Quantifying vegetation biophysical variables from the Sentinel-3/ FLEX tandem mission: Evaluation of the synergy of OLCI and FLORIS data sources

Charlotte De Grave[a,*], **Jochem Verrelst**[a], **Pablo Morcillo-Pallarés**[a], **Luca Pipia**[a], **Juan Pablo Rivera-Caicedo**[b], **Eatidal Amin**[a], **Santiago Belda**[a], **José Moreno**[a]

[a]Image Processing Laboratory (IPL), Parc Científic, Universitat de València, 46980, Paterna, València, Spain

[b]CONACyT-UAN, Secretaría de Investigación y Posgrado, Universidad Autónoma de Nayarit, Ciudad de la Cultura Amado Nervo, CP. 63155 Tepic, Nayarit, Mexico

Abstract

The ESA's forthcoming FLuorescence EXplorer (FLEX) mission is dedicated to the global monitoring of the vegetation's chlorophyll fluorescence by means of an imaging spectrometer, FLORIS. In order to properly interpret the fluorescence signal in relation to photosynthetic activity, essential vegetation variables need to be retrieved concomitantly. FLEX will fly in tandem with Sentinel-3 (S3), which conveys the Ocean and Land Colour Instrument (OLCI) that is designed to characterize the atmosphere and the terrestrial vegetation at a spatial resolution of 300 m. In this work we present the retrieval models of four essential biophysical variables: (1) Leaf Area Index (LAI), (2) leaf chlorophyll content (Cab), (3) fraction of absorbed photosynthetically active radiation (fAPAR), and (4) fractional vegetation cover (FCover). These variables can be operationally inferred by hybrid retrieval approaches, which combine the generalization capabilities offered by radiative transfer models (RTMs) with the flexibility and computational efficiency of machine learning methods. The RTM SCOPE (Soil Canopy Observation, Photochemistry and Energy fluxes) was used to generate a database of reflectance spectra corresponding to a large variety of canopy realizations, which served subsequently as input to train a Gaussian Process Regression (GPR) algorithm for each targeted variable. Three sets of GPR models were developed, based on different spectral band settings: (1) OLCI (21 bands between 400 and 1040 nm), (2) FLORIS (281 bands between 500 and 780 nm), and (3) their synergy. Their respective performances were assessed based on simulated reflectance scenes. Regarding the retrieval of Cab, the OLCI model gave good model performances (R^2: 0.91; RMSE: 7.6 $\mu g.\ cm^{-2}$), yet superior accuracies were achieved as a result of FLORIS' higher spectral resolution (R^2: 0.96; RMSE: 4.8 $\mu g.\ cm^{-2}$). The synergy of both datasets did not further enhance the variable retrieval. Regarding LAI, the improvement of the model performances by using only FLORIS spectra (R^2: 0.87; RMSE: 1.05 $m^2.m^{-2}$) rather than only OLCI spectra (R^2: 0.86; RMSE: 1.12 $m^2.m^{-2}$) was less evident but merging both data sets was more beneficial (R^2: 0.88; RMSE:

[*]Corresponding author. charlotte.degrave@uv.es (C. De Grave).

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1.01 $m^2.m^{-2}$). Finally, the three data sources gave good model performances for the retrieval of fAPAR and Fcover, with the best performing model being the Synergy model (fAPAR: R^2: 0.99; RMSE: 0.02 and FCover: R^2: 0.98; RMSE: 0.04). The ability of the models to process real data was subsequently demonstrated by applying the OLCI models to S3 surface reflectance products acquired over Western Europe and Argentina. Obtained maps showed consistent patterns and variable ranges, and comparison against corresponding Sentinel-2 products (coarsened to a 300 m spatial resolution) led to reasonable matches (R^2: 0.5–0.7). Altogether, given the availability of the multiple data sources, the FLEX tandem mission will foster unique opportunities to quantify essential vegetation properties, and hence facilitate the interpretation of the measured fluorescence levels.

1 Introduction

The increase in atmospheric carbon dioxide (CO_2) due to terrestrial emissions, and the corresponding global warming and associated climate change (Cox et al., 2000), are highly concerning issues, and it becomes urgent to take actions to mitigate the extent of the already occurring adverse effects (e.g. increase in rainfall intensity and associated flooding events, increase in drought frequency and severity, sealevel rise, permafrost thawing) (IPCC, 2019). Vegetation dynamics modulate the CO_2 balance through carbon assimilation during the photosynthetic process (Dixon et al., 1994; McLeod et al., 2011). Improved knowledge of global vegetation photosynthesis becomes therefore clearly a priority in research about the Earth system. The forthcoming FLuorescence EXplorer (FLEX) mission (Drusch et al., 2017) of the European Space Agency (ESA) is dedicated to the global monitoring of chlorophyll fluorescence, which is a sensitive probe for the photosynthetic performance of vegetation, and thus for the related carbon sequestration (Baker, 2008; Mohammed et al., 2019), also known as gross primary production (GPP) at ecosystem level (Zhang et al., 2016). However, the measurement of fluorescence alone is not enough to quantitatively determine vegetation photosynthesis as a function of environmental conditions. Additional measurements about canopy temperature, solar irradiance at the surface, and vegetation conditions, are crucial for a proper interpretation of fluorescence levels (Moreno et al., 2006). As a matter of fact, by the time the emitted fluorescence reaches the satellite sensor, it has been subjected to the influences of diverse drivers in the vegetation, environment, and atmosphere, and disentangling the effects of these various factors is needed to use fluorescence as an effective proxy for photosynthesis and associated stress effects (Mohammed et al., 2019). For this reason, the following four essential vegetation variables need to be retrieved together with the fluorescence signal: (1) Leaf Area Index (LAI), (2) leaf chlorophyll content (Cab), (3) fraction of Absorbed Photosynthetically Active Radiation (fAPAR), and (4) fractional vegetation cover (FCover). LAI is defined as half the total radiation intercepting area per unit ground surface area (Chen and Black, 1992), accounting for the amount of vegetation that absorbs or scatters solar radiation. Because

the leaf surfaces are the primary sites of energy and mass exchange with the atmosphere, important processes such as evapotranspiration and photosynthesis are strongly related to LAI (Fang and Liang, 2014). The chlorophylls, Chl a and Chl b, are pigments that are responsible for the conversion of light energy into stored chemical energy. Some of the solar photons absorbed by the pigments are re-emitted as fluorescent photons (Porcar-Castell et al., 2014). The amount of solar radiation absorbed by a leaf is a direct function of the photosynthetic pigment content and thus, chlorophyll content determines the photosynthetic potential and primary production of vegetation (Gitelson et al., 2003). Solar radiation in the spectral range from 400 to 700 nm, known as Photosynthetically Active Radiation (PAR), provides the energy required by terrestrial vegetation to produce organic materials from mineral components. The part of this PAR that is effectively absorbed by plants is the fAPAR. Hence, also this variable plays a critical role in the primary productivity of canopies, in the associated fixation of atmospheric CO_2, and in the energy balance of the surface (GCOS, 2011). FCover is the fraction of green vegetation as seen from nadir, and is used to decouple vegetation and soil contribution in energy balance processes such as evapotranspiration (Weiss et al., 2000). These four essential variables can be retrieved by optical sensors since they all impact the spectral signature of the observed surface in the visible and infrared part of the spectrum. The FLEX mission will fly in tandem formation with Sentinel-3 (S3) and as such, together they will cover almost the entire optical domain with an unprecedented spectral resolution. FLEX will carry a push-broom hyperspectral imager (FLORIS), which has been optimized for the discrimination of the fluorescence signal in terrestrial vegetation, providing images with a 150 km swath and a 300 m pixel size. By combining two imaging spectrometers, FLORIS will measure the radiance between 500 and 780 nm with high spectral sampling (0.1–2 nm) and with a very high spectral resolution of 0.3–0.7 nm (Full Width Half Maximum) within the Oxygen-A (755–780 nm) and –B bands (677–697 nm) (Coppo et al., 2017; Moreno et al., 2016). Meanwhile, S3 will provide the land surface temperature and the ancillary data for the atmospheric correction, and will contribute to retrieve vegetation information (Coppo et al., 2017). Its Ocean and Land Colour Instrument (OLCI) is an imaging spectrometer with a spectral definition (21 bands between 400 and 1020 nm) that is primarily designed to measure ocean colour over open ocean and coastal zones but is also optimized to characterize the atmosphere (aerosol composition, water vapour content and illumination conditions) and the terrestrial vegetation (ESA, 2013). The official Level-2 products from OLCI include the OLCI Global Vegetation Index (OGVI), which is equivalent to fAPAR, and the OLCI Terrestrial Chlorophyll Index (OTCI), as a proxy for Cab (ESA, 2017). The other instrument on S3, the Sea and Land Surface Temperature Radiometer (SLSTR), provides the surface temperature at a very high accuracy (1-2K) (ESA, 2013), which is also a necessary prerequisite to construe the observed fluorescence signal.

This paper aims at describing the vegetation retrieval models that are integrated into the Level-2 (L2) Retrieval Module of the FLEX End-to-End Mission Performance Simulator (E2ES) (Tenjo et al., 2018; Vicent et al., 2016) in order to generate biophysical products from FLEX/S3 reflectances. Among the various methods suitable for the quantitative extraction of terrestrial vegetation variables using optical remote sensing imagery, probably the most promising ones are the hybrid regression methods (Verrelst et al., 2015a, 2019).

These methods combine the generalization capabilities of physically-based radiative transfer models (RTMs) with the flexibility and computational efficiency of machine learning regression algorithms (MLRA) (Verrelst et al., 2015a). The core concept is to apply inverse mapping by means of a MLRA trained with simulated data coming from a RTM. This approach has been used earlier to develop operational algorithms for various satellite missions (e.g. ENVISAT/MERIS, SPOT/VEGETATION, Sentinel-2/MSI) (Bacour et al., 2006; Baret et al., 2007; Weiss and Baret, 2016), consisting of an artificial neural network (ANN) (Hornik et al., 1989) trained by simulations of the coupled leaf-canopy RTM PROSAIL (PROSPECT + SAIL) (Jacquemoud and Baret, 1990; Verhoef, 1984). One-dimensional (1D) canopy RTMs such as SAIL run very fast because they are based on a simple description of canopy architecture, i.e. approximated by a turbid medium, and assume homogeneity in horizontal direction (Verhoef, 1984; Van Der Tol et al., 2009b). Conversely, complex three-dimensional (3D) RTMs, although being more realistic, require extensive parameterizations and are computationally demanding (Gastellu-Etchegorry et al., 1996; North, 1996). They are therefore less suitable for the generation of large databases for the MLRA training. Also, because of the frequent lack of knowledge on prior information on the statistical distribution and co-distributions of most land surface features, turbid medium RTMs are often preferred for operational purposes (Bacour et al., 2006; Baret et al., 2007; Verger et al., 2011; Weiss and Baret, 2016). Furthermore, the directional reflectances simulated by 1D and 3D models tend to converge at coarser spatial resolutions, typically beyond 100 m (Widlowski et al., 2005), which is valid for the FLEX mission.

Among existing MLRAs, the kernel-based methods are known to deal well with low numbered, potentially high-dimensional, training samples, as well as with outliers and noise in the data, which makes them perfectly suited to extract information related to vegetation properties from imaging spectroscopy data (Gehler and Schölkopf, 2009). In earlier experimental studies, Gaussian process regression (GPR) proved to be particularly promising for the retrieval of biophysical variables, thereby outperforming other MLRAs such as ANNs and random forests (Verrelst et al., 2012, 2015c; Rivera-Caicedo et al., 2017). In Rivera Caicedo et al. (2014) for example, GPR outperformed the majority of other tested MLRA's for the prediction of Cab and LAI from various spectroscopic datasets. In García-Haro et al. (2018), GPR was compared with Kernel Ridge Regression and ANN for the retrieval of LAI, fAPAR and FCover from the AVHRR (Advanced Very High Resolution Radiometer) sensor on board of the MetOp (Meteorological-Operational) satellites and revealed as the most robust and stable regression method. Similarly, Verrelst et al. (2012) concluded that GPR yielded most accurate estimations of Cab, LAI and FCover from different S3 and Sentinel-2 (S2) band settings, as compared to three other MLRAs. Furthermore, GPR offers two main advantages compared to the widely used ANN: (1) it gives an indication of which bands are more relevant to the constructed model, making it more transparent in terms of model development; (2) it provides an uncertainty (or confidence) level for each per-pixel prediction. This is appealing for mapping applications, as it enables to assess the model transferability in space and time (Verrelst et al., 2015a). Hence, in view of optimizing the extraction of the variables Cab, LAI, fAPAR and FCover from FLEX/S3 images, we developed hybrid retrieval models, combining the 1D RTM SCOPE (Soil Canopy Observation, Photochemistry and Energy fluxes) (Van Der Tol et

al., 2009b) with the GPR algorithm. SCOPE is based on SAIL (Verhoef, 1984, 1985) regarding the calculation of radiative transfer within a canopy. The model has however been extended with various sub-models, some of which have specifically been developed within FLEX preparatory studies, such as calculation modules for fluorescence and photosynthesis (Van Der Tol et al., 2009a, 2014). While these modules are necessary for the retrieval of higher-level products related to photosynthetic activity and vegetation stress indicators, only the SCOPE reflectance outputs and some fluxes such as fAPAR are required for the here developed products. Yet, SCOPE has been selected since it is the benchmark model in all FLEX studies (e.g. Cogliati et al., 2015; Sabater et al., 2017; Celesti et al., 2018). SCOPE has earlier been used to explore the retrieval of biophysical variables from FLEX/S3 Top-Of-Atmosphere (TOA) radiance spectra by model inversion through numerical optimization (Verhoef et al., 2018). This method has however a high computational cost due to per-pixel iterations and as such, the authors used only forty simulated cases for validation, with relatively little variation in the observation conditions. Hybrid retrieval methods should be preferred for the development of operational algorithms (Verrelst et al., 2019). In this respect, the objectives of the present study are the following: (1) to develop generic retrieval models based on the hybrid SCOPE-GPR concept for four variables that are essential within the context of monitoring photosynthesis from space; (2) to evaluate, based on simulated reflectance scenes, the benefit of the synergistic use of both FLEX/S3 tandem mission data sources on the retrieval performances; and, (3) to generate variable maps from S3 imagery and to compare them against equivalent S2 vegetation products in order to demonstrate the validity of the models using real data.

The remaining of this paper is structured as follows. Section 2 describes the retrieval method, including the database generation and the used algorithms. Section 3 presents the validation results obtained with simulated reference scenes and shows the retrieval maps resulting from the application of the OLCI models (i.e. the GPR models based on OLCI reflectances) on real S3 images. We finally close this paper with a discussion (see Section 4) on the achieved model performances and how they can be further improved in preparation for operational use.

2 Methodology

The pursued strategy for the development of the retrieval models forms part of the FLEX E2ES framework (Vicent et al., 2016) and is shown schematically in Fig. 1. SCOPE simulates surface reflectances in the viewing direction by means of the spectral response function of the requested sensor, which are then used to train the GPR algorithm. After the algorithms development, the first stage is their testing in a controlled environment in order to avoid increasing the uncertainty and to enable a proper model performance evaluation. For this purpose, the models were validated using SCOPE simulated reflectance scenes, to which some noise was added to account for different uncertainty sources. The trained GPR models were then applied to these simulated scenes, and the obtained retrieval maps were compared against reference images. The following sections describe the used algorithms and elaborate on the subsequent steps of the retrieval procedure.

