The TROPOSIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission

Luis Guanter[1], Cédric Bacour[2], Andreas Schneider[3,a], Ilse Aben[3], Tim A. van Kempen[3], Fabienne Maignan[4], Christian Retscher[5], Philipp Köhler[6], Christian Frankenberg[6], Joanna Joiner[7], and Yongguang Zhang[8]

[1]Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, Valencia, Spain
[2]NOVELTIS, Labège, France
[3]SRON Netherlands Institute for Space Research, Utrecht, The Netherlands
[4]Laboratoire des Sciences du Climat et de l'Environnement (LSCE) / Institut Pierre Simon Laplace (IPSL), Gif-sur-Yvette, France
[5]European Space Agency (ESA/ESRIN), Frascati, Italy
[6]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
[7]National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC), Greenbelt, MD, USA
[8]International Institute for Earth System Science, Nanjing University, Nanjing, Jiangsu, China
[a]Now at: Finnish Meteorological Institute, Sodankylä, Finland

Correspondence: Luis Guanter (lguanter@fis.upv.es)

Abstract.

The first satellite-based global retrievals of terrestrial sun-induced chlorophyll fluorescence (SIF) were achieved in 2011. Since then, a number of global SIF datasets with different spectral, spatial and temporal sampling characteristics have become available to the scientific community. These datasets have been useful to monitor the dynamics and productivity of a range of vegetated areas worldwide, but the coarse spatio-temporal sampling and low signal-to-noise ratio of the data hamper their application over small or fragmented ecosystems. The recent advent of the Copernicus Sentinel-5P TROPOMI mission and the high quality of its data products promise to alleviate this situation, as TROPOMI provides daily global measurements at a much denser spatial and temporal sampling than earlier satellite instruments. In this work, we present a global SIF dataset produced from TROPOMI measurements within the TROPOSIF project funded by the European Space Agency. The current version of the TROPOSIF dataset covers the time period between May 2018 and April 2021. Baseline SIF retrievals are derived from the 743–758 nm window. A secondary SIF dataset derived from an extended fitting window (735–758 nm window) is included. This provides an enhanced signal-to-noise ratio at the expense of a higher sensitivity to atmospheric effects. Spectral reflectance spectra at seven 3-nm windows devoid of atmospheric absorption within the 665–785 nm range are also included in the TROPOSIF dataset as an important ancillary variable to be used in combination with SIF. The methodology to derive SIF and ancillary data as well as results from an initial data quality assessment are presented in this work. The TROPOSIF dataset is available through the following Digital Object Identifier (DOI) https://doi.org/10.5270/esa-s5p_innovation-sif-20180501_20210320-v2.1-202104 (Guanter et al., 2021).

1 Introduction

The sun-induced fluorescence (SIF) signal emitted by the chlorophyll-a of terrestrial vegetation has been shown to be a closer indicator of vegetation functioning than other variables traditionally derived from optical remote sensing data (Mohammed et al., 2019). Global retrievals of SIF from space were first achieved in late 2011 from the GOSAT mission (Frankenberg et al., 2011b; Joiner et al., 2011; Guanter et al., 2012). Since then, a number of global SIF datasets have been produced from spaceborne spectrometers originally intended for atmospheric research, such as GOME-2 (e.g. Joiner et al., 2013; Köhler et al., 2014; van Schaik et al., 2020), SCIAMACHY (Köhler et al., 2014; Khosravi et al., 2015; Joiner et al., 2016), OCO-2 (e.g. Frankenberg et al., 2014; Sun et al., 2018), and TanSat (e.g. Du et al., 2018; Yao et al., 2021). The SIF data from those missions have been typically used to investigate the spatial and temporal patterns of the gross primary production (GPP) of different ecosystems (e.g. Frankenberg et al., 2011b; Guanter et al., 2014; Yoshida et al., 2015; Sun et al., 2017; Luus et al., 2017; Walther et al., 2017; Jeong et al., 2017; Smith et al., 2018; Zuromski et al., 2018; Wu et al., 2018), and also ecosystem transpiration and water limitation (Pagán et al., 2019; Maes et al., 2020; Shan et al., 2021).

The main limitations of those SIF datasets for scientific work are the high precision errors and the low spatial resolution and/or sparse spatial sampling. ESA's FLuorescence EXplorer mission, scheduled for launch in >2025, will enable a breakthrough in the spatial resolution of space-based SIF datasets (Drusch et al., 2017). Before that, the advent of the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Copernicus Sentinel-5P mission in October 2017 is already helping to substantially reduce those limitations in spatio-temporal sampling and signal-to-noise ratio. TROPOMI combines a global continuous spatial sampling with a $3.5 \times 7.5 \, km^2$ pixel size at nadir in the near-infrared ($3.5 \times 5.5 \, km^2$ since August 2019) with a daily revisit time, which leads to a large increase in the number of clear-sky measurements per day in comparison to earlier missions. In addition, it measures with a high signal-to-noise ratio, a spectral resolution of 0.37 nm, and a wide spectral coverage in the near-infrared window. These all enable high-performance SIF retrievals (see Fig. 1). The high potential of TROPOMI for SIF monitoring was first anticipated by the sensitivity analysis performed by Guanter et al. (2015), and later demonstrated by the first SIF retrievals from real TROPOMI data recently published by Köhler et al. (2018) and Köhler et al. (2020) (hereinafter referred to as the "Caltech product"). The first publications exploiting TROPOMI SIF data for scientific applications are confirming those high expectations on the use of TROPOMI for vegetation monitoring (Turner et al., 2020; Doughty et al., 2019; Zhang et al., 2019; Yin et al., 2020; He et al., 2020).

Regarding the exploitation of SIF data for scientific applications, the so-called canopy structure effect has to be taken into account to either disentangle the physiological information in SIF from other confounding variables, to normalize the impact of illumination-observation geometries, or to quantify GPP from SIF. The structure of the canopy defines wavelength-dependent multiple scattering and absorption processes, which determine the amount of SIF photons escaping the vegetation cover and hence the SIF flux measured at the satellite level (Joiner et al., 2020). A number of studies are proposing the use of spectral surface reflectance in the red visible (especially, wavelengths in the red) to near-infrared parts of the spectrum to help model those radiative transfer processes (Yang and van der Tol, 2018; Liu et al., 2019; Zhang et al., 2020; Badgley et al., 2017; Dechant et al., 2020) (Yang and van der Tol, 2018; Liu et al., 2019; Zhang et al., 2020; Badgley et al., 2017; Dechant et al., 2020; Yang et al., 2020)
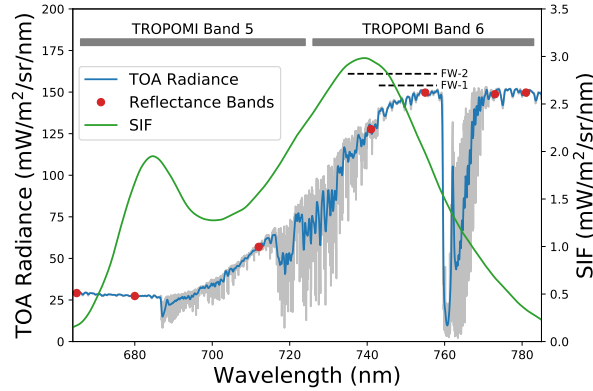
Figure 1. Top-of-atmosphere (TOA) radiance spectrum and top-of-canopy SIF spectrum from a green vegetation surface in the 665–785 nm spectral range covered by TROPOMI's bands 5 and 6. The TOA radiance spectrum at high spectral resolution is plotted in gray, and the result of the convolution with TROPOMI's spectral resolution is shown in blue. The horizontal dashed lines show the span of the baseline SIF retrieval fitting window (~~FT-1~~FW-1, 743–758 nm) and the secondary fitting window (~~FT-2~~FW-2, 735–758 nm). The spectral locations of the 3 nm macro-channels at which spectral reflectance data are provided in the TROPOSIF product are marked in red. A typical SIF spectrum is depicted in green.

. In addition, spectral reflectance from the entire 675–775 nm range can be used to derive indicators of the fraction of absorbed photosynthetically active radiation (FAPAR), leaf-area index (LAI) and other biophysical parameters and spectral vegetation indices. These can complement SIF for the characterization of the vegetation condition and functioning. For example, it has been recently shown that the near-infrared reflectance of vegetation (NIRv) index, or its variation resulting from the multiplication by incoming sunlight (NIRvP), are good proxies for SIF (Badgley et al., 2017; Dechant et al., 2020, 2021). NIRv and NIRvP could then ~~represent a powerful variable~~ provide useful information for the calculation of the SIF yield, which is ~~the variable that can be compared to photosynthetic efficiency in studies dealing with vegetation stress~~ a top-of-canopy variable providing information on both fluorescence yield (leaf-level variable) and any remaining canopy structure effects not accounted for by NIRv and NIRvP. The fact that 675–775 nm spectral reflectance can be derived from the same TROPOMI spectra as SIF (see Fig. 1), and that this reflectance data are provided at a sufficiently high spatial resolution (e.g. similar to some MODIS global vegetation products) is another strength of using TROPOMI data for vegetation monitoring (Guanter et al., 2015).

In this paper, we describe the SIF and spectral reflectance data product derived from TROPOMI in the frame of the TROPOSIF project funded by the Sentinel-5p+ Innovation activity of the European Space Agency (ESA)[1]. Our work is aimed at developing a TROPOMI-based SIF processor which can be implemented at ESA's data processing facilities for the operational generation and distribution of the data product to users. In addition to SIF, reflectance spectra from each input radiance spectrum are also included in the product for combination with the SIF retrievals.

---

[1]https://eo4society.esa.int/projects/sentinel-5pinnovation/

2 Methodology

2.1 SIF retrieval approach

70 The different SIF retrieval schemes implemented for spaceborne measurements in the last years have relied on the Fraunhofer line in-filling principle proposed by Plascyk and Gabriel (1975). This principle establishes that the fractional depth of solar Fraunhofer lines, which are included in the reflected solar light spectrum, decreases when it is combined with an additive signal such as SIF. The feasibility of this concept for satellite-based SIF retrieval was first shown by simulations (Sioris et al., 2003; Frankenberg et al., 2011a) and then applied to all satellite missions for which global SIF data products have been derived.

