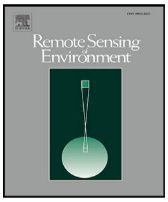
Remote Sensing of Environment 328 (2025) 114865

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse



On the potential of principal component analysis for the reconstruction of full-spectrum SIF emission and emulated airborne-to-satellite upscaling

Miguel Morata [a],*, Bastian Siegmann [b], José Luis García-Soria [a], Juan Pablo Rivera-Caicedo [c], Jochem Verrelst [a]

[a] IPL - University of Valencia, Catedrático Agustín Scardino Benlloch 9, Paterna, 46980, Spain
[b] Institute of Bio- and Geosciences, Plant Sciences (IBG-2), Forschungszentrum Jülich GmbH, 52428, Germany
[c] SECIHTI-SIP, Universidad Autónoma de Nayarit, Tepic, 63155, Mexico

ARTICLE INFO

Edited by Jing M. Chen

Keywords:
Sun-Induced Fluorescence (SIF)
Full-spectrum SIF
Principal Component Analysis (PCA)
Emulation
Spectral Fitting Method (SFM)
Hyperspectral
HyPlant
PRecursore IperSpectrale de la Missione Applicativa (PRISMA)
FLuorescence EXplorer (FLEX)

ABSTRACT

Solar-induced fluorescence (SIF) emitted by plants as a byproduct of photosynthesis provides critical insights into vegetation health and climate regulation. However, detecting the weak SIF signal from small telluric oxygen absorption features remains challenging. ESA's upcoming Fluorescence Explorer (FLEX) mission will retrieve full-spectrum SIF data at 300 m spatial resolution. In the meantime, we propose an alternative approach to reconstruct full-spectrum SIF from O_2A and O_2B bands using Principal Component Analysis (PCA) and the Soil Canopy Observation, Photochemistry, and Energy fluxes (SCOPE) model. Based on 100,000 SCOPE simulations (640–850 nm at 1 nm resolution), the SIF signals in the O_2A (760 nm) and O_2B (687 nm) bands showed high correlations with adjacent spectral regions and the full spectrum (R^2 > 0.89). From this data, we derived linear regression functions linking SIF at the O_2A (760 nm) and O_2B (687 nm) bands to the first two principal components (PCs), enabling inverse PCA transformation to reconstruct full-spectrum SIF with R^2 > 0.98 and RMSE < 0.12 mW m^{-2} nm^{-1} sr^{-1}. Applying the functions to HyPlant airborne SIF maps (1.7 m resolution) in northeastern Spain, and subsequent PC transformation, successfully reconstructed full-spectrum SIF including peaks and total emitted flux (SIF_{Tot}) with propagated uncertainties. To transfer this airborne full-spectrum SIF data to the satellite scale, we then trained an emulator with PRecursore IperSpectrale de la Missione Applicativa (PRISMA) Bottom of Atmosphere (BOA) reflectance spectra as input to produce spaceborne synthetic full-spectral SIF maps at 30 m resolution, and resampled to the nominal 300 m FLEX resolution. Despite PRISMA's lower spectral resolution for SIF retrieval, the emulator reliably produced full-spectrum SIF with R^2 of 0.69 and 0.52 for SIF_{760} and SIF_{687} bands, enabling FLEX-like SIF products (e.g., peaks, SIF_{Tot}). This reconstruction and upscaling approach demonstrates its utility for generating FLEX-compatible SIF datasets, supporting FLEX' mission preparation and cal/val activities.

1. Introduction

Photosynthesis sustains life on Earth and plays a vital role in regulating the planet's climate. As a byproduct of photosynthesis, the emission of solar-induced fluorescence (SIF) is a plant strategy developed to release excessive energy (Porcar-Castell et al., 2014; Mohammed et al., 2019; Yang et al., 2021). SIF has become crucial for studying vegetation photosynthesis and its response to environmental stressors. It is an important indicator of vegetation photosynthesis and can provide insight into plant productivity and health (Celesti et al., 2018; Mohammed et al., 2019). Space-based quantification of the full-spectrum SIF emission signal at a high spatial resolution of 300 m as planned with the future FLuorescence EXplorer (FLEX) satellite mission thus provides a direct linkage to the canopy's photosynthetic activity, opening insights into terrestrial vegetation health and productivity (Drusch et al., 2017). However, the SIF emission signal is relatively weak compared to the reflected radiance of vegetation, making it challenging to detect and separate it from background noise (Sabater et al., 2018). For this reason, retrieving the SIF emission spectrum typically relies on methods that only allow retrieving the SIF signal in narrow spectral absorption features in which almost no light is reflected (Cogliati et al., 2015; Köhler et al., 2015; Liu et al., 2015; Cendrero-Mateo et al., 2019). The most widely used SIF retrieval methods are briefly outlined below.

The Fraunhofer Line Discriminator (FLD) method (Plascyk, 1975) uses Fraunhofer lines for SIF estimation, which include telluric oxygen

---

* Corresponding author.
E-mail address: miguel.morata@uv.es (M. Morata).

https://doi.org/10.1016/j.rse.2025.114865
Received 13 June 2024; Received in revised form 30 May 2025; Accepted 5 June 2025

absorption lines among other spectral features present in the solar spectrum. In these absorption lines, SIF accounts for a relatively larger portion of the total upwelling radiance of the canopy (Theisen, 2002). The FLD method is based on the assumption that canopy reflectance and SIF remain constant within and outside the absorption line under consideration. Yet, this is an unrealistic assumption, which results in the retrieval of less accurate values of SIF. One of the most commonly applied approaches to retrieve SIF in the O_2A and O_2B absorption features is the Spectral Fitting Method (SFM) (Cogliati et al., 2015, 2019). In contrast with FLD, SFM assumes that the canopy reflectance and SIF can be described by smooth mathematical functions around the absorption line (see a detailed description in Supplementary Material). For instance, the SFM has been applied to airborne HyPlant imagery to study the SIF response of various vegetation types, including crops, forests, and grasslands, and can provide valuable insights into vegetation photosynthesis and stress (Meroni et al., 2009; Bandopadhyay et al., 2019; Rascher et al., 2022). Going beyond the standard SFM method, over the years several alternative spectral fitting methods have been proposed, and some of them provide the possibility to reconstruct the full SIF emission spectrum (640–850 nm) from top-of-canopy (TOC) reflected radiance measurements (Zhao et al., 2018). These methods involve the Fluorescence Spectrum Reconstruction (FSR) method (Zhao et al., 2014), the Full-spectrum Spectral Fitting Method (F-SFM) (Liu et al., 2015), an advanced spectral fitting method called SpecFit (Cogliati et al., 2015), the advanced Fluorescence Spectrum Reconstruction (aFSR) method (Zhao et al., 2018) and the Fluorescence Inference (SIFFI) (Kukkurainen et al., 2025). Despite the merits and limitations of the above methods, most of them rely on TOC radiance data. While Zhao et al. (2024) demonstrated the potential for full-spectrum SIF reconstruction using satellite reflectance data, such as TROPOMI, challenges related to atmospheric correction and spectral resolution remain significant barriers to widespread application (Cendrero-Mateo et al., 2019). With the launch of the FLEX mission on the horizon in 2026, there is a pressing need for full-spectrum SIF reconstruction methods tailored to imaging spectrometer satellite missions. FLEX will carry the high-resolution Fluorescence Imaging Spectrometer (FLORIS), which will acquire data in the 500–780 nm spectral range. FLORIS will offer an ultrahigh spectral sampling of 0.1 nm in the oxygen absorption bands (759-–769 nm and 686-–697 nm), and 0.5—2.0 nm in the red edge, chlorophyll absorption, and Photochemical Reflectance Index (PRI) bands (Drusch et al., 2017). FLEX aims for precise and physically rigorous SIF retrieval within the oxygen absorption bands using spectral high-resolution data. In anticipation of its launch, alternative solutions are needed that can interpret coarser-resolution spectra as recorded from currently operational imaging spectrometer missions and translate them into full-spectrum SIF estimates. Accordingly, by developing a workflow that can generate surrogate FLEX-like full-spectrum SIF data, we can pave the way for prototyping and interpreting advanced satellite-based photosynthesis products. This preparatory work ensures that sound processing pipelines are ready for FLEX immediately post-launch.

A fleet of new-generation imaging spectroscopy missions with high spatial resolution (i.e., 30 m) has recently been designed and partly launched, possessing hundreds of spectral bands in the visible and near-infrared (VNIR) regions of the electromagnetic spectrum that provide a vast amount of contiguous spectral data. These missions include the already launched PRecursore IperSpectrale de la Missione Applicativa (PRISMA) (Loizzo et al., 2019) and the Environmental Mapping and Analysis Program (EnMAP) (Guanter et al., 2015) missions, and the upcoming European Copernicus Hyperspectral Imaging Mission for the Environment (CHIME) (Celesti et al., 2022) and NASA's Surface Biology and Geology (SBG) (National Academies of Sciences and Medicine, 2018) mission. In addition, from low Earth orbit, the German Earth-sensing imaging spectrometer DESIS (Krutz et al., 2019) and the EMIT imaging spectrometer developed by NASA-JPL (Green and Thompson, 2020), both on board the International Space Station, deliver hyperspectral image data of the Earth's surface.

Launched in March 2019, the PRISMA mission paved the way for the rapid distribution and utilization of hyperspectral satellite imagery within the scientific community. PRISMA's spectral range overlaps with FLEX's target regions and covers the spectral region where SIF is emitted (i.e., 640–850 nm), making it a potentially powerful data source for synthesizing FLEX-like SIF products at 30 m resolution. Thanks to its high spectral resolution, PRISMA has been successfully exploited for the retrieval of numerous vegetation traits (e.g., Veraverbeke et al., 2018; Verrelst et al., 2021). The question now arising is whether PRISMA data can be exploited as a basis for reproducing the full-spectrum SIF signal. Since PRISMA has not been designed to retrieve SIF, given the sensor's relatively broad bandwidth (≤12 nm), it requires the development of alternative methods capable of approximating SIF signals that can handle lower-resolution spectral image data. One possibility to tackle this problem and derive SIF from PRISMA is the combination of physically-based vegetation radiative transfer models (RTMs) providing SIF as output with the flexibility of competitive machine learning algorithms. Moving along this line, earlier experimental studies suggested that strong and reliable links between two sources of spectral data can be established through statistical learning, i.e., *emulation* (Verrelst et al., 2019; Morata et al., 2021).

The concept of emulation involves approximating input–output relationships using a statistical learning model, also known as an *emulator* or a surrogate model (Verrelst et al., 2019). An emulator is essentially a trained machine learning regression algorithm (MLRA) that uses a dataset consisting of input–output pairs for model training. Once an accurate emulator is developed, it can approximate original models, such as advanced RTMs, at a much faster speed. Thus, the emulator can infer statistical relationships between pairs of spectral data and process them at a low computational cost. For example, the emulation technique can reconstruct a robust statistical relationship between PRISMA's bottom-of-atmosphere (BOA) reflectance spectra and full-spectrum SIF data. This emulation approach converts PRISMA reflectance into full-spectrum SIF and can thus generate synthetic datasets that replicate FLEX's anticipated spectral and spatial resolution. These surrogate products are critical for pre-launch algorithm validation and mission planning. However, the key challenge in emulating hyperspectral data lies in simultaneously predicting a vast number of spectral bands. Fortunately, not all bands carry unique information; there is usually strong collinearity present in spectral data. Capitalizing on this redundancy, known as the Hughes phenomenon, allows us to condense the data into a lower-dimensional space using dimensionality reduction (DR) techniques (Hughes, 1968). Among the most straightforward DR methods is Principal Component Analysis (PCA) (Wold et al., 1987). As PCA applies linear transformations, it goes along with the advantage that the principal components (PCs) can be transposed again backward into the original data. Consequently, applying PCA to spectral data followed by the transposition of its components back to the full spectrum has proven effective for reconstructing hyperspectral data. Combining this approach with a statistical model that predicts PCA components enables the rapid emulation of RTM-like spectral output data, as demonstrated in recent emulation studies (Vicent et al., 2018; Morata et al., 2021, 2023). Progressing along this line, a potentially interesting application of this PCA reconstruction technique lies in the fact that the SIF-spectrum signal is a spectrally smooth double-peaked curve with peaks around 685 nm and 740 nm. This means it can be easily compressed into a few PCA components without losing key information. Consequently, PCA reduction and subsequent transformation can similarly serve as an attractive method to reconstruct the full-spectrum SIF signal. This approach opens opportunities to develop a robust linkage between SIF retrieval and subsequent PCA components to enable the reconstruction of the full-spectrum SIF signal. To achieve this, SIF simulations generated by a vegetation RTM offer a potential data source.

Altogether, given this outlined framework, this study aims to develop a PCA-based model for reconstructing the full spectral SIF response from airborne-retrieved SIF data in the two oxygen absorption

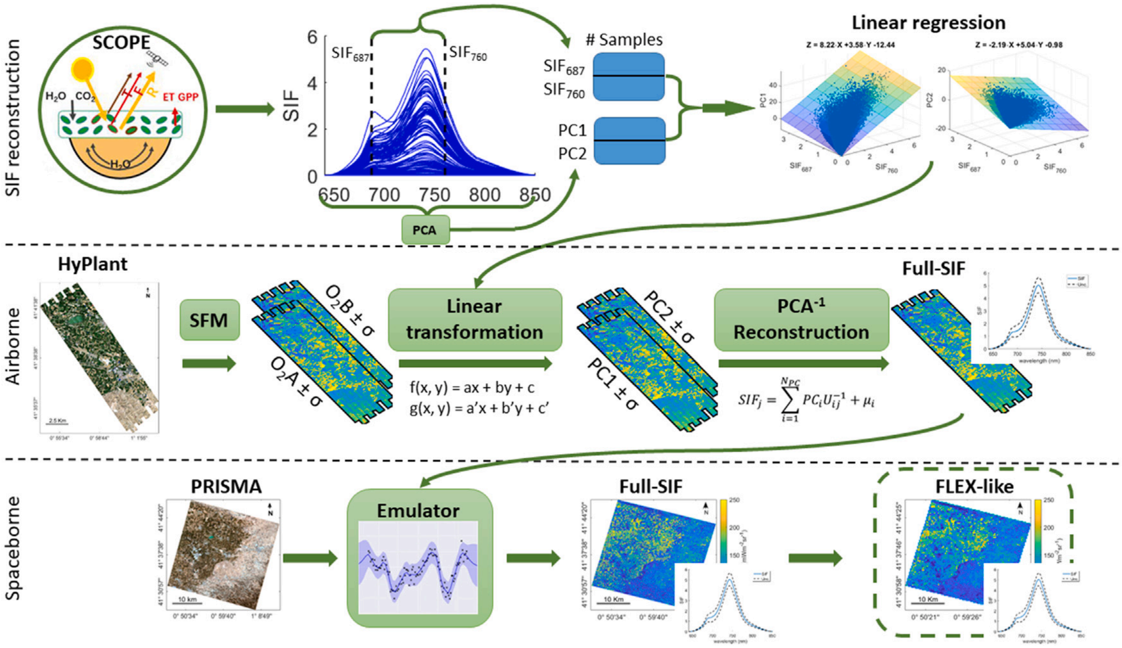


Fig. 1. Graphical representation of the proposed full-SIF reconstruction and upscaling workflow.

bands that can be upscaled to the satellite level. To achieve this main goal, we divided the presented approach into four steps associated with specific objectives: (1) to evaluate the role of simulated SIF data to enable a full-spectrum SIF emission spectrum reconstruction from HyPlant airborne data through PCA transformation; (2) to propagate SIF uncertainty derived from the two HyPlant oxygen absorption bands through the full-spectrum SIF reconstruction procedure; (3) to validate the full-spectrum SIF signal reconstruction at the airborne scale against reference data measured on the ground; and, (4) to develop an emulator that enables the upscaling of the full-spectrum SIF model from the airborne to the satellite level to finally produce SIF data with associated uncertainties from PRISMA image data.