2.1 FLEX/S3 tandem mission data

The FLEX tandem mission is composed of multiple sensors, each having their own specific spectral configuration (see Fig. 2). The retrieval algorithms are developed based on surface reflectances and hence, atmospheric correction is assumed. Within the FLEX L2 processing chain, atmospheric correction is conducted at the prior level L2B, fluorescence is retrieved at level L2C, and the retrieval of the biophysical variables is carried out at level L2D (Vicent et al., 2016). Regarding the S3 OLCI instrument, only 16 out of the 21 bands were used for the retrieval of the biophysical variables, since the five remaining bands were specifically designed for the characterization of the atmosphere (retrieval of cloud top pressure, water vapour and aerosols) (ESA, 2013). As such, these bands are not included in the Level-2 Surface Directional Reflectance variable (SDR) of the S3 SYNERGY product (Henocq et al., 2018), which was used to derive the retrieval maps of the biophysical variables (see Section 2.4). The S3 SDR product includes also five optical bands of the S3 SLSTR sensor, which are measured by two independent scanners, one pointing at nadir and one pointing backwards at a viewing zenith angle of 55° (ESA, 2013). Only the SLSTR bands measured in nadir view were considered in the present study since viewing angles until 25° were used in SCOPE (see Section 2.3). Two of these bands are located in the SWIR (short-wave-infrared) region of the spectrum, which can be beneficial for the estimation of LAI (Asner, 1998) and FCover (Wang et al., 2018). Their potential added value was therefore assessed but initial tests using simulated data showed negligible improvement and for some variables, even poorer retrievals were observed. Hence, these bands were discarded for further analysis. For consistency purposes, the S3 SDR product is referred to as "OLCI SDR" in the remainder of this paper. Finally, while FLORIS will cover the spectrum from 500 to 780 nm with a spectral resolution between 0.3 and 3 nm, the spectra were resampled at 1 nm according to the Bottom-Of-Atmosphere reflectance product, which will be provided in the previous level of the processing chain (i.e. L2C).

2.2 Hybrid method

2.2.1 SCOPE—SCOPE (used version: v.1.70) is a soil-vegetation-atmosphere scheme that includes soil-leaf-canopy RTMs along with a micro-meteorological model for simulating turbulent heat exchange, and a plant physiological model for photosynthesis (Van Der Tol et al., 2014). The radiative transfer scheme is based on SAIL but is extended with emitted radiation, namely chlorophyll fluorescence and thermal radiation. Leaf radiative transfer is calculated with Fluspect (Vilfan et al., 2016), which also provides emitted fluorescence radiation. Within this study, the SCOPE code was extended in order to enable the calculation of FCover. In SCOPE, the canopy is modeled as a turbid medium, which means that variations of macroscopic properties in the horizontal plane as well as clumping of twigs and branches are neglected (Van Der Tol et al., 2009b). FCover is therefore approximated empirically from the gap fraction at nadir. The gap fraction can be expressed mathematically as $P = exp\,(-k \times LAI)$, where k is the extinction coefficient (Nilson, 1971; Campbell, 1986; Zheng and Moskal, 2009). In SCOPE, k is calculated based on the leaf inclination distribution and the viewing angle (Verhoef, 1998).

2.2.2 Gaussian process regression and dimensionality reduction—GPR is a nonparametric Bayesian method that learns the relationship between the inputs, i.e. the

surface reflectances, and the outputs, i.e. the biophysical variable to retrieve, by fitting a flexible nonlinear function directly from the data pairs (Rasmussen and Williams, 2006). In this method, the variable to be retrieved is actually a Gaussian Process (GP), which is defined by mean and covariance information, calculated with a kernel function. The kernel function essentially encodes prior assumptions about the underlying function, such as its smoothness or periodicity, by means of hyperparameters that are optimized during the training phase. Here, the radial basis function (RBF) kernel with an adaptive lengthscale and a diagonal noise covariance matrix was used. This composite kernel assumes that the correlation between two input points decays proportionally to their distance. This means that the closer the input points, the more similar the corresponding predictions (Schulz et al., 2018). The process starts with a prior GP distribution, which gets constrained by the observed data points, producing the posterior GP distribution over functions. The more data points, the more the distribution is constrained, thus reducing the uncertainty. From the posterior GP distribution, the posterior predictive distribution is then calculated. Therefore, instead of a simple point prediction for each test point, it is possible to use the mean of the predictive distribution and compute confidence intervals for the prediction, using its variance. The utility of providing such uncertainty intervals has yet been demonstrated (e.g. Verrelst et al., 2012, 2013); it allows to evaluate whether the estimation is reliable relative to the data that were presented during the training phase.

Because of the high spectral resolution of FLORIS, an additional step was required to deal with band multicollinearity when training the FLORIS and the Synergy models with GPR (Rivera-Caicedo et al., 2017). Here, Principal Component Analysis (PCA) (Jolliffe, 1986) was used to reduce the number of dimensions in the dataset by combining the bands into twenty components. This transformation makes the GPR models better adapted to capture the full spectral information content while reducing considerably the processing time. Twenty components were evaluated as an acceptable trade-off between accuracy and processing time (not shown), which is in accordance with what is stated in Verrelst et al. (2017). With this number of components, PCA is an attractive dimensionality reduction method because it performs as well as other, more complex, methods while being more robust to spectral noise (Rivera-Caicedo et al., 2017).

2.3 Development of the training database

As conducting accurate ground measurements at medium spatial resolution (300 m) over a large range of surface types is difficult, if not impossible, RTM simulations enable us to generate the database needed to train the GPR algorithm. It comes however inevitably with inherent empirical assumptions on the distributions of the variables (Bacour et al., 2006). SCOPE was used to simulate surface directional reflectances based on the combination of values of different input variables characterizing the vegetation canopies, the background soil and the sun-sensor geometry. The training database (TDB) was built initially at the SCOPE spectral resolution (1 nm from 400 to 2400 nm) and then resampled according to the spectral response functions of the targeted sensor (OLCI, FLORIS or their synergy). Subsequently, the resampled TDB was used to train the GPR retrieval models. The next sections elaborate further on how the TDB was constructed.

2.3.1   Defining SCOPE inputs—Since SCOPE consists of multiple sub-models, it requires to define a large quantity of input variables (Van Der Tol et al., 2009b). In order to identify the variables, which primarily contribute to the variation of the spectral signal simulated by SCOPE (from 400 to 2400 nm at 1 nm resolution), a preliminary sensitivity analysis was applied using the global sensitivity analysis (GSA) toolbox that is integrated in the Automated Radiative Transfer Models Operator (ARTMO) Graphic User Interface (Verrelst, 2015; Verrelst et al., 2015b). As Verrelst et al. (2015b) already demonstrated, the SCOPE meteorological and leaf biochemical variables (e.g. maximum carboxylation capacity), which are necessary for the the calculation of the photosynthetic rate and the fluorescence (Van Der Tol et al., 2009a, 2009b), have no impact on the reflectance. The GSA was therefore only conducted on the remaining SCOPE inputs, which are parameters characterizing the sun-sensor geometry, the soil, the canopy structure and the leaf structure and biochemistry. The method implemented in the toolbox is a modified version of the Sobol' method proposed by Saltelli et al. (2010) and its validity has yet been ascertained in numerous studies (e.g. Morcillo-Pallarés et al., 2019; Prikaziuk and Tol, 2019; Verrelst, 2015; Verrelst et al., 2015b). This method integrates both the first-order effects (the contribution of each input variable to the variance of the model output) and the total sensitivity effects (the first-order effects plus the interactions between input variables (Song et al., 2012)). The total effect sensitivity index defined by Saltelli et al. (2008) quantifies the relative contribution of each input variable on the variability of the spectral reflectances. The sample size, N, which is used to estimate the effect of each variable (k), determines the number, $N(k + 2)$, of generated simulations. Here, the sample size was fixed at 2000, as recommended by Morcillo-Pallarés et al. (2019) to ensure stable GSA results. The authors used PROSAIL for their analysis but since it is similar to SCOPE, taking the same number of samples is considered as an acceptable approach. The contributing variables were then ranged according to the ranges found in existing databases (OPTICLEAF; http://opticleaf.ipgp.fr/) and in previous studies (Berger et al., 2018; García-Haro et al., 2018; Verhoef et al., 2018; Rivera-Caicedo et al., 2017; Weiss and Baret, 2016; Croft et al., 2015; Houborg et al., 2015; Verrelst et al., 2015c; Houborg and Boegh, 2008; Lauvernet et al., 2008; Bacour et al., 2006; Combal et al., 2003; Baret and Fourty, 1997; Curran and Milton, 1983), and in order to cover all geometrical configurations and canopy realizations (see Table 1). The leaf anthocyanin content (Cant) and the leaf senescent material content (Cs) were set to 0 as green canopies are key in the context of FLEX. Most variables were varied according to a Gaussian distribution as it is typically the distribution used for describing parameters occurring in nature (Frank, 2009; Gordon, 2006). The Gaussian distributions are actually truncated as minimum and maximum values were provided. However, for some variables, the whole range has the same probability to occur and therefore a uniform distribution was rather selected (e.g. geometrical variables). The latter distribution was also chosen for the retrieved variables LAI and Cab, as a model predicting the whole existing range of values needs to be developed. LIDFa and LIDFb are the two parameters characterizing the leaf inclination distribution function (LIDF), LIDFa controlling the average leaf inclination and LIDFb, the bi-modality of the distribution (Verhoef, 1984). For these parameters, all possible combinations within the range [-1;1] were used with equal prevalence, although with one restriction, i.e. that the sum of the absolute values of both parameters is smaller than or equal to 1, in accordance with Verhoef

(1998). Also, the leaf water content (Cw) and the leaf dry matter content (Cdm) are usually tied in such a way that the values for the relative water content (H = Cw/(Cw + Cdm)) are kept close to 0.8, which is the typical value for green leaves (Lauvernet et al., 2008; Bacour et al., 2006; Combal et al., 2003). In the OPTICLEAF database, however, a larger range of values can be found for H and accordingly, a range going from 0.45 to 0.93, with a mean of about 0.7 (± 0.1) was selected.

2.3.2   Size of the training database—The size of the training database (TDB) is an additional element to be determined. GPR does not require being trained over many simulations, as opposed to other MLRAs such as neural networks, but adding samples increases the model's computational load, and thus processing speed (Camps-Valls et al., 2016; Rasmussen and Williams, 2006). An acceptable trade-off must be found between a sufficient sampling of the parameter space while keeping moderate processing times. The impact of the TDB size was therefore analyzed by building retrieval models based on TDB's of 1000, 2000, 3000, 5000, and 10,000 simulations. Four-fold cross-validation (Snee, 1977) was used to test the trained GPR models and calculate their performances, based on the usual goodness-of-fit statistics: the Coefficient of determination (R^2), the Root Mean Squared Error (RMSE) and the relative RMSE (RRMSE, in % relative to the mean value of the observed/reference data). Further, the GPR processing time was recorded, including the training and testing phases of the models.

2.3.3   Adding noise to the reflectance spectra—Simulated spectra are usually injected with noise in order to account for different uncertainty sources such as RTM assumptions, sensor noise and data processing including radiometric calibration, atmospheric and geometric corrections (Verger et al., 2011; García-Haro et al., 2018; Baret et al., 2007). Adding noise helps also to generalize the model and to prevent over-fitting on the pure RTM outputs (Brede et al., 2020). As such, a Gaussian white noise was added to the Top-Of-Canopy reflectances according to a generic model, which was also used for simulated S2 spectra in Weiss and Baret (2016):

$$R'(\lambda) = R(\lambda)[1 + (MD(\lambda) + MI)/100] + AD(\lambda) + AI, \tag{1}$$

where R(λ) is the raw simulated reflectance at wavelength λ, R'(λ) is the reflectance contaminated with noise, MD is the Multiplicative wavelength Dependent factor, MI is the Multiplicative wavelength Independent factor, AD is the Additive wavelength Dependent factor, and AI is the Additive wavelength Independent factor. The values for these four factors have to be determined for each sensor individually. The chosen level of 3% uncertainty (averaged over the whole spectrum; see Table 2 for specific values selected for the different noise factors) is of the same order of magnitude as other values adopted for other sensors (e.g. global uncertainty of 1.5% in García-Haro et al. (2018); 3% in Lauvernet et al. (2008); 4% in Bacour et al. (2006); 2.5% in Weiss et al. (2000)). Weiss and Baret (2016) used identical values for all wavelengths (0.01 for AD and AI and 2% for MD and MI) to add noise to the S2 spectra. If these values were applied to the OLCI spectral configuration, it would result in a global uncertainty of 23% due to the large impact of the additive terms of the noise model on the lower reflectances at smaller wavelengths. Therefore, a wavelength-dependent additive noise was more suited for our case.

2.4 Validation of the retrieval models based on synthetic scenes

Because FLEX is yet to be launched (expected by 2023), we rely on synthetic scenes to assess the performance of the FLORIS and the Synergy models. S3 is already in orbit and OLCI scenes are publicly available for testing the models. Nevertheless, the main objective of this work was to carry out a fair comparison of the performance of the three different models and therefore synthetic scenes were simulated for both instruments as well as for their synergistic product. The simulated scenes consist of surface reflectances without atmospheric effects. Some level of noise was added to make the scenes more realistic (see Section 2.3.3). In the scene generator toolbox (Rivera et al., 2014) from ARTMO, SCOPE was used to generate scenes with a solar zenith angle (SZA) of 30° and a viewing angle (OZA) of 0° (nadir view). Flat topography was assumed in the scene definition. The scenes were made up of two different vegetation classes in order to introduce some spatial heterogeneity while keeping a reasonable computational time. Table 3 shows the selected ranges for Cab and LAI for the two vegetation classes, while the same ranges as for the training database generation were chosen for the other SCOPE input parameters (see Table 1). The reference images of Cab and LAI were generated using a uniform sampling distribution within the selected variable ranges and a random spatial distribution inside each of two vegetation classes composing the image (see Fig. 3). The fAPAR and the FCover reference images, on the other hand, were generated by SCOPE, along with the surface reflectance scenes, based on the defined inputs. The GPR models, i.e. one for each band configuration (OLCI, FLORIS and their synergy), were then applied to the reflectance scenes and the resulting retrieval maps, one for each variable of interest, were evaluated against the reference maps by calculating the goodness-of-fit statistics (R^2, RMSE and RRMSE (%)).

2.5 Assessment of the retrieval models on S3 OLCI images

In order to demonstrate the models' applicability to real data, atmospherically corrected OLCI SDR products were processed into maps of estimated LAI, Cab, fAPAR and FCover. To do so, one more final step was required. Since satellite images also contain areas of non-vegetated surfaces such as bare soil, 418 non-vegetated spectra were added to the training of the OLCI models. It includes 310 bare soil spectra that come from two datasets gathered by the French Aerospace Lab, ONERA (datasets were published in Bablet et al. (2018); Fabre et al. (2015)). The remaining 108 were collected directly from OLCI images, mostly over man-made surfaces. The OLCI SDR products were obtained over Western Europe (image from 20/06/2018) and Argentina (image from 20/02/2019) and downloaded from the Copernicus Open Access Hub[1]. The image over Western Europe (see Fig. 4) represents a heterogeneous area including various land covers (e.g. forests, croplands, urban areas, coastal regions), where the retrieval of vegetation properties at moderate resolution can be challenging. By contrast, the Province of Buenos Aires in Argentina (see Fig. 4) offers a more homogeneous land cover (mainly grasslands and crops) and hence, the retrieval should be facilitated. The retrieval maps over Argentina were compared against the corresponding S2 MultiSpectral Instrument (MSI) vegetation products. The S2 products

---

[1]Last accessed on July 31th, 2019.

were generated from multiple S2 tiles using the Sentinel Application Platform (SNAP) biophysical processor (Weiss and Baret, 2016). Afterwards, the tiles were grouped by dates and merged into mosaics, which were subsequently upscaled from a 20-m to a 300-m spatial resolution. To facilitate a pixel-to-pixel comparison, a common reference grid projection (WGS 84 / UTM zone 21S) was used.

3 Results

3.1 GSA and variable ranges

Because a representative training dataset is crucial in a hybrid retrieval approach, first the SCOPE parameters driving the variability of the reflectance spectra have been identified by means of a GSA (see Fig. 5). In the visible domain and at the red edge (400–750 nm), the main drivers of the variation in reflectances are LAI and Cab, while at the NIR plateau covered by OLCI (750–1040 nm), reflectance is mostly affected by the leaf dry matter content (Cdm) and the average leaf angle (LIDFa). On the other hand, the contribution of the leaf water content (Cw) becomes important after 950 nm. The impact of the soil and geometrical parameters is more regular across the whole spectrum. Further, the relative contribution of LAI increases from 1400 nm onwards. The only parameters that do not have any impact on the reflectance are the leaf width (LW) and the vegetation height (VH). The contributing variables were consequently ranged while the unin-fluential variables were given the SCOPE default value (LW: 0.1 m; VH: 1 m). The parameter space was then randomly generated according to the specified statistical distribution for each variable (see Table 1).