75 We can group SIF retrieval strategies into physically-based and data-driven methods. The first group of methods are based on a physical formulation of the radiative transfer and the instrument's response for the modeling of top-of-atmosphere (TOA) radiance measurements. These types of methods have been applied to instruments with a high spectral resolution, such as GOSAT and OCO-2 (Frankenberg et al., 2011b; Köhler et al., 2015). In turn, data-driven methods model the input TOA radiance spectra as a linear combination of spectral functions derived from statistical analysis of SIF-free training sets. The

80 performance of these methods is conditioned by the arbitrary selection of the training set and several model parameters, but they are fast and very effective to account for both atmospheric and instrumental effects. Data-driven methods have been used with both high and medium spectral resolution instruments (Guanter et al., 2012; Joiner et al., 2013; Köhler et al., 2014; Joiner et al., 2016; Sanders et al., 2016; Guanter et al., 2015; Köhler et al., 2018; van Schaik et al., 2020).

In this work, we have adapted the data-driven retrieval scheme described in Guanter et al. (2015) for the processing of real

85 TROPOMI data. The forward model takes the form

$$\boldsymbol{F}(\boldsymbol{a}, \boldsymbol{\alpha}, F_{\mathrm{s}}) = \boldsymbol{v}_1 \sum_{i=0}^{n_{\mathrm{p}}} a_i \boldsymbol{\lambda}^i + \sum_{j=2}^{n_{\mathrm{v}}} \alpha_j \boldsymbol{v}_j + F_{\mathrm{s}} \boldsymbol{h}_{\mathrm{F}}, \tag{1}$$

where $\boldsymbol{\lambda}$ is the array of measurement wavelengths used for the representation of spectrally-smooth terms such as surface reflectance and atmospheric scattering, $\boldsymbol{v}$ is the basis of principal components describing the variability in solar irradiance and atmospheric transmittance, $\boldsymbol{a}$ are the coefficients of a polynomial in wavelength, $\boldsymbol{\alpha}$ are the weights of the singular vectors, F_{s}

90 is SIF at the reference wavelength of 740 nm, $\boldsymbol{h}_{\mathrm{F}}$ is a fixed spectral function normalized at 740 nm, which accounts for the spectral shape of SIF, n_{p} is the order of the polynomial used to represent spectrally-smooth terms, and n_{v} is the number of principal components representing high spectral frequency variations. The first two terms on the right hand side of Eq. 1 are a simplification of the product of $\sum_{i=0}^{n_{\mathrm{p}}} a_i \boldsymbol{\lambda}^i$ and $\sum_{j=1}^{n_{\mathrm{v}}} \alpha_j \boldsymbol{v}_j$, which are terms representing contributions of low and high spectral frequencies to the reflected solar radiation, respectively (Guanter et al., 2013, 2015). The third term on the right hand side of

95 Eq. 1 is the SIF contribution to the TOA radiance measurement. The absorption of SIF by the atmosphere between the ground and the TOA can be neglected for fitting windows devoid of strong atmospheric absorption lines such as the ones selected in this work (see Sect. 2.2), which was justified in Guanter et al. (2015). The effect of atmospheric absorption on SIF retrievals at far-red wavelengths had been previously evaluated by means of simulation in Guanter et al. (2012); Frankenberg et al. (2012). The effect should be in the range ∼3–6% for a typical aerosol optical thickness of 0.2 and observation angles between 0° and

100 45°.
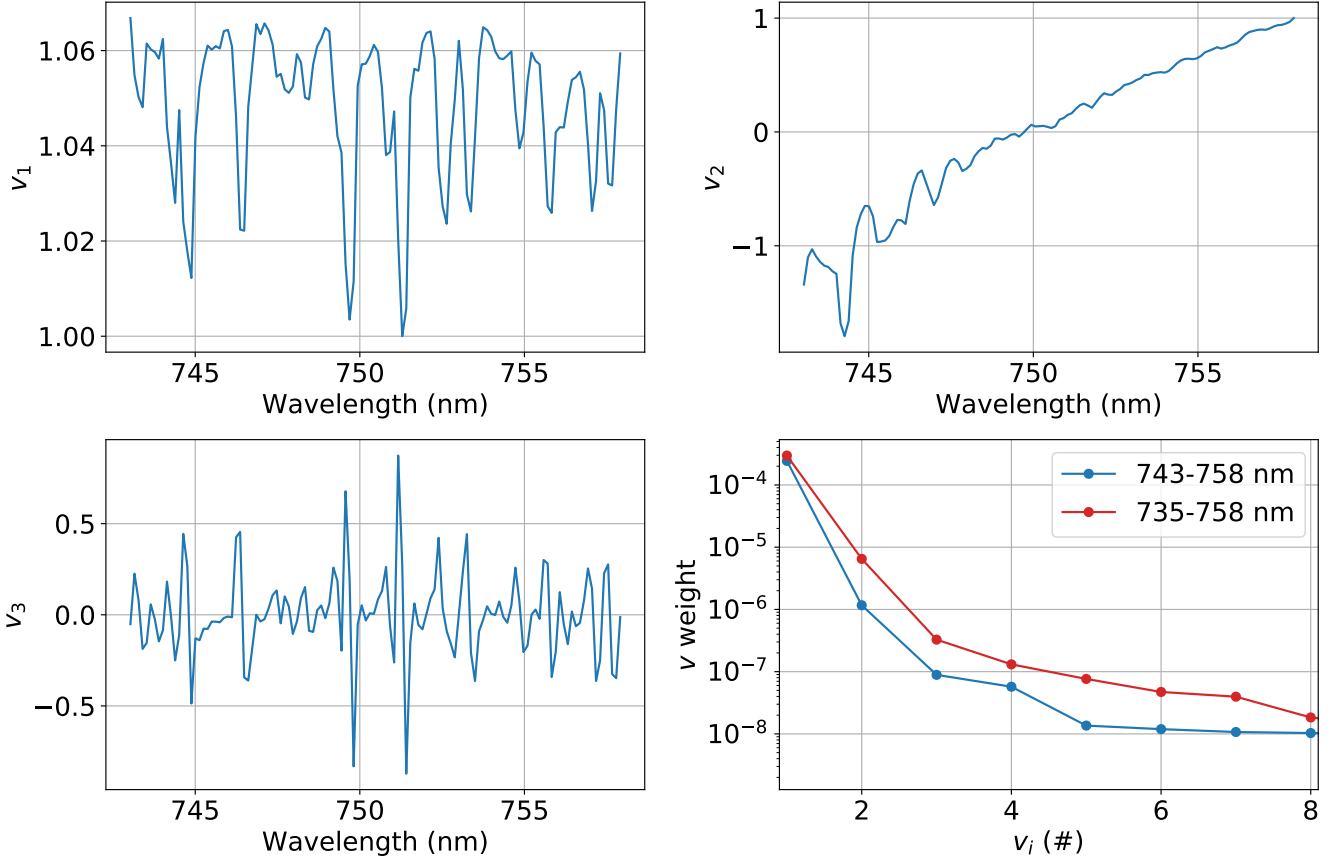
Figure 2. First 3 singular vectors (v_i) and weights of the first ~~10~~ 8 v_i for the 743–758 fitting window from the singular vector decomposition of a training set consisting of TROPOMI spectra over non-vegetated areas. The weights of the first ~~10~~ 8 v_i are also shown for the 735–758 nm fitting window.

The state vector elements to be estimated in the retrieval process are a_i, α_j, and F_s, whereas $\boldsymbol{\lambda}$, $\boldsymbol{v}$, and $\boldsymbol{h}_F$ are model parameters. Regarding these, $\boldsymbol{\lambda}$ is known for each input TOA spectrum. The singular vectors $\boldsymbol{v}$ are calculated through singular vector decomposition of a training set consisting of TOA radiance spectra extracted from measurements over non-vegetated areas (mostly, the Sahara desert and Arctic and Antarctic surfaces). The training is done on a per-column basis in order to account for slight variations in the spectral and radiometric response of different detectors in the across-track direction. Finally, the spectral function $\boldsymbol{h}_F$ is extracted from spectral libraries (Guanter et al., 2013). The forward model is inverted for each TOA radiance spectrum using ordinary least squares.

2.2 Retrieval fitting windows

We have chosen to run the TROPOSIF retrieval for two fitting windows, 743–758 nm and 735–758 nm. The first one is purely based on solar Fraunhofer lines, whereas the second is also slightly affected by water vapor lines in the 735–743 nm range (see Fig. 1). These two windows are selected as good compromises between retrieval noise and sensitivity to cloud contamination, as it was demonstrated in the sensitivity analysis by Guanter et al. (2015): the 735–758 nm window leads to smaller precision errors than the narrower 743–758 nm window because of the larger number of spectral points, whereas the 743–758 nm window is more robust against atmospheric effects due to the absence of atmospheric lines. The 743–758 nm window is therefore a better choice for applications not requiring strict clear-sky observations, as the impact of sub-pixel clouds is lower than for the other window and the greater number of observations available compensate for the higher retrieval random errors; for applications only relying on clear-sky data, the 735–758 nm window would be a better option, as the lower precision errors would compensate for the lower number of available measurements and the higher sensitivity to sub-pixel clouds would not be relevant.

The number of singular vectors n_v is set to 4 for the 743–758 nm fitting window and 7 for 735–758 nm. This choice is based on the results from the sensitivity analysis by Guanter et al. (2015) and recent tests with real TROPOMI data. Results from those tests can be seen in Fig. 2, which shows an abrupt drop in the weight of the v after 4 for the 743–758 nm window and 7 for the 735–758 nm window. On the other hand, n_p is set to 3 in both cases. We have not found major differences in the retrieval accuracy and precision when using a greater n_v or n_p.