Tackling these four objectives will eventually lead to a novel and easily applicable workflow that enables the emulation of the full-spectrum SIF signal from PRISMA space-borne data, which can serve to prototype SIF-related products derived from FLEX data.

2. Material & methods

The workflow, describing the reconstruction of the full SIF emission spectrum based on hyperspectral satellite data, is outlined in Fig. 1. A detailed description of the key steps of the workflow is provided in the following sections: (1) used simulated and experimental data; (2) principle of PCA transformation to reconstruct the full-spectrum SIF emission; and (3) upscaling to satellite data by using the emulation technique. Additionally, an uncertainty propagation scheme is explained, that consists of two parts: (i) uncertainty propagation through the PCA transformation, and (ii) uncertainty estimation through the emulation process. Integrating all steps, this novel and largely automated workflow enables the generation of realistic FLEX-like full-spectrum SIF image data.

2.1. SCOPE simulations

To achieve the main objectives, we exploited a combination of data sources and techniques to realize satellite-based full-spectrum SIF emission reconstruction. The first step consists of the generation of a synthetic dataset of full-spectrum SIF using the RTM Soil Canopy Observation, Photochemistry, and Energy fluxes (SCOPE) (Van der Tol et al., 2014).

SCOPE is a powerful RTM that simulates the radiative and non-radiative energy fluxes in the soil-vegetation-atmosphere continuum. While the SCOPE sub-module RTMo describes the radiative transfer of solar radiation within the leaf and canopy, the sub-module RTMf simulates the SIF emission of vegetation at the photosystem, leaf, and canopy scale. Together with the Brightness–Shape–Moisture (BSM) soil reflectance model, SCOPE forms a modular framework of multiple RTMs that allows the simulation of optical and SIF data of vegetation (Van der Tol et al., 2009, 2014). A range of inputs are needed to run SCOPE, such as information about leaf properties, photochemistry, canopy structure, sun-observer illumination geometry, and soil properties, to simulate the radiative transfer within vegetation canopies. Within this study, the ability of SCOPE to provide the full double-peak SIF spectrum spanning from 640 to 850 nm (211 spectral bands) with a spectral resolution of 1 nm for each combination of input variables is of vital importance to derive the full-spectrum SIF emission curve from airborne and satellite data.

With the purpose of exploiting the correlation and thus the redundancy in the double-peak SIF curve between all spectral bands, we used SCOPE to create a spectral database consisting of 100,000 simulated SIF spectral profiles. This was realized by varying the input variables within predefined ranges. These variables were carefully selected based on findings of a global sensitivity analysis (GSA) study in which SCOPE key variables were identified that drive the SIF emission spectrum at canopy scale (Verrelst et al., 2015). Based on the results of the GSA, 16 input variables required to run the weather, leaf, biochemistry, soil, canopy, and geometry sub-models of SCOPE were identified and varied to generate the SIF spectral database. The specific boundaries of each variable are displayed in Table 1. The variables are used to have a wide representation of the vegetation properties. By introducing broad variability across these variables, it is assumed that a wide variety of real-world leaf and canopy representations are covered. We employed a Latin Hypercube Sampling (LHS) strategy to sample the input variable space within the specified variable boundaries. Using LHS offers distinct advantages over systematic gridded sampling because the LHS strategy allows for a comprehensive representation of the input parameter space, capturing its full variability (Mckay et al., 1979). Therefore, LHS sampling ensures the generation of a SCOPE-simulated SIF database that represents the full SIF variability, presumably as it is found in nature.

Table 1
Range of SCOPE input variables.

Type	Biophysical parameter	Min.	Max.
Weather	Incoming shortwave radiation (W/m^2)	0	1000
Weather	Air temperature (°C)	0	40
Weather	Atmospheric vapor pressure (hPa)	0	100
Weather	CO_2 concentration in the air (ppm)	100	800
Leaf	Chlorophyll (μg/cm^2)	0	50
Leaf	Dry matter (g/cm^2)	0	0.05
Leaf	Fraction of senescent material	0	0.3
Biochemistry	Maximum carboxylation capacity (μmol m^{-1} s^{-1})	0	150
Biochemistry	Ball-Berry stomatal conductance	5	15
Soil	BSM model parameter for soil brightness	0	0.9
Canopy	Leaf area index (m^2/m^2)	0	7
Canopy	Vegetation height (m)	0.05	100
Canopy	Leaf width (m)	0.01	2
Canopy	Leaf inclination distribution factor a	−1	1
Canopy	Leaf inclination distribution factor b	−1	1
Geometry	Solar zenith angle (°)	20	60

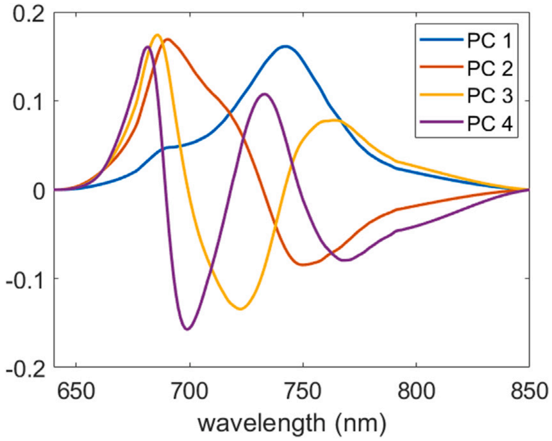


Fig. 2. The first four principal components (PCs) are represented by their respective spectra. The PCs exhibit distinct spectral characteristics, including the overall SIF level (PC1), the difference between the peaks (PC2), and the depth of the valley between the peaks (PC3).

2.2. Principal component analysis (PCA)

This work builds upon PCA at two critical stages within the entire workflow (info on PCA principles in Supplementary Material). First, PCA is applied in the reconstruction procedure of the full-spectrum SIF emission. We used PCA to reduce the dimensionality of the full-spectrum SIF and thus obtain a linear transformation between SIF_{760} and SIF_{687} with the first PCs. The four most significant eigenvector spectra are illustrated in Fig. 2. These vectors are particularly well-suited for use as basis functions for the generation of SIF spectra, with the application of weights in a linear combination. The shape of these functions indicates that the first PC is similar to that of the SIF spectrum, thereby emphasizing total fluorescence. The second PC represents the differences between the two peaks, while the third PC represents the valley's depth between the two peaks. These findings are in alignment with those previously reported by Zhao et al. (2014), Verhoef et al. (2018).

Through the employment of a limited number of PCs, we can reconstruct the 100,000 simulated SIF spectral profiles. Fig. 3 illustrates the residuals of the spectra obtained by only one to four PCs by plotting the percentiles ranging from 2 to 98% over the full spectral range. Specifically, in the case of 2 PCs, the percentiles 25% and 75% (red dashed lines) are < 0.03mW m^{-2}nm^{-1}sr^{-1} and the percentiles 2% and 98% (colored areas) are < 0.2mW m^{-2}nm^{-1}sr^{-1} for the entire spectral range, which is below the expected FLEX SIF retrieval error of 0.2 mW m^{-2} nm^{-1} sr^{-1}. The maximum and minimum values (black dashed lines) are in line with those obtained by Verhoef et al. (2018).

Second, PCA is also an essential step in the emulation procedure to transfer the full-spectrum SIF emission from the airborne to the satellite scale. After a description of the used data in Section 2.3, the role of PCA in the reconstruction of the full-spectrum SIF signal is detailed in Section 2.4 and the error propagation is described in Section 2.5, while the role of PCA in emulation is outlined in Section 2.6.

2.3. Image data

The SCOPE-simulated 640–850 nm SIF dataset formed the basis for reconstructing the full-spectrum SIF emission from experimental HyPlant SIF data derived from the O_2A and O_2B oxygen absorption regions located at the far-red 760 (SIF_{760}) and red 687 nm (SIF_{687}), respectively. HyPlant and corresponding PRISMA imagery were acquired during an intensive field campaign in the frame of the LIAISE (Land surface Interactions with the Atmosphere over the Iberian Semi-arid Environment) project in the northeastern part of Spain, covering the area around the city of Lleida on 17th July 2021. Additionally, we used calibration coefficients determined for SIF_{760} and SIF_{687} derived by comparing HyPlant and FloX data recorded over Selhausen area in western Germany for SIF validation purposes during the FLEXSense campaign on 26th and 27th June 2018. Then, we applied them to the HyPlant LIAISE SIF data recorded in Spain. The reason for applying the coefficients derived from the FLEXSense campaign data to the HyPlant data recorded in Spain was that the German HyPlant data set was validated on a wide variety of crops. The used datasets are briefly outlined below.

2.3.1. HyPlant. LIAISE campaign, Spain

The airborne and spaceborne hyperspectral imagery recorded during the LIAISE campaign were exploited to develop the full-SIF reconstruction workflow. The nine HyPlant flight lines and the PRISMA scene were acquired with a time offset of 28 to 80 min for the first and last HyPlant flight lines, respectively. The airborne and spaceborne datasets are illustrated in Fig. 4. The HyPlant airborne sensor system consists of the two modules DUAL and FLUO (Siegmann et al., 2019). The FLUO module has been specially designed to measure SIF. It covers the spectral range from 669.5 to 781.9 nm with a spectral sampling interval (SSI) of 0.11 nm and an FWHM of 0.31 nm, resulting in 1024 spectral bands (Alonso et al., 2015).

Nine adjacent flight lines covering an area of 81.79 km^2 were recorded by HyPlant's FLUO module. The single flight lines were acquired from 1150 m above ground level on 17th July 2021 between 13:27 and 14:19 local time (CEST). The airborne image data have a spatial resolution of 1.7 m, which were aggregated and resampled to 30 m to match the size and position of corresponding PRISMA pixels. Each of the nine flight lines has an overlap of 20% with its neighboring flight lines, which allowed the generation of a mosaiced single image. A modified version of the SFM containing an atmospheric correction as an additional step was applied to the HyPlant FLUO at-sensor radiance data to derive TOC SIF_{760} and SIF_{687} along with corresponding uncertainty intervals (Siegmann et al., 2019; Cogliati et al., 2019, 2015). It is important to note that while the SFM successfully retrieves SIF over vegetation, it can produce artifacts, including negative values, when applied to non-vegetated surfaces such as artificial structures, bare soil, or water bodies (Rascher et al., 2021, 2022).

2.3.2. PRISMA data

The PRISMA image provides hyperspectral information in 231 spectral bands covering the range from 402 to 2497 nm. The image has a spatial resolution of 30 m. The used PRISMA image was acquired at 12:58 local time (CEST) on 17th July 2021. It covers the area around the city of Lleida in the northeastern part of Spain (Fig. 4). We tested both PRISMA top-of-atmosphere (TOA) radiance (L1) and BOA reflectance data to evaluate the capability of the emulator to obtain a relationship between PRISMA spectra and reconstructed SIF.

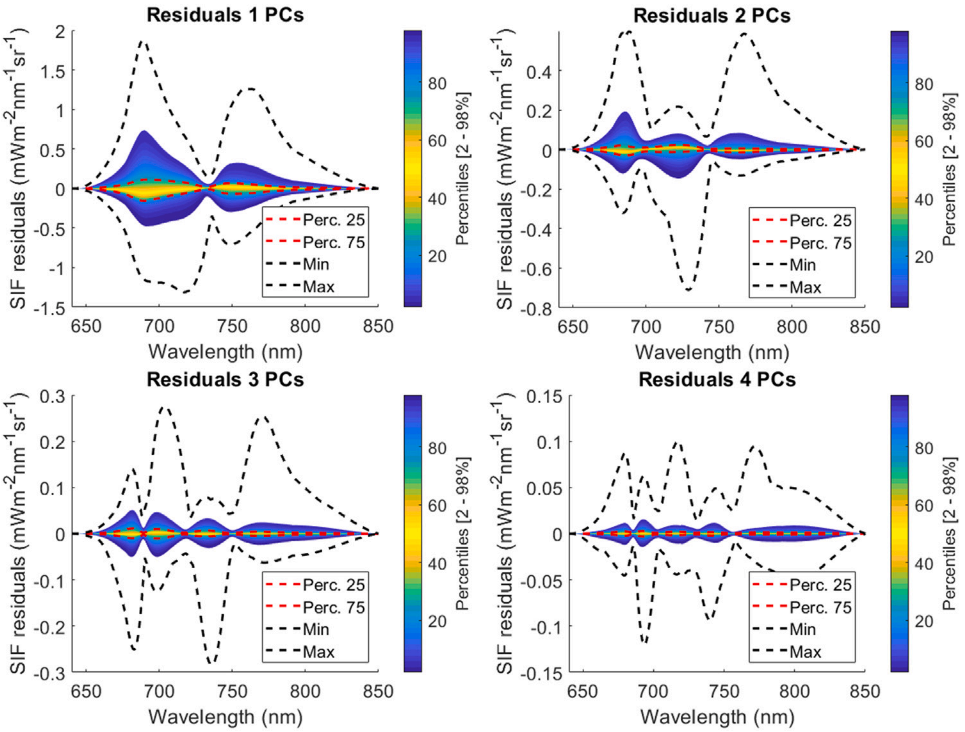


Fig. 3. The residuals resulting from approximating the original spectra with one to four principal components (PC). Note that the Y-axis SIF range shrinks as more PCs are added.