3.2 Size of the training database (TBD)

After the identification of the key SCOPE input variables to vary and their respective ranges and distributions, the optimal number of training samples was determined. A TDB size of 3000 turns out to be the best option with respect to the processing time and the used configuration of sensor band settings, parameter space, and machine learning algorithm (see Fig. 6). Increasing further the size of the TDB only leads to a minor increase in the model performances. At the same time, the processing time rises exponentially with increasing TDB size as the computation involves the inversion of a N x N matrix, where N is the number of simulations (Rasmussen and Williams, 2006). Hence, a TBD of 3000 was considered as a fair trade-off between accuracy and processing time, and used in further processing.

3.3 Validation of the retrieval models against synthetic scenes

3.3.1 Validation of Cab models—Regarding the retrieval of Cab, the model based on OLCI reflectance data alone demonstrates quite good performances (R^2: 0.91; RMSE: 7.6 $\mu g. \ cm^{-2}$; RRMSE: 16.9%; see Fig. 7). A relatively small number of pixels shows large errors (depicted as dark pixels on the error map). The performances improve remarkably when the model is based on FLORIS reflectances (R^2: 0.96; RMSE: 4.8 $\mu g. \ cm^{-2}$; RRMSE: 10.7%). This improvement was expected since FLORIS covers the chlorophyll *a*bsorption bands at very high spectral resolution (1 nm). Finally, the synergy between OLCI and FLORIS does not increase further the model performances. This agrees with the results of

the conducted sensitivity analysis (see Fig. 5) indicating that Cab influences mostly the part of the spectrum located between 550 and 750 nm, which is entirely covered by FLORIS. Our results suggest that the bands measured by OLCI, outside this range, hardly provide supplementary information that could contribute to a more accurate retrieval of Cab.

3.3.2 Validation of LAI models—Despite satisfactory validation statistics (R^2: 0.86; RMSE: 1.12 $m^2.m^{-2}$; RRMSE: 28.0%; see Fig. 8), the OLCI model for LAI prediction shows saturation at high values (LAI > 6). This saturation is a well-known effect and is due to the insensitivity of reflectance to LAI changes in vegetated areas with dense canopies (Bacour et al., 2006; García-Haro et al., 2018). Consequently, the vegetation class composed of larger LAI values contains a rather high amount of pixels with large errors. The retrieval of this variable somewhat improves with the FLORIS model (R^2: 0.87; RMSE: 1.05 $m^2.m^{-2}$; RRMSE: 26.3%) and a small offset of the saturation effect can be observed, which occurs around LAI values of 7 instead of 6. Yet, the error map still indicates a substantial amount of pixels with large errors. The synergy enables to improve further the model performances (R^2: 0.88; RMSE: 1.01 $m^2.m^{-2}$; RRMSE: 25.3%). This can be attributed to the fact that the OLCI model performs better at low LAI values (LAI < 2), while the FLORIS model is more suited at high LAI values (LAI > 6), and the Synergy model combines the advantages of both. As such, the synergy of both sensors enables to lower the estimation error by 10% compared to OLCI but given the relatively high RRMSE associated to the estimated LAI values, this decrease cannot be considered significant.

3.3.3 Validation of fAPAR models—Also fAPAR benefits noticeably from the high spectral resolution of FLORIS (1 nm) for its retrieval (R^2: 0.99; RMSE: 0.021; RRMSE: 2.8%; see Fig. 9) as opposed to the results obtained with OLCI's spectral configuration (R^2: 0.99; RMSE: 0.026; RRMSE: 3.4%). This benefit is reflected in the scatterplot of the reference values against estimated values showing a smaller dispersion (see second column in Fig. 9). This was expected as FLORIS covers a large part of the PAR spectral region at a high spectral resolution. The Synergy model decreases further the RMSE value by 5% as opposed to the results obtained with the FLORIS model. This accuracy improvement is still appreciable considering the high model accuracies (RRMSE around 3% for all models).

3.3.4 Validation of FCover models—All three models perform well with respect to the retrieval of FCover (see Fig. 10). The synergy of both sensors enables to diminish the estimation error by 8% compared to OLCI. Again, as the relative errors lie around 5% for this variable, this diminution is considered meaningful.

3.4 Application of the OLCI models to real OLCI images

As a demonstration case, the developed OLCI models were applied to a OLCI SDR product coming from a S3 image acquired in June 2018 over Western Europe (see Fig. 4). The retrieval maps for the four variables and their associated uncertainties are shown in Fig. 11. Overall, the patterns of the image are well captured, and the values are within consistent ranges. The uncertainty associated to each pixel estimation is approximated by the standard deviation around the predictive mean. It can be observed that the uncertainties are relatively low, except for 2 areas, one situated along the coast of Northern France and the other in the

Pyrenees. These two areas are precisely where the image quality is the lowest and where a speckle effect can be observed due to masked pixels (see Fig. 4). We suspect that the higher uncertainties associated to these pixels are coming from residual errors of the cloud screening process. LAI exhibits the highest uncertainty levels, which is in agreement with our findings regarding the validation of our models using simulated data.

3.5 Comparison of OLCI maps against upscaled S2 vegetation products

The developed OLCI models were subsequently applied to a segment (410×210 km) of an OLCI SDR product acquired over Argentina on the 20th of February 2019. The S2 vegetation products were retrieved from S2 tiles acquired on the 18th of February (see Fig. 4). It was not possible to obtain S3 and S2 data from the exact same date because of cloud contamination. However, the four studied variables are not expected to vary a lot in two days' time. The visual inspection of the retrieval maps suggests that the products of both sensors are consistent (see Fig. 12). This is confirmed by the scatterplots between the products (see Fig. 13). Especially fAPAR and FCover show notable matches with R^2 values above 0.7 and relatively small errors (RMSE of 0.09 and 0.07 respectively), although our Fcover product shows a systematic underestimation. Cab also presents a small error (RMSE of 7.1 $\mu g.\ cm^{-2}$) but the consistency is lower with a correlation of $R^2 = 0.58$. This is due to the fact that most of the pixels of the image are concentrated in a rather narrow range, i.e. between 40 and 60 $\mu g.\ cm^{-2}$ and this, for both sensors. Again, LAI exhibits poorer results, with a R^2 of 0.54 and a RMSE of 0.61 $m^2\ m^{-2}$, which is rather high considering the fact that most pixels have LAI values between 0 and 2.5 (see Fig. 13). The same comparison exercise was repeated over a segment of the more heterogeneous Western Europe image and similar patterns were observed, confirming the consistency between OLCI and S2 vegetation products (results not shown for brevity).

4 Discussion

4.1 Assessment of the retrieval models

This study presents the prototype version of the biophysical retrieval models that will be integrated into the FLEX processing chain for the estimation of the vegetation products that are essential for a proper understanding of the fluorescence signal. Although hybrid retrieval models have been earlier developed for various satellite missions (see e.g. review in Verrelst et al. (2015a)), our methodology introduces some attractive novelties, such as: (1) the presentation of the first vegetation algorithms from OLCI, FLEX and their synergy, (2) the firsttime use of the advanced RTM SCOPE combined with the GPR method for the development of operational retrieval algorithms, and (3) the first-time provision of uncertainty estimates associated to the retrieved values. Moreover, in contrary to earlier Land missions, the produced vegetation products in the FLEX/S3 tandem mission context are not the end products, but serve in support of a proper interpretation of the concurrent fluorescence retrieval and the subsequent calculation of photosynthesis products, vegetation stress indicators and relevant inputs to global carbon models (Moreno et al., 2006), the so-called higher level products. Hence, the vegetation properties need to be derived from the FLEX/S3 data itself so that the vegetation status is quantified concomitantly with the retrieved fluorescence. Furthermore, SCOPE was used for the development of the retrieval

models in order to meet the accuracy requirements of the higher level products, which are also derived from SCOPE, and as such, consistency in the processing chain is ensured.

The validation of the developed models using simulated data showed good performances, at least for Cab, FAPAR and FCover (relative RMSE under 20% for Cab and under 10% for FAPAR and FVC). FLORIS' band settings led to substantial improvements of the variable estimations. Particularly Cab and fAPAR benefited from the high spectral resolution of the sensor with prediction errors (RMSE) decreasing by 37% and 19% respectively, compared to their retrieval based on OLCI data. The improvements obtained with FLORIS for the two other variables, LAI and FCover, are below the estimation accuracies and therefore not considered significant. The synergy of both sensors enabled to improve slightly the retrieval of LAI, fAPAR and FCover (decrease of the RMSE by 4%, 5% and 5% respectively, as opposed to using FLORIS data alone). Considering the estimation accuracies, only the improvement for fAPAR is deemed relevant. OLCI includes bands in the blue region of the spectrum, which, provide useful information for the retrieval of fAPAR, even if these bands are generally more prone to atmospheric effects and hence are noisier due to residual errors (Bodhaine et al., 1999; Martins et al., 2017; Wang and Gordon, 2018). Besides the small gain in accuracy for three of the four studied variables, there is little doubt that the synergy will be advantageous once FLEX data will become available. For example, the synergy concept is expected to be beneficial for observations over specific biomes, e.g. over highly structured canopies such as forests (Lutz et al., 2008), or with other sun-sensor geometries, or in time-series analyses, when cloud cover restricts the number of available observations (Mousivand et al., 2015). Also, the higher temporal resolution of S3 compared to FLEX (around two or three days instead of two weeks) opens opportunities to be exploited in future synergy studies. The algorithms will ultimately be provided through ESA's FLEX/S3 L2 processing chain and synergy studies will be initiated as soon as FLEX data become available to the community.

From all the retrieval models, the LAI models performed poorest with the FLORIS and the Synergy models achieving only slight improvements (respective decrease of RRMSE by 6 and 10% compared to the OLCI model). Consequently, the relative error remained rather high, between 25 and 28% depending on the model. High LAI values are known to be difficult to retrieve from remote sensing data. It is a variable that gives information over the whole canopy while the satellite only sees the top layer of the tree crown in dense canopies. This leads to the well-known saturation effect that results in an underestimation of high LAI values (LAI > 6–7) (García-Haro et al., 2018; Zheng and Moskal, 2009; Bacour et al., 2006). On the one hand, in such dense canopies, the upper layer carries out most of the photosynthesis (Ellsworth and Reich, 1993). Within the scope of the FLEX mission, which seeks to quantify the photosynthetic activity of vegetation, a saturation at this level of LAI can therefore be tolerated. On the other hand, LAI estimates of 6 to 7 at the level of a pixel with moderate spatial resolution (300 m), represent already quite high values and going beyond these numbers is rather uncommon, except maybe for large areas of dense evergreen forests (Tum et al., 2016; Canisius and Fernandes, 2012; Canisius et al., 2010; Satalino et al., 2009). As a confirmation, a global LAI time series from 2002 to 2012 based on data from the ENVISAT MERIS sensor, which is OLCI's predecessor and has the same spatial resolution, only displays values below 6, although a large collection of vegetation types is

studied, among which evergreen broad-leaved forests (Tum et al., 2016). The validation of the MERIS LAI product with in situ data seems, however, to indicate that the used algorithm (BEAM MERIS vegetation processor) underestimates LAI values above 4 (see Fig. 5 in Tum et al. (2016)), which was already suggested in previous work (Canisius et al., 2010; Satalino et al., 2009). Nevertheless, the field data used to validate moderate resolution LAI products generally do not exceed LAI values of 6 or 7 (Tum et al., 2016; Canisius and Fernandes, 2012; Canisius et al., 2010). Our LAI maps presented in Figs. 11 and 12 also show maximum values of 6, which could indicate that our models saturate at this level of LAI. Comparison of results against S2 LAI product enables, however, to assert with relatively good confidence that the considered area actually encompasses LAI values lower than 6. Even though, it is important to develop models capable of predicting LAI ranges going from 0 to 10 as this range is recommended by the Land Product Validation (LPV) sub-group of the Committee on Earth Observation Satellites (CEOS) for the validation and the sensitivity analyses of satellite based LAI products (Fernandes et al., 2014).

The developed GPR models delivered low uncertainties over the S3 pixels (see Fig. 11), which should indicate that they are robust enough to be applied to other data sets. The uncertainties enable evaluating whether all pixels have been properly retrieved and if necessary, to adjust the training data e.g. by accounting for more variation in land cover. For instance, some of the high uncertainties observed in the area of the Pyrenees are from pixels covered with snow, which is a land cover that was not included in the model training. Furthermore, the OLCI maps and the S2 official vegetation products showed generally good agreement (see Fig. 13). The relatively small but systematic discrepancy that was observed between our FCover product and the corresponding S2 product demands for further analysis. Although an underestimation of our FCover product is suspected, in order to gauge which one is closer to the actual value, validation with in situ data is required. One reason for the mismatch may be attributed to the OLCI model lacking a band in the SWIR region, since the S2 SWIR band at 2190 nm was demonstrated to be important for the estimation of FCover (Wang et al., 2018).

It must be remarked that the assessment of our GPR models might be biased as both training and testing data sets are from SCOPE. At this initial stage in the mission developing phase however, a controlled environment is essential in order to avoid increasing the uncertainty and enable a proper model performance evaluation. This is standard procedure in developing an operational processing chain, the first step being the development of the algorithms and their testing in supervised conditions (Kerekes and Landgrebe, 1989). In a second step, they will be tested within the FLEX E2ES, which serves to consolidate the mission specifications and to test the accuracy of the data processing algorithms (Vicent et al., 2016). The E2ES is also based on SCOPE but coupled with an atmospheric RTM to simulate TOA radiance scenes. Yet, even though the presented models are solely a prototype, their application to real OLCI images and the comparison with official S2 products demonstrate their validity for further development.

4.2 Limitations and future improvements

Needless to say that the presented models will continue to be improved during the further development of the mission. Also, just like for any Earth Observation mission, algorithms will undergo multiple optimizations subsequent to validation studies after the launch, when FLEX data will be at disposal. Some limitations and future improvements are worth discussing.

An inconvenience of the GPR technique is its high computational cost and memory requirements for training (Camps-Valls et al., 2018), which grows cubically and quadratically with the number of training points, respectively (Hensman et al., 2013; Moore et al., 2016). Nevertheless, as for any MLRA, once the model is trained, the prediction step is highly performant. For instance, GPR runs hundreds times faster than RTM inversion methods (e.g. based on look-up tables; Verrelst et al. (2015a, 2015c)). Its fast processing speed has also been reported within the FLEX/S3 L2 processing chain. Relative to the geometric and atmospheric correction or the fluorescence retrieval, the GPR-based retrieval of biophysical variables takes only a small fraction ($< 1\%$) of the total processing time. From this point of view, GPR manifests as a competitive candidate for operational processing. When comparing GPR against the commonly used ANN, however, it must be noted that the latter deals easier with larger training samples than GPR, and once trained, it is also much faster in image processing. Yet, GPR does not need as many training samples as ANN and, hence, the training and the prediction process can still be accelerated by optimizing the number of training samples. Moving along this line, intelligent subsampling procedures such as active learning techniques, whereby the most informative samples are selected for optimized regression accuracy, would be a possible strategy to further reduce the dimensionality of the trained models (Verrelst et al., 2016).