An example of the three first v_i and their weights for the 743–758 nm fitting window derived from the singular vector decomposition of a particular training set are displayed in Fig. 2. It can be observed that v_1 carries most of the weight (i.e. reproduces most of the variance of the training set), whereas the others may account for low spectral frequency variations in albedo (e.g. v_2) or measurement artifacts like spectral shifts (e.g. v_3). This justifies the choice of only multiplying v_1 by the polynomial in wavelength accounting for low-frequency spectral variations in Eq. 1.

2.3 Retrieval random error

As it was described in Guanter et al. (2015), the retrieval error covariance matrix $\mathbf{S}_e$ is given by

$$\mathbf{S}_e = \left(\mathbf{J}^T \mathbf{S}_0^{-1} \mathbf{J} \right)^{-1}, \tag{2}$$

where $\mathbf{S}_0$ is the measurement error covariance matrix and $\mathbf{J}$ the Jacobian matrix, which in the particular case of the retrieval forward model given in Eq. 1 consists of the terms $\mathbf{J}(a_i) = v_1 \lambda^i$, $\mathbf{J}(\alpha_j) = v_j$ and $\mathbf{J}(F_s) = h_F$. The last element in the diagonal of the $\mathbf{S}_e$ matrix contains the squared random error of the retrieved SIF value.

As a result, the retrieval random error for F_s only depends on measurement noise (i.e. the noise in the input radiance spectra), which in turn is driven by incoming radiance through photon noise. For the robust and computationally-efficient calculation of retrieval errors, empirical radiance-σ(SIF) curves have been produced for each fitting window. The 1-σ retrieval errors were first calculated by means of Eq. 2 for several orbits and grouped into TOA radiance bins of $5\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$. A 3rd-order
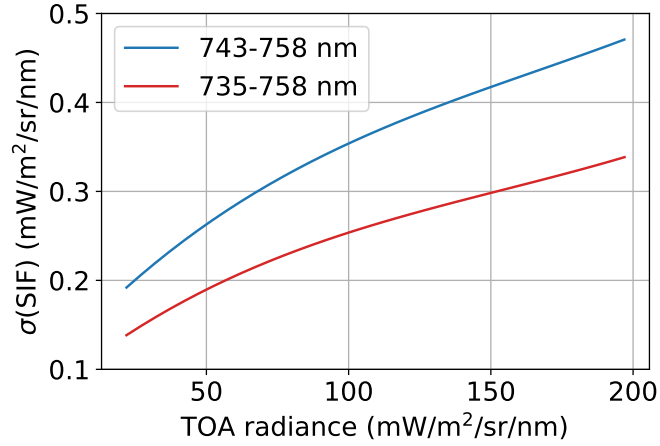
Figure 3. Empirical radiance-σ(SIF) curves used for the calculation of 1-σ SIF retrieval errors in the operational processing.

140 polynomial was then fitted to the binned radiance-σ(SIF) data. The result of this process is shown in Fig. 3. The resulting polynomial coefficients are used to predict σ(SIF) as a function of average radiance on a per-pixel basis. With this strategy we avoid the inversion of the measurement error covariance matrix in Eq. 2, which had proved to be unstable and computationally-expensive. It must be mentioned that the measurement noise data provided in TROPOMI's Level 1B (L1B) files suffer from float number truncation errors from the storage as byte data. This makes that the σ(SIF) that we calculate from the measurement

145 noise data can affected by the same effect. This manifest as a underestimation of the σ(SIF) values of a varying magnitude (from 0 to $\sim$15%) along the radiance-σ(SIF) curve.

2.4 Day-length scaling

The variability in SIF retrievals due to instantaneous illumination conditions must be accounted for when SIF retrievals are to be compared with other SIF datasets acquired with a different satellite overpass time, or with daily GPP estimates (Zhang

150 et al., 2018).

We have followed the approach proposed by Frankenberg et al. (2011b) to generate daily-equivalent SIF (SIF_d) estimates in TROPOSIF. This is based on the calculation of a day-length scaling factor DL as

$$DL(t_0) = \frac{\int_{t_{sr}}^{t_{ss}} \mu_s(t)\,\mathrm{d}t}{\mu_s(t_0)}, \tag{3}$$

where t_0 is the time of the measurement, t_{sr} is the time of sunrise, t_{ss} is the time of sunset, and $\mu_s(t)$ is the cosine of the sun

155 zenith angle. This approach assumes that SIF scales linearly with incoming solar radiation, and that the entire day is cloud-free.

2.5 Quality value for SIF retrievals

An empirical quality value (`qa_value`) indicating the reliability of each SIF retrieval is included in the TROPOSIF product. This value is a score between 0 and 1 (from lowest to highest quality) calculated as the combination of a series of empirical penalty factors applied in those conditions which are expected to potentially degrade the retrieval quality.

Starting from a maximum `qa_value` of 1.0, penalty factors are applied on the view zenith angle (VZA), the sun zenith angle (SZA), and the reduced χ^2 of the fit (χ_r^2) as follows:

1. VZA$> 60° \rightarrow$ `qa_value=qa_value`-0.5

2. SZA$> 70° \rightarrow$ `qa_value=qa_value`-0.5

3. average TOA radiance $\notin [20, 200]\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}} \rightarrow$ `qa_value=qa_value`-0.5

4. $\chi_r^2 \notin [0.6, 2] \rightarrow$ `qa_value=qa_value`-1.0

5. SIF $\notin [-10, 10]\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}} \rightarrow$ `qa_value=qa_value`-1.0

If the final `qa_value` becomes <0, it is reset to 0.

Cloud contamination is not part of the computation of the `qa_value`. Cloudy observations are included in the generation of the singular vectors used by the forward model, so the latter is generally expected to properly represent cloudy spectra. In those cases in which clouds do degrade the fit quality, this will be captured in the χ_r^2 value. Further filtering by cloud fraction is possible for those applications in which only cloud free acquisitions are used.

Only SIF retrievals with `qa_value=1.0` (equivalently, `qa_value>0.5`) are recommended for use. In practice, this means that the data point should be discarded if any of the previous conditions applies.

2.6 Spectral reflectance

Spectral reflectance spectra at 7 spectral points in the 665–785 nm range covered by TROPOMI's band 5 and 6 (namely, 665, 680, 712, 741, 755, 773 and 781 nm) are included in the TROPOSIF product as an ancillary dataset. Spectral reflectance spectra are calculated as

$$\rho_{\mathrm{TOA},i} = \frac{\pi \langle L_{\mathrm{TOA}} \rangle_i}{\mu_{\mathrm{s}} \langle I_{\mathrm{sc}} \rangle_i}, \tag{4}$$

where L_{TOA} refers to the TOA radiance measurement (L1B radiance spectra), I_{sc} is an extraterrestrial solar irradiance spectrum corrected by the Sun-to-Earth distance at the day of the overpass, and $\langle \rangle_i$ refers to the spectral convolution operator, which in this case corresponds to a boxcar spectral response function with a 3 nm width, applied at the ith spectral point. The *Kurucz 2005* high resolution irradiance spectrum[2] (from 300 to 1000 nm, with $\sim$0.001 nm sampling at 740 nm) resampled at the 7 spectral points with the same 3-nm boxcar function is used for $\langle I_{\mathrm{sc}} \rangle$. This approach is computationally much simpler than the

---

[2]http://kurucz.harvard.edu/sun/irradiance2005/irradthu.dat

alternative of using TROPOMI's solar irradiance product, and still yields a sufficient accuracy because of the reduce spectral
resolution of the $\rho_{\mathrm{TOA},i}$ calculations.

The 7 spectral points are selected so that they provide both a proper sampling of the red, red-edge and near-infrared parts of
the spectrum and a low sensitivity to atmospheric effects (see Fig. 1). This is important because no atmospheric correction is
performed on the spectral reflectance spectra.

2.7 Processing flow

TROPOSIF's processing chain starts by reading in TOA radiance data from TROPOMI band 5 and 6 L1B orbit files as
well as cloud fraction data from the TROPOMI L2 Cloud product. Data acquired over water bodies (identified using the
MODIS MCD12C1 2018 land cover product), with a high cloud fraction (> 0.8) or with a lower quality according to the
`quality_level` flag attached to the L1B data (threshold of 80) are screened out from the processing. Spectral band #179
in TROPOMI's channel 6 is also removed from the processing, since spectral spikes were found for that band in pixels at the
vicinity of clouds.

The first step in the processing is the generation of the singular vectors. For that, spectra over barren areas worldwide as
identified in the MODIS MCD12C1 2018 dataset are used. Next, TOA radiance data from either a continuous fitting window
(for SIF retrieval) or at certain "macro-channels" at atmospheric windows within the far-red region (for reflectance retrieval)
are read in. SZA and latitude and longitude fields are also extracted for the computation of the DL factor. SIF and reflectance
retrievals are run separately for each column in order to account for column-dependent changes in TROPOMI's spectral and
radiometric responses. The output SIF, spectral reflectance and DL are written to a NetCDF file together with other variables
of interest for later data visualization and processing. This NetCDF file has the same structure as the official TROPOMI L2
products (see Appendix A).