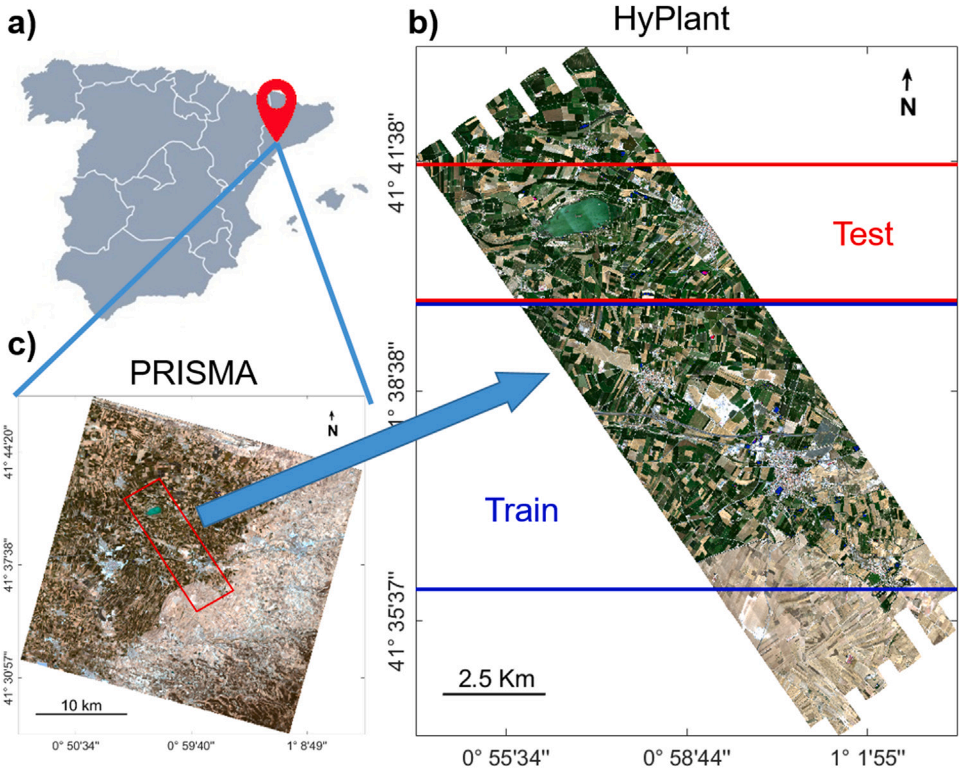


Fig. 4. Location of the study area (red marker) in the northeastern part of Spain (**a**). RGB composite (639.7/550.35/459.86 nm) of the HyPlant DUAL mosaic consisting of nine single flight lines (**b**). The HyPlant image was divided into a training (blue box) and a test subset (red box). PRISMA image of the same region overlaid with a red rectangle indicating the area covered by HyPlant (**c**). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Although PRISMA's spectral resolution is not ideal for retrieving SIF given bandwidths in the order of 10 nm, an emulator can establish a relationship between what is the most probable SIF spectrum for a given PRISMA reflectance spectrum (Pato et al., 2023).

2.3.3. *Calibration. FLEXSense campaign, Germany*

HyPlant SIF retrieval can slightly deviate from field measurements; for this reason, field SIF recordings acquired by the fluorescence box (FloX) were used to calibrate the airborne SIF data (Rascher et al., 2021). The FloX is a field instrument specifically designed for continuous in-situ monitoring of long and short-term fluctuations in SIF (Julitta et al., 2016). During the FLEXSense campaign in the Selhausen area (western Germany) that took place on June 26th and 27th 2018, SIF values derived from HyPlant data were validated against in-situ FLoX measurements. This SIF validation activity was employed on a variety of four different vegetation types: sugar beet, corn, lawn, and winter wheat. Strong relationships were obtained between the airborne and ground measurements for the SIF_{760} ($R^2 = 0.777$; $y = 1.24x - 0.37$) and SIF_{687} bands ($R^2 = 0.850$; $y = 1.3x + 0.01$), respectively. Details of the campaign results can be found in Rascher et al. (2022).

SIF_{687} strongly correlates with wavelengths lower than 710 nm. However, its retrieval from airborne sensors can be challenging (Qi et al., 2023). This discrepancy can be mitigated through calibration of the original HyPlant SIF-SFM product to obtain SIF values that are more realistic and conducive to the application of the full-spectrum SIF PCA technique. By applying the necessary calibration procedures, we aimed to finetune the HyPlant SIF product, aligning it more closely to ground-truth measurements and thus increasing its fidelity. For this reason, we applied empirical equations for SIF_{760} (y = 1.24x - 0.37) and SIF_{687} bands (y = 1.3x + 0.01) that are based on in-situ FloX and corresponding HyPlant SIF data from a former campaign in Germany. This allowed us to calibrate the HyPlant data acquired during the LIAISE campaign. Considering that the calibration equations have the form $x_{cal} = (x - n)/m$, the propagated error can be encoded using the error propagation formula for independent variables (Eq. (3)) as $\varepsilon_{cal} = \varepsilon/m$.

2.4. Full-spectrum SIF reconstruction technique

First, we created a large SCOPE dataset consisting of 100,000 distinct SIF spectra covering the spectral range from 640 to 850 nm in 211 bands (1 nm bands). We extracted the simulated SIF emission values at SIF_{760} and SIF_{687} from the full spectra. Additionally, owing to the high correlation between all the SIF spectral bands, we applied a PCA to reduce the simulated SIF spectra to 2 PCs that contain most of the original spectral information. Subsequently, we calculated a linear regression using SIF_{760} and SIF_{687} as independent variables and the PCs as the dependent variables and obtained a linear transformation for each PC. After determining the linear regression functions that convert the SIF_{760} and SIF_{687} emissions into PCs, we applied the same regression functions to the HyPlant SIF_{760} and SIF_{687} maps to transfer into the two PCs that contain most of the spectral variation. As the next step, an inverse PCA was conducted to reconstruct the full-spectrum SIF emission for each pixel from the PCs. Since PCA transformation represents a linear projection of the original data to other feature spaces and given that the entire process is a sequence of linear transformations, uncertainties can be analytically propagated throughout the process using the variance formula for uncertainty propagation for a multivariate function.

Executing the described approach, we can reconstruct the full-spectrum SIF emission at the nominal 1 nm spectral resolution of SCOPE from the two SIF products derived at 760 and 687 nm retrieved from HyPlant airborne data, together with the corresponding uncertainties for each waveband of the full-spectrum SIF emission. Having reconstructed the full-spectrum SIF signal, subsequently, key SIF products can be extracted, such as: (1) the maximum SIF values at the peak positions of the SIF spectrum (685 and 740 nm), and (2) the valley dip between the two curves at 720 nm. Moreover, (3) the integral under the emission curve, i.e., the total SIF emission flux, can also be calculated. These metrics are critical and will become products of the future FLEX mission (Drusch et al., 2017). To calculate the total SIF flux (SIF_{Tot}), we considered the 1 nm bands as infinitesimal pieces and obtained the sum of SIF values. The uncertainty was obtained as the square root of the sum of squared uncertainties for all bands (see next section). An overview of the entire full-spectrum SIF reconstruction workflow is presented in Fig. 5.

2.5. Uncertainty propagation of full-SIF reconstruction

Along with the workflow presented in the previous section, it is likewise possible to propagate pixel-wise the original SIF_{760} and SIF_{687} uncertainty intervals to the full-spectrum SIF emission map. Uncertainties provide vital information about the range of variation between the predicted and the real SIF value. The benefit of using linear transformations is that they allow consistent error propagation throughout the entire process. Since PCA is a linear transformation applied to the spectral feature space of the data used in this study, we can analytically propagate the original uncertainties through the PCA reconstruction procedure and assign them to each waveband and pixel in the final SIF product (Morata et al., 2023).

To derive the full-spectrum SIF signal, a linear transformation is initially carried out using the functions obtained through the linear regression. The PCs are derived using a two-dimensional linear transformation. Thus, the equation of the principal components adopts the structure presented in Eq. (1). The reconstructed full-spectrum SIF is then obtained by using the PCA matrix for the inverse projection by applying Eq. (2) to the PCs. For a better understanding, we define the used variables with the following notation: $A \equiv SIF_{760}$, $B \equiv SIF_{687}$ and $\mathbf{U}^{-1} \equiv PCA^{-1}$.

$$PC_i = a_i A + b_i B + c_i \tag{1}$$

$$SIF_j = \sum_{i=1}^{N_{PC}} PC_i \mathbf{U}_{ij}^{-1} + \mu_j \tag{2}$$

where PC_i represents the ith Principal Component, a_i, b_i and c_i represent the linear transformation coefficients for each PC_i obtained by the multi-variable linear regression. SIF_j is the jth band of the reconstructed SIF signal in the hyperspectral domain, N_{PC} is the maximum number of PCs used for the reconstruction, $\mathbf{U}_{ij}^{-1}$ is the inverse matrix of the PCA projection, and μ_j is the mean of the jth SIF value obtained by SCOPE.

To ensure consistent propagation of uncertainties throughout the SIF spectral range, we have applied the variance formula for uncertainty propagation for a multivariate function $Y(X_k)$ (3):

$$\varepsilon_Y = \sqrt{\sum_{k=1}^{N} \left(\frac{\partial Y}{\partial X_k}\right)^2 \varepsilon_{X_k}^2} \tag{3}$$

where ε_Y represents the uncertainty of the function $Y(X_k)$, X_k are the input variables, ε_{X_k} represents the uncertainty of kth input variable X_k and N is the number of variables.

Applying the error propagation (3) in the linear transformation of the PCs (1) and in the PCA model inversion (2), we deduced the uncertainty relations (4) and (5), respectively:

$$\varepsilon_{PC_i} = \sqrt{(A\varepsilon_{a_i})^2 + (a_i\varepsilon_A)^2 + (B\varepsilon_{b_i})^2 + (b_i\varepsilon_B)^2 + \varepsilon_{c_i}^2} \tag{4}$$

$$\varepsilon_{SIF_j} = \sqrt{\sum_{i=1}^{N_{PC}} \left(\mathbf{U}_{ij}^{-1}\varepsilon_{PC_i}\right)^2} \tag{5}$$

where ε_{PC_i} represents the uncertainty of the PCs used to reconstruct the full-SIF spectrum, ε_A and ε_B are the uncertainties determined for the retrieved SIF values SIF_{760} and SIF_{687}, obtained by the SFM process succeeding the HyPlant campaign (Rascher et al., 2022), and ε_{a_i}, ε_{b_i} and ε_{c_i} are the uncertainties of the linear transformation coefficients obtained from the linear regression models. ε_{SIF_j} represents the uncertainty of the SIF values in the reconstructed spectral domain.

2.6. Upscaling full-spectrum SIF from hyplant to PRISMA through emulation

Emulation is a statistical learning technique to approximate complex systems (e.g., costly RTMs) using a trained machine learning

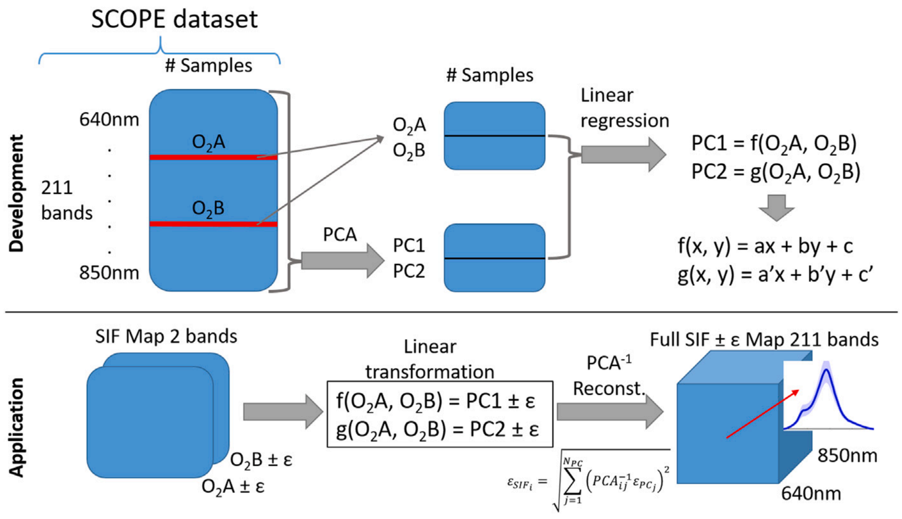


Fig. 5. Workflow showing the procedure of full-SIF reconstruction using PCA.

model for efficient output predictions (e.g., Verrelst et al., 2017, 2019; Morata et al., 2021). This is valuable for computationally expensive simulations, allowing faster analysis with approximate responses. Here, we explore this technique to model nonlinear relationships between multi-dimensional input spectral data (e.g., reflectance) and desired multi-dimensional output spectral data (e.g., SIF). Specifically, we implemented an emulator to upscale full-spectrum SIF data from airborne to spaceborne imagery. To this end, various emulators were trained and evaluated using experimental data derived from subsets of two simultaneously recorded image datasets. While spectra derived from the spaceborne PRISMA hyperspectral image served as input to build the emulator, the output consisted of reconstructed full-spectrum SIF emission signals obtained from the airborne-based HyPlant SIF maps. However, emulating hyperspectral data poses significant challenges in dealing with the high dimensionality and redundancy of hyperspectral data. To address this, we employed PCA on both input (PRISMA TOA radiance and BOA reflectance) and output (reconstructed full-SIF emission spectra from HyPlant) to create a lower-dimensional feature space that allows for more efficient training of machine learning regression algorithms (MLRAs). To evaluate the role of the MLRAs in reconstructing the full spectral SIF range, we selected three competitive MLRAs: Kernel Ridge Regression (KRR), Neural Networks (NN) and Gaussian Processes Regression (GPR). See a detailed description in Supplementary Material. Briefly, both input and output data were compressed by PCA, followed by MLRA training to capture nonlinear relationships between both input and output components. The predicted PCs were then transformed back to the full hyperspectral feature space using an inverse PCA. This statistical learning method enables the effective transformation of one type of hyperspectral data into another one in a nonlinear and computationally efficient way (see details in Morata et al., 2021). Using this approach, we can emulate full-SIF emission spectra from PRISMA radiance and reflectance data with high computational efficiency. A subsequent validation against the test subset data was conducted to assess the accuracy of the generated spectral output.

Practically, to spatially match the reconstructed full-spectrum SIF image derived from HyPlant with the PRISMA subset covering the same area, the HyPlant image was aggregated and resampled to 30 m spatial resolution. Then, random samples were collected from corresponding locations from subsets of the PRISMA radiance and reflectance and the reconstructed HyPlant full-spectrum SIF images and used to train the emulator (see Fig. 4). The datasets were used to evaluate different aspects of the emulation strategy systematically: (1) number of PCs in the DR-PCA, (2) comparative analysis of PRISMA radiance and reflectance spectra used as input and (3) performance comparison of competitive MLRAs and size of the training dataset. The dataset consisting of corresponding PRISMA input and HyPlant output spectra was divided into five equally sized subsets that were used for cross-validation.

We gradually increased the size of the randomly sampled dataset used for MLRA training from 250 to 30,000 samples to analyze the impact of the sample size on MLRA emulation performance. Five-fold cross-validation was applied to determine model accuracy, and the times required for model training and testing were recorded. Finally, the best-performing emulator was applied to a PRISMA test subset and cross-compared to the corresponding 30 m resampled HyPlant test subset to validate per-wavelength the full-SIF emission signal image derived from PRISMA data (see Fig. 4).

All processing and evaluation steps were executed on a personal computer (Windows 10 Enterprise v.19041.572 64-bit OS, Intel i7-9700K CPI 3.60 GHz, 32 GB RAM) within the in-house developed ARTMO (Automated Radiative Transfer Models Operator) software framework (Verrelst et al., 2012). ARTMO is a scientific modular package developed in Matlab that provides tools and toolboxes for running a suite of leaf, canopy, and atmosphere RTMs, and for post-processing applications such as the emulator toolbox (Rivera et al., 2015). As part of the ARTMO software package, the emulator toolbox allows the evaluation of MLRAs for their ability to approximate (e.g., RTM) spectral outputs as a function of input variables or spectra (Rivera et al., 2015; Verrelst et al., 2016). Furthermore, we implemented the new SIF reconstruction technique as a new ARTMO tool, named "SIF reconstruction". This tool facilitates the user's ability to reconstruct full-spectrum SIF emission from retrieved SIF values in the O_2A and O_2B regions. The tool can generate full-spectrum SIF signals from corresponding O_2A and O_2B products. It also gives access to the configuration of the SIF-reconstruction parameters, i.e., the linear transformation equations or the SIF-LUT used to generate the PCA matrix. The ARTMO toolboxes are freely downloadable at www.artmotoolbox.com.