Synthetic scenes are undoubtedly a simplification of reality and using them for validation imposes therefore some limitations. Firstly, a perfect co-registration between FLEX and S3 is assumed, and accuracies could decrease when co-registration mismatches appear. This effect should however be abated by the fact that we are dealing with 300 m pixels and that the validation with field measurements, or with products from other sensors with higher spatial resolution, will involve an upscaling (Morisette et al., 2006; Fernandes et al., 2014). Secondly, the validation was done considering only specific illumination conditions (SZA of 30° and OZA of 0°). Considering that FLEX is a nadir-viewing mission (OZA mainly around zero) with a sun-synchronous orbit, the solar-observation geometry is quite restricted, and not all angle combinations are possible. Nevertheless, directional effects will be further addressed in the next version of the retrieval algorithms by including the three angles characterizing the geometry of image acquisition (SZA, OZA and RAA) as supplementary layers in the models, similarly to what was done for S2 (Weiss and Baret, 2016). This should allow to exploit the directional information more explicitly into the retrieval models rather than solely integrating simulations with different geometries in the training database (Weiss and Baret, 2016). Another way to improve our model performances would be to further reduce the parameter space of the training data by using prior information on the typical distribution of vegetation biophysical variables (Verger et al., 2011). Some unrealistic parameter combinations were already filtered out based on

the relationships between Cw and Cdm and between the two variables characterizing the leaf inclination distribution function (LIDFa and LIDFb), but others may still be present and could lead to inconsistent reflectance spectra. In fact, the four retrieved variables, Cab, LAI, fAPAR and FCover, are all correlated and their correlation could be implicitly included by making use of a covariance matrix before performing the random sampling within the parameter space. Also, since the variables are correlated, their estimation errors are not independent from each other and therefore, the overall error might decrease if the correlations among the variables are taken into account in the retrieval scheme. In addition, once the orbit configuration of FLEX is fully defined, the possible sun-sensor geometries will be more restricted.

A limitation linked to SCOPE is that it assumes a random spatial distribution of the canopy foliage while clumped distributions of leaves are neglected. However, leaves are often grouped to form foliage clumps such as tree crowns, branches, and shoots in forests, or rows in crops, which all cause some light obstruction and impact the surface spectral signature (Chen, 2017). This is why indirect methods that quantify LAI based on the inversion of Beer's law generally underestimate LAI values in clumped canopies, meaning they measure an effective quantity rather than the true LAI (Yan et al., 2019). While these clumping effects are especially important for observations at high spatial resolutions, i.e. below 100 m (Hardiman et al., 2018), they likely also impact the LAI retrievals at the pixel scale of the FLEX mission and should therefore be adressed in the next model development stages. Furthermore, we are well aware that the presented FCover product derived from SCOPE simulations is only an approximation, and is not fully consistent in terms of radiative transfer. The use of a 3-D RTM might be more appropriate for this purpose but these models are complex and less suited for the production of large TDBs. This is why many earlier FVC products were actually based on the 1D RTM SAIL (e.g (Bacour et al., 2006; Baret et al., 2007; Weiss and Baret, 2016).

Another important consideration is the level of noise that is injected in the reflectance spectra, which can be further optimized in order to make it more representative of real noise affecting S3/FLORIS spectra. Once the FLEX instruments are fully characterized, we will rely on the FLEX E2ES (Tenjo et al., 2018; Vicent et al., 2016) to provide us with adjusted noise levels. In fact, other operational retrieval algorithms integrated a somewhat random level of Gaussian noise to the training simulations (Weiss and Baret, 2016; García-Haro et al., 2018; Bacour et al., 2006). In recent studies, the adding of noise was demonstrated to be an effective approach in the hybrid model development using real data for crop nitrogen mapping (Berger et al., 2020) and forest LAI mapping (Brede et al., 2020). In any case, the level of noise that can be expected from instrument failures or imperfections in the data processing is likely negligible compared to the effect of pixel heterogeneity. Pixels of 300 m often encompass multiple land cover types, which all contribute to the observed signal. With SCOPE, only vegetated pixels can be simulated with more or less contribution of soil depending on the defined LAI value, although, values as low as 0.0001 can be chosen, which nearly corresponds to bare soil. A fair amount of soil variability can be introduced by ranging the three parameters of the integrated BSM (Brightness - Shape - Moisture) spectral soil model (Verhoef et al., 2018) and the Soil Moisture Content. To account for non-vegetated pixels, spectra corresponding to bare soil surfaces were added to the training

database. Yet, our models will likely work less well over mixed pixels. A common way to address this mixing effect is to create synthetic mixed spectra composed of a linear combination of pure vegetated spectra and pure bare soil spectra (García-Haro et al., 2018; Bacour et al., 2006). This is based on the assumption that reflectance has a linear behaviour, i.e. that the reflectances from the different reflecting surfaces add up linearly to form the signal emitted from the observed 300-m pixel, which is obviously a simplification of what happens in reality. This topic will be further explored when in situ data is available for validation.

Finally, initial tests with simulated data, analyzing the benefit of using the SLSTR optical bands, showed no improvement of the retrieval performances. This was unexpected since two bands are located in the SWIR domain, which is highly sensitive to structural variables such as LAI (Asner, 1998), as also supported by our GSA results. Only the nadir bands were used while the oblique bands were left out since the generated training database did not include the concerned viewing angle, although Verhoef et al. (2018) suggested that these bands can contribute to the retrieval of canopy structure. Some inconsistencies in the S3 processing chain were however reported, such as coregistration issues between OLCI and SLSTR and calibration residual errors in the SLSTR SWIR bands (Sentinel-3 MPC, 2019), which might hamper the added value of these bands at this point in time.

5 Conclusion

In support of the photosynthesis mission FLEX, prototype retrieval models were developed for the operational delivery of essential vegetation products (Cab, LAI, fAPAR and FCover) from three FLEX/S3 tandem mission data sources: FLORIS, OLCI and their synergy. A hybrid approach was pursued based on SCOPE model simulations and the GPR machine learning method, which has the evident advantage of delivering associated uncertainties. A simulation study based on synthetic scenes evidenced that: (1) the high spectral resolution of FLORIS enables to retrieve the studied biophysical variables with high accuracy levels, especially for Cab and fAPAR, which are essential for the estimation of photosynthesis and carbon uptake; and (2) the synergy of both data sources is beneficial for the retrieval of fAPAR, FCover and LAI, attesting the relevance of the tandem mission for the extraction of vegetation properties from spectroscopic data. The retrieval models were subsequently applied to two contrasting OLCI images over Western Europe and Argentina, composed of various vegetated and non-vegetated land covers. The consistency of the four vegetation products with corresponding S2 products corroborates that the developed models produce reliable retrieval products from FLEX/S3 surface reflectance data. Because FLEX is yet to be launched by 2023, improved versions of the retrieval models will continue to be developed. Follow-up research will be necessary to fine-tune the models for real FLEX data, including multi-temporal analysis and validation with ground truth data over multiple vegetation types.

Acknowledgement

References

Asner G. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sens Environ. 1998; 64 (3) 234–253.

Bablet A, Vu P, Jacquemoud S, Viallefont-Robinet F, Fabre S, Briottet X, Sadeghi M, Whiting M, Baret F, Tian J. MARMIT: A multilayer radiative transfer model of soil reflectance to estimate surface soil moisture content in the solar domain (400-2500 nm). Remote Sens Environ. 2018; 217: 1–17.

Bacour C, Baret F, Béal D, Weiss M, Pavageau K. Neural network estimation of LAI, fAPAR, fCover and LAIxCab, from top of canopy MERIS reflectance data: principles and validation. Remote Sens Environ. 2006; 105 (4) 313–325.

Baker N. Chlorophyll fluorescence: A probe of photosynthesis in vivo. Annu Rev Plant Biol. 2008; 59: 89–113. [PubMed: 18444897]

Baret F, Fourty T. Estimation of leaf water content and specific leaf weight from reflectance and transmittance measurements [estimation du contenu en eau et de la masse seche surfacique des feuilles a partir de spectres de reflectance et de transmittance]. Agronomie. 1997; 17 (9–10) 455–464.

Baret F, Hagolle O, Geiger B, Bicheron P, Miras B, Huc M, Berthelot B, Niño F, Weiss M, Samain O, Roujean J, et al. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: principles of the algorithm. Remote Sens Environ. 2007; 110 (3) 275–286.

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 Sens. 2018; 10 (1)

Berger K, Verrelst J, Féret J-B, Wang Z, Wocher M, Strathmann M, Danner M, Mauser W, Hank T. Crop nitrogen monitoring: recent progress and principal developments in the context of imaging spectroscopy missions. Remote Sens Environ. 2020; 242

Bodhaine B, Wood N, Dutton E, Slusser J. On rayleigh optical depth calculations. J Atmos Ocean Technol. 1999; 16 (11 PART 2) 1854–1861.

Brede B, Verrelst J, Gastellu-Etchegorry J-P, Clevers JGPW, Goudzwaard L, den Ouden J, Verbesselt J, Herold M. Assessment of workflow feature selection on forest LAI prediction with sentinel-2A MSI, Landsat 7 ETM+ and Landsat 8 OLI. Remote Sens. 2020; 12 (6)

Campbell G. Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution. Agric For Meteorol. 1986; 36 (4) 317–321.

Camps-Valls G, Verrelst J, Muñoz J, Laparra V, Mateo F, Gomez-Dans J. A survey on Gaussian processes for earth-observation data analysis: a comprehensive investigation. IEEE Geosci Remote Sens Magaz. 2016; 4: 58–78.

Camps-Valls G, Martino L, Svendsen DH, Campos-Taberner M, Noz Marí JM, Laparra V, Luengo D, García-Haro FJ. Physics-aware gaussian processes in remote sensing. Appl Soft Comput. 2018; 68: 69–82.

Canisius F, Fernandes R. Evaluation of the information content of Medium Resolution Imaging Spectrometer (MERIS) data for regional leaf area index assessment. Remote Sens Environ. 2012; 119: 301–314.

Canisius F, Fernandes R, Chen J. Comparison and evaluation of medium resolution imaging spectrometer leaf area index products across a range of land use. Remote Sens Environ. 2010; 114 (5) 950–960.

Celesti M, van der Tol C, Cogliati S, Panigada C, Yang P, Pinto F, Rascher U, Miglietta F, Colombo R, Rossini M. Exploring the physiological information of sun-induced chlorophyll fluorescence through radiative transfer model inversion. Remote Sens Environ. 2018; 215: 97–108.

Chen, J. Remote Sensing of Leaf Area Index and Clumping Index. 2017.

Chen J, Black T. Defining leaf area index for non-flat leaves. Plant Cell Environ. 1992; 15 (4) 421–429.

Cogliati S, Verhoef W, Kraft S, Sabater N, Alonso L, Vicent J, Moreno J, Drusch M, Colombo R. Retrieval of sun-induced fluorescence using advanced spectral fitting methods. Remote Sens Environ. 2015; 169: 344–357.

Combal B, Baret F, Weiss M, Trubuil A, Macé D, Pragnère A, Myneni R, Knyazikhin Y, Wang L. Retrieval of canopy biophysical variables from bi-directional reflectance using prior information to solve the ill-posed inverse problem. Remote Sens Environ. 2003; 84 (1) 1–15.

Coppo P, Taiti A, Pettinato L, Francois M, Taccola M, Drusch M. Fluorescence imaging spectrometer (FLORIS) for ESA FLEX mission. Remote Sens. 2017; 9: 649.

Cox P, Betts R, Jones C, Spall S, Totterdell I. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature. 2000; 408 (6809) 184–187. [PubMed: 11089968]

Croft H, Chen J, Zhang Y, Simic A, Noland T, Nesbitt N, Arabian J. Evaluating leaf chlorophyll content prediction from multispectral remote sensing data within a physically-based modelling framework. ISPRS J Photogramm Remote Sens. 2015; 102: 85–95.

Curran P, Milton E. The relationships between the chlorophyll concentration, lai and reflectance of a simple vegetation canopy. Int J Remote Sens. 1983; 4 (2) 247–255.

Dixon R, Brown S, Houghton R, Solomon A, Trexler M, Wisniewski J. Carbon pools and flux of global forest ecosystems. Science. 1994; 263 (5144) 185–190. [PubMed: 17839174]

Drusch M, Moreno J, Del Bello U, Franco R, Goulas Y, Huth A, Kraft S, Middleton E, Miglietta F, Mohammed G, Nedbal L, et al. The FLuorescence EXplorer mission concept-ESA's earth explorer 8. IEEE Trans Geosci Remote Sens. 2017; 55 (3) 1273–1284.

Ellsworth D, Reich P. Canopy structure and vertical patterns of photosynthesis and related leaf traits in a deciduous forest. Oecologia. 1993; 96 (2) 169–178. [PubMed: 28313412]

ESA. Sentinel-3 User Handbook. 1.0 ed. Sentinel-3 Team; 2013. September, Source. https://sentinel.esa.int

ESA. , editor. S3 Product Notice-OLCI. 1st ed. ESA Mission Management; 2017. November, 1.0 ed. (product Notice ID: S3A.PN-OLCI-L2L.02)

Fabre S, Briottet X, Lesaignoux A. Estimation of soil moisture content from the spectral reflectance of bare soils in the 0.4-2.5 μm domain. Sensors (Switzerland). 2015; 15 (2) 3262–3281.

Fang, H, Liang, S. Reference Module in Earth Systems and Environmental Sciences. Elsevier; 2014.

Fernandes, R, Plummer, S, Nightingale, J, Frederic, B, Camacho de Coca, F, Fang, H, Garrigues, S, Gobron, N, Lang, M, Lacaze, R, Leblanc, S. , et al. CEOS Global LAI Product Validation Good Practices. 2014.

Frank S. The common patterns of nature. J Evol Biol. 2009; 22 (8) 1563–1585. [PubMed: 19538344]

García-Haro F, Campos-Taberner M, Muñoz Marí J, Laparra V, Camacho F, Sánchez-Zapero J, Camps-Valls G. Derivation of global vegetation biophysical parameters from eumetsat polar system. ISPRS J Photogramm Remote Sens. 2018; 139: 57–74.

Gastellu-Etchegorry J, Demarez V, Pinel V, Zagolski F. Modeling radiative transfer in heterogeneous 3-D vegetation canopies. Remote Sens Environ. 1996; 58 (2) 131–156.

GCOS. Systematic observation requirements for satellite-based products for climate 2011 update: supplemental details to the satellite-based component of the implementation plan for the global observing system for climate in support of the unfccc (2010 update). GCOS Rep. 2011; 154

Gehler, PV, Schölkopf, B. An Introduction to Kernel Learning Algorithms 2. John Wiley Sons, Ltd; 2009. 25–48.

Gitelson A, Gritz Y, Merzlyak M. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J Plant Physiol. 2003; 160 (3) 271–282. [PubMed: 12749084]

Gordon, S. Mathematics Learning Centre. The University of Sydney; 2006.

Hardiman B, LaRue E, Atkins J, Fahey R, Wagner F, Gough C. Spatial variation in canopy structure across forest landscapes. Forests. 2018; 9 (8)

Henocq, C; North, P; Heckel, A; Ferron, S; Lamquin, N; Dransfeld, S; Bourg, L; Tote, C; Ramon, D. OLCI/SLSTR SYN L2 algorithm and products overview; Int Geosci Remote Sens Sympos (IGARSS); 2018. 8723–8726.

Hensman J, Fusi N, Lawrence ND. Gaussian Processes for Big Data. CoRR abs/1309.6835. 2013. arXiv:1309.6835

Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw. 1989; 2 (5) 359–366.

Houborg R, Boegh E. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. Remote Sens Environ. 2008; 112 (1) 186–202.

Houborg R, McCabe M, Cescatti A, Gao F, Schull M, Gitelson A. Joint leaf chlorophyll content and leaf area index retrieval from Landsat data using a regularized model inversion system (REGFLEC). Remote Sens Environ. 2015; 159: 203–221.

IPCC. IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems. Summary for Policymakers - Approved Draft. 2019. Available online. https://www.ipcc.ch/site/assets/uploads/2019/08/Fullreport-1.pdf

Jacquemoud S, Baret F. PROSPECT: A model of leaf optical properties spectra. Remote Sens Environ. 1990; 34 (2) 75–91.

Jolliffe, I. Principal Component Analysis. Springer-Verlag New York, Inc; 1986.

Kerekes JP, Landgrebe DA. Simulation of optical remote sensing systems. IEEE Trans Geosci Remote Sens. 1989; 27 (6) 762–771.