2.8 L2B processing

In order to facilitate data download and processing, a final processing step extracts all valid retrievals (`qa_value`>0.5) from
the L2 orbit files in a given day and combines them into single daily files, referred to as L2B files. The application of the
`qa_value` filter substantially reduces the volume of data in the L2B file with respect to the combined L2 dataset. In the case
of spectral reflectance data, only spectra acquired under a cloud fraction smaller than 0.2 are included. For the sake of reducing
the data volume, some information fields in the L2 files are not included in L2B files, such as the χ_r^2 values, the `qa_value`,
the day-length factor and the illumination and observation azimuth angles (which are converted into a single relative azimuth
angle).
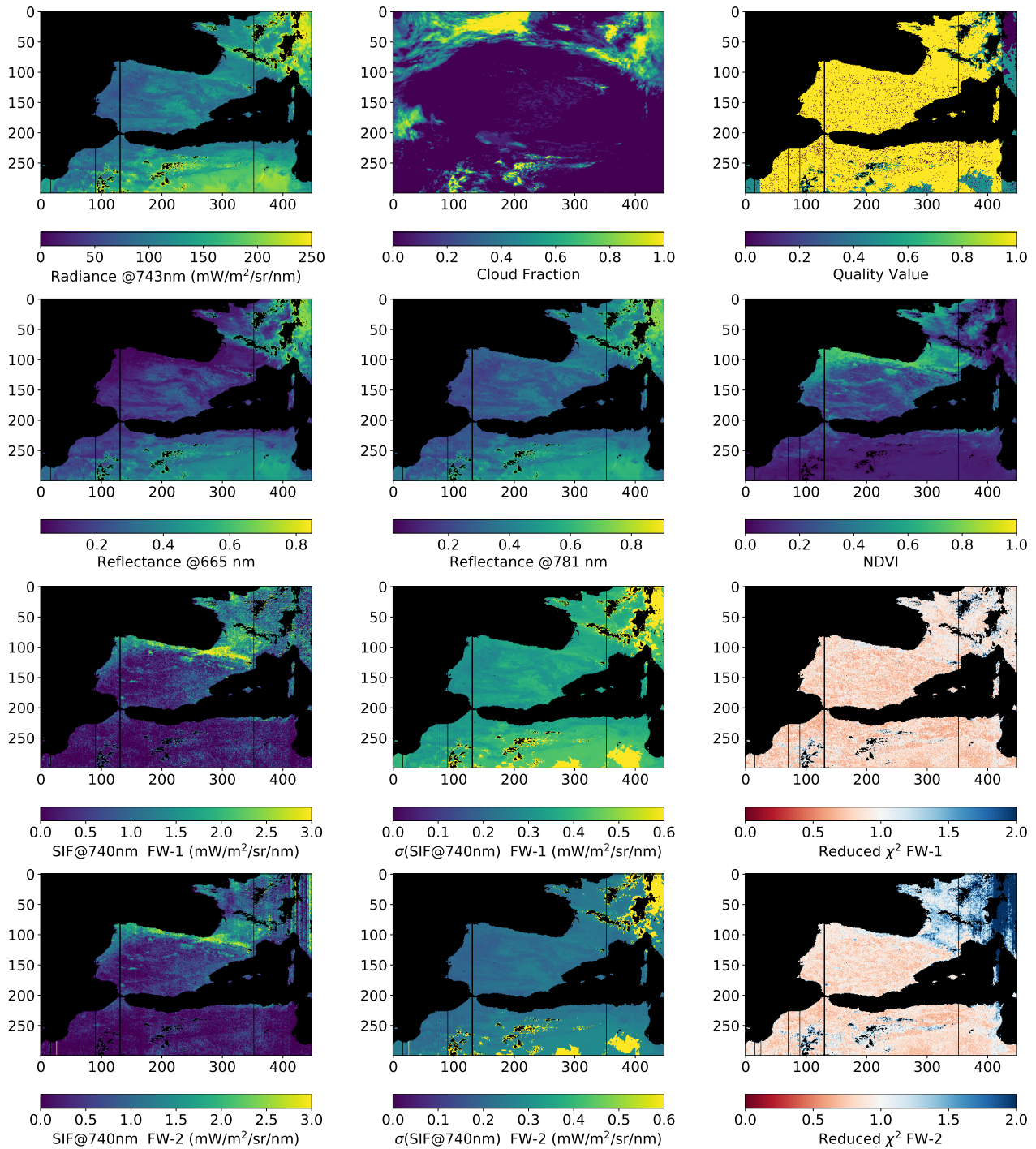
Figure 4. Subsets of the input data (TOA radiance at 743 nm and cloud fraction) and the results obtained from the processing of TROPOMI's 09025 orbit (11 July 2019) with the TROPOSIF SIF processor. FW-1 refers to the 743–758 nm fitting window, and FW-2 to 735–758 nm.
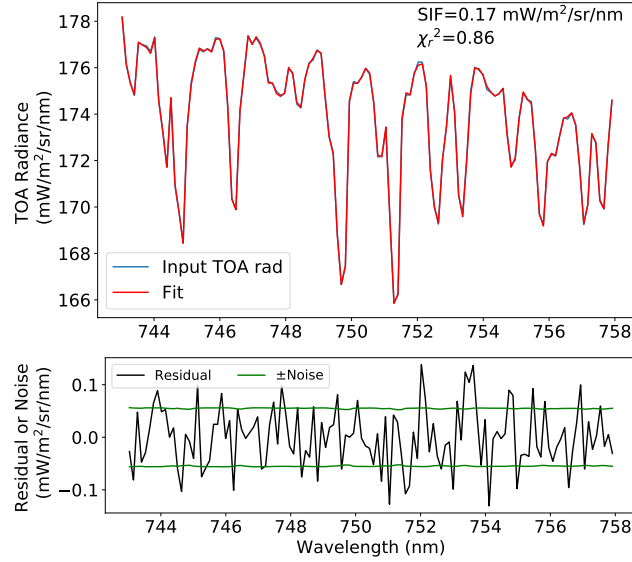
Figure 5. Example of spectral fit for the 743–758 nm fitting window. The fit residual is compared to the noise level included in the input L1B dataset for the same spectrum. The resulting SIF at 740 nm and χ_r^2 of the fit are shown in the upper panel.

3   Results

3.1   Consistency of SIF retrievals

Maps from spatial subsets of different variables derived from the processing of one TROPOMI orbit dataset are displayed in Fig. 4. The different panels show the spatial distribution of TOA radiance at 743 nm, the cloud fraction in TROPOMI's L2 Cloud product, `qa_value`, the TOA reflectance at 665 nm and 781 nm, the normalized difference vegetation index (NDVI) derived from those two channels, as well as the retrieved SIF, 1-σ error and the χ_r^2 of the fit for both 743–758 and 735–758 nm fitting windows. χ_r^2 values larger (smaller) than 1 indicate an underestimation (overestimation) of the model-data error variance.

The following specific points can be highlighted from Fig. 4:

- The areas with the highest abundance of green vegetation (e.g. the northern Iberian Peninsula, South West France, Brittany and Normandy) are clearly visible in the SIF and NDVI maps. In the case of SIF, the range of values is within the expectation for both fitting windows. Slightly higher values are found for the 743-758 nm fitting window in clear-sky regions.

- Single-retrieval uncertainties are also within the expected range (0.3–0.6 mW m^{-2} sr^{-1} nm^{-1}). They are substantially lower for the 735-758 nm window.

- For both fitting windows, the χ_r^2 is typically around 0.8-0.9. This deviation from an ideal case of $\chi_r^2{=}1$ would indicate that the model is slightly over-fitting the data ($\chi_r^2 > 1$ implies that the model does not model the full variance in the spectrum). However, the 0–15% truncation errors in the measurement noise input data discussed in Sect. 2.3 could also explain that χ_r^2 is in general smaller than 1. On the other hand, the sensitivity to clouds of the 735-758 nm retrieval can also be noticed in the χ_r^2 maps. Relatively high χ_r^2 values ($>$1.5) are found for retrievals over cloudy areas for 735-758 nm retrievals, which is explained by the existence of water vapor absorption lines in the 735–743 nm range. No relevant differences in χ_r^2 between clear and cloudy areas are found for the 743-758 nm window. More details on the fit performance can be found in Fig. 5.

- The quality value map indicates that the retrievals are in general performed under the best retrieval conditions (`qa_value=1.0`). Lower quality retrievals are found for the edges of the subset where VZA$> 60°$, for some patches in the Sahara desert where the mean TOA radiance is greater than the $200\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$ threshold, and for the cloudy areas on the top-right of the sampled area for which $\chi_r^2 > 2$.

- Even though the processing is performed on a per-column basis, detector artifacts in the form of striping are present in the SIF maps, which can be explained by an across-track dependence of TROPOMI's spectral and radiometric calibration. In particular, the processing has not run over some across-track columns which did not pass the `quality_level`> 80 filter, and higher uncertainty values are found at the swath edges.

3.2 Spatio-temporal composites of the all-sky SIF product

A global composite of SIF data from the 743–758 nm fitting window in a 0.2° latitude-longitude grid for the 8–15 July 2019 period and a cloud fraction lower than 0.5 is presented in Fig. 6. A total of 102 orbit files have been used for the composite. Prior to cloud filtering, retrievals are filtered using the recommended `qa_value`>0.5, which corresponds to VZA$< 60°$, average TOA radiance between 20 and $200\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$, and χ_r^2 between 0.6 and 2.5 for consistency with the filtering scheme proposed by Köhler et al. (2018). The number of retrievals used in each 0.2° gridbox and the resulting standard error of the mean (σ), calculated as described in Guanter et al. (2015), are shown in the smaller panels of Fig. 6.

The resulting 8–15 July 2019 SIF map shows the expected SIF global maximum at the U.S. Corn Belt (Guanter et al., 2014). High SIF values are also found in Central America, Central Asia and China. The data gap in the southern part of Greenland is due to the incoming radiance being higher than the $200\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$ threshold used for quality filtering, and the one in Southern Argentina is due to SZA$>70°$. The effect of cloud filtering on the standard errors of SIF composites is shown in the σ(SIF) map. It is especially noticeable in regions with persistent cloud cover such as the tropics and the northern part of Europe and America. Overall, typical standard errors are in the range 0.1–$0.2\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$ for this weekly composite.

