIL typically achieves a speed of 150 pixel s^{-1} (7 ms per pixel). Therefore, this would be achieved with ease by 246 threads of OptiSAIL. While this leaves not much room for re-processing, it demonstrates, that OptiSAIL puts PROSAIL inversion within the performance limits of current super-computers. The computational speed per thread could be improved by exploiting parameter co-variances, reduction of the number of parameters, e.g. by choosing a different version of PROSPECT, and by improving the first guess in the inversion by, either recycling a previous or neighbouring result, or using the guess of an Artificial Neural Network (ANN) or similar.

Parameters which are expected to have a slow temporal evolution, as for instance the soil EOF coefficients at non-agricultural sites, could use a stronger prior from previous inversions, in order to increase the stability of the retrieval and reduce the effective dimension of the parameter space.

Increased global performance could be achieved by porting the code to a GPGPU (General Purpose Graphics Processing Unit) or a similar device, since the processing of all pixels is strictly parallel without inter-dependencies, which is in line with the design goals of this type of devices.

The present study has focused on processing data from single overpasses of the Sentinels. However, it is common practice to aggregate observations in time over at least a few days (e.g. MODIS 4-day LAI/fPAR product), in order to get better angular sampling and to fill gaps. While the processing of more observations per inversion will take slightly longer, there would only be needed one inversion every n days, such that a net speed-up

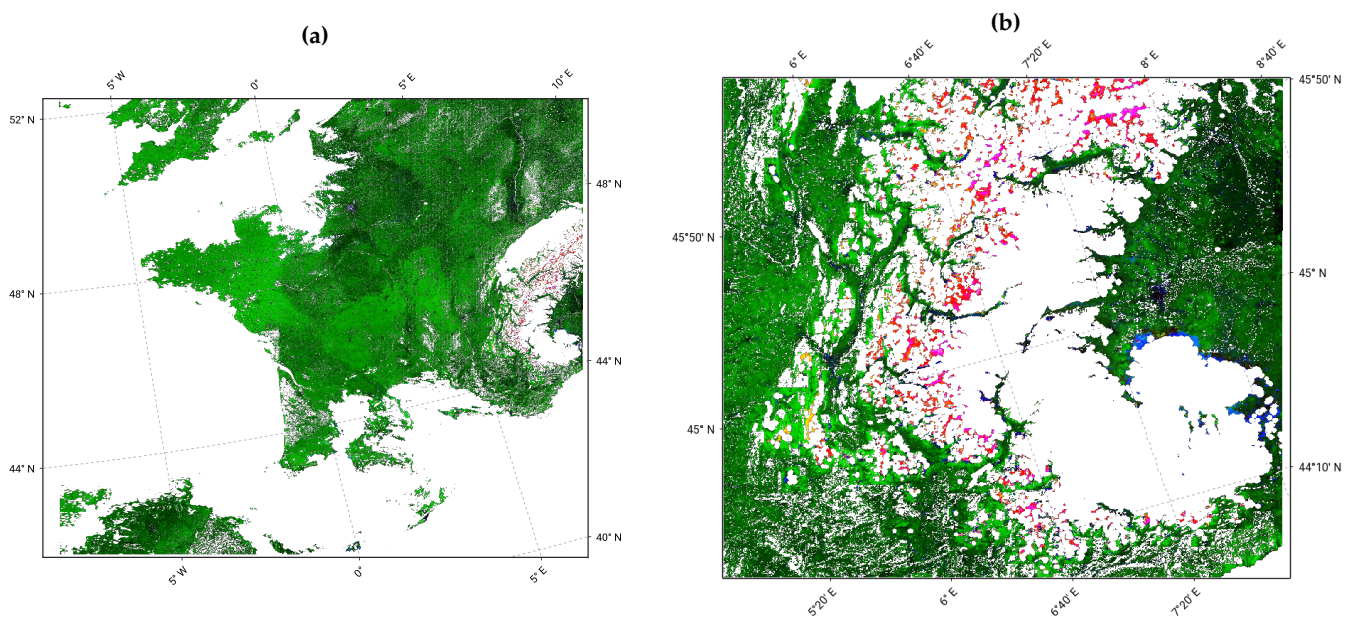


Figure 16. False colour image using Sentinel-3A data of 2021-02-28 10:16:52 UTC, with OptiSAIL, (a) centred over France and (b) detail over the Alps, with snow height (red [1mm-..]), fAPAR (green [0,1]), and cloud thickness (blue [10m-..]). Missing values are white.