Lauvernet C, Baret F, Hascoët L, Buis S, Le Dimet F-X. Multitemporal-patch ensemble inversion of coupled surface-atmosphere radiative transfer models for land surface characterization. Remote Sens Environ. 2008; 112 (3) 851–861.

Lutz D, Washington-Allen R, Shugart H. Remote sensing of boreal forest biophysical and inventory parameters: a review. Can J Remote Sens. 2008; 34: S286–S313.

Martins V, Barbosa C, de Carvalho L, Jorge D, Lobo F, de Moraes Novo E. Assessment of atmospheric correction methods for sentinel-2 MSI images applied to Amazon floodplain lakes. Remote Sens. 2017; 9 (4)

McLeod E, Chmura G, Bouillon S, Salm R, Björk M, Duarte C, Lovelock C, Schlesinger W, Silliman B. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO_2. Front Ecol Environ. 2011; 9 (10) 552–560.

Mohammed GH, Colombo R, Middleton EM, Rascher U, van der Tol C, Nedbal L, Goulas Y, Pérez-Priego O, Damm A, Meroni M, Joiner J, et al. Remote sensing of solar–induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. Remote Sens Environ. 2019; 231 111177 [PubMed: 33414568]

Moore C, Chua A, Berry C, Gair J. Fast methods for training gaussian processes on large datasets. R Soc Open Sci. 2016; 3 (5)

Morcillo-Pallarés P, Rivera-Caicedo J, Belda S, De Grave C, Burriel H, Moreno J, Verrelst J. Quantifying the robustness of vegetation indices through global sensitivity analysis of homogeneous and forest leaf-canopy radiative transfer models. Remote Sens. 2019; 11 (20)

Moreno J, Asner G, Bach H, Belenguer T, Bell A, Buschmann C, Calera A, Calpe J, Campbell P, Cecchi G, Colombo R, et al. Fluorescence Explorer (FLEX): An Optimised Payload to Map Vegetation Photosynthesis from Space. 2006; 3: 2065–2074.

Moreno J, Goulas Y, Huth A, Middleton E, Miglietta F, Mohammed G, Nedbal L, Rascher U, Verhoef W, Drusch M. Very High Spectral Resolution Imaging Spectroscopy: The Fluorescence Explorer (FLEX) Mission. 2016. 264.

Morisette J, Baret F, Privette J, Myneni R, Nickeson J, Garrigues S, Shabanov N, Weiss M, Fernandes R, Leblanc S, Kalacska M, et al. Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup. IEEE Trans Geosci Remote Sens. 2006; 44 (7) 1804–1814.

Mousivand A, Menenti M, Gorte B, Verhoef W. Multi-temporal, multi-sensor retrieval of terrestrial vegetation properties from spectral-directional radiometric data. Remote Sens Environ. 2015; 158: 311–330.

Nilson T. A theoretical analysis of the frequency of gaps in plant stands. Agric Meteorol. 1971; 8 (C) 25–38.

North P. Three-dimensional forest light interaction model using a Monte Carlo method. IEEE Trans Geosci Remote Sens. 1996; 34 (4) 946–956.

Porcar-Castell A, Tyystjärvi E, Atherton J, Van Der Tol C, Flexas J, Pfündel E, Moreno J, Frankenberg C, Berry J. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges. J Exp Bot. 2014; 65 (15) 4065–4095. [PubMed: 24868038]

Prikaziuk E, Tol C. Global sensitivity analysis of the scope model in sentinel-3 bands: thermal domain focus. Remote Sens. 2019; 11 2424

Rasmussen, CE, Williams, CKI. Gaussian Processes for Machine Learning. 2006th ed. the MIT Press; 2006. www.GaussianProcess.org/gpml

Rivera Caicedo J, Verrelst J, Muñoz Marí J, Moreno J, Camps-Valls G. Toward a semiautomatic machine learning retrieval of biophysical parameters. IEEE J SelectTop Appl Earth Observ Remote Sens. 2014; 7 (4) 1249–1259.

Rivera, JP; Sabater, N; Tenjo, C; Vicent, J; Alonso, L; Moreno, J. Synthetic scene simulator for hyperspectral spaceborne passive optical sensors. Application to esa's flex/sentinel-3 tandem mission; 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS); 2014 June. 1–4.

Rivera-Caicedo J, Verrelst J, Muñoz Marí J, Camps-Valls G, Moreno J. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. ISPRS J Photogramm Remote Sens. 2017; 132: 88–101.

Sabater N, Vicent J, Alonso L, Cogliati S, Verrelst J, Moreno J. Impact of atmospheric inversion effects on solar-induced chlorophyll fluorescence: exploitation of the apparent reflectance as a quality indicator. Remote Sens. 2017; 9 (6)

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S. Global sensitivity analysis. Primer. 2008; 304

Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput Phys Commun. 2010; 181 (2) 259–270.

Satalino G, Mattia F, Ruggieri S, Rinaldi M. LAI Estimation of Agricultural Crops from Optical Data at Different Spatial Resolution. 2009; 4: IV797–IV800.

Schulz E, Speekenbrink M, Krause A. A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. J Math Psychol. 2018; 85: 1–16.

Sentinel-3 MPC. [last accessed on 20 April 2020] Sentinel-3 Mission Performance Center Optical Annual Performance Report-Year 2018. Tech Rep ACRI-ST. 2019. April, Available online. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-olci/document-library/-/asset_publisher/hkf7sg9Ny1d5/content/sentinel-3-optical-annual-performance-report-year-2

Snee RD. Validation of regression models: methods and examples. Technometrics. 1977; 19 (4) 415–428.

Song X, Bryan B, Paul K, Zhao G. Variance-based sensitivity analysis of a forest growth model. Ecol Model. 2012; 247: 135–143.

Tenjo C, Rivera-Caicedo JP, Sabater N, Vicent Servera J, Alonso L, Verrelst J, Moreno J. Design of a Generic 3-D scene generator for passive optical missions and its implementation for the ESA's FLEX/Sentinel-3 tandem mission. IEEE Trans Geosci Remote Sens. 2018; 56 (3) 1290–1307.

Tum M, Günther K, Böttcher M, Baret F, Bittner M, Brockmann C, Weiss M. Global gap-free MERIS LAI time series (2002-2012). Remote Sens. 2016; 8 (1)

Van Der Tol C, Verhoef W, Rosema A. A model for chlorophyll fluorescence and photosynthesis at leaf scale. Agric For Meteorol. 2009a; 149 (1) 96–105.

Van Der Tol C, Verhoef W, Timmermans J, Verhoef A, Su Z. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. Biogeosciences. 2009b; 6 (12) 3109–3129.

Van Der Tol C, Berry J, Campbell P, Rascher U. Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence. J Geophys Res Biogeosci. 2014; 119 (12) 2312–2327. [PubMed: 27398266]

Verger A, Baret F, Camacho F. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: evaluation over an agricultural area with CHRIS/PROBA observations. Remote Sens Environ. 2011; 115 (2) 415–426.

Verhoef W. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. Remote Sens Environ. 1984; 16 (2) 125–141.

Verhoef W. Earth observation modeling based on layer scattering matrices. Remote Sens Environ. 1985; 17: 165–178.

Verhoef, W. Theory of Radiative Transfer Models Applied in Optical Remote Sensing of Vegetation Canopies. Ph.D. thesis. Wageningen University; 1998.

Verhoef W, van der Tol C, Middleton E. Hyperspectral radiative transfer modeling to explore the combined retrieval of biophysical parameters and canopy fluorescence from FLEX-Sentinel-3 tandem mission multi-sensor data. Remote Sens Environ. 2018; 204: 942–963.

Verrelst J. ARTMO's global sensitivity analysis (GSA) toolbox to quantify driving variables of leaf and canopy radiative transfer models. EARSeL eProc. 2015; 14: 1–11.

Verrelst J, Muñoz J, Alonso L, Delegido J, Rivera J, Camps-Valls G, Moreno J. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. Remote Sens Environ. 2012; 118: 127–139.

Verrelst J, Rivera J, Moreno J, Camps-Valls G. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. ISPRS J Photogramm Remote Sens. 2013; 86: 157–167.

Verrelst J, Camps-Valls G, Muñoz Marí J, Rivera J, Veroustraete F, Clevers J, Moreno J. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review. ISPRS J Photogramm Remote Sens. 2015a; 108: 273–290.

Verrelst J, Rivera J, van der Tol C, Magnani F, Mohammed G, Moreno J. Global sensitivity analysis of the SCOPE model: what drives simulated canopy-leaving sun-induced fluorescence? Remote Sens. Environ. 2015b; 166: 8–21.

Verrelst J, Rivera J, Veroustraete F, Muñoz Marí J, Clevers J, Camps-Valls G, Moreno J. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods - A comparison. ISPRS J Photogramm Remote Sens. 2015c; 108: 260–272.

Verrelst J, Dethier S, Rivera J, Munoz-Mari J, Camps-Valls G, Moreno J. Active learning methods for efficient hybrid biophysical variable retrieval. IEEE Geosci Remote Sens Lett. 2016; 13 (7) 1012–1016.

Verrelst J, Rivera Caicedo J, Muñoz Marí J, Camps-Valls G, Moreno J. SCOPE-based emulators for fast generation of synthetic canopy reflectance and sun-induced fluorescence spectra. Remote Sens. 2017; 9 (927)

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. Surv Geophys. 2019; 40 (3) 589–629.

Vicent J, Sabater Medina N, Tenjo C, Ramon Acarreta J, Manzano M, Rivera Caicedo J, Jurado Lozano P, Franco R, Alonso L, Verrelst J, Moreno J. FLEX end-to-end Mission performance simulator. IEEE Trans Geosci Remote Sens. 2016; 54: 1–9.

Vilfan N, van der Tol C, Muller O, Rascher U, Verhoef W. Fluspect-B: a model for leaf fluorescence, reflectance and transmittance spectra. Remote Sens Environ. 2016; 186: 596–615.

Wang M, Gordon H. Sensor performance requirements for atmospheric correction of satellite ocean color remote sensing. Opt Express. 2018; 26 (6) 7390–7403. [PubMed: 29609295]

Wang B, Jia K, Liang S, Xie X, Wei X, Zhao X, Yao Y, Zhang X. Assessment of Sentinel-2 MSI spectral band reflectances for estimating fractional vegetation cover. Remote Sens. 2018; 10 (12)

Weiss, M; Baret, F. S2 ToolBox Level 2 Products : LAI, FAPAR, FCOVER Version 1.1. 2016.

Weiss M, Baret F, Myneni R, Pragnère A, Knyazikhin Y. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. Agronomie. 2000; 20 (1) 3–22.

Widlowski J-L, Pinty B, Lavergne T, Verstraete M, Gobron N. Using 1-D models to interpret the reflectance anisotropy of 3-D canopy targets: issues and caveats. IEEE Trans Geosci Remote Sens. 2005; 43 (9) 2008–2017.

Yan G, Hu R, Luo J, Weiss M, Jiang H, Mu X, Xie D, Zhang W. Review of indirect optical measurements of leaf area index: recent advances, challenges, and perspectives. Agric For Meteorol. 2019; 265: 390–411.

Zhang Y, Xiao X, Jin C, Dong J, Zhou S, Wagle P, Joanna J, Guanter L, Zhang Y, Zhang G, Qin Y, et al. Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America. Remote Sens Environ. 2016; 183: 154–169.

Zheng G, Moskal LM. Retrieving leaf area index (LAI) using remote sensing: theories, methods and sensors. Sensors (Basel, Switzerland). 2009; 9: 2719–2745.
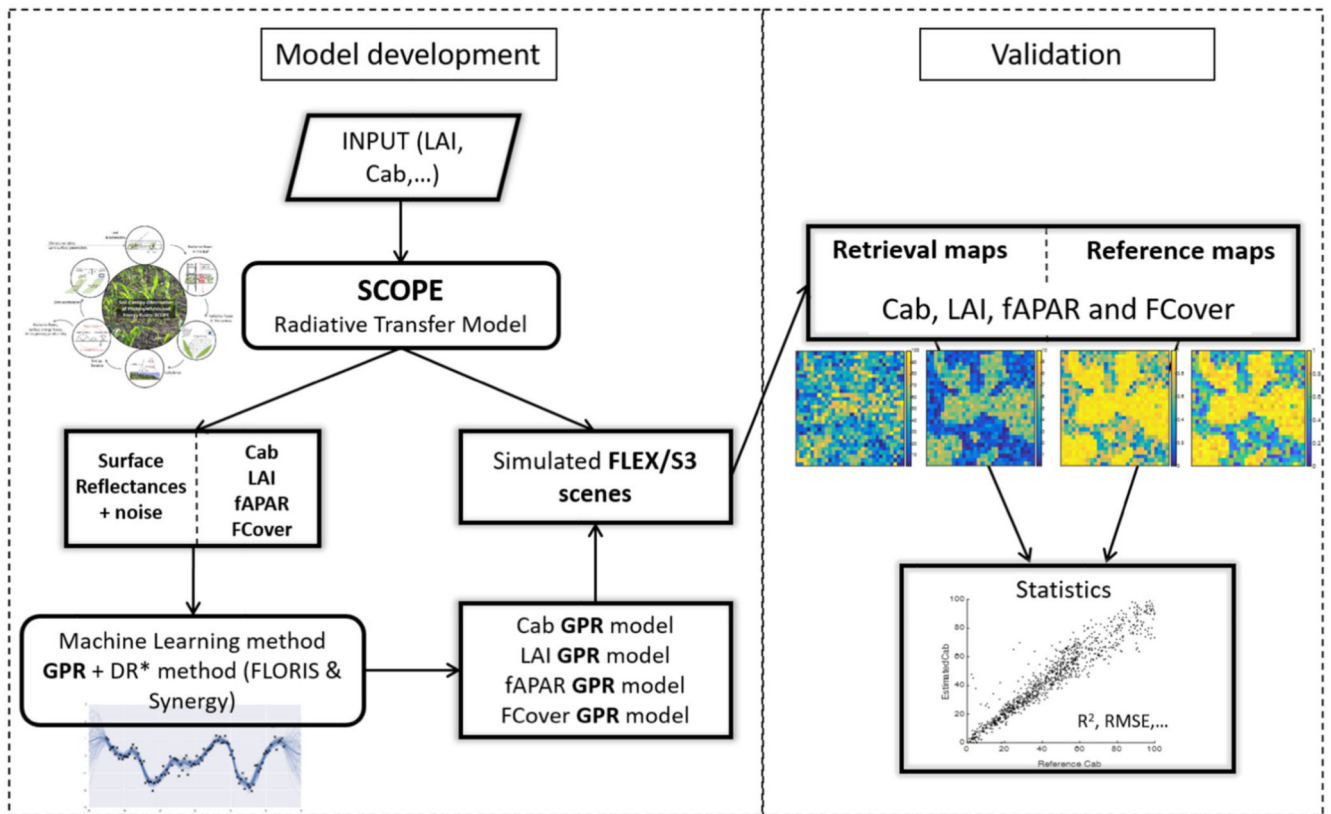
Fig. 1.