3. Results

3.1. Full-spectrum SIF reconstruction with SCOPE simulations

Using the SCOPE database of 100,000 simulated spectra (Section 2.1), correlations between all SIF spectral bands were first explored. To assess the strength of these correlations, i.e., the similarity between two bands, we computed the R^2 correlation matrix encompassing all the wavelengths associated with the SIF emissions (Fig. 6). The correlation matrix revealed the following noteworthy findings: Firstly, if we look at the correlation of all bands with SIF_{687} and SIF_{760},

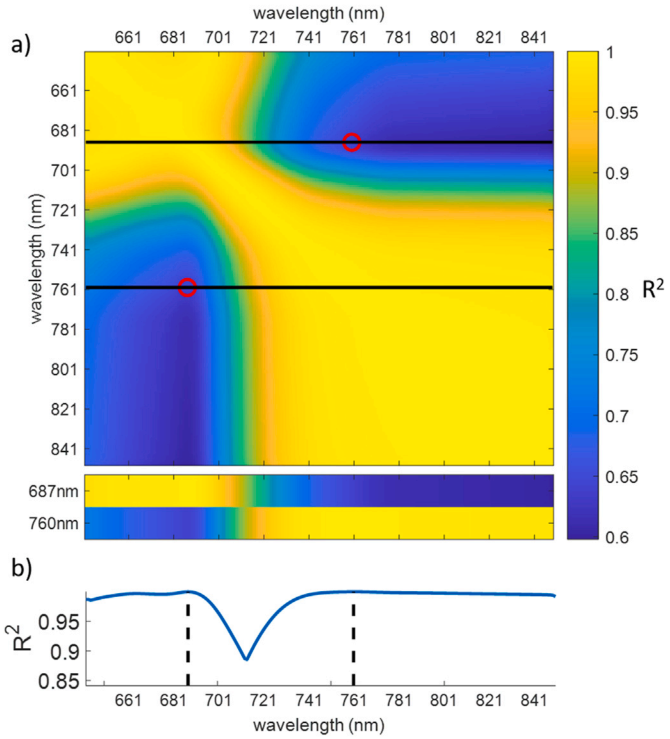


Fig. 6. Correlation matrix (R^2) for all SIF spectral bands from the SCOPE dataset. The black lines represent the SIF_{760} and SIF_{687} bands, with a red circle at the intersection between these two bands. The two bottom correlation bars represent the correlation of the two SIF_{760} and SIF_{687} bands across the spectral bands (**a**). Graph of the maximum correlation between SIF_{760} and SIF_{687}, the dashed lines correspond to the bands 687 and 760 nm (**b**).

we observed a strong correlation between the SIF_{687} band with the SIF bands contained in the range from 640 to 710 nm (which refers to the first peak), with an R^2 higher than 0.9. Similarly, the SIF_{760} band provided an R^2 value higher than 0.9 with the bands in the spectral region spanning from 715 to 850 nm, which refers to the second peak. This highlights the potential of the SIF_{760} and SIF_{687} bands to capture most of the information contained within the full SIF spectrum. Inspecting the correlation matrix of Fig. 6, the presence of a distinct region that exhibits a remarkably low correlation is of interest. Specifically, in the correlation zone between the wavelength 686 nm with the range 739–850 nm, the R^2 decreases until the lowest correlation is reached at 713 nm with a R^2 value of 0.598. Notably, the wavelength 686 nm maintains a considerably high R^2 of 0.999 with the SIF_{687} band. Thanks to its proximity to SIF_{687}, we can leverage the information captured by the SIF_{687} band to extract the values within this seemingly uncorrelated region effectively. It highlights the importance of accurate SIF retrievals from the O_2B absorption feature. The results also demonstrate that SIF_{687} and SIF_{760} are linearly independent, while each captures the majority of the information from the respective left and right peaks. Note in Fig. 6(b) that the region with the lowest correlation with the SIF_{760} and SIF_{687} bands is around 713 nm, which exhibits a comparatively lower R^2 of 0.89. This divergence indicates that this particular wavelength region possesses distinct characteristics and may contribute unique information beyond what is captured by the SIF_{760} and SIF_{687} bands. Yet, the fact that the R^2 in the full spectrum is higher than 0.95, with the lowest correlation at 713 nm with 0.89, indicates that SIF_{760} and SIF_{687} are highly correlated across the bands, containing most of the full-spectrum information.

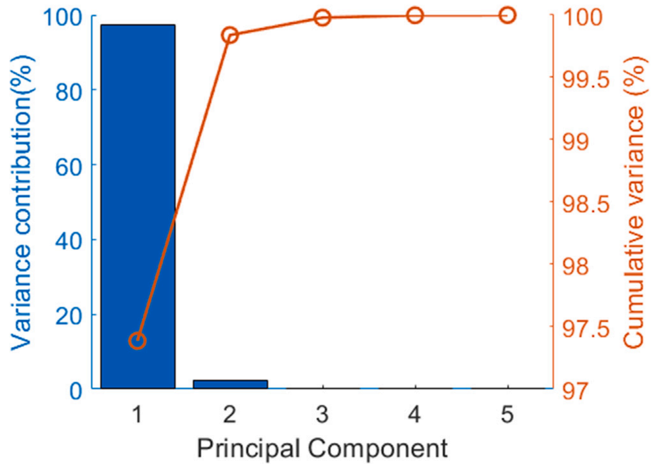


Fig. 7. Explained variance of PCs (blue bars) and cumulative explained variance of the PCs (orange line). The cumulative explained variance by the number of PCs is 97.3870% for 1PC, 99.8385% for 2PCs, 99.9789% for 3PCs, 99.9977% for 4PCs and 99.9995% for 5PCs.

By exploiting this correlation, we can effectively reduce the dimensionality of the dataset thanks to applying PCA with minimal loss of information. Upon examining Fig. 6, it becomes evident that the SIF response exhibits strong linear correlations among its bands. To measure the impact of dimensionality reduction through PCA, we evaluated the cumulative explained variance and showed it in Fig. 7. The first PC already accounts for 97.39% of the total variance. With solely two PCs, we achieve an exceptional cumulative explained variance of 99.84%. This signifies that we can significantly reduce the dimensionality of the dataset with only two PCs without losing key information.

As the next step, we split the simulated SIF spectral dataset into two subsets, 70% for model training and 30% for model testing. Furthermore, we extracted the SIF_{760} and SIF_{687} values from all simulated spectra, in addition to the PCs obtained through PCA applied to the full SIF spectra (Fig. 5). We then established linear relationships between the oxygen bands and the PCs by applying linear regression techniques. In multiple linear regression, the predictors must be independent, i.e., are poorly correlated. Here, the R^2 for SIF_{760} and SIF_{687} has a moderate value of 0.64, one of the lowest values in the determined correlation matrix.

To support the argument that SIF_{760} and SIF_{687} are effectively independent variables, we constructed the same correlation matrix using the same LUT created before, but this time fixing the weather, soil, and geometry variables (Incoming shortwave radiation: 600 W/m^2, Air temperature: 20 °C, Atmospheric vapor pressure: 15 hPa, CO$_2$ concentration: 380 ppm, BSM model parameter: 0.5 and Solar zenith angle: 30°), ranging only the vegetation variables. In this way, we obtained a reduced R^2 for SIF_{760} and SIF_{687} with a distinctly lower value of 0.41 (not shown), indicating that non-vegetated factors add the same variability to SIF_{760} and SIF_{687} and thus affect the full spectrum homogeneously. In contrast, vegetation-related variables have a more localized impact on specific parts of the spectrum. Therefore, the two bands SIF_{760} and SIF_{687} are influenced differently by vegetation parameters, confirming that they can be treated as independent variables. We also analyzed the independence of the predictors by evaluating the correlation between SIF_{760} and SIF_{687} from the HyPlant-retrieved product. In this case, the R^2 is reduced to 0.28, considering the HyPlant SFM retrieval for the entire mosaic as a good starting point for applying linear transformations.

This low correlation supports the assumption of independent predictors to achieve robust linear regression functions for the SIF values and the PCs. By applying linear regression, we can then derive linear equations that underline the connection between the SIF values in the oxygen bands and the PCs for all spectra in our dataset. The linear regression functions obtained for the first four components are shown in Fig. 8. The figure shows that the first two components behave linearly

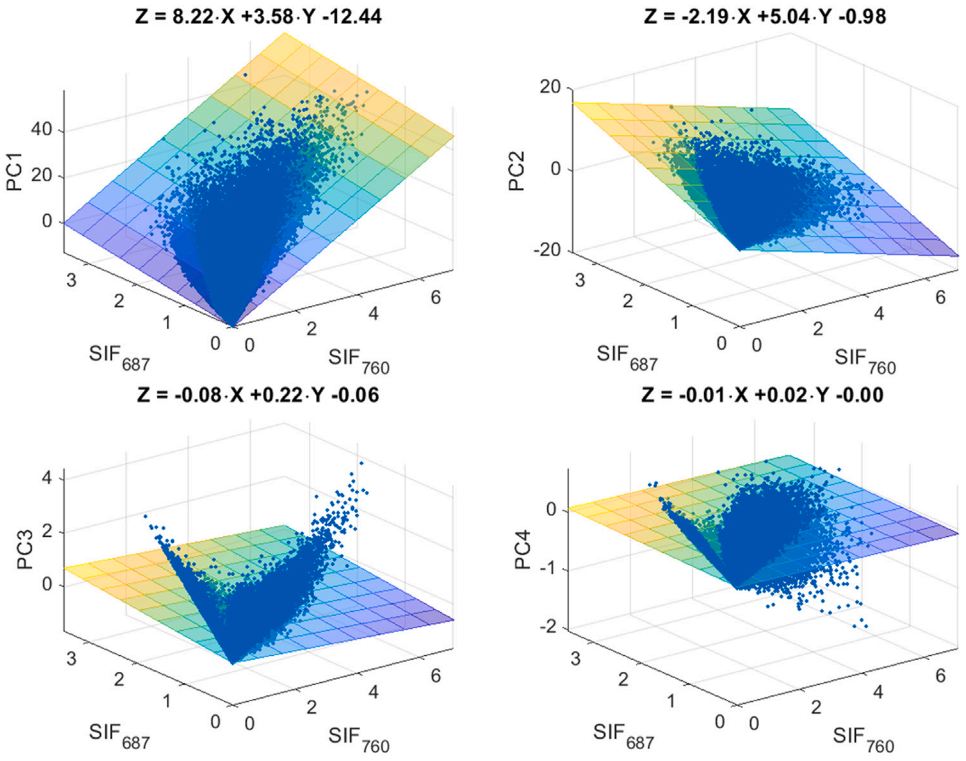


Fig. 8. Scatterplots of the multiple linear regression functions for the first four PCs with SIF_{760} and SIF_{687}, and the corresponding regression equations.

Table 2
Statistics for the scatterplots from the four PCs evaluated.

PCs	RMSE	R^2	NRMSE (%)
PC1	0.47	0.998	0.78
PC2	0.27	0.969	1.43
PC3	0.37	0.031	6.00
PC4	0.14	0.002	4.89

Table 3
Least-squares adjustment coefficient values for each regression with the associated uncertainty. In parentheses is the relative error of each coefficient.

PCs	SlopeA ± $\varepsilon_{abs}(\varepsilon_r$ (%))	SlopeB ± $\varepsilon_{abs}(\varepsilon_r$ (%))	Intercept ± $\varepsilon_{abs}(\varepsilon_r$ (%))
PC1	8.221 ± 0.006 (0.07)	3.583 ± 0.012 (0.33)	−12.441 ± 0.006 (0.05)
PC2	−2.188 ± 0.003 (0.15)	5.037 ± 0.007 (0.13)	−0.982 ± 0.004 (0.36)
PC3	−0.075 ± 0.005 (6.01)	0.215 ± 0.009 (4.21)	−0.064 ± 0.005 (7.45)
PC4	−0.009 ± 0.002 (17.81)	0.021 ± 0.003 (16.26)	−0.004 ± 0.002 (50.55)

with minimal dispersion outside the plane, while components 3 and 4 exhibit clear nonlinear behaviors. Consequently, two linear regression functions can be constructed with PC1 and PC2, for which uncertainties can be propagated linearly.

We also calculated the linear regression statistics between the SIF values (SIF_{760} and SIF_{687}) and the first four PCs. Table 2 shows the R^2 and the NRMSE of the linear regression functions for each PC to evaluate their prediction performance. In addition, in Table 3 we report the slope and intercept values together with the uncertainty estimated for each component. The slope and intercept values were used to transform SIF_{760} and SIF_{687} to the PCs.

As shown in Table 2, the first two components provide R^2 values of 0.998 and 0.969, respectively. Nevertheless, the individual regression functions for the PCs beyond the second one led to distinctly lower R^2 values of 0.031 and lower. Thus, the feasibility of estimating values through multiple linear regressions using PCs higher than the second one is limited. Furthermore, the results depicted in Table 3 show the regression coefficients and the associated absolute and relative uncertainties (in parentheses). While the relative errors determined for the first two components were below 0.4%, the same errors of PCs 3 and 4 were higher than 4%. For this reason, we decided to reconstruct the full-spectrum SIF signal only using the first two components in the linear transformation.

To evaluate the reconstruction accuracy, we applied the PCA-based method to the test subset of the SCOPE simulations using SIF_{760} and SIF_{687} as inputs. The resulting PC1 and PC2 scores were converted back to the original spectral domain via inverse PCA and compared to the original spectra. The reconstruction results are shown in Fig. 9. As expected, the bands with higher R^2 and lower RMSE values were found at 687 and 760 nm. This is because the reconstruction method is based on a projection to the original feature space, so the SIF values at 687 and 760 nm will remain mainly unaltered. The remainder of the full spectrum distribution is obtained as the most frequent shape of the spectrum based on the PCA projection. It led to an R^2 near to 1 and an RMSE near to 0 in the left (640–690 nm) and right (735–850 nm) parts of the spectral range, and a slightly lower correlation is found around 707 nm with a R^2 superior to 0.98 and an NRMSE below 0.12 mW m^{-2} nm^{-1} sr^{-1}. Those results are consistent with the 0.2 mW m^{-2} nm^{-1} sr^{-1} for the expected SIF retrieval error for FLEX. We can therefore consider that the reconstruction method is robust and allows for the reconstruction of the full-spectrum SIF emission with high accuracy.