The impact of cloud fraction thresholds on SIF composites (systematic and random errors) has been further investigated. Fig. 7 displays the differences in SIF and σ(SIF) composites (8-15 July 2019, 743–758 nm fitting window) using either a strict or a relaxed cloud fraction threshold (0.2 and 0.8, respectively). As expected, the absolute SIF values increase with the strictness of the cloud filter, as clouds have a shielding effect for SIF photons travelling from the surface to the sensor. At
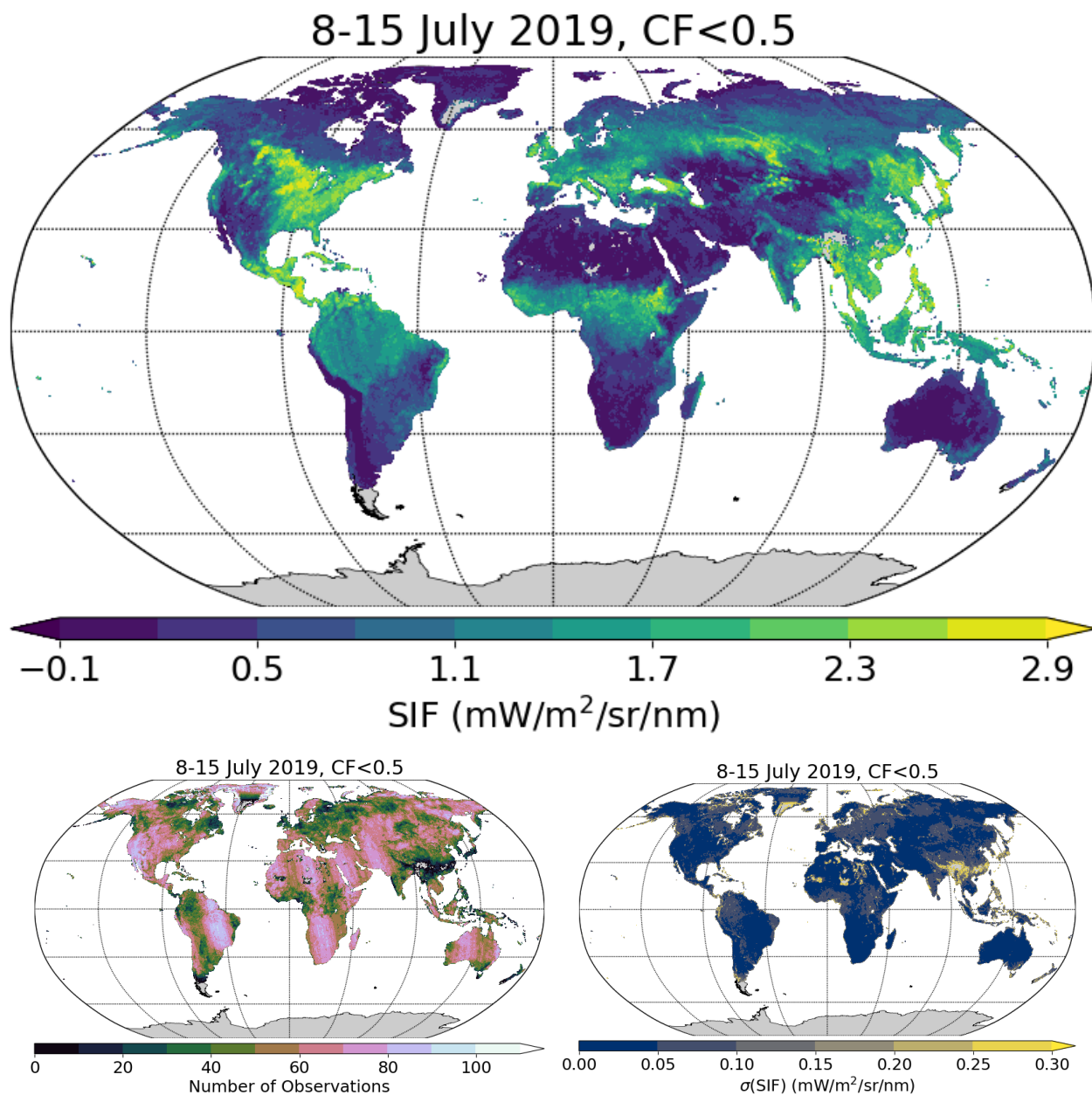
Figure 6. Global composites of SIF at 740 nm for the time period 8-15 July 2019 using retrievals from the 743–758 nm fitting window and a cloud fraction (CF) threshold of 0.5. Data are gridded in $0.2°$ gridboxes. The number of observations and the resulting standard error of the mean per gridbox are shown in the lower panels.
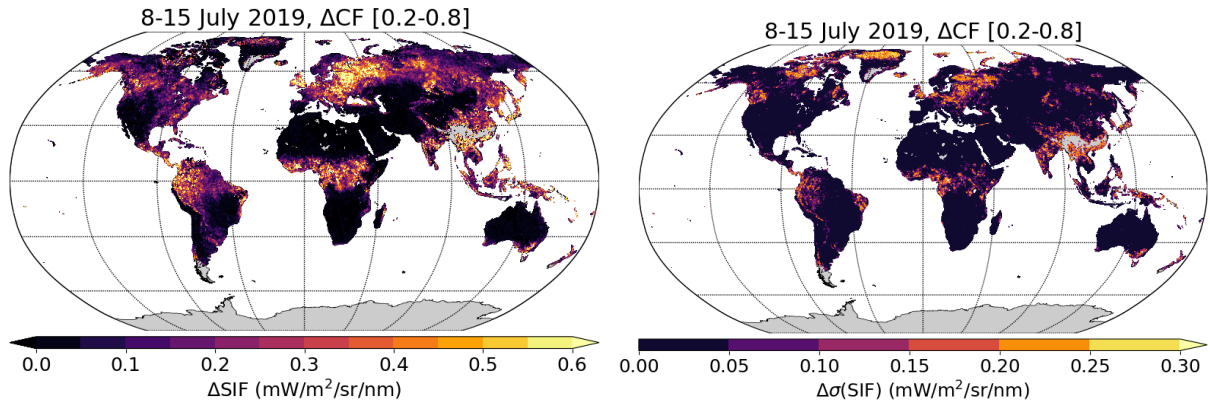
Figure 7. Differences of average SIF (left) and σ(SIF) (right) for 0.2 and 0.8 cloud fraction thresholds. Data are from the time period 8-15 July 2019 and the 743–758 nm fitting window. ~~Panel Size?~~

260 the same time, the standard errors of SIF composites decrease with the cloud fraction threshold following the increase in the number of retrievals per gridbox. For the selected 8-15 July 2019 time period, the greatest impact of clouds in SIF and σ(SIF) is found in Western Europe. In this region, the change from 0.8 to 0.2 cloud fraction thresholds implies an increase in the SIF average of about 0.4–0.6 mW m^{-2} sr^{-1} nm^{-1} at the gridbox level, which can be understood as a systematic error in the SIF composite due to clouds. For the same region, the change of the standard error due to the lower number of retrievals in

265 the stricter cloud filtering is about 0.25 mW m^{-2} sr^{-1} nm^{-1}. This illustrates that the selection of the optimal cloud fraction threshold depends on the particular use of the data. A strict cloud filtering (e.g. cloud fraction lower than 0.2) should be applied if an accurate SIF mean value is needed, whereas a more relaxed cloud filter (e.g. cloud fraction lower than 0.8) could be applied if smooth spatial and/or temporal signals are needed. The second case is often preferred for SIF, but it must be taken into account that unfiltered clouds can also introduce noise-like changes in SIF time series. Including cloud fraction data in the

270 analysis would be needed for a proper interpretation of the SIF signals if a relaxed cloud filtering was applied.

The seasonal variation of TROPOSIF SIF data is illustrated in Fig. 8, which shows daily and 0.1° composites over a set of instrumented SIF sites (Parazoo et al., 2019) and the entire time period covered by the TROPOSIF product. Two configurations of fitting window and cloud fraction thresholds are used: 743–758 nm with cloud fraction <0.8 ("all-sky") and 735–758 nm with a cloud fraction <0.2 ("clear-sky") The higher number of daily estimates available for the all-sky case is a result of less

275 strict cloud filtering. Figure 8 illustrates the lower retrieval uncertainty associated with the clear-sky case (lower error bars and less scattering) compared to the all-sky SIF. SIF data derived from the 743–758 nm fitting window are usually greater than those derived from the broader 735–758 nm window, as indicated by the slope of the linear fit in the scatter plots on the right hand side of Fig. 8. On the other hand, negative SIF values are found for the two fitting windows. These are caused by noise in the data being propagated to random errors in the retrieved SIF. Removing negative SIF retrievals from the processing should

280 be avoided, as it would lead in positively-biased averages.

It must be stated that the slopes of the scatter plots in Fig. 8 cover the minimum and maximum slope values found in the analysis of 13 different sites (roughly, 1.0 to 1.5). Further research is needed to understand whether the variations in the slopes are due to retrieval biases over some vegetation types or to leaf/canopy radiative transfer effects making the shape of the SIF emission to depend on the leaf and canopy type. A dependence of SIF retrievals on the fitting window and atmospheric absorption is also reported in Parazoo et al. (2019). Regarding potential canopy-dependent retrieval biases, we have observed that the 735–758 nm retrieval produces lower quality fits ($\chi_r^2 > 2$) much more often than the 743–758 nm fitting window. This is especially the case for densely vegetated areas with fully developed plants, such as the U.S. Corn Belt in July. This effect accounts for the lower density of 735-758 nm SIF points in the Coles corn field during the summer season (first panel of Fig. 8). The reason might be in the limitation of the retrieval forward model to represent the steep reflectance spectra for high LAIs. Further tests are needed to fully characterize and solve this issue.

3.3 Comparison to Caltech's TROPOMI and OCO-2 SIF products

The quality of the TROPOSIF SIF product has been investigated through the comparison to the Caltech TROPOMI SIF product (Köhler et al., 2018) (hereinafter, "Caltech") and SIF retrievals derived from OCO-2 (Sun et al., 2018). Using the latter products as benchmark datasets, TROPOSIF SIF data have been evaluated in terms of absolute SIF values, systematic errors (biases) and random errors (noise).

Comparisons between the TROPOSIF (743-758 nm), Caltech TROPOMI, and OCO-2 SIF estimates were performed at the global scale for the evaluation of SIF absolute values. The instantaneous SIF retrievals were binned at $0.1°$ and daily resolutions, with a cloud fraction below 0.2 for those derived from TROPOMI observations. The pixel-to-pixel comparison between each pair of products shown in Fig. 9 highlights their good agreement. Over the entire year 2019, the highest agreement is found between the two TROPOMI products (TROPOSIF and Caltech), with the highest correlation (greater than 0.93), and a root mean square deviation of about $0.09 \, \mathrm{mW \, m^{-2} \, sr^{-1} \, nm^{-1}}$, while the root mean square deviation between either Caltech or TROPOSIF with OCO-2 is $0.17 \, \mathrm{mW \, m^{-2} \, sr^{-1} \, nm^{-1}}$.