Schematic workflow of the presented methodology for the retrieval of the vegetation variables (Cab, LAI, fAPAR, and FCover) from FLEX/S3 reflectance data.* DR: dimensionality reduction (see Section 2.2.2).
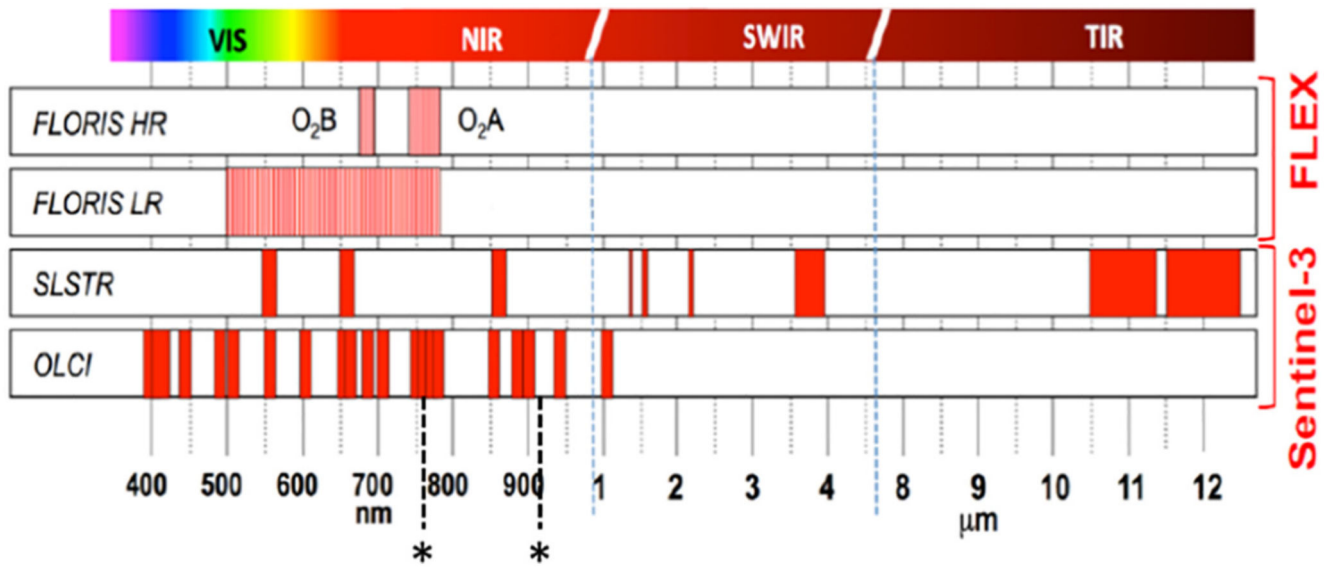
Fig. 2.

Spectral coverage of the instruments from the FLEX mission. The two "*" indicate the position of the 5 OLCI bands that are used for atmospheric characterization - Oa13: 761 nm, Oa14: 764 nm, Oa15: 768 nm, Oa19: 900 nm and Oa20: 940 nm (image credit: ESA).
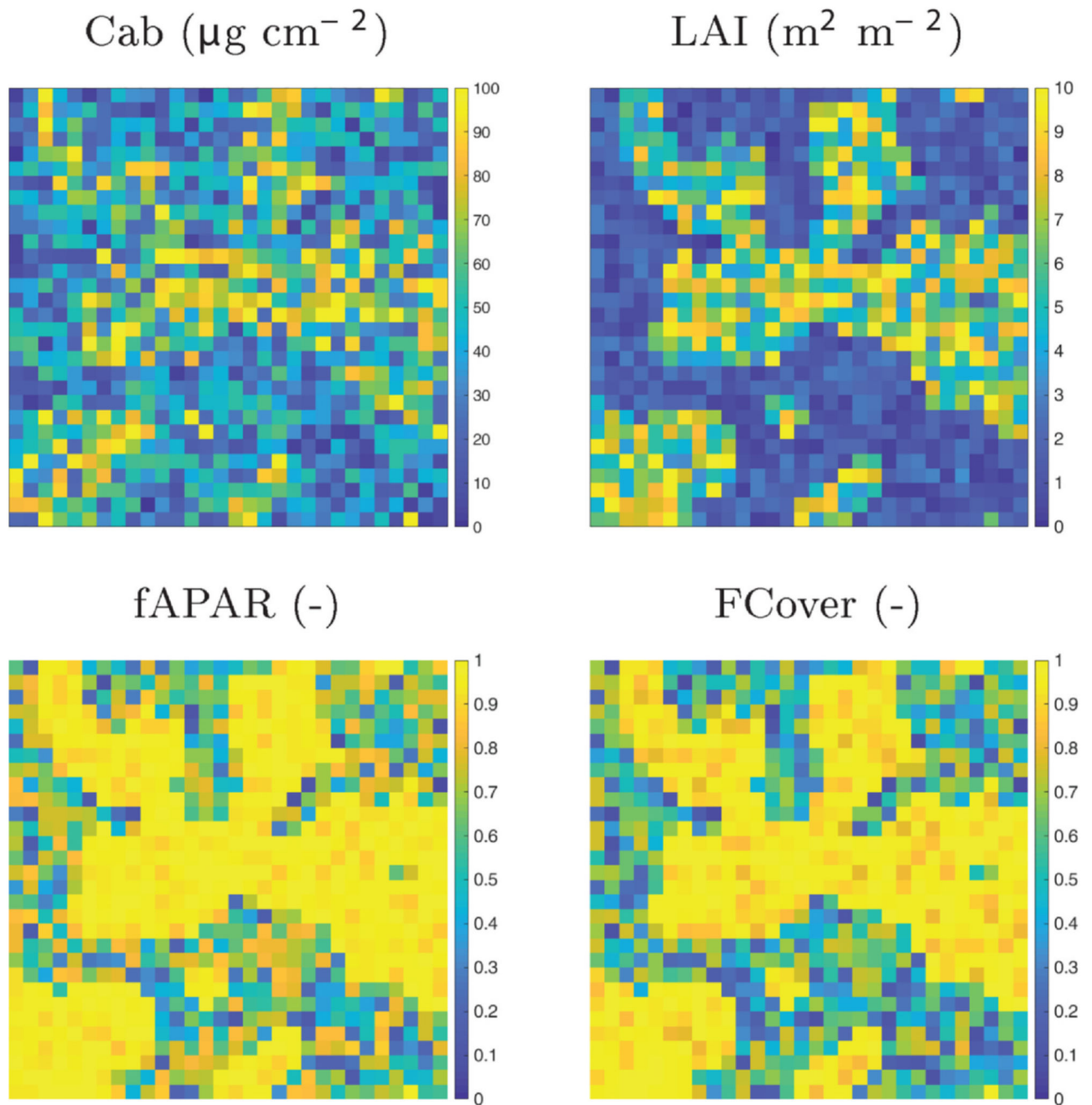
Fig. 3.
Reference maps for the four retrieved variables: Cab, LAI, fAPAR and FCover. The Cab and LAI maps were generated using a uniform sampling distribution within the selected variable ranges and a random spatial distribution inside each of two vegetation classes composing the image. The fAPAR and FCover maps were generated by the RTM SCOPE based on the defined model inputs.
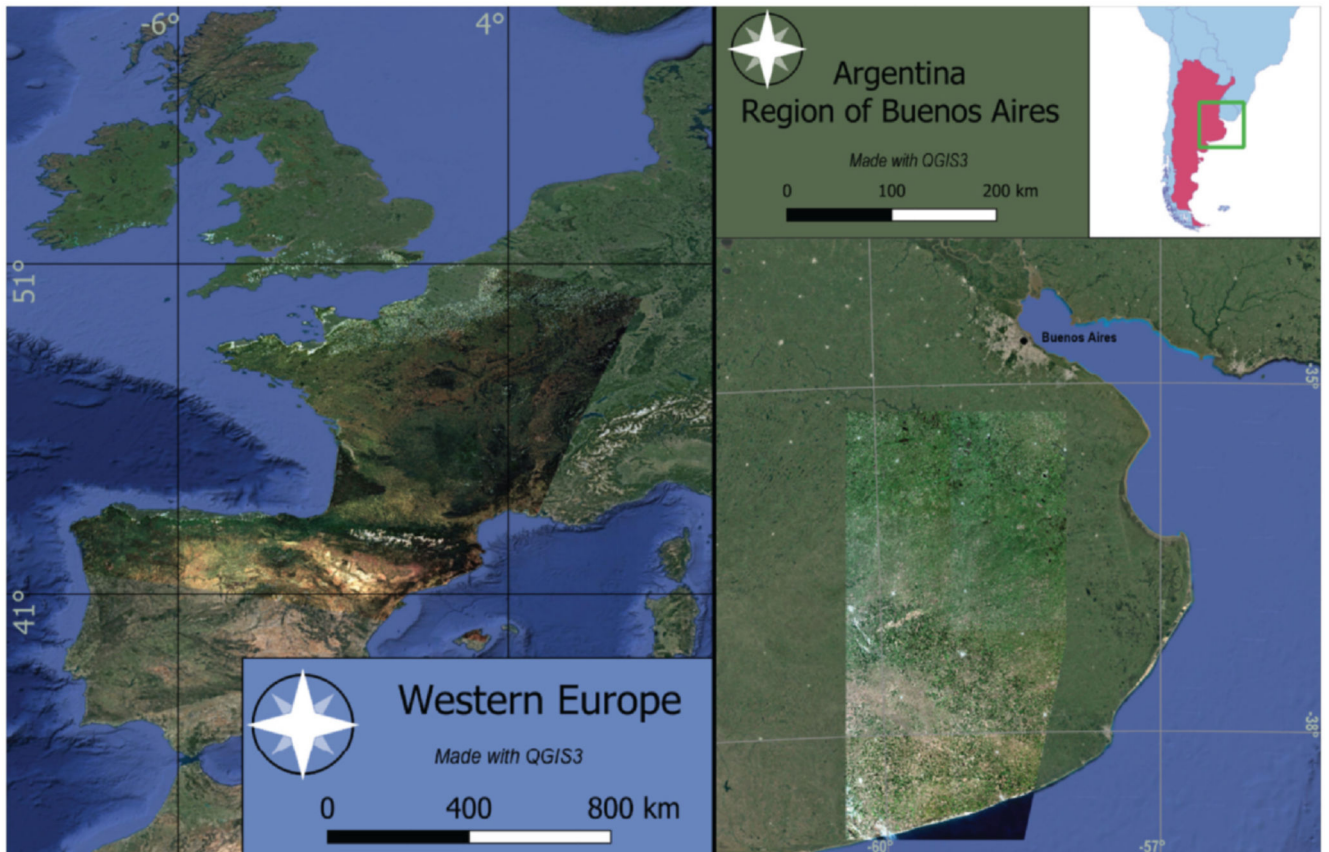
Fig. 4.
Used satellite images and their localization. Left: True-color composite (RGB: 665; 560; 490 nm) of the OLCI SDR product above Western Europe from 20/06/2018. Right: segment (410 × 210 km) of the S2 RGB (665; 560; 490 nm) mosaic above Argentina from 18/02/2019.
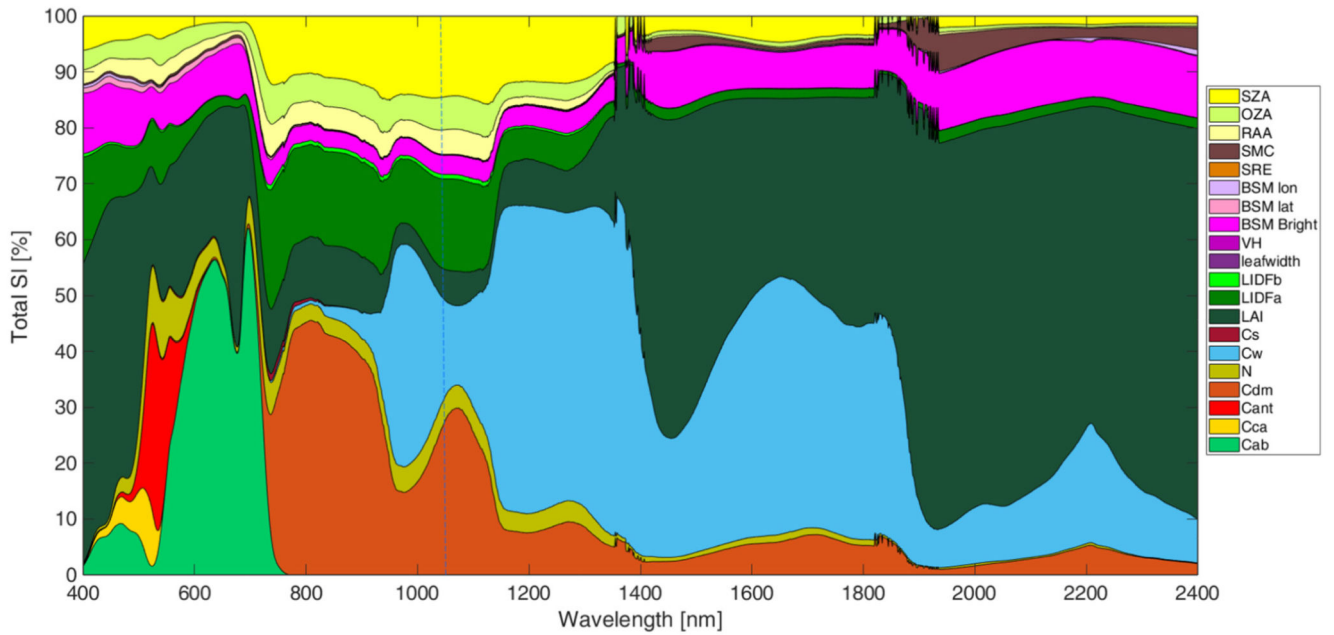
Fig. 5.
Global sensitivity analysis of the SCOPE input variables (parameters characterizing the leaf, the canopy, the soil and the geometry) along the visible and infrared spectral domain (400–2400 nm; 1 nm resolution). The blue dashed line indicates the extent of OLCI's spectral range. See Table 1 for the list of acronyms. Supplementary acronyms: SRE: Soil Resistance for Evaporation (s m^{-1}); VH: Vegetation Height (m); Cs: Leaf senescent material content (−); Cant: Leaf anthocyanin content (μg cm^{-2}). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
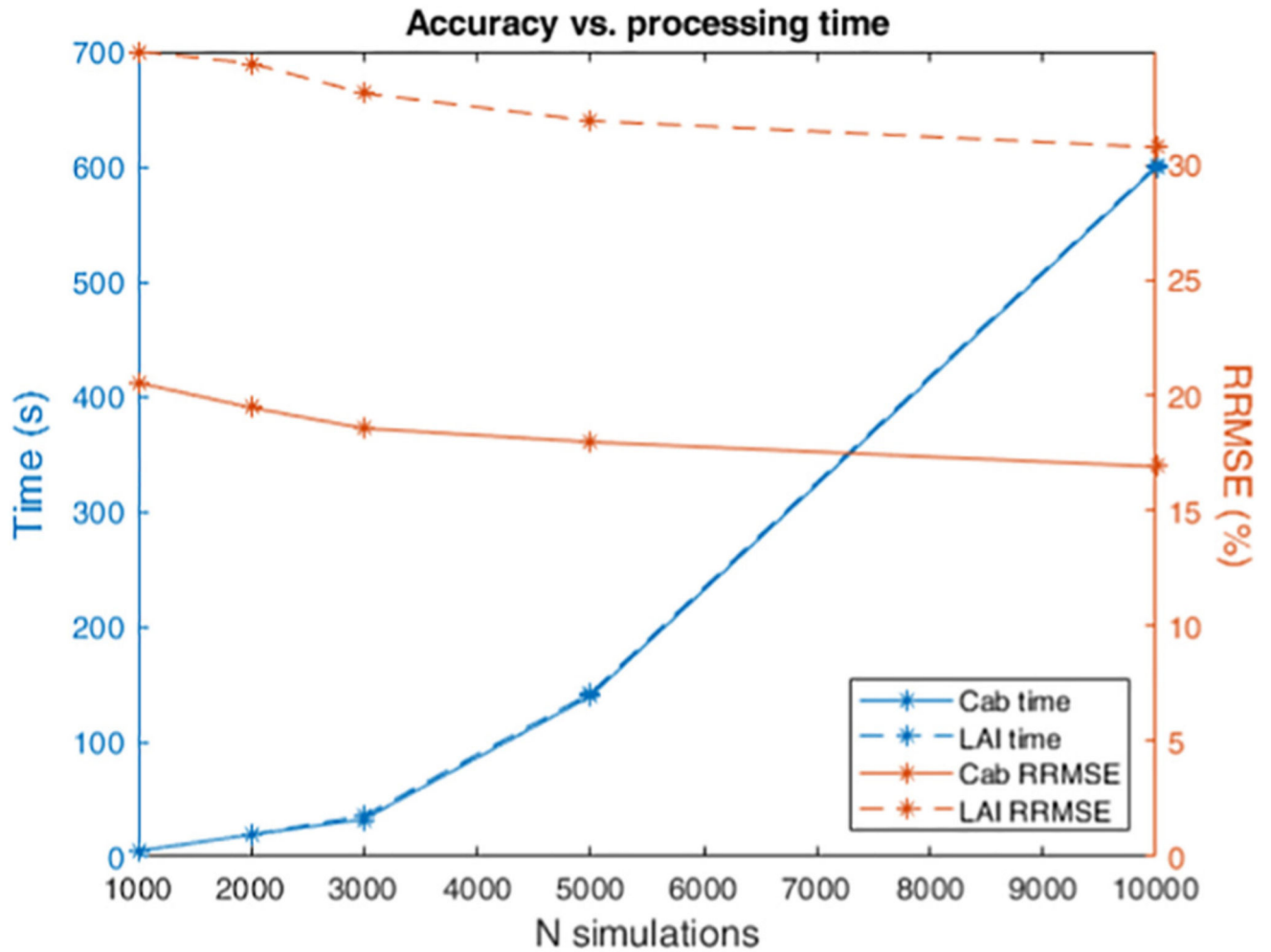
Fig. 6.
Impact of the size of the training database (number of simulations, N) on the model performances (RRMSE, %) and the processing time (including training and testing of the GPR algorithm, in seconds). As an example, the given performances are from OLCI models for the retrieval of LAI and Cab.
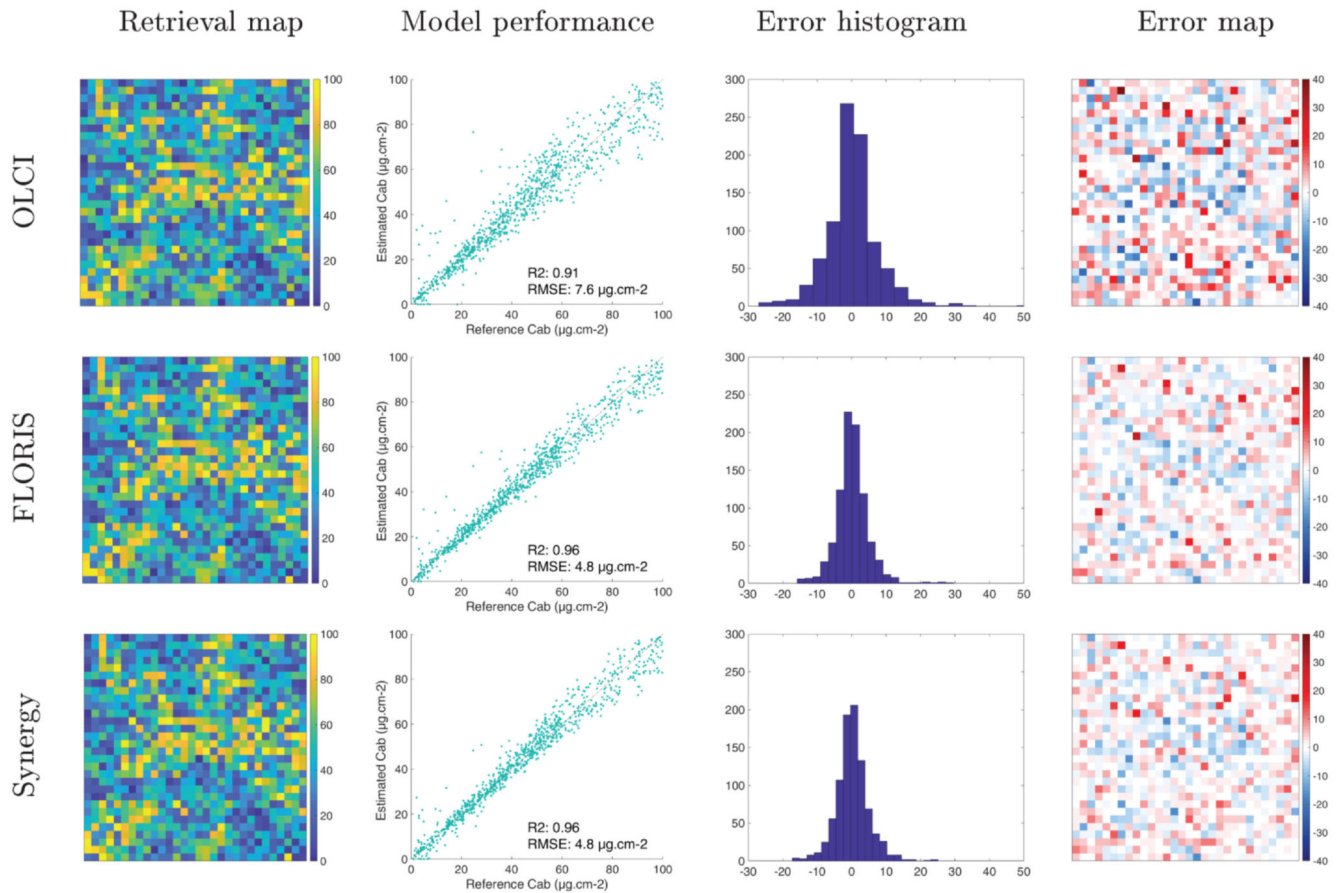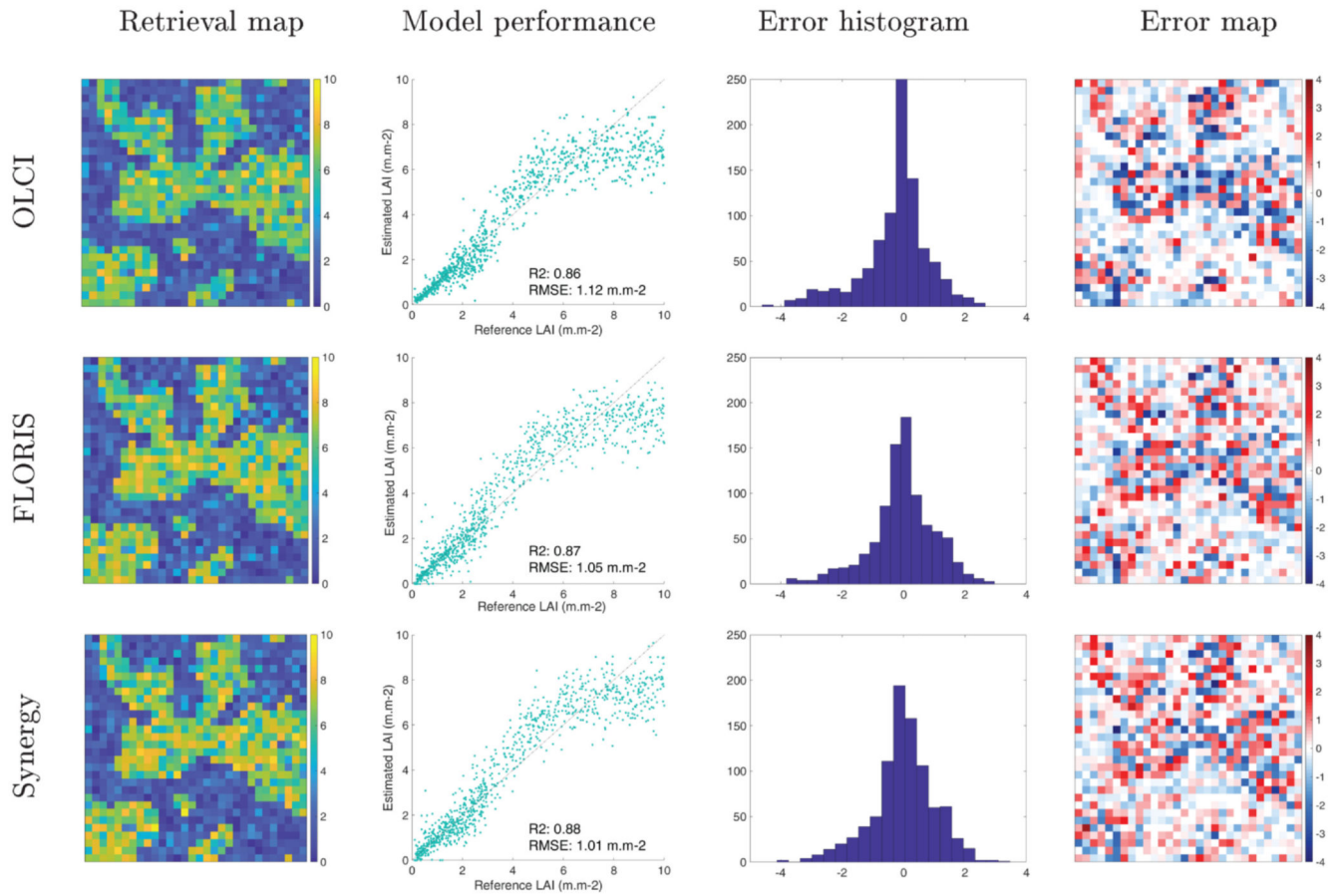
Fig. 7.