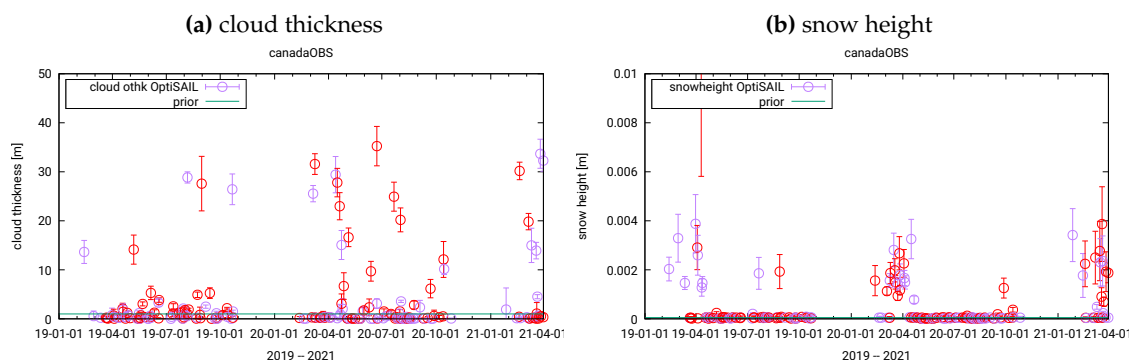


Figure 17. (a) Cloud contamination and (b) sub canopy snow height at mixed evergreen and deciduous needle leaf site canadaOBS. These results are instantaneous, without temporal smoothing. Different marker colours distinguish between Sentinel-A (purple) and Sentinel-B (red).

can be expected for the processing.

The system seamlessly extends to multi-sensor and hyper-spectral retrievals. Other potential applications of the OptiSAIL system are sensitivity analyses in the hyper-spectral context. With the aid of automatic differentiation, the relative importance of individual bands for the retrieval can be quantified efficiently for many different situations.

The direct coupling of the system with an atmospheric correction model for a joint retrieval from top-of-atmosphere radiances is a viable and logical next step.

Author Contributions: Conceptualisation, S.B. and R.G.; methodology, S.B. and R.G.; software, S.B.; validation, S.B.; formal analysis, S.B. and R.G.; investigation, S.B. and R.G.; resources, R.G.; data curation, S.B.; writing—original draft preparation, S.B.; writing—review and editing, S.B. and R.G.; visualisation, S.B.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to their preliminary status.

Acknowledgements

For the PhenoCam network: We thank our many collaborators, including site PIs and technicians, for their efforts in support of PhenoCam. The development of PhenoCam has been funded by the Northeastern States Research Cooperative, NSF's Macrosystems Biology program (awards EF-1065029 and EF-1702697), and DOE's Regional and Global Climate Modeling program (award DE-SC0016011). We acknowledge additional support from the US National Park Service Inventory and Monitoring Program and the USA National Phenology Network (grant number G10AP00129 from the United States Geological Survey), and from the USA National Phenology Network and North Central Climate Science Center (cooperative agreement number G16AC00224 from the United States Geological Survey). Additional funding, through the National Science Foundation's LTER program, has supported research at Harvard Forest (DEB-1237491) and Bartlett Experimental Forest (DEB-1114804). We also thank the USDA Forest Service Air Resource Management program and the National Park Service Air Resources program for contributing their camera imagery to the PhenoCam archive.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BHR	Bi-Hemispheric Reflectance
BRDF	Bi-directional Reflectance Distribution Function
OLCI	Ocean and Land Colour Instrument
PROSPECT	optical leaf PROPERTIES SPECTra (model)
SAIL	Scattering from Arbitrarily Inclined Leaves (model)
SLSTR	Sea and Land Surface Temperature Radiometer
TARTES	Two-streAm Radiative TransfEr in Snow (model)
TOC	Top Of Canopy