The distribution of the differences between Caltech and TROPOSIF data is slightly skewed towards negative values, indicating that TROPOSIF estimates tend to be greater than Caltech SIF ones. The typical bias between TROPOSIF and Caltech is about $0.008 \, \mathrm{mW \, m^{-2} \, sr^{-1} \, nm^{-1}}$ (for the daily-corrected product). Both TROPOSIF and Caltech estimates are slightly lower than OCO-2 at 740 nm data: The typical bias between TROPOSIF and OCO-2 is about $0.029 \, \mathrm{mW \, m^{-2} \, sr^{-1} \, nm^{-1}}$, which is lower than the mean bias between Caltech and OCO-2 ($0.036 \, \mathrm{mW \, m^{-2} \, sr^{-1} \, nm^{-1}}$). The difference between the Caltech and TROPOSIF SIF estimates is due not only to the difference in the estimation algorithm, but also to the different cloud products that are used as input: concomitant cloud fraction data from TROPOMI's L2 Cloud product for TROPOSIF, when Caltech relies on cloud data derived from VIIRS (Köhler et al., 2018).

The existence of potential biases in the TROPOSIF SIF product has firstly been investigated using retrievals over large non-vegetated areas where it can be assumed that no SIF emission is present (i.e. Sahara, Antarctica and Greenland). The analysis has been conducted for the TROPOSIF products inferred from the two fitting windows, as well as for the Caltech SIF product and for SIF retrievals derived from OCO-2, which both act as benchmark datasets. For each SIF product, the
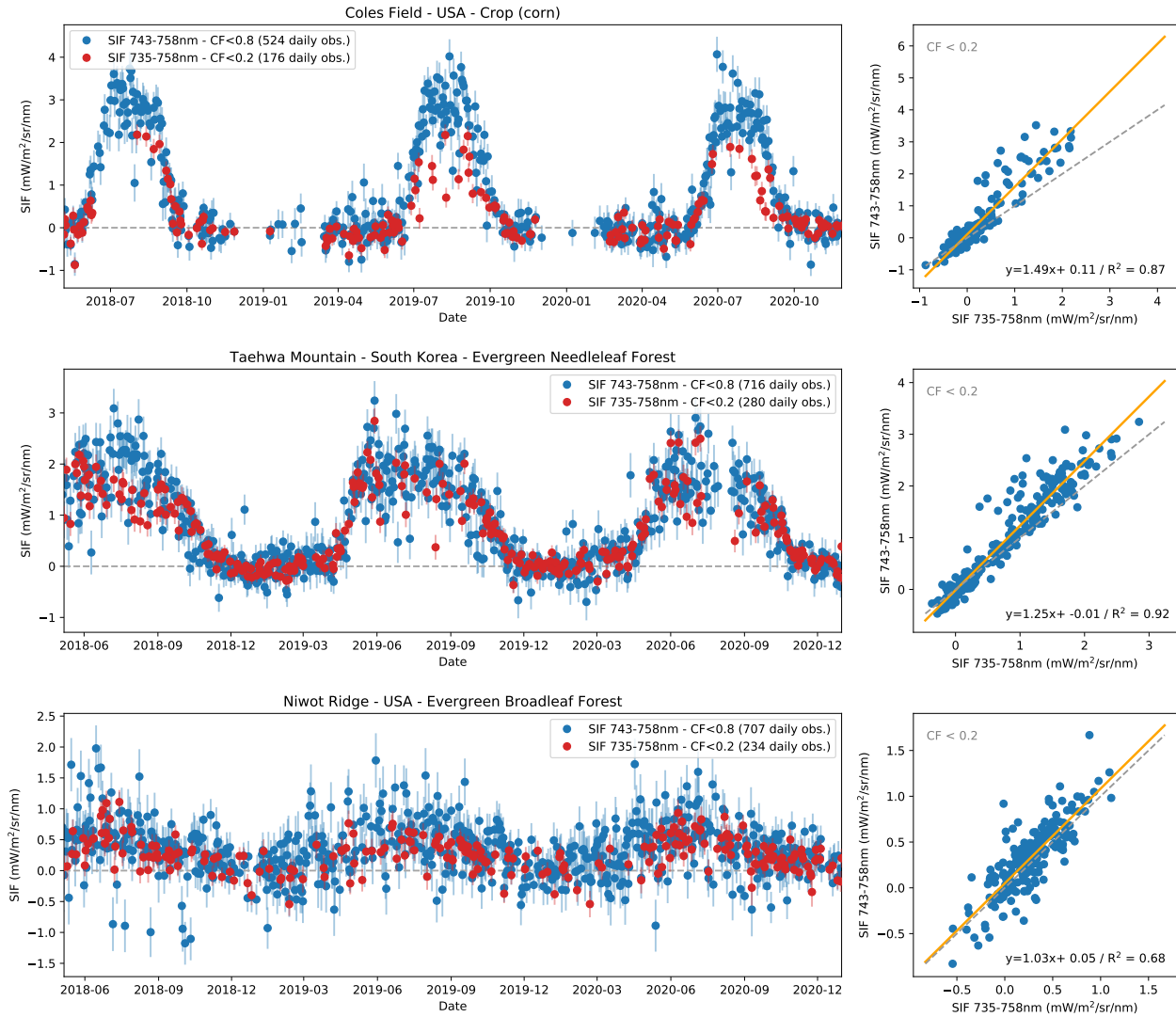
Figure 8. Comparison of SIF retrievals from the 743-758 nm and 735–758 nm fitting windows and 0.2 and 0.8 cloud fraction (CF) thresholds at three sites (0.1° radius) corresponding to different biomes (Parazoo et al., 2019). Left: time series of the daily SIF retrievals for two combinations of fitting window and cloud screening (743–758 nm data with CF<0.8 and 735–758 nm with CF<0.2) from 05-2018 to 12-2020. The vertical bars are the averaged 1-σ retrieval errors. Right: Scatterplot between 735–758 nm and 743–758 nm retrievals for CF<0.2. The coefficient of determination R^2 and the linear regression equation are provided.

315   retrievals for year 2019 were binned at 0.1° at a daily scale, screening measurements with cloud fraction above 0.05, and extracted over the regions of interest. OCO-2 SIF data were scaled at 740 nm using the approach proposed by Köhler et al. (2018). The distributions of the different SIF product values agree well over the regions considered albeit a lesser consistency
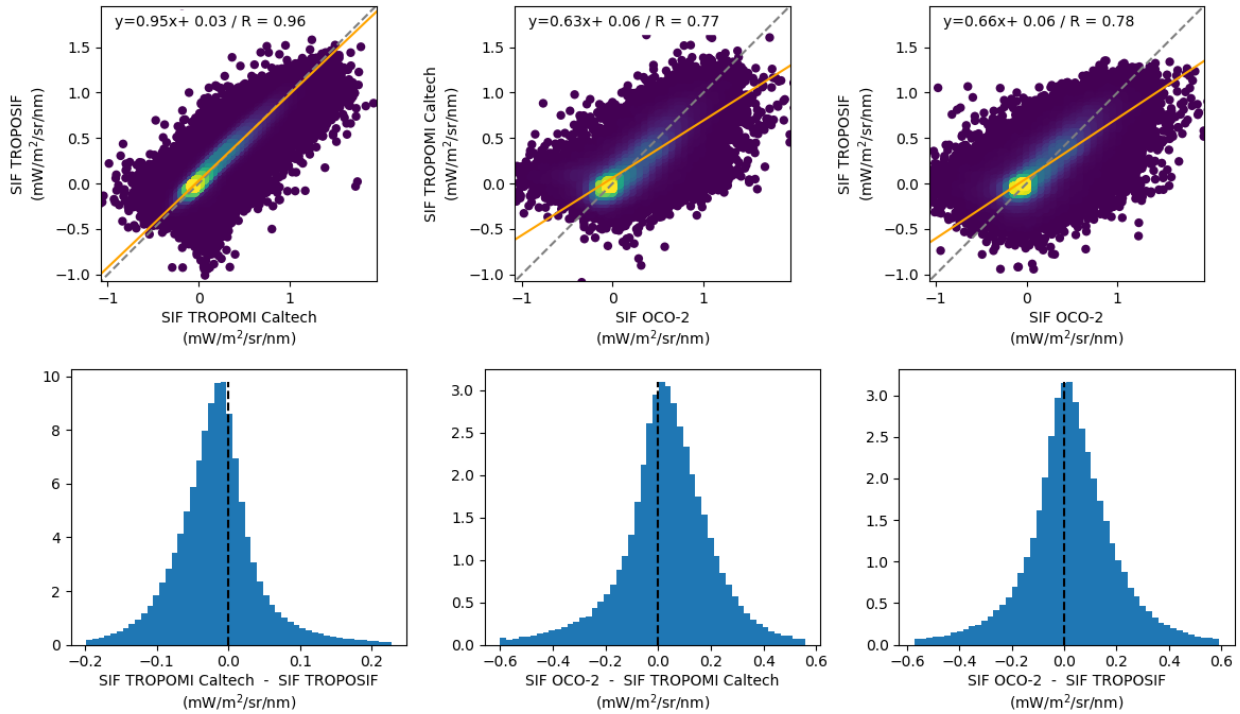
Figure 9. Scatterplots (top row) and histograms (bottom row) of the difference between each pair of SIF dataset, namely TROPOSIF (743–758 nm window), Caltech TROPOMI and OCO-2 for August 2019 (daily/0.1° data, CF<0.2). The linear regression line between two datasets is shown in orange. The linear regression equations and the correlation coefficient R are provided. A daylength-correction has been applied to all three datasets.