Validation of the Cab retrieval models. First line: model based on OLCI reflectance data; second line: model based on FLORIS reflectance data; third line: model based on the combined datasets of both sensors. The error is the difference between the estimated and the reference values for each pixel and is expressed in μg. cm^{-2}.
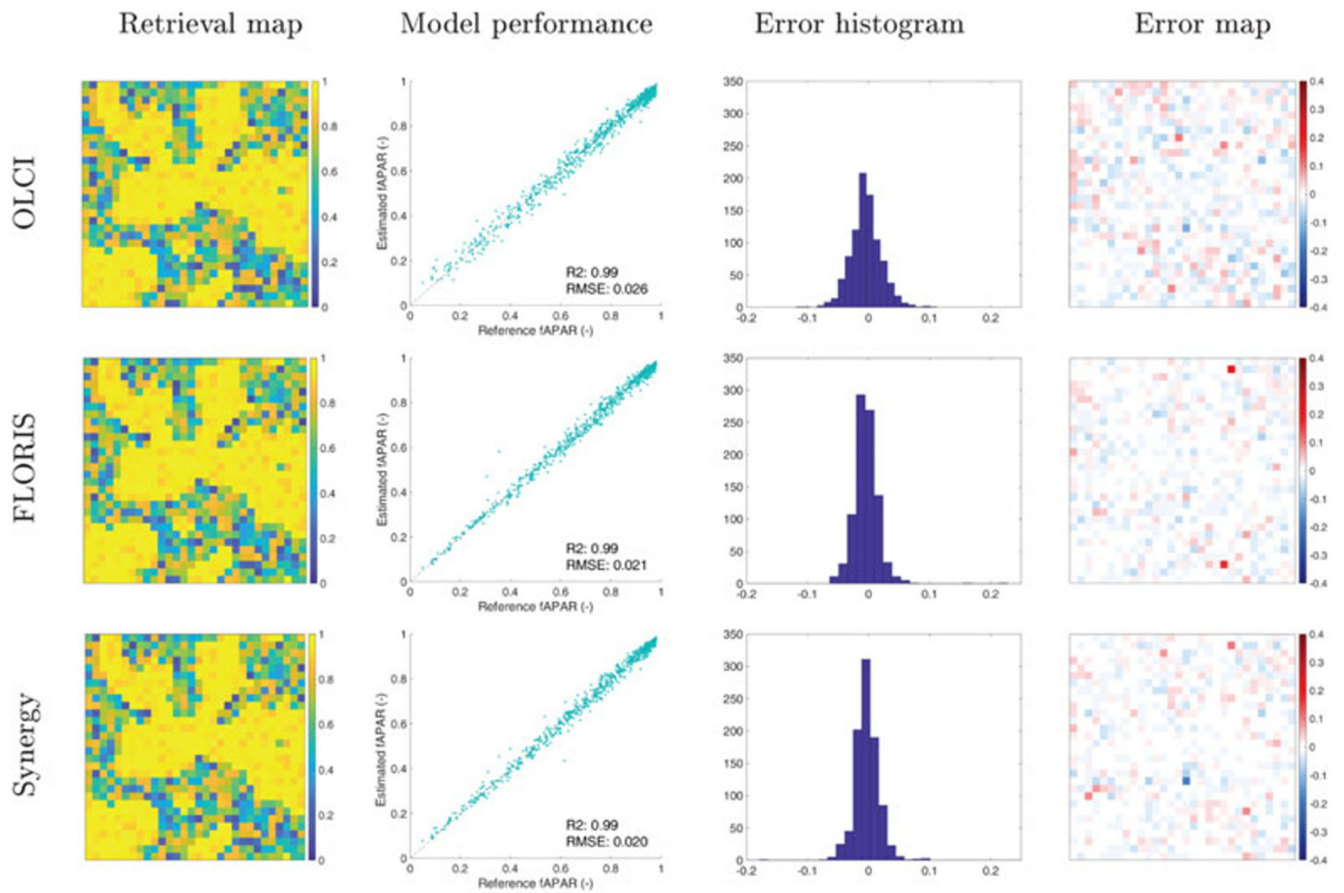
Fig. 8.
Validation of the LAI retrieval models. First line: model based on OLCI reflectance data; second line: model based on FLORIS reflectance data; third line: model based on the combined datasets of both sensors. The error is the difference between the estimated and the reference values for each pixel and is expressed in $m^2 \cdot m^{-2}$.

Fig. 9.
Validation of the fAPAR retrieval models. First line: model based on OLCI reflectance data; second line: model based on FLORIS reflectance data; third line: model based on the combined datasets of both sensors. The error is the difference between the estimated and the reference values for each pixel (units: -).
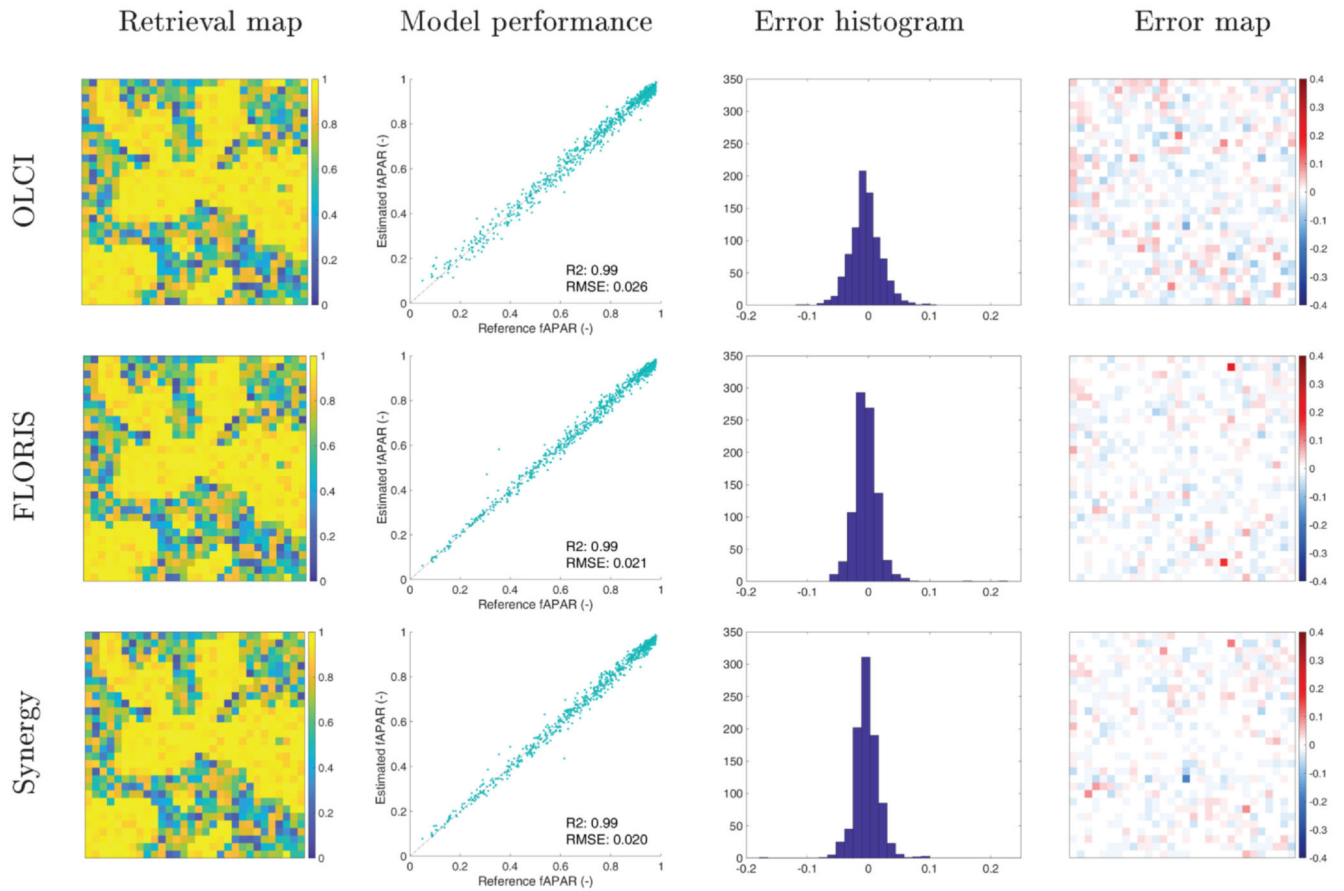
Fig. 10.