3.2. Generation of full SIF emission spectrum HyPlant data cube

Having successfully developed a full-spectrum SIF reconstruction technique, the next step is to apply the developed technique to produce a HyPlant-like SIF data cube. To this end, we used the HyPlant SIF product from the LIAISE campaign that provides retrieved SIF and associated uncertainties at 760 and 687 nm. Following the calibration and uncertainty propagation outlined in Section 2.3.3, the calibrated HyPlant SIF_{760} and SIF_{687} values align more closely with those generated by SCOPE (see the supporting figure in supplementary Material). As the next step, we applied the full-spectrum SIF reconstruction process

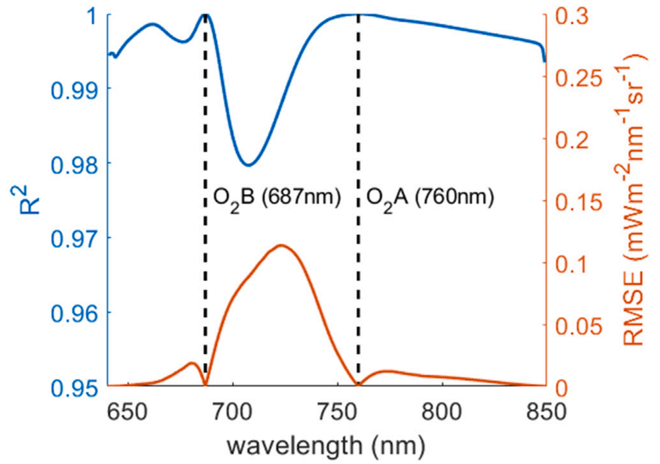


Fig. 9. Statistics of the spectral reconstruction validation. The R^2 is displayed with the blue line (left y-axis) and the RMSE is represented by the orange line (right y-axis). The dashed black lines correspond to the wavelengths 687 and 760 nm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and analytically propagated the original HyPlant SIF_{760} and SIF_{687} uncertainties through all the processing steps. As such, we produced a hyperspectral SIF data cube with 1.7 m spatial resolution covering the full SIF emission range together with associated uncertainty intervals for each of the 211 SIF bands. From this SIF data cube, which has a spectral sampling interval of 1 nm, we subsequently extracted the maximum SIF values at the peak positions of the SIF spectrum (685 and 740 nm) and calculated the total full-spectrum SIF emission flux, i.e., spectrally integrated over 640–850 nm: SIF_{Tot}. Fig. 10 shows the SIF_{760}, SIF_{740} and SIF_{Tot} results with three pixels selected to show the variability in SIF. We also show the uncertainty maps for the 740 nm (peak A) and the 685 nm band (peak B). The full-spectrum SIF dataset opens up new opportunities to study specific wavelengths of interest, such as the SIF spectrum peak positions and absolute values, and the full integrated SIF signal (SIF_{Tot}). Note that the original SIF_{760} and SIF_{687} depicted in black circles in Fig. 10 perfectly match the reconstructed values. This is because we used the first two components, capturing 99.84% of the full spectra variability, to reconstruct the full spectrum. Consequently, we can assume that the precision obtained for SIF_{760} and SIF_{687} can be extrapolated to all wavelengths.

3.3. Full-spectrum SIF PRISMA emulation

Given the successful HyPlant full-spectrum SIF reconstruction, the next step is to build an emulator to upscale the airborne SIF spectra to the satellite scale. To this end, the HyPlant reconstructed full-SIF image was first coarsened to 30 m spatial resolution and divided into two parts, a training and a test subset, as shown in Fig. 4. Similarly, the PRISMA tile was also split into the same two sections to enable the comparison of the images. We evaluated key emulation parameters that can affect the performance: (1) dimensionality reduction using PCA; (2) comparative use of top-of-atmosphere (TOA) radiance and BOA reflectance from PRISMA as input; (3) the used MLRA and the dataset size.

3.3.1. Dimensionality reduction by PCA

The first analysis was conducted to evaluate the PCA dimensionality reduction of the PRISMA input data to create the full-SIF emulator. To accomplish this, we first extracted 1000 spectra from the BOA reflectance PRISMA, then performed a dimensionality reduction analysis, and finally determined the cumulative variance of the PCs. The 99.0% of the variance of the PRISMA reflectance data is kept in the first 27 PCs. Consequently, we decided to use the first 30 PCs to train the emulator. Regarding the output SIF spectrum used to train the emulator, it comprises the full-SIF reconstructed map, which has a total of 211 bands. Since these SIF profiles are reconstructed using only two PCs, 100% of the variability of these reconstructed spectra will be found in the first two PCs, and the remaining PCs will not contain any variability. For this reason, the emulator is trained with only two components as output, which are then transformed back into the full-spectrum SIF.

3.3.2. TOA radiance vs. BOA reflectance

The second analysis evaluates the applicability of TOA radiance and BOA reflectance spectra from PRISMA as input to generate the full-SIF emulation. A dataset of 10,000 samples was created by applying random sampling to the Train subset created from PRISMA and the full-SIF HyPlant data cube. 10-fold cross-validation was applied with 30 PCs as input and 2 PCs as output to train a KRR-based emulator. KRR was chosen for its fast performance, but in the following Section 3.3.3, we also evaluate alternative MLRAs to identify the most accurate emulator. In Fig. 11, the accuracy of the KRR emulator is represented by the mean R2, NRMSE and RMSE, together with the corresponding standard deviations across the spectral range.

As illustrated in the figure, TOA radiance and BOA reflectance of wavelengths beyond 750 nm provide R^2 values higher than 0.7 and 0.78, and NRMSE values lower than 9.23% and 8.18%, respectively. For wavelengths smaller than 700 nm the R^2 values decrease to 0.44 and 0.49, while NRMSE increases to 12.59% and 11.18% for TOA radiance and BOA reflectance, respectively. The RMSE is similar for TOA and BOA with a maximum error of 0.57 and 0.5 mW m^{-2} nm^{-1} sr^{-1} respectively. Both emulators effectively relate PRISMA input spectra to the O_2A region of SIF spectra but struggle with the O_2B region. Using PRISMA BOA reflectance for SIF emulation led to slightly superior and more consistent results, justifying its selection as input for the final emulator.

3.3.3. MLRA and dataset size

To evaluate the role of the MLRAs in reconstructing the full spectral SIF range, we selected three algorithms (KRR, NN, GPR) and trained each with 500 random samples. The statistics for the mean of the five-fold cross-validation, along with its standard deviation, are presented in Fig. 12(a) and (b). As expected, the statistics follow the same pattern as in Fig. 11. All the MLRAs achieved better reconstruction performances for the right side of the SIF spectrum compared to the left side. KRR outperformed GPR and NN across the entire spectral range.

Next, we evaluated the number of samples used for GPR, KRR, and NN model training (see Fig. 12(c)) and recorded the runtime for model training and testing (for $4{\times}10^4$ pixels) (Fig. 12(d)). As the figure reveals, models built with kernel-based MLRAs (GPR and KRR) emulate SIF spectra more accurately and quickly when trained with only up to a few thousand samples. Instead, when more samples are included in model training, kernel-based MLRAs become computationally costly as the number of samples scales cubically with the processing time owing to matrix inversion of the training and validation datasets. Yet, in this case, where we used a maximum of 30,000 training samples, KRR remains the best-performing model at an acceptable computational cost. Based on these findings, we developed a final emulator based on KRR using 30,000 random samples with 30 input PCs and 2 output PCs.

3.3.4. Application to test subset

This KRR emulator was applied to the PRISMA test subset, and the emulation result was subsequently evaluated. Subsequently, to compare the emulated SIF product based on PRISMA data with the original calibrated HyPlant Test image coarsened at PRISMA resolution (30 m), we produced two scatterplots for SIF_{760} and SIF_{687}, respectively, which are shown in Fig. 13. In contrast to SIF_{687}, the SIF_{760} appears

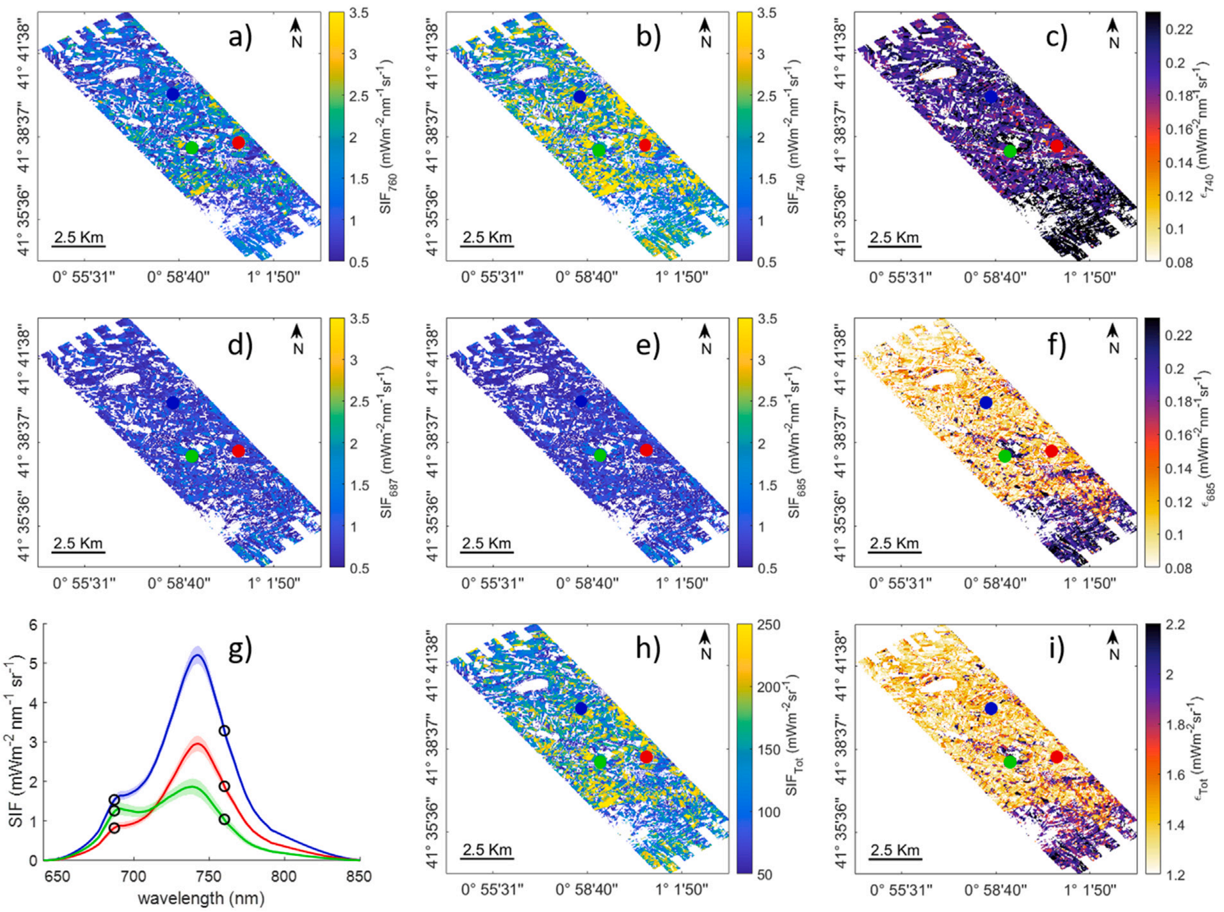


Fig. 10. HyPlant maps with masked negative values (white) extracted from the full-spectrum reconstructed SIF data cube and propagated uncertainty image from uncertainty data cube (ε). SIF_{760} map (mW m^{-2} nm^{-1} sr^{-1}) (O_2A) (**a**). SIF_{740} map (mW m^{-2} nm^{-1} sr^{-1}) (peak A) (**b**). ε_{740} map (peak A) (mW m^{-2} nm^{-1} sr^{-1}) (**c**). SIF_{687} map (mW m^{-2} nm^{-1} sr^{-1}) (O_2B) (**d**). SIF_{685} map (mW m^{-2} nm^{-1} sr^{-1}) (peak B) (**e**). ε_{685} map (peak B) (mW m^{-2} nm^{-1} sr^{-1}) (**f**). SIF spectra of three pixels covering distinct vegetation types (LAI: blue = 3.69, red = 1.92, green = 1.49) with shaded areas representing the corresponding uncertainties and black circles indicating the original HyPlant SIF values at 760 and 687 nm (**g**). Total integrated SIF flux (SIF_{Tot}, in mW m^{-2}sr^{-1}) (**h**). ε_{Tot} map (mW m^{-2}sr^{-1}) (**i**).

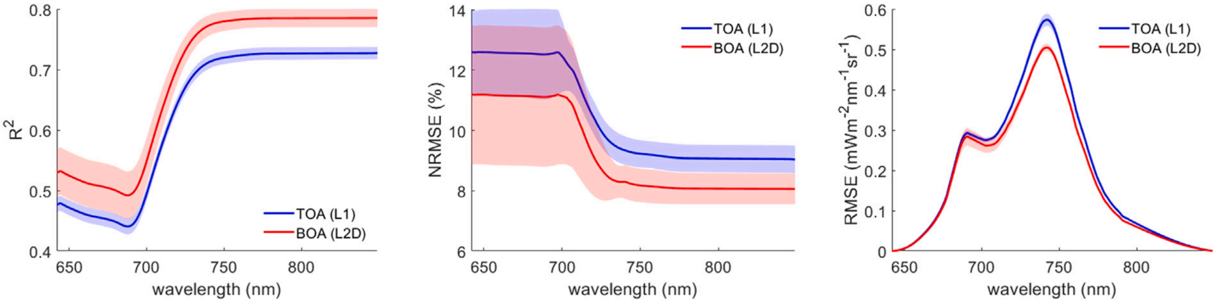


Fig. 11. Comparison of mean R^2 (left), NRMSE (center) and RMSE (right) with corresponding standard deviations (shaded areas) of TOA radiance (blue) and BOA reflectance (red) across the spectral range (650–850 nm) using 10-fold cross-validation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to be reasonably well emulated at satellite scale despite the introduced limitations inherent to PRISMA and the coarsening of the HyPlant data.

We then evaluated the quality of the emulated full-spectrum SIF dataset using the Test subset of the HyPlant dataset, enabling performance assessment across the entire spectral range. The PRISMA-emulated SIF spectra were cross-correlated with the coarsened Hyplant-reconstructed SIF spectra map (Fig. 14). Overall, the figure shows an R^2 ranging between 0.4 and 0.65 and a RMSE lower than 0.8 mW m^{-2} nm^{-1} sr^{-1} across the observed spectral range. The goodness-of-fit metrics are in agreement with the above comparison against the original Hyplant SIF data (see Fig. 13). The KRR emulator exhibits consistent R^2 values across both the left and right regions of the SIF spectrum, likely due to the high correlation along these regions (see also Fig. 6). Similar to what was already observed in the airborne data, we achieved superior results for the right (O_2A) compared to the left part of the spectrum(O_2B).