of the OCO-2 data with the TROPOSIF and Caltech estimates is observed over Greenland. It is attributed to fewer available OCO-2 observations (hence a lower decrease of the random error for each binned data compared to TROPOMI products, and an increased mismatch in spatio-temporal sampling between OCO-2 and TROPOMI observations). Over the Sahara, where a higher number of measurements are available, we estimated the mean yearly bias for all products. The mean bias for the various TROPOMI SIF estimates is slightly negative: -0.006 mW m^{-2} sr^{-1} nm^{-1} for the daily-corrected TROPOSIF data derived from the 735-758 nm window (-0.017 for the "raw" SIF data), -0.028 (-0.080) for the 743–758 nm TROPOSIF product, and -0.036 (-0.105) for the Caltech product. The mean bias for OCO-2 was found to be slightly positive: 0.010 (0.031). All SIF products exhibit a seasonal variation of the mean bias over the different regions (note there are no OCO-2 observations over Antarctica). Over Sahara, the magnitude of these seasonal variations is about 0.1 mW m^{-2} sr^{-1} nm^{-1}. The origin of this seasonality is still unclear but it could partly be attributed to directional effects. We discarded that it was the result of a temporal drift by checking that the level of the SIF biases estimated for June 2018 and June 2020 was similar to the one estimated for June 2019.
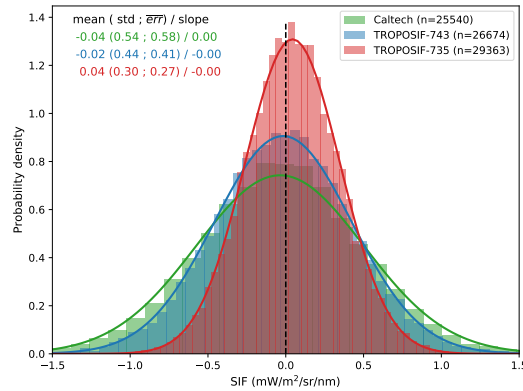
Figure 10. Histogram of all available TROPOMI-based SIF estimates at ten PICS over the 2018-2020 period, for Caltech SIF and TROPOSIF SIF estimates from the 743–758 nm and 735-758 nm fitting windows. The distributions were fitted using a Gaussian function; the corresponding mean (bias) and standard deviation (std) are provided for each SIF product. The mean slope of the SIF temporal variations over the ten PICS is given in brackets.

We completed this analysis by characterizing the distribution of the SIF estimates from TROPOSIF and Caltech at a set
330   of Pseudo Invariant Calibration Sites (PICS). These sites are selected because of their radiometric temporal stability, spatial homogeneity, and high reflectance, and are used for vicarious calibration of space-borne sensors. Ten sites were considered: the six endorsed by the CEOS/WGCV/IVOS as Calibration/Validation test sites (Algeria3, Mauritania1, Libya4, Mauritania2, and Algeria5, which where identified by Cosnefroy et al., 1998), plus four new PICS located in Algeria, Saudi Arabia, Namibia and Sudan (Bacour et al., 2019a). The distribution of the daily SIF values for the period 2018-2020 over these sites (data within
335   the vicinity of 0.1°) is seen in Fig. 10 for the Caltech product and the two TROPOSIF products. The mean bias is close to $0\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$ for the three products. The smaller width of the histogram for the TROPOSIF product from the 735—758 nm window is due to the smaller single-retrieval errors. No temporal drift was detected in the SIF 2018–2020 time series for any of the sites.

Over Sahara, we quantified the "averaged value" of the 1-σ retrieval error (square root of the quadratic mean) for the two
340   TROPOSIF products and the Caltech retrievals (0.1°/daily). The retrieval error for TROPOSIF is typically $0.5\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$ for the 743-758 nm fitting window estimates and $0.4\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$ for the product derived from the 735–758 nm window. These empirical values are consistent with the theoretical retrieval errors, although on the higher end. We found an typical retrieval error of $0.6\,\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$ for the Caltech product.
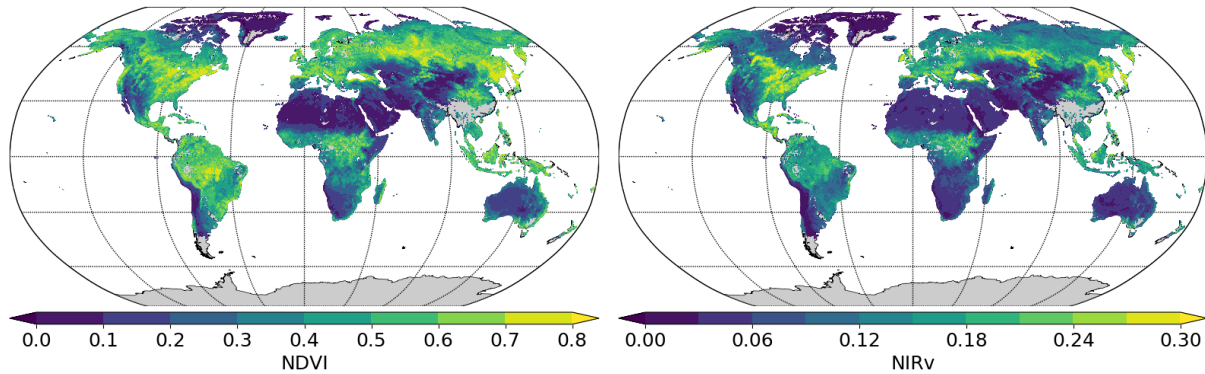
Figure 11. Global composites of NDVI and NIRv derived from the TROPOSIF product for the 8–15 July 2019 time period. Data are gridded in 0.2° gridboxes. A cloud fraction threshold of 0.1 is applied for cloud screening.

3.4 Evaluation of reflectance-based indices

345 One of the assets of the TROPOSIF product is the 665–785 nm spectral reflectance spectra attached to the SIF retrievals. Spectral reflectance data can be used to derive empirical indices and biophysical variables which are complementary to SIF, as discussed earlier in this paper.

Global composites of two of such empirical indices (NDVI and NIRv) for the 8–15 July 2019 time period are shown in Fig. 11. NDVI is calculated using the 665 and 781 nm channels as red and near-infrared bands, respectively. NIRv is calculated 350 as the product of NDVI and near-infrared reflectance (781 nm channel) (Badgley et al., 2017). The comparison of the spatial patterns in these two maps with those of SIF in Fig. 6 confirms the high correlation between SIF and NIRv that has already been reported in recent studies (Badgley et al., 2017; Zeng et al., 2019; Mengistu et al., 2020). In contrast, the dynamic range of NDVI appears to be smaller than that of NIRv and SIF, since several densely vegetated areas around the world reach the highest NDVI values.

355 These patterns are further illustrated in Fig. 12, which shows a comparison between SIF and vegetation indices for an ensemble of 0.5° pixels representative of several biomes. The pixel selection is described in Bacour et al. (2019b). The NIRvP is added to the analysis. NIRvP is calculated as the product of NDVI and the average TOA radiance in the SIF retrieval fitting window (also attached to the TROPOSIF product), which is a good proxy for the product of PAR and near-infrared reflectance in the original NIRvP formulation (Dechant et al., 2021). NIRvP is designed so that it mimics SIF dependencies on FAPAR, 360 PAR and canopy escape fraction, whereas NDVI would only indicate FAPAR, and NIRv the product between FAPAR and escape fraction. The different plots in Fig. 12 indeed illustrate that SIF compares better to NIRv and NIRvP than to NDVI, with the latter saturating with respect to SIF. This effect is especially visible for the Boreal Deciduous Broadleaf Forest biome (third column of Fig. 12). On the other hand, a slightly more linear relationship is found between SIF and NIRv than between SIF and NIRvP despite the theoretically higher compatibility between SIF and NIRvP (Dechant et al., 2021). Further exploration
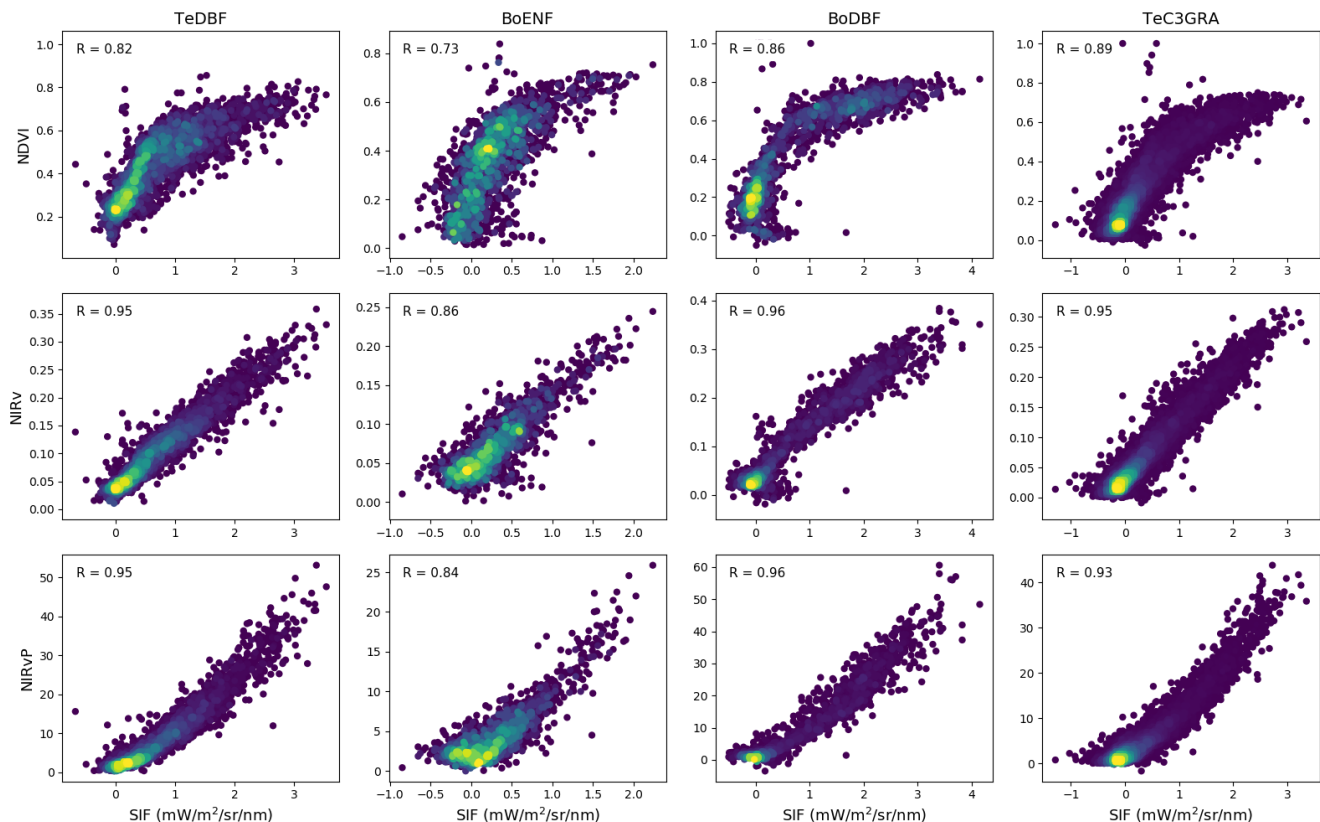
Figure 12. Comparison of weekly SIF data (743–758 nm, cloud fraction <0.2) with NDVI, NIRv and NIRvP derived from TROPOSIF spectral reflectance estimates for an ensemble of 0.5° pixels representative of Temperate Deciduous Broadleaf Forest (TeDBF), Boreal Evergreen Needleleaf Forest (BoENF), Boreal Deciduous Broadleaf Forest (BoDBF), and Temperate C3 grass (TeC3GRA) biomes.