Validation of the FCover retrieval models. First line: model based on OLCI reflectance data; second line: model based on FLORIS reflectance data; third line: model based on the combined datasets of both sensors. The error is the difference between the estimated and the reference values for each pixel (units: -).
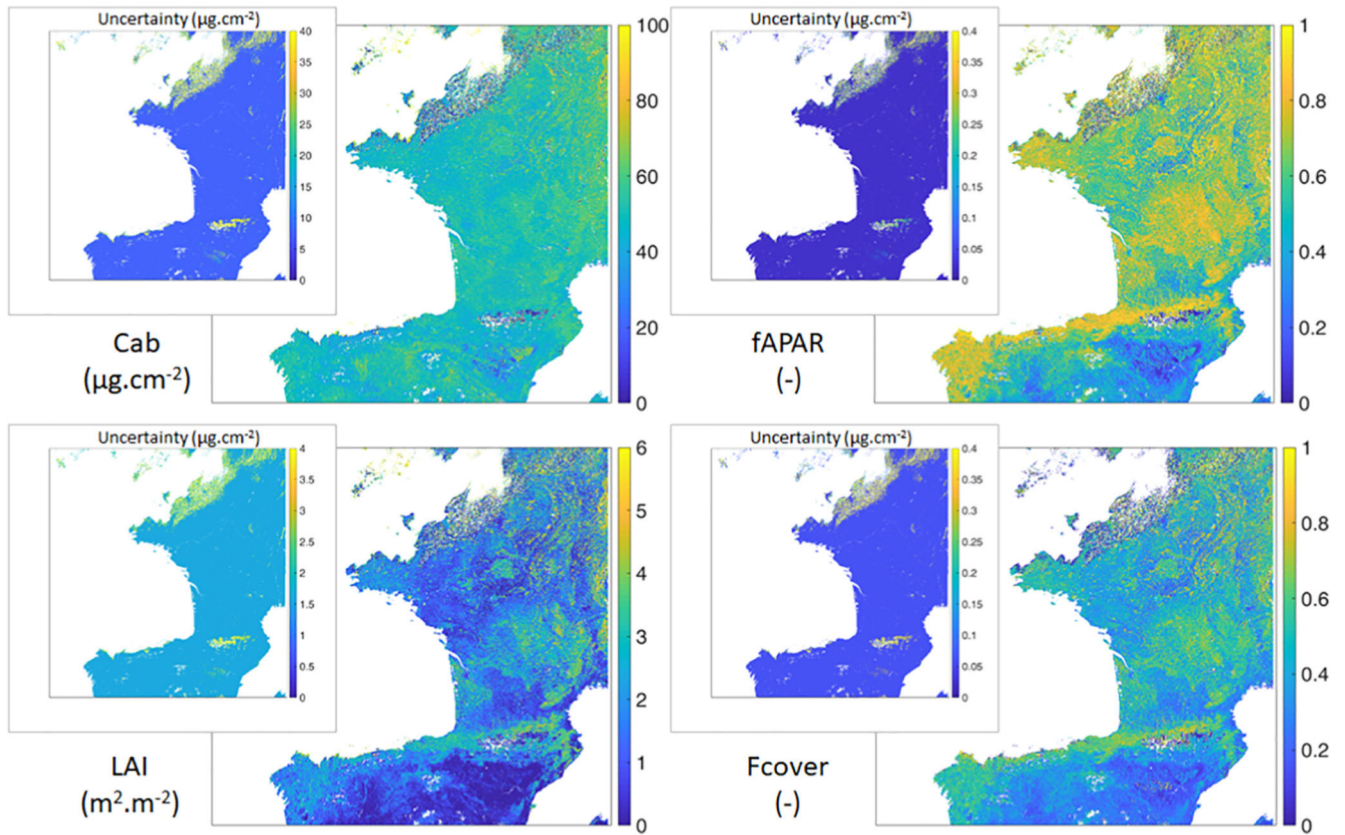
Fig. 11.
Retrieval maps for the four studied variables (larger maps) and associated uncertainties, approximated by the standard deviation around the mean estimate per pixel (smaller maps). These maps result from the application of the OLCI models on a OLCI SDR product of the 20th of June 2018 (see Fig. 4).
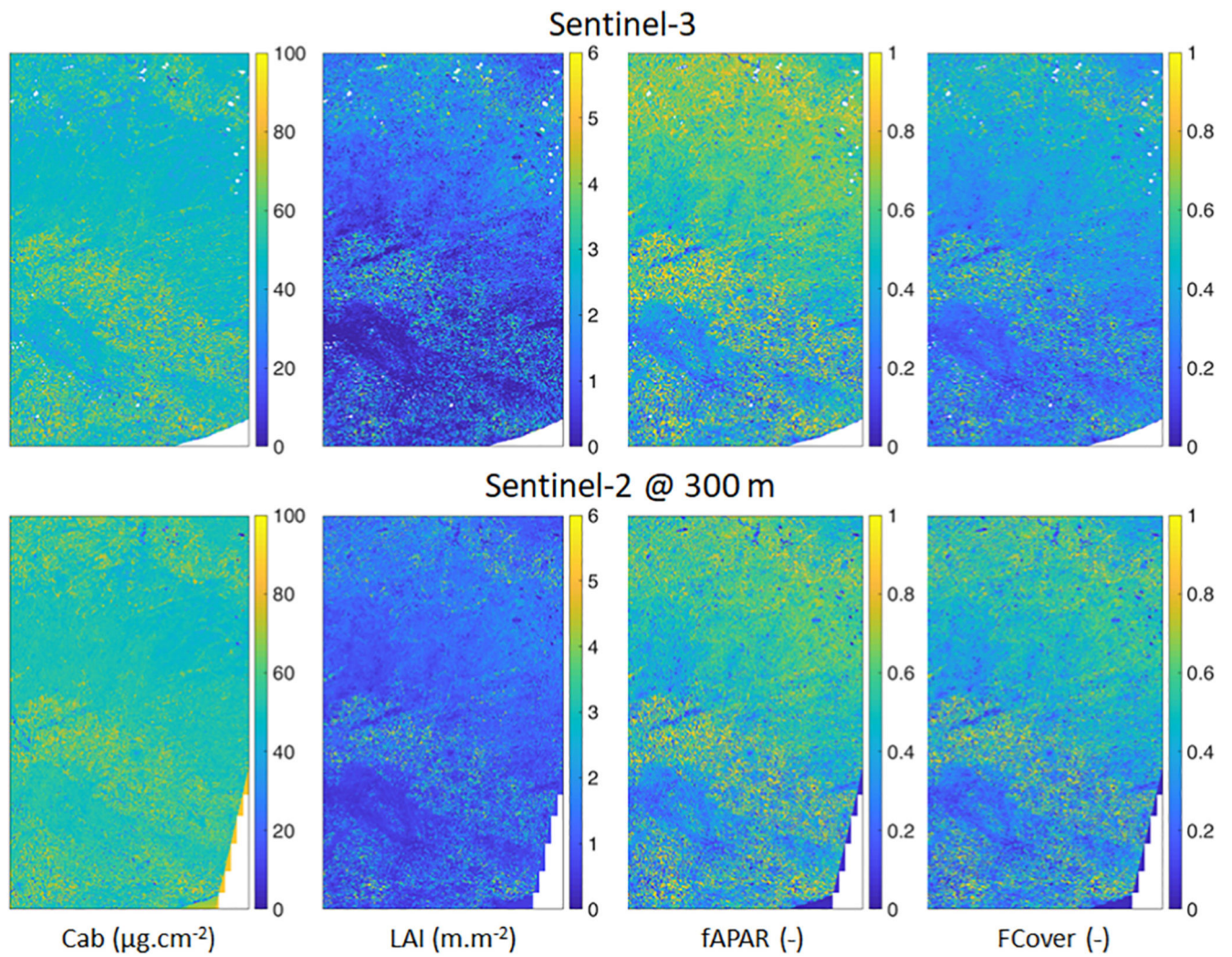
Fig. 12.
Comparison between the vegetation products (Cab, LAI, fAPAR and FCover) derived from
the developed models based on S3/OLCI reflectances and the corresponding S2 products,
which were rescaled to a 300-m spatial resolution (Argentina data set; see Fig. 4).
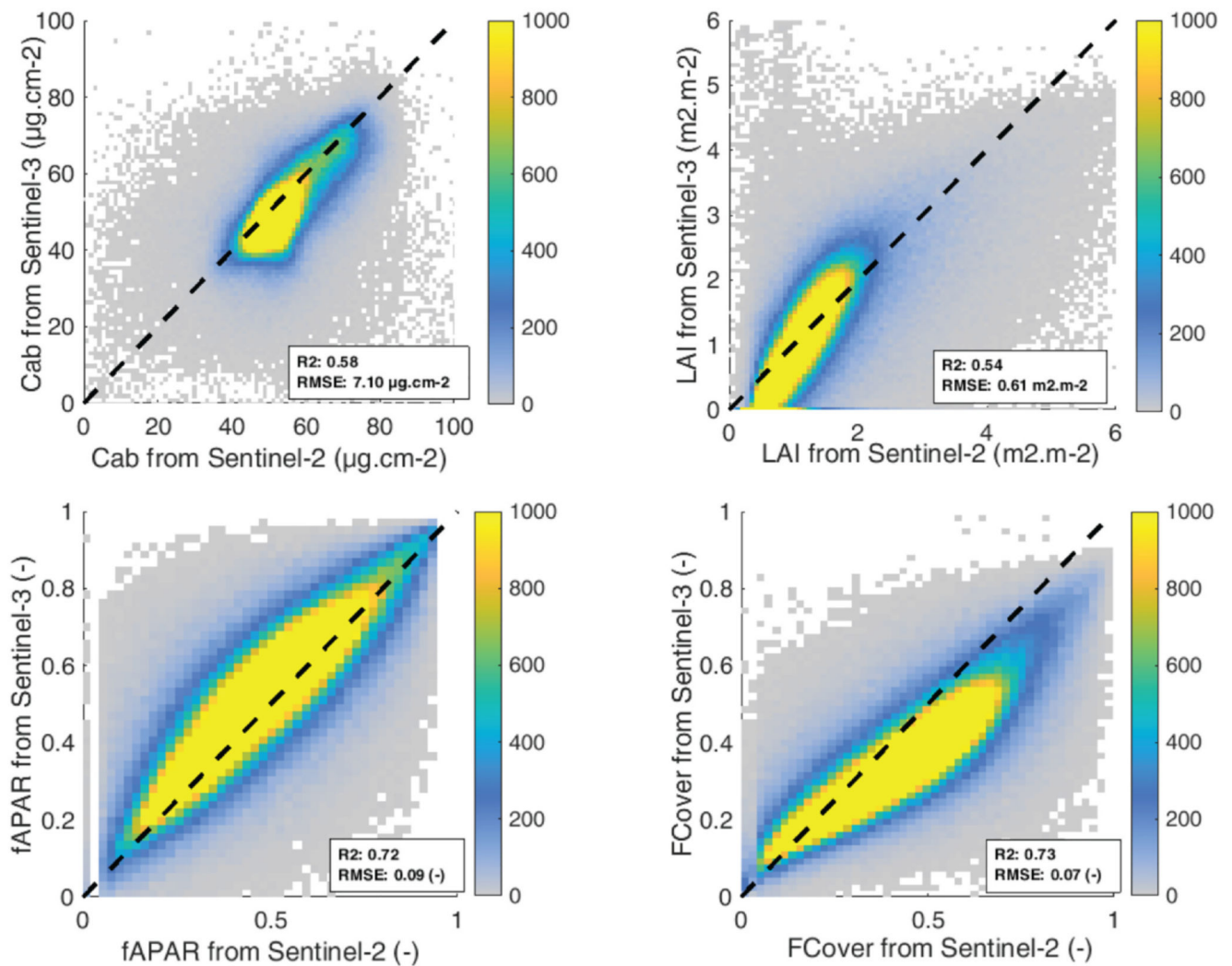
Fig. 13.
Scatterplots between the vegetation products (Cab, LAI, fAPAR and FCover) derived from the developed models based on S3/OLCI reflectances and the corresponding S2 products (Argentina data set, see Fig. 4). The colour legend specifies the number of bins (pixel counts), a higher number meaning a higher density of data points.

Table 1

The used ranges for the input variables of the SCOPE model. N: leaf mesophyll structure; Cab: leaf chlorophyll content; Cca: leaf carotenoid content; Cdm: leaf dry matter content; Cw: leaf water content; LAI: Leaf Area Index; LIDFa & LIDFb: parameters characterizing the leaf inclination distribution function (LIDF); SMC: Soil Moisture Content; BSM Bright, BSM lat and BSM lon are the three parameters characterizing the BSM (Brightness - Shape - Moisture) spectral soil model (for more information on the model, see (Verhoef et al., 2018)); SZA: Solar Zenith Angle; OZA: Observer Zenith Angle; RAA: Relative Azimuth Angle. * truncated Gaussian; ** constraint: Cw/(Cw + Cdm) between 0.45 and 0.93; *** constraint: |LIDFa| + |LIDFb|

  1.

Variable type	Variable	Distribution	Min	Max	Mean	SD
	N	Gaussian*	1	2.7	1.5	0.5
	Cab ($\mu g.\ cm^{-2}$)	Uniform	1	100		
Leaf structure	Cca ($\mu g.\ cm^{-2}$)	Gaussian*	0	30	10	5
	Cdm (g. cm^{-2})**	Gaussian*	0.002	0.02	0.005	0.003
	Cw (g. cm^{-2})**	Gaussian*	0.005	0.035	0.012	0.006
	LAI ($m^2.\ m^{-2}$)	Uniform	0.1	10		
Canopy structure	LIDFa ***	Uniform	−1	1		
	LIDFb ***	Uniform	−1	1		
	SMC (%)	Gaussian*	5	55	25	12.5
	BSM Brightness	Gaussian*	0.01	0.9	0.5	0.25
Soil	BSM lat (°)	Gaussian*	20	40	25	12.5
	BSM long (°)	Gaussian*	45	65	50	10
	SZA (°)	Uniform	0	80		
Geometry	OZA (°)	Uniform	0	25		
	RAA (°)	Uniform	0	180		

Table 2

Selected values for the four factors of the model (see Eq. (1)) used to inject Gaussian white noise into the simulated reflectance spectra of OLCI and FLORIS. MD: Multiplicative wavelength Dependent factor, MI: Multiplicative wavelength Independent factor, AD: Additive wavelength Dependent factor, AI: Additive wavelength Independent factor.

MI (%)	AI	MD (%)	AD
1	0.001	1 (for all bands)	Bands < 560 nm: 0 560 nm Bands < 709 nm: 0.001 709 nm Bands < 754 nm: 0.005 Bands 754 nm: 0.02

Table 3

Cab and LAI ranges for the two vegetation classes composing the simulated reference scenes, which were used for the validation of the retrieval models. The other SCOPE input variables were given the same ranges as for the training database generation (see Section 2.3.1) and are identical for both vegetation classes.

	Class 1	Class 2
Cab (μg cm $^{-2}$)	0–60	15–100
LAI (m^2 m^{-2})	0–3	3–10