References

1. Verrelst, J.; Malenovský, Z.; Van der Tol, C.; Camps-Valls, G.; Gastellu-Etchegorry, J.P.; Lewis, P.; North, P.; Moreno, J. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surveys in Geophysics* **2019**, *40*, 589–629. doi:10.1007/s10712-018-9478-y.
2. Fang, H.; Baret, F.; Plummer, S.; Schaepman-Strub, G. An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications. *Reviews of Geophysics* **2019**, *57*, 739–799, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018RG000608> doi:10.1029/2018RG000608.
3. Gómez-Dans, J.; Lewis, P.; Disney, M. Efficient Emulation of Radiative Transfer Codes Using Gaussian Processes and Application to Land Surface Parameter Inferences. *Remote Sensing* **2016**, *8*, 119. doi:10.3390/rs8020119.
4. Chernetskiy, M.; Gómez-Dans, J.; Gobron, N.; Morgan, O.; Lewis, P.; Truckenbrodt, S.; Schmullius, C. Estimation of FAPAR over Croplands Using MISR Data and the Earth Observation Land Data Assimilation System (EO-LDAS). *Remote Sensing* **2017**, *9*, 656. doi:10.3390/rs9070656.
5. Bacour, C.; Frederic, B.; Béal, D.; Weiss, M.; Pavageau, K. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sensing of Environment* **2006**, *105*, 313–325. doi:10.1016/j.rse.2006.07.014.
6. Verhoef, W.; Bach, H. Coupled soil-leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. *Remote Sensing of Environment* **2007**, *109*, 166 – 182. doi:10.1016/j.rse.2006.12.013.
7. Yang, P.; Prikaziuk, E.; Verhoef, W.; van der Tol, C. SCOPE 2.0: A model to simulate vegetated land surface fluxes and satellite signals. *Geosci. Model Dev. Discuss. [preprint]* **2020**. in review, doi:10.5194/gmd-2020-251.
8. Berger, K.; Atzberger, C.; Danner, M.; D'Urso, G.; Mauser, W.; Vuolo, F.; Hank, T. Evaluation of the PROSAIL Model Capabilities for Future Hyperspectral Model Environments: A Review Study. *Remote Sensing* **2018**, *10*, 26. doi:10.3390/rs10010085.
9. Féret, J.B.; Gitelson, A.; Noble, S.; Jacquemoud, S. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sensing of Environment* **2017**, *193*, 204 – 215. doi:10.1016/j.rse.2017.03.004.