3.3.5. Application to PRISMA imagery

As a final mapping application, the KRR emulator was executed on the full PRISMA BOA reflectance tile to produce a 30 m resolution full-spectrum SIF reconstructed image (Fig. 15(a)). Three representative SIF profiles illustrate spectral responses across vegetation density gradients

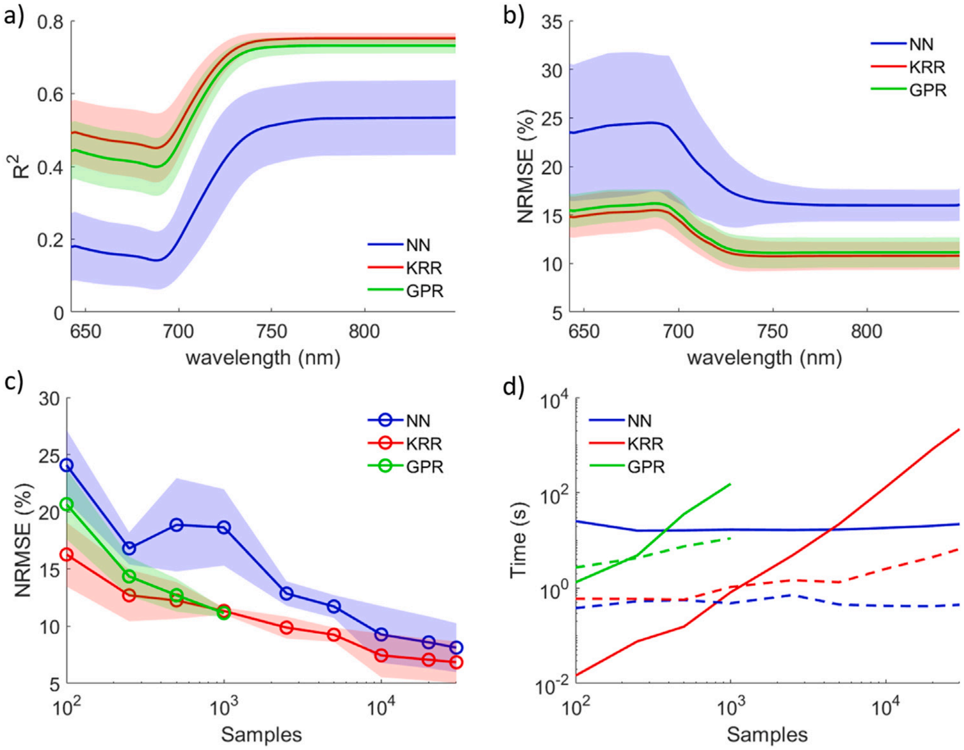


Fig. 12. Comparison of mean R^2 (**a**) and NRMSE (**b**) with corresponding standard deviations (shaded areas) using three MLRAs for model training based on 500 samples using five-fold cross-validation. NRMSE and associated standard deviations (shaded areas) were determined for the three MLRA emulators built with an increasing number of training samples using five-fold cross-validation (**c**). The processing time required for emulator training (solid line) and testing (4×10^4 validation pixels, dashed line) considering increasing training sample sizes (**d**).

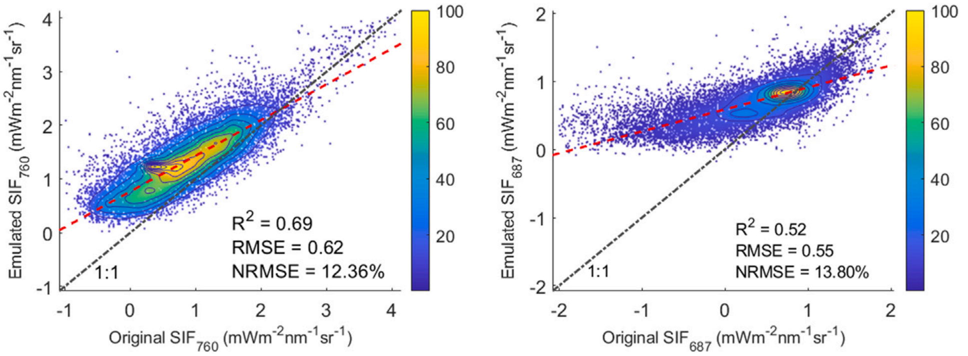


Fig. 13. Density scatterplot of original HyPlant coarsened at PRISMA resolution (30 m) and emulated PRISMA SIF_{760} and SIF_{687}. The black dashed line represents the 1:1-line, the red dashed line represents the linear regression and the colorbar the points density. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 15b), showing slightly reduced values compared to HyPlant due to spatial coarsening. Processing an entire PRISMA tile consisting of more than 10^6 pixels took less than 3 min with a velocity of 1.59 sec/10000 pixels in the case of processing without uncertainty estimation. When including the uncertainty determination, the processing took around 2 h, with a 75 sec/10000 pixels velocity. Lastly, the same KRR emulator was applied to the PRISMA image resampled by aggregation at the FLEX spatial resolution (300 m) to assess the impact of heterogeneity on the full-spectrum SIF product (Fig. 15(c)). Same location spectral profiles are presented in Fig. 15(d), highlighting resolution-dependent SIF suppression. This reduction was caused by pixel heterogeneity in our study area, small agricultural plots (< 300 m) mixed with bare soil, causing reduced yet realistic full-spectrum SIF values in composite pixels. Critically, this emulation scheme delivers scalable, FLEX-like datasets to validate retrieval algorithms prelaunch while quantifying resolution-induced biases. For both full-spectrum SIF products, model-based uncertainties were determined through bootstrapping. Since the emulator was trained with a large number of samples, the uncertainty interval is negligible and difficult to visualize together with the SIF profiles in Fig. 15 (right). Moreover, the reconstruction allows calculating the total emitted SIF flux, i.e., SIF_{Tot}, displayed in Fig. 15 (left). With the upcoming launch of the FLEX mission, the presented workflow can produce FLEX-like SIF images at a similar resolution to support pre-launch activities, such as algorithm development and validation.

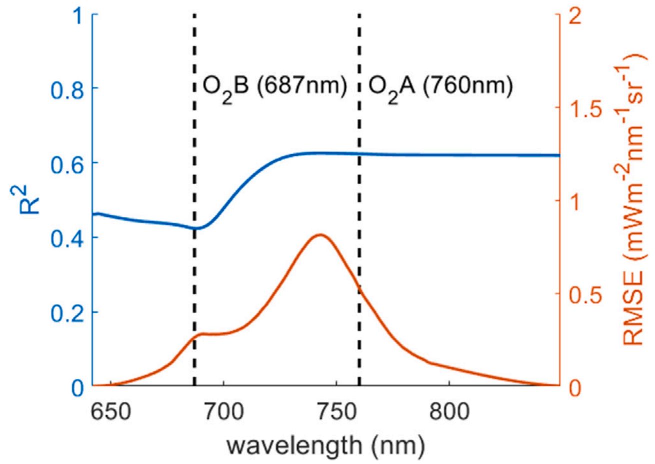


Fig. 14. Performance of the PRISMA SIF emulation when compared to the full-SIF reconstructed HyPlant dataset coarsened at PRISMA resolution (30 m). R^2 values are shown on the left axis, while RMSE values are displayed on the right axis. The vertical dashed lines correspond to the wavelengths 687 and 760 nm.

4. Discussion

4.1. PCA-based full-spectrum SIF reconstruction

The quantification of the subtle canopy-leaving full-spectrum SIF radiant flux from space is an emerging field of research (e.g., Cendrero-Mateo et al., 2019; Mohammed et al., 2019). Given the critical role of accurately capturing full-spectrum SIF in establishing direct links to the plant's actual photosynthetic activity, recent research addressed the challenge of full-spectrum SIF reconstruction (Scodellaro et al., 2022; Oehl and Damm, 2023; Zhao et al., 2024). However, so far no initiatives have been presented that achieve full-spectrum SIF reconstruction from satellite data at a high spatial resolution, e.g., as will be targeted by the upcoming FLEX mission (300 m). Although Zhao et al. (2024) proposed a reconstruction method using TROPOMI at 5.5 km spatial resolution, instead, here, by using PRISMA images at 30 m resolution, our method allows us to explore how SIF heterogeneity across scales can affect sensor recordings with a coarser spatial resolution (e.g., FLEX sensor). We explored an analytical technique of full-spectrum SIF reconstruction through PCA transformation alongside uncertainty propagation. Exploiting PCA transformation for capturing the complete SIF spectral variability, followed by the implementation of two linear regression functions, has shown remarkable accuracy in reconstructing simulated SIF spectra. Using only the first two PCs enables simple and fast processing while preserving the two most informative components and reducing the increasing uncertainties and low correlations that come with higher PCs (Tables 2 and 3). The first two PCs display a linear correlation with SIF_{760} and SIF_{687}, and form the basis of an effective reconstruction model. This choice not only streamlines the reconstruction method but also corresponds with the overarching objective of balancing accuracy and computational efficiency (e.g., Wold et al., 1987; Liu et al., 2006; Jollife and Cadima, 2016). Although the reconstruction validation is limited to SIF derived from the two oxygen absorption bands, the strong representation of spectral information by the two selected PCs gives confidence in the reliability of the associated uncertainty estimates across the entire spectrum.

As an additional advantage as opposed to data-driven, statistical reconstruction approaches (e.g., as proposed by Zhao et al., 2024), the presented full-spectrum SIF reconstruction method offers a fully analytical solution. Previous studies likewise relied on SCOPE-generated spectral databases and presented analytical solutions (e.g., Zhao et al., 2018; Cogliati et al., 2015; Liu et al., 2015; Zhao et al., 2014). Yet, a key distinction of our method is its demonstrated applicability to real airborne data, and the workflow has already been built into a user-friendly tool as part of the ARTMO software framework. This practical deployment ensures broader applicability to SIF_{760} and SIF_{687} imaging products without dependency on training data or the associated risks of overfitting often found in machine learning-based approaches (see reviews in Verrelst et al., 2019; Chen et al., 2023).

The reconstruction method uses only 2 PCs, which implies that the reconstructed full-SIF emission spectrum is a second-order spectrum and only derives two linearly independent spectral features. Nevertheless, Fig. 6 illustrates that the spectral bands are not independent and contain spectral redundancy. In the case of the valley dip, this region is more independent than other spectral regions. However, as demonstrated in Fig. 6(b), the independence with SIF_{760} and SIF_{687} is minimal. This is evidenced by the R^2 of 0.98 and NRMSE of 2% obtained when comparing the reconstructed SIF spectra with the original SCOPE spectra in Fig. 9. The figure indicates that the method is capable of estimating SIF values in the valley dip with high accuracy. Even so, the complexity of the valley region necessitates the inclusion of additional PCs for more precise retrieval.

4.2. Uncertainty propagation of PCA-based full-spectrum SIF reconstruction

Uncertainty propagation through PCA for spectral reconstruction approaches presents an opportunity to enhance the fidelity of the reconstruction method. In case uncertainties are present and this noise has been captured in the original HyPlant SIF_{760} and SIF_{687} uncertainty products, this was reflected in the propagation of the errors to the full-spectrum SIF. As demonstrated throughout this work, analytical uncertainty propagation through PCA transformations acts as an elegant solution for dealing with high-output dimensionality. This approach enables us to trace uncertainties in multi-output models at a marginal computational cost (Morata et al., 2023). For instance, the propagated uncertainty ranges of SIF_{760} and SIF_{687} values from the reconstructed full-spectrum SIF overlap with the original SIF Hyplant uncertainty ranges (histograms not shown), which means that the procedure and the achieved results are reliable.

The pursued uncertainty propagation implies that the reconstruction approach enables a more nuanced understanding of the fidelity associated with the models' SIF predictions. Nonetheless, while uncertainty propagation through PCA transformation is presented as an effective strategy, it is essential to acknowledge that its validity relies on certain assumptions. For instance, it is assumed that the original features are highly correlated. As the PCA transformation is based on the first two PCs, the remaining information from the removed PCs may be lost, particularly if the correlation between the original features is low.

4.3. Limitations and further developments of PCA-based SIF full-spectrum reconstruction

The exploration of full-spectrum SIF reconstruction through PCA and subsequent emulation upscaling has uncovered promising avenues for upcoming research and development. In principle, it only requires the HyPlant SIF_{760} and SIF_{687} products and additional hyperspectral satellite imagery to upscale SIF to the satellite scale. Moving ahead, we concentrate now on the following key aspects to further improve and extend the applicability of the proposed workflow.

As this work demonstrated, full-spectrum SIF can be reconstructed using only two SIF bands, i.e., SIF_{760} and SIF_{687}. However, this methodology is highly susceptible to the quality of these two measurements, so we highlight the necessity of accurate retrievals with low uncertainties. Of particular significance is the area around 713 nm, where the correlation with SIF_{760} and SIF_{687} decreases to an R^2 of 0.89. This region is subject to significant influence from water vapor, which presents a considerable challenge to the retrieval of SIF from aerial or satellite sensors, resulting in inherently imprecise measurements.

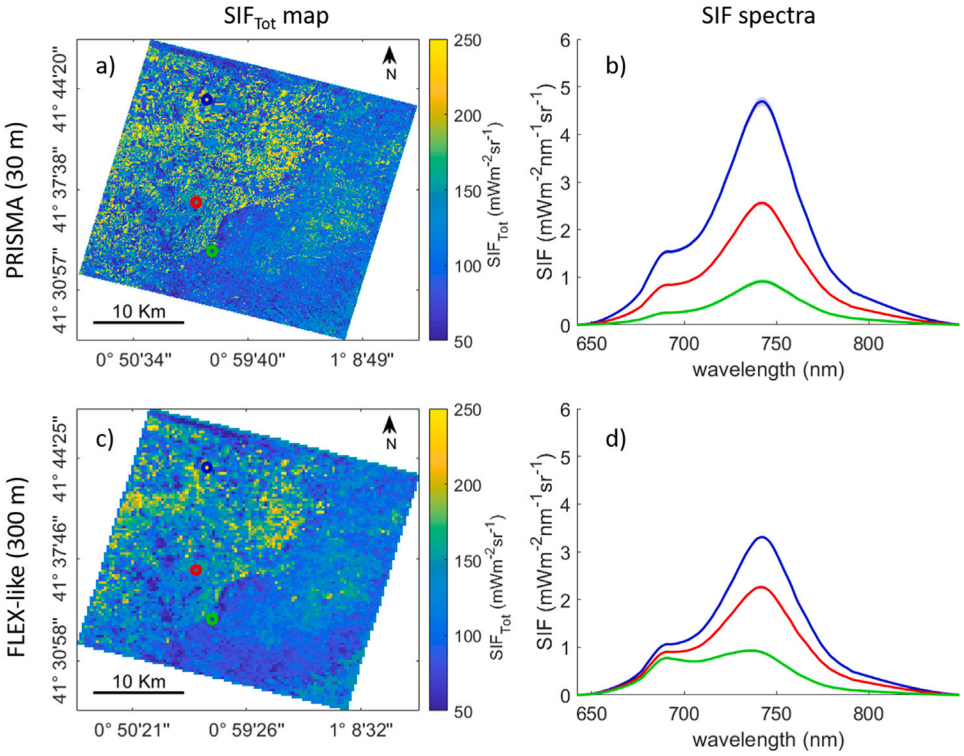


Fig. 15. PRISMA-emulated (30 m) (a) and FLEX-like (300 m) (c) SIF data cubes containing information about the full SIF emission spectrum in each pixel (SIF_{Tot} in mW m^{-2}sr^{-1}). Extracted SIF spectra of three pixels from the PRISMA (b) and FLEX-like (d) SIF data cubes representing SIF_{Tot} of three distinct vegetation types with associated uncertainties (LAI: blue = 3.83, red = 2.7, green = 0.33). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Unlike other reconstruction methods, such as FSR (Zhao et al., 2014, 2024), which employs five values for full-spectrum SIF retrieval, our reconstruction method is constrained by atmospheric conditions, which can be challenging. For instance, the Fraunhofer Line for water vapor (719 nm) presents a considerable challenge when applying the retrieval to aircraft or satellite data. This is attributable to the atmospheric heterogeneity and variability of water vapor, introducing additional noise, which in turn complicates the retrieval in this spectral region and affects the subsequent reconstruction. Moreover, the SIF value in the hydrogen Fraunhofer Line (656 nm) is extremely low, resulting in high uncertainties in the data for this spectral band. For this reason, we focused on spectral regions that provide more stability and are less affected by atmospheric conditions. Our approach stands out for its practicality, targeting feasibility and reliability in airborne applications. Unlike other methods that may be theoretically similar (Zhao et al., 2014, 2018), we prioritize addressing these specific challenges in a real-world context.