365 of these relationships will be subject of future work, since the identification of practical approaches to disentangle structural and physiological contributions to the SIF signal is key to the proper use of SIF data.

4 Discussion

This paper has described the first version of a processing chain for the generation of far-red SIF product from the S5P-TROPOMI mission which has been developed within the framework of the ESA TROPOSIF project.

370 The retrieval is based on a linear forward model fitting TOA radiances in the far-red spectral region. It relies on a data driven forward model similar to the one initially proposed by Guanter et al. (2012) for GOSAT, later adapted to GOME-2 by Joiner et al. (2013), and finally applied to simulated and real TROPOMI data by Guanter et al. (2015) and Köhler et al. (2018), respectively. In these data-driven approaches, a series of orthogonal spectral vectors derived from a so-called training set are

used to model high spectral frequency features in the spectrum from both solar and atmospheric lines. SIF is estimated through the inversion of the linear forward model, together with the weights of the singular vectors and the coefficients of a 3rd-order polynomial used to model low spectral frequency variations in the spectrum, such as those from varying surface reflectance.

Two fitting windows are selected for the retrieval, 743–758 nm and 735–758 nm. The first has shown to be very robust against atmospheric effects (especially cloud contamination), whereas the second gives smaller precision errors due to the greater number of spectral points. The 743–758 nm window has been selected as the baseline retrieval window, as it provides the best compromise between retrieval precision and sensitivity to clouds. This fitting window is consistent with the one used for the existing Caltech SIF product (Köhler et al., 2018), which allows a direct comparison between the two datasets. In addition, we have decided to include retrievals from the 735–758 nm fitting window as well. Retrievals from this fitting window have shown a higher sensitivity to cloud contamination, but also lower precision errors because of the higher number of spectral points in the fitting window. Retrievals from the 735–758 nm fitting window can thus be advantageous for those applications using only clear-sky data.

The first assessment of the consistency of the TROPOSIF SIF data has relied on the comparison with the accurate SIF retrievals from the OCO-2 mission and with the well-established TROPOMI SIF product from Caltech. A very high similarity has been found between the two TROPOMI SIF datasets (R usually greater than 0.93, 0.008 mW m^{-2} sr^{-1} nm^{-1} bias), whereas the evaluation of precision errors over sand deserts and PICS has shown slightly smaller random errors for the TROPOSIF product. The retrieval error is typically 0.5 mW m^{-2} sr^{-1} nm^{-1} in the 743–758 nm fitting window and 0.4 mW m^{-2} sr^{-1} nm^{-1} in the 735–758 nm window. The mean bias for the raw SIF estimates are -0.080 (-0.017) mW m^{-2} sr^{-1} nm^{-1} in the 743–758 nm (735–758 nm) fitting window.

In addition to SIF, spectral TOA reflectance at 7 spectral points in atmospheric windows within the 665–781 nm window of TROPOMI's channels 5 and 6 are derived and included as meta-data in the output files. These reflectance data can be useful to derive spectral vegetation indices and biophysical variables that can help interpret and exploit SIF data. This has been illustrated in this work by the comparison of SIF data to NDVI, NIRv and NIRvP indices, which has shown the different information content carried by each of them.

5 Data availability

The dataset is available through the following Digital Object Identifier (DOI) https://doi.org/10.5270/esa-s5p_innovation-sif-20180501_20210320-v2.1-202104 (Guanter et al., 2021). Additional informative materials are also provided on the project website at https://s5p-troposif.noveltis.fr/.

6 Conclusions

The TROPOSIF product presented in this document is intended to become a reference data stream for scientific work dealing with the global monitoring of vegetation, especially in those studies focused on the dynamics and productivity of terrestrial ecosystems.

Our evaluation of the TROPOSIF product in this work has shown that it has an overall similar quality to the well-established Caltech TROPOMI SIF product (Köhler et al., 2018). Two main additional features are included in TROPOSIF: a secondary SIF dataset from the 735–758 nm fitting window with an enhanced signal-to-noise ratio, and spectral reflectance spectra. In the case of 735–758 nm SIF retrievals, however, the first analysis of the data has found potential difficulties of the retrieval to deal with spectrally-steep radiance spectra, such as in the case of some crops. This issue will be investigated during future TROPOSIF consolidation activities. Users are recommended to use the baseline 743–758 nm SIF product for the moment. Approaches for the combination of the two SIF data streams into a single one will also be evaluated in future research.

Feedback from the user community after the first experience with the data is also expected to guide future developments. In this respect, the TROPOSIF processing chain is being implemented in the Sentinel-5P Product Algorithm Laboratory (S5P-PAL)[3], which is an ongoing project funded by the European Commission to allow fast and cost-efficient Sentinel-5P prototype product development. It is expected that the TROPOSIF product will be generated and distributed to users by S5P-PAL in an operational manner in the near-future.

Appendix A: L2 and L2B file description

The NetCDF4 data files contaning the TROPOSIF product (either L2 or L2B) are structured as follows:

1. `METADATA/ALGORITHM_SETTINGS`: It contains information about the processor's configuration variables, such as the spectral fitting window or the thresholds used for data filtering.

2. `PRODUCT`: It contains the SIF product itself (`SIF`), the daily-average corrected SIF (`SIF_Corr`) and the estimated 1-σ uncertainty (`SIF_ERROR`) for the two fitting windows.

3. `PRODUCT/SUPPORT_DATA/DETAILED_RESULTS`: It contains ancillary data fields for the exploitation or interpretation of the SIF retrievals, such as day-length correction factor (`DayLength_fac`), the reduced χ^2 (χ^2_r) of the fit (`redCHI2`) (in the case of L2 files), the average TOA radiance in the fitting window (`TOA_RAD`), the quality value (`QA_value`), and the spectral TOA reflectance and corresponding wavelengths at several spectral points (`TOA_RFL` and `WVL_RFL`) (in the case of L2 and L2B "clear-sky" products).

4. `PRODUCT/SUPPORT_DATA/GEOLOCATIONS`: It contains data fields describing the acquisition location (latitude and longitude bounds for each retrieval) (L2 files) and geometry (viewing and solar angles).

---

[3]https://maps.s5p-pal.com/

5. `PRODUCT/SUPPORT_DATA/INPUT_DATA`: It contains geospatial data used as input for the retrieval, such as the cloud fraction from TROPOMI's L2 product (`cloud_fraction_L2`) and the land cover mask used to separate land and water pixels (`LC_mask`)

Author contributions. Conceptualization and Methodology: L.G. and C.B. Formal analysis: L.G., C.B., A.S., T.A.v.K., F.M., P.K. Investigation: All authors. Resources: L.G. C.B., A.S., F. M. Supervision: L.G., C.B., I.A., C.R. Writing-original draft: L.G. and C.B. Writing-review and editing: All authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The TROPOSIF project is funded by ESA's Sentinel-5p+ Innovation activity (ESA Contract Nº 4000127461/19/I-NS). C. Frankenberg and P. Köhler acknowledge funding through NASA's Earth Science U.S. Participating Investigator grant NNX15AH95G. The authors thank Léo Grignon (NOVELTIS) for the development of TROPOSIF website (https://s5p-troposif.noveltis.fr/), Alexandru Dandocsi from ESA for his support to obtain a DOI for the dataset, and Elena Sánchez-García (UPV) for her earlier data quality assessment assessment.

References

Bacour, C., Briottet, X., Bréon, F.-M., Viallefont-Robinet, F., and Bouvet, M.: Revisiting Pseudo Invariant Calibration Sites (PICS) Over Sand Deserts for Vicarious Calibration of Optical Imagers at 20 km and 100 km Scales, Remote Sensing, 11, https://doi.org/10.3390/rs11101166, 2019a.

Bacour, C., Maignan, F., Peylin, P., MacBean, N., Bastrikov, V., Joiner, J., Köhler, P., Guanter, L., and Frankenberg, C.: Differences Between OCO-2 and GOME-2 SIF Products From a Model-Data Fusion Perspective, Journal of Geophysical Research: Biogeosciences, 124, 3143–3157, https://doi.org/https://doi.org/10.1029/2018JG004938, 2019b.

Badgley, G., Field, C. B., and Berry, J. A.: Canopy near-infrared reflectance and terrestrial photosynthesis, Science Advances, 3, https://doi.org/10.1126/sciadv.1602244, 2017.

Dechant, B., Ryu, Y., Badgley, G., Zeng, Y., Berry