10. Libois, Q.; Picard, G.; France, J.; Arnaud, L.; Dumont, M.; Carmagnola, C.; King, M.D. Influence of grain shape on light penetration in snow. *The Cryosphere* **2013**, *7*, 1803–1818. doi:10.5194/tc-7-1803-2013.
11. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P.J.; Asner, G.P.; François, C.; Ustin, S.L. PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment* **2009**, *113*, S56 – S66. Imaging Spectroscopy Special Issue, doi:10.1016/j.rse.2008.01.026.
12. Jacquemoud, S.; Baret, F. PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment* **1990**, *34*, 75–91. doi:https://doi.org/10.1016/0034-4257(90)90100-Z.
13. Preisendorfer, R.W.; Mobley, C. *Principal component analysis in meteorology and oceanography*; Vol. 17, *Developments in Atmospheric Science*, Elsevier: Amsterdam – Oxford – New York – Tokyo, 1988.
14. von Storch, H.; Zwiers, F.W. *Statistical Analysis in Climate Research*; Cambridge University Press, 1999; p. 494.
15. World Agroforestry Centre. The ICRAF/ISRIC spectral library. online 2014-07-11, 2014.
16. Aitkenhead, M.J.; Black, H.I. Exploring the Impact of Different Input Data Types on Soil Variable Estimation Using the ICRAF-ISRIC Global Soil Spectral Database. *Appl. Spectrosc.* **2018**, *72*, 188–198.
17. A, S.; M, N.; G, T.; L, M.; van Wesemael B. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS ONE* **2013**, *8*. doi:10.1371/journal.pone.0066409.
18. Philpot, W. Spectral Reflectance of Wetted Soils. Symposium: “Art, Science and Applications of Reflectance Spectroscopy (ASARS)”, 2010. doi:10.13140/2.1.2306.0169.
19. Sun, Y.; Zheng, X.; Qin, Q.; Meng, Q.; Gao, Z.; Ren, H.; Wu, L.; Wang, J.; Wang, J. Modeling Soil Spectral Reflectance with Different Mass Moisture Content. *Spectroscopy and Spectral Analysis* **2015**, *35*, 2236–2240. doi:10.3964/j.issn.1000-0593(2015)08-2236-05.
20. Lucht, W.; Schaaf, C.B.; Strahler, A.H. An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing* **2000**, *38*, 977–998. doi:10.1109/36.841980.
21. Yang, Q.; Liu, X.; Wu, W. A Hyperspectral Bidirectional Reflectance Model for Land Surface. *Sensors* **2020**, *20*. doi:10.3390/s20164456.
22. Doctor, K.Z.; Bachmann, C.M.; Gray, D.J.; Montes, M.J.; Fusina, R.A. Wavelength dependence of the bidirectional reflectance distribution function (BRDF) of beach sands. *Appl. Opt.* **2015**, *54*, F243–F255. doi:10.1364/AO.54.00F243.
23. Sandmeier, S.R.; Middleton, E.M.; Deering, D.W.; Qin, W. The potential of hyperspectral bidirectional reflectance distribution function data for grass canopy characterization. *Journal of Geophysical Research: Atmospheres* **1999**, *104*, 9547–9560, [https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/1999JD900094]. doi:10.1029/1999JD900094.
24. Kokhanovsky, A. Optical properties of terrestrial clouds. *Earth-Science Reviews* **2004**, *64*, 189 – 241. doi:10.1016/S0012-8252(03)00042-4.
25. Kokhanovsky, A.A.; Rozanov, V.V.; Zege, E.P.; Bovensmann, H.; Burrows, J.P. A semianalytical cloud retrieval algorithm using backscattered radiation in 0.4–2.4 μm spectral region. *Journal of Geophysical Research: Atmospheres* **2003**, *108*, "AAC 4–1"–"AAC 4–19", [https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2001JD001543]. doi:10.1029/2001JD001543.
26. Myneni, R.; Knyazikhin, Y.; Park, T. MOD15A3H MODIS/Combined Terra+Aqua Leaf Area Index/FPAR Daily L4 Global 500m SIN Grid. Technical report, Boston University and MODAPS SIPS - NASA, NASA LP DAAC, 2015.
27. ORNL DAAC. MODIS Collection 6 Land Products Fixed Sites Subsetting and Visualization Tool, 2018. doi:10.3334/ORN LDAAC/1567.
28. Richardson, A.D.; Hufkens, K.; Milliman, T.; Aubrecht, D.M.; Chen, M.; Gray, J.M.; Johnston, M.R.; Keenan, T.F.; Klosterman, S.T.; Kosmala, M.; Melaas, E.K.; Friedl, M.A.; Frolking, S. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Scientific Data* **2018**, *5*, 180028. doi:10.1038/sdata.2018.28.
29. Liu, D.C.; Nocedal, J. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Program.* **1989**, *45*, 503–528. doi:10.1007/BF01589116.
30. Giering, R.; Kaminski, T. Recipes for Adjoint Code Construction. *ACM Transactions on Mathematical Software* **1998**, *24*, 437–474. doi:10.1145/293686.293695.
31. Giering, R.; Kaminski, T. Applying TAF to generate efficient derivative code of Fortran 77-95 programs. Proceedings of GAMM 2002, Augsburg, Germany, 2002.
32. Giering, R.; Kaminski, T. Recomputations in Reverse Mode AD. In *Automatic Differentiation: From Simulation to Optimization*; Corliss, G.; Faure, C.; Griewank, A.; Hascoët, L.; Naumann, U., Eds.; Computer and Information Science, Springer: New York, 2002; chapter 33, pp. 283–291. doi:10.1007/978-1-4613-0075-5.
33. Giering, R.; Kaminski, T.; Slawig, T. Generating Efficient Derivative Code with TAF. *Future Gener. Comput. Syst.* **2005**, *21*, 1345–1355. doi:10.1016/j.future.2004.11.003.
34. Giering, R.; Voßbeck, M. Increasing Memory Locality by Executing Several Model Instances Simultaneously. Recent Advances in Algorithmic Differentiation; Forth, S.; Hovland, P.; Phipps, E.; Utke, J.; Walther, A., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; Vol. 87, pp. 93–101. doi:10.1007/978-3-642-30023-3_9.