While PCA-based dimensionality reduction offers significant advantages in simplifying high-dimensional datasets, it also introduces limitations. The rank-2 approximation, while effective for capturing the primary spectral features, may overlook subtler variations present in the full-spectrum data, particularly under conditions affected by atmospheric interference, sensor noise, or anisotropy in SIF and reflectance. Furthermore, although PCA efficiently compresses the spectral information, it assumes linear relationships between the PCs and the original bands, which may not fully represent complex nonlinear phenomena inherent in the data. Consequently, users of full-spectrum SIF reconstructed data should be aware that this approximation is best suited for cases where the dominant spectral variance is sufficient for their analyses.

This SIF reconstruction technique combines PCs and linear regression to achieve full-spectrum SIF retrieval with robust uncertainty propagation. The use of linear regression ensures a mathematically consistent propagation of uncertainties from the retrieved values to all reconstructed SIF bands, assuming uncertainty in both the model and the inputs (see Eq. (4)). However, while our uncertainty propagation captures sensor errors and the PCA transformation, it does not account for systematic biases in the empirical calibration or limitations of the SCOPE model.

As shown in Fig. 8 and Tables 2 and 3, the first two PCs exhibit a strong linear correlation with SIF_{760} and SIF_{687}, enabling accurate reconstruction of the full SIF spectrum using linear regression. While PCA is inherently a linear transformation, higher-order principal components primarily capture residual variance and noise, which do not maintain a direct or interpretable relationship with SIF_{760} and SIF_{687}. Although including these components could increase the explained variance from 99.84% (with 2 PCs) to 99.9995% (with 4 PCs, see Fig. 7), enhancing the representativeness of the reconstructed SIF spectrum, they tend to exhibit more complex relationships with the SIF bands, ultimately reducing interpretability and complicating uncertainty quantification. Conventional multivariate linear regression allows for straightforward uncertainty propagation via partial derivatives, assuming uncorrelated errors between SIF_{760} and SIF_{687}. In principle, nonlinear adjustments could be exploited to assess the relationships of higher-order PCs greater than 3 with the independent variables SIF_{760} and SIF_{687}. Yet, incorporating nonlinear models lacks physical interpretability, complicates error estimation, and introduces bias without meaningful gains. Although advanced nonlinear MLRAs such as GPR can estimate model-based uncertainties within a Bayesian framework, most MLRAs lack built-in uncertainty quantification. They would require computationally intensive alternatives like bootstrapping (García-Soria et al., 2024). Given these considerations, our method favors a simpler, transparent linear approach applicable to the first two PCs, ensuring traceable error estimates across all reconstructed SIF bands.

4.4. Emulation airborne-to-spaceborne full-spectrum SIF

As a secondary objective, the work demonstrated the effectiveness of emulation in upscaling reconstructed full-spectrum SIF imagery based on satellite imaging spectroscopy reflectance data. Given the corresponding airborne full-spectrum SIF and PRISMA reflectance data, we evaluated the best-performing MLRA and optimized the training data settings. The KRR emulator adequately reconstructed the right side (R^2: 0.65) and moderately well the left side of the SIF spectrum (R^2: 0.42) (Fig. 14). The rationale behind this is that the emulator predicts the two output PCs and afterwards reconstructs the original full SIF spectrum. We expect that the accuracy achieved when predicting SIF_{760} and SIF_{687} can be extrapolated to the other spectral bands forming the right and left parts of the full SIF spectrum, respectively. The higher correlation for the O_2A peak suggests that this part of the SIF spectrum is strongly correlated with vegetation reflectance. At the same time, the poorer correlation of the O_2B peak suggests more complex, irregular behaviors that are not fully captured by the vegetation reflectance spectrum, which can be explained as follows. On the one hand, SIF in O_2B is related to photosynthetic activity while at the same time being susceptible to reabsorption within the canopy (Qi et al., 2023), meaning that the canopy-leaving O_2B flux reaching a sensor primarily originates from the top-of-canopy, making it less sensitive to variations in canopy structural variability. Thus, the O_2B flux tends to be more independent of reflectance variability but the canopy-leaving signal is smaller (Bandopadhyay et al., 2020; Verrelst et al., 2015). On the other hand, while the canopy-leaving O_2A flux is also related to photosynthetic activity, at 760 nm this flux is less reabsorbed but more prone to scattering dynamics. This translates to the O_2A flux being more influenced by canopy structure, similar to reflectance. Consequently, the variability in the O_2A region exhibits a stronger correlation with reflectance variability compared to the O_2B region (Verrelst et al., 2015). Altogether, the O_2B flux is influenced by more complex processes than the O_2A flux, leading to greater challenges in accurately retrieving SIF from vegetation reflectance. This suggests the need for a hyperspectral sensor with an enhanced signal-to-noise ratio and narrower bandwidth compared to PRISMA.

Additionally, emulation from BOA reflectance achieved superior reconstruction accuracy compared to TOA radiance. This improvement arises because TOA radiance inherently includes atmospheric effects, forcing the emulator to simultaneously account for both vegetation-driven SIF signals and atmospheric perturbations. In contrast, BOA reflectance (i.e., after atmospheric correction) isolates vegetation-related radiative processes, reducing the emulator's burden to model complex interactions between surface and atmosphere. He et al. (2021), Patadia et al. (2018).

Apart from SIF prediction accuracy, one crucial new development in emulation strategies is the ability to generate emulation-based uncertainty maps. The application of uncertainty propagation by PCA reconstruction and a bootstrapping technique, which in principle can be applied to any trained multi-output model, e.g., an emulator, enables the development of emulators that provide predicted full-spectrum SIF along with confidence intervals. KRR with bootstrapping uncertainty was selected as the optimal option, with the drawback that bootstrapping can potentially increase the processing time from 1.59 sec/10000 pix to 75 sec/10000 pix. Therefore, in cases where computation time is more critical than performance and a large training dataset is available, instead a NN together with the dropout method can be used to determine uncertainties with lower computation time requirements (Fig. 12) (Morata et al., 2023; Gal and Ghahramani, 2016).

Finally, the emulator's effectiveness is demonstrated by upscaling the airborne full SIF emission signal to spaceborne PRISMA BOA reflectance data. The emulated SIF maps demonstrate the versatility of our method, which is in principle applicable to images from any terrestrial imaging spectroscopy mission. Although trained with HyPlant data, the emulation method is flexible and can be adapted to high-resolution hyperspectral image data such as DLR Earth Sensing Imaging Spectrometer (DESIS) with retraining. The finer spectral resolution of this sensor (3.5 nm) and the SNR of about 200 mounted on the International Space Station (ISS) could be beneficial to improve spaceborne SIF emulation (as also demonstrated by Alonso et al. (2019), Pato et al. (2023)). This work lays the foundation for scaling landscape full-spectrum SIF retrieval, bridging the gap between high-resolution airborne and coarser satellite SIF data. At the same time, we know that the canopy-leaving SIF signal is the result of multiple highly complex processes involving numerous internal interactions within the leaf/canopy and atmosphere (Verrelst et al., 2015). Although emulation paves the way for the approximation of the full SIF emission spectrum from hyperspectral reflectance data, it is important to be aware of its limitations. While emulators can effectively relate nonlinear relationships between spectral features using statistical methods, they are limited in their ability to fully capture the underlying photochemical and biophysical mechanisms (e.g. diurnal cycles driven by solar irradiance) and the inherent complexities governing SIF emission. After all, the emulated SIF maps are statistical predictions that vary as a function of the BOA reflectance. Therefore, the emulator is not designed to capture changes in SIF that occur before detectable changes in reflectance, nor to simulate diurnal variations driven by solar irradiance.

4.5. Opportunities and limitations of emulated full-spectrum SIF signal in support of FLEX

Since the workflow emulates realistic full-spectrum SIF imagery with FLEX-like characteristics at a high spatial resolution, it can serve as a preparatory tool to generate surrogate full-spectrum SIF data in anticipation of real-world FLEX data. The FLEX mission concept is unique in targeting global, full-spectrum SIF retrieval at a spatial resolution of 300 meters (Drusch et al., 2017). With the expected launch in 2026, the mission is currently under development, and among other preparatory activities, a FLEX End-to-End simulator (FLEX-E) was developed (Vicent et al., 2016). Yet, simulated scenes cannot fully reproduce the complexities of terrestrial surfaces, highlighting the need for real-world data in algorithm development and evaluation. Although space agencies already offer global single-band SIF data at coarse spatial resolutions (in the order of kms, e.g., see Frankenberg et al., 2014; Köhler et al., 2018), the here presented advancements can expedite the familiarization of the broader community with high-resolution, full-spectrum SIF data streams.

Apart from FLEX, it is worth remarking that some coarse-resolution atmospheric satellites also provide two or more SIF bands (e.g., Sentinel-5P TROPOMI). Yet, those data are mainly restricted to the far-red region, meaning that the PCA-based reconstruction method would be limited, as the first two PCs are strongly correlated to the O_2B and O_2A absorption bands. Retrieving red SIF is more challenging than far-red SIF, mainly because the signal is smaller and the red region of the spectrum is more affected by Rayleigh scattering, which is stronger at shorter wavelengths (Guanter et al., 2013). This scattering adds a high complexity to separating the SIF signal from the atmospheric noise. This fact was demonstrated by anomalies observed in the global maps of red SIF datasets in the studies by Köhler et al. (2020), Zhao et al. (2022). For these data, it is worth analyzing to what extent the PCA transformation method can reconstruct the full-spectrum SIF signal. It is however important to note that the accuracy of the reconstructed SIF is inherently linked to the accuracy of the input SIF retrievals, with red SIF being a limiting factor. At the same time, these coarse-resolution pixels cover heterogeneous surfaces, likely including non-vegetated areas within a pixel. SCOPE simulations represent homogeneously covered, vegetated areas, which suggests a potential mismatch because the heterogeneity in land cover of real-world surfaces cannot be taken into account (Jantol et al., 2023).

As a closing remark, exploiting the full-spectrum SIF signal as a direct measurable reporter of the photosynthetic machinery in plants requires considering multiple factors such as chlorophyll content, vegetation structure, and non-photosynthetic quenching (NPQ) (De Grave et al., 2020; Van Wittenberghe et al., 2021). FLEX is planned to orbit in tandem with Copernicus' Sentinel-3 satellite (S3). As such, the tandem FLEX-Sentinel-3 mission concept is dedicated to quantifying these key factors, enabling a more comprehensive interpretation of SIF data and photosynthesis mechanisms that lead to carbon fixation (Drusch et al., 2017; Van Wittenberghe et al., 2021). The scientific framework and algorithms for processing FLEX-S3 data stream into reliable real-world vegetation productivity estimates are actively being developed (Van Wittenberghe et al., 2024; Reyes-Muñoz et al., 2024).

5. Conclusions and outlook

This study explored an analytical reconstruction method to produce full SIF emission spectra given the availability of SIF data retrieved from the O_2A and O_2B absorption regions. By employing PCA to SCOPE-simulated SIF data, we found that most of the variability of the full SIF spectrum is captured with only two principal components. By using these two PCs, more than 99.84% of the cumulative variance of the original data is preserved. This dimensionality reduction allowed us to establish two linear relationships between the two PCs and readily available airborne HyPlant SIF data derived from the O_2A and O_2B absorption features, followed by the inverse PCA back to the full-spectrum SIF domain. Cross-comparing the generated full-spectrum SIF against the original values revealed an almost perfect reconstruction of SIF at the position of the O_2A and O_2B bands. Linear functions also allowed for the analytical propagation of uncertainties along the full-spectrum SIF signal, which enabled the generation of full-spectrum SIF maps with associated uncertainties of an entire landscape. The overall precision of the reconstructed full-wavelength SIF spectrum, yet, ultimately depends on the accuracy of the SIF measurements obtained by HyPlant in the O_2A and O_2B bands.

We subsequently presented an upscaling technique to transfer the airborne full-spectrum SIF signal to the satellite scale to accommodate the generation of satellite-based, high-resolution, full-spectrum SIF maps and related products (e.g., integrated total full-spectrum SIF flux: SIF_{Tot}). To this end, we developed an airborne-to-spaceborne emulator trained based on the produced HyPlant full-spectrum SIF data and corresponding PRISMA hyperspectral BOA reflectance. The emulator adequately reconstructed the full SIF spectrum at the satellite scale with associated uncertainties by applying bootstrapping combined with KRR. Coarsening the PRISMA full SIF signal data from 30 to 300 m spatial resolution paves the way for generating FLEX-like full-spectrum SIF imagery.

In a broader context, the analytical full-spectrum SIF reconstruction combined with the emulation upscaling technique holds promising implications for a better understanding of the terrestrial SIF radiant flux as derived from satellite imaging spectrometers. Firstly, the workflow generates simulated SIF products at a resolution similar to that of the FLEX mission, thus providing a valuable resource for preparatory and optimization activities in advance of the FLEX mission. Secondly, the ability to emulate full-spectrum SIF spectra from readily available satellite hyperspectral reflectance data opens doors for producing meaningful, high-resolution full-spectrum SIF imagery over any terrestrial surface. Altogether, the presented workflow can provide surrogate datasets for developing improved full-spectrum SIF retrieval methods and downstream models for tracking the plant's photosynthetic activity from space, thereby accounting for the role of other drivers that govern the canopy-leaving SIF radiant flux, as is foreseen with the tandem FLEX-Sentinel-3 mission concept. Eventually, this path progresses towards global real-time monitoring of the vegetation's actual productivity dynamics.

CRediT authorship contribution statement

Miguel Morata: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bastian Siegmann:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **José Luis García-Soria:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis. **Juan Pablo Rivera-Caicedo:** Software. **Jochem Verrelst:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

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.

Acknowledgments

This research was funded by the European Research Council (ERC) under the project FLEXINEL (#101086622). The views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. We also gratefully acknowledge the financial support by the European Space Agency (ESA), France for airborne data acquisition and data analysis in the frame of the LIAISE campaign (ESA Contract No. 4000125402/18/NL/NA). We thank the four reviewers for their constructive comments.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used DeepLWrite and BARD in order to improve the language and readability. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.rse.2025.114865.

Data availability

Data will be made available on request.

References

Alonso, K., Bachmann, M., Burch, K., Carmona, E., Cerra, D., de los Reyes, R., Dietrich, D., Heiden, U., Hölderlin, A., Ickes, J., Knodt, U., Krutz, D., Lester, H., Müller, R., Pagnutti, M., Reinartz, P., Richter, R., Ryan, R., Sebastian, I., Tegler, M., 2019. Data products, quality and validation of the DLR earth sensing imaging spectrometer (DESIS). Sensors 19 (20).

Alonso, L., Burkart, A., et al., 2015. Sun-induced fluorescence - a new probe of photosynthesis: First maps from the imaging spectrometer HyPlant. Global Change Biol. 21.

Bandopadhyay, S., Rastogi, A., Juszczak, R., 2020. Review of top-of-canopy sun-induced fluorescence (SIF) studies from ground, UAV, airborne to spaceborne observations. Sensors 20 (4).

Bandopadhyay, S., Rastogi, A., et al., 2019. HyPlant-derived sun-induced fluorescence—A new opportunity to disentangle complex vegetation signals from diverse vegetation types. Remote. Sens. 11 (14).

Celesti, M., Rast, M., Adams, J., Boccia, V., Gascon, F., Isola, C., Nieke, J., 2022. The copernicus hyperspectral imaging mission for the environment (CHIME): Status and planning. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium. pp. 5011–5014.

Celesti, M., Van der Tol, C., et al., 2018. Exploring the physiological information of Sun-induced chlorophyll fluorescence through radiative transfer model inversion. Remote Sens. Environ. 215, 97–108.

Cendrero-Mateo, M., Wieneke, S., et al., 2019. Sun-induced chlorophyll fluorescence III: Benchmarking retrieval methods and sensor characteristics for proximal sensing. Remote. Sens. 11 (8).

Chen, Y.-N., Fan, K.-C., Chang, Y.-L., Moriyama, T., 2023. Special issue review: Artificial intelligence and machine learning applications in remote sensing. Remote. Sens. 15 (3).

Cogliati, S., Celesti, M., et al., 2019. A spectral fitting algorithm to retrieve the fluorescence spectrum from canopy radiance. Remote. Sens. 11, 1840.

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

De Grave, C., Verrelst, J., Morcillo-Pallarés, P., Pipia, L., Rivera-Caicedo, J.P., Amin, E., Belda, S., Moreno, J., 2020. Quantifying vegetation biophysical variables from the sentinel-3/FLEX tandem mission: Evaluation of the synergy of OLCI and FLORIS data sources. Remote Sens. Environ. 251, 112101.

Drusch, M., Moreno, J., et al., 2017. The fluorescence explorer mission concept-ESA's earth explorer 8. IEEE Trans. Geosci. Remote Sens. 55 (3), 1273–1284, Cited by: 230.

Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., Pollock, R., Taylor, T.E., 2014. Prospects for chlorophyll fluorescence remote sensing from the orbiting carbon observatory-2. Remote Sens. Environ. 147, 1–12.

Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning. ICML '16, vol. 48, JMLR.org, pp. 1050–1059.

García-Soria, J.L., Morata, M., Berger, K., Pascual-Venteo, A.B., Rivera-Caicedo, J.P., Verrelst, J., 2024. Evaluating epistemic uncertainty estimation strategies in vegetation trait retrieval using hybrid models and imaging spectroscopy data. Remote Sens. Environ. 310, 114228.

Green, R.O., Thompson, D.R., 2020. An earth science imaging spectroscopy mission: The earth surface mineral dust source investigation (EMIT). In: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. pp. 6262–6265.

Guanter, L., Kaufmann, H., et al., 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. Remote. Sens. 7 (7), 8830–8857.

Guanter, L., Rossini, M., Colombo, R., Meroni, M., Frankenberg, C., Lee, J.-E., Joiner, J., 2013. Using field spectroscopy to assess the potential of statistical approaches for the retrieval of sun-induced chlorophyll fluorescence from ground and space. Remote Sens. Environ. 133, 52–61.

He, Q., Fang, Z., Shoshanim, O., Brown, S.S., Rudich, Y., 2021. Scattering and absorption cross sections of atmospheric gases in the ultraviolet–visible wavelength range (307–725 nm). Atmos. Chem. Phys. 21 (19), 14927–14940.

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inform. Theory 14 (1), 55–63.

Jantol, N., Prikaziuk, E., Celesti, M., Hernandez-Sequeira, I., Tomelleri, E., Pacheco-Labrador, J., Van Wittenberghe, S., Pla, F., Bandopadhyay, S., Koren, G., Siegmann, B., Legović, T., Kutnjak, H., Cendrero-Mateo, M.P., 2023. Using sentinel-2-based metrics to characterize the spatial heterogeneity of FLEX sun-induced chlorophyll fluorescence on sub-pixel scale. Remote. Sens. 15 (19).

Jollife, I.T., Cadima, J., 2016. Principal component analysis: A review and recent developments. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 374 (2065).

Julitta, T., Corp, L.A., et al., 2016. Comparison of sun-induced chlorophyll fluorescence estimates obtained from four portable field spectroradiometers. Remote. Sens. 8 (2).

Köhler, P., Behrenfeld, M., Landgraf, J., Joanna, J., Magney, T., Frankenberg, C., 2020. Global retrievals of solar-induced chlorophyll fluorescence at red wavelengths with TROPOMI. Geophys. Res. Lett. 47.

Köhler, P., Frankenberg, C., Magney, T.S., Guanter, L., Joiner, J., Landgraf, J., 2018. Global retrievals of solar-induced chlorophyll fluorescence with TROPOMI: First results and intersensor comparison to OCO-2. Geophys. Res. Lett. 45 (19), 10,456–10,463.

Köhler, P., Guanter, L., Joiner, J., 2015. A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data. Atmospheric Meas. Tech. 8 (6), 2589–2608.

Krutz, D., Müller, R., et al., 2019. The instrument design of the DLR earth sensing imaging spectrometer (DESIS). Sensors 19 (7).

Kukkurainen, A., Lipponen, A., Kolehmainen, V., Arola, A., Cogliati, S., Sabater, N., 2025. SIFFI: Bayesian solar-induced fluorescence retrieval algorithm for remote sensing of vegetation. Remote Sens. Environ. 318, 114558.

Liu, X., Liu, L., Zhang, S., Zhou, X., 2015. New spectral fitting method for full-spectrum solar-induced chlorophyll fluorescence retrieval based on principal components analysis. Remote. Sens. 7 (8), 10626–10645.

Liu, X., Smith, W.L., Zhou, D.K., Larar, A., 2006. Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. Appl. Opt. 45 (1), 201–209.

Loizzo, R., Daraio, M., Guarini, R., Longo, F., Lorusso, R., Dini, L., Lopinto, E., 2019. PRISMA mission status and perspective. In: IGARSS 2019. pp. 4503–4506.

Mckay, M., Beckman, R., Conover, W., 1979. A comparison of three methods for selecting vales of input variables in the analysis of output from a computer code. Technometrics 21, 239–245.

Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., Moreno, J., 2009. Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. Remote Sens. Environ. 113 (10), 2037–2051.

Mohammed, G.H., Colombo, R., et al., 2019. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. Remote Sens. Environ. 231, 111177.

Morata, M., Siegmann, B., Morcillo-Pallarés, P., Rivera-Caicedo, J., Verrelst, J., 2021. Emulation of sun-induced fluorescence from radiance data recorded by the HyPlant airborne imaging spectrometer. Remote. Sens. 13 (21).

Morata, M., Siegmann, B., Pérez-Suay, A., García-Soria, J.L., Rivera-Caicedo, J.P., Verrelst, J., 2023. Neural network emulation of synthetic hyperspectral sentinel-2-like imagery with uncertainty. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 16, 762–772.

National Academies of Sciences, Medicine, E., 2018. Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space. The National Academies Press, Washington, DC.

Oehl, V., Damm, A., 2023. WAFER: A new method to retrieve sun-induced fluorescence based on spectral wavelet decompositions. Remote Sens. Environ. 298, 113786.

Patadia, F., Levy, R.C., Mattoo, S., 2018. Correcting for trace gas absorption when retrieving aerosol optical depth from satellite observations of reflected shortwave radiation. Atmos. Meas. Tech. 11 (6), 3205–3219.

Pato, M., Alonso, K., et al., 2023. Fast machine learning simulator of at-sensor radiances for solar-induced fluorescence retrieval with DESIS and HyPlant. 2023-July, pp. 7563–7566,

Plascyk, J.A., 1975. The MK II fraunhofer line discriminator (FLD-II) for airborne and orbital remote sensing of solar-stimulated luminescence. Opt. Eng., Bellingham 14 (4), 144339.

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

Qi, M., Liu, X., Du, S., Guan, L., Chen, R., Liu, L., 2023. Improving the estimation of canopy fluorescence escape probability in the near-infrared band by accounting for soil reflectance. Remote. Sens. 15 (18).

Rascher, U., Baum, S., et al., 2022. FLEXSense: Technical assistance for airborne measurements during the FLEX sentinel tandem experiment.

Rascher, U., Siegmann, B., et al., 2021. FLEXSense 2018 final report: FLEX sentinel tandem campaign - technical assistance for airborne measurements.

Reyes-Muñoz, P., D.Kovács, D., Berger, K., Pipia, L., Belda, S., Rivera-Caicedo, J.P., Verrelst, J., 2024. Inferring global terrestrial carbon fluxes from the synergy of sentinel 3 & 5P with Gaussian process hybrid models. Remote Sens. Environ. 305, 114072.

Rivera, J.P., Verrelst, J., Gómez-Dans, J., Muñoz-Marí, J., Moreno, J., Camps-Valls, G., 2015. An emulator toolbox to approximate radiative transfer models with statistical learning. Remote. Sens. 7 (7), 9347–9370.

Sabater, N., Vicent, J., Alonso, L., Verrelst, J., Middleton, E.M., Porcar-Castell, A., Moreno, J., 2018. Compensation of oxygen transmittance effects for proximal sensing retrieval of canopy–leaving sun–Induced chlorophyll fluorescence. Remote. Sens. 10 (10).

Scodellaro, R., Cesana, I., et al., 2022. A novel hybrid machine learning phasor-based approach to retrieve a full set of solar-induced fluorescence metrics and biophysical parameters. Remote Sens. Environ. 280, 113196.

Siegmann, B., Alonso, L., et al., 2019. The high-performance airborne imaging spectrometer HyPlant—From raw images to top-of-canopy reflectance and fluorescence products: Introduction of an automatized processing chain. Remote. Sens. 11, 2760.

Theisen, A.F., 2002. Detecting chlorophyll fluorescence from orbit: The fraunhofer line depth model. In: Muttiah, R.S. (Ed.), From Laboratory Spectroscopy To Remotely Sensed Spectra of Terrestrial Ecosystems. Springer Netherlands, Dordrecht, pp. 203–232.

Van der Tol, C., Berry, J., Campbell, P., 2014. Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence. J. Geophys. Res.: Biogeosciences 119.

Van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., Su, Z., 2009. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. Biogeosciences 6, 12, 6.

Van Wittenberghe, S., Amin, E., Pascual-Venteo, A.B., Pérez-Suay, A., Tenjo, C., Sabater, N., Van der Tol, C., Drusch, M., Moreno, J., 2024. Retrieval of leaf-level fluorescence quantum efficiency and NPQ-related xanthophyll absorption through spectral unmixing strategies for future VIS-NIR imaging spectroscopy. Remote Sens. Environ. 300, 113879.

Van Wittenberghe, S., Sabater, N., et al., 2021. Towards the quantitative and physically-based interpretation of solar-induced vegetation fluorescence retrieved from global imaging. Photosynthetica 59 (SPECIAL ISSUE), 438–457.

Veraverbeke, S., Dennison, P., et al., 2018. Hyperspectral remote sensing of fire: State-of-the-art and future perspectives. Remote Sens. Environ. 216, 105–121.

Verhoef, W., Van der Tol, C., Middleton, E.M., 2018. 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. 204, 942–963.

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

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

Verrelst, J., Rivera-Caicedo, J.P., Reyes-Muñoz, P., Morata, M., Amin, E., Tagliabue, G., Panigada, C., Hank, T., Berger, K., 2021. Mapping landscape canopy nitrogen content from space using PRISMA data. ISPRS J. Photogramm. Remote Sens. 178, 382–395.

Verrelst, J., Rivera Caicedo, J., Vicent, J., Morcillo Pallarés, P., Moreno, J., 2019. Approximating empirical surface reflectance data through emulation: Opportunities for synthetic scene generation. Remote. Sens. 11 (2).

Verrelst, J., Romijn, E., Kooistra, L., 2012. Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data. Remote. Sens. 4 (9), 2866–2889.

Verrelst, J., Sabater, N., Rivera, J., Muñoz-Marí, J., Vicent, J., Camps-Valls, G., Moreno, J., 2016. Emulation of leaf, canopy and atmosphere radiative transfer models for fast global sensitivity analysis. Remote. Sens. 8 (8).

Vicent, J., Sabater, N., et al., 2016. FLEX end-to-end mission performance simulator. IEEE Trans. Geosci. Remote Sens. 54 (7), 4215–4223.

Vicent, J., Verrelst, J., Rivera-Caicedo, J.P., Sabater, N., Muñoz-Marí, J., Camps-Valls, G., Moreno, J., 2018. Emulation as an accurate alternative to interpolation in sampling radiative transfer codes. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 11 (12), 4918–4931.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemometr. Intell. Lab. Syst. 2 (1), 37–52, Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

Yang, P., Van der Tol, C., Campbell, P.K.E., Middleton, E.M., 2021. Unraveling the physical and physiological basis for the solar-induced chlorophyll fluorescence and photosynthesis relationship using continuous leaf and canopy measurements of a corn crop. Biogeosciences 18 (2), 441–465.

Zhao, F., Guo, Y., Verhoef, W., Gu, X., Liu, L., Yang, G., 2014. A method to reconstruct the solar-induced canopy fluorescence spectrum from hyperspectral measurements. Remote. Sens. 6 (10), 10171–10192.

Zhao, F., Li, R., Verhoef, W., Cogliati, S., Liu, X., Huang, Y., Guo, Y., Huang, J., 2018. Reconstruction of the full spectrum of solar-induced chlorophyll fluorescence: Intercomparison study for a novel method. Remote Sens. Environ. 219, 233–246.

Zhao, F., Ma, W., Köhler, P., Ma, X., Sun, H., Verhoef, W., Zhao, J., Huang, Y., Li, Z., Ratul, A.K., 2022. Retrieval of red solar-induced chlorophyll fluorescence with TROPOMI on the sentinel-5 precursor mission. IEEE Trans. Geosci. Remote Sens. 60, 1–14.

Zhao, F., Ma, W., Zhao, J., Guo, Y., Tariq, M., Li, J., 2024. Global retrieval of the spectrum of terrestrial chlorophyll fluorescence: First results with TROPOMI. Remote Sens. Environ. 300, 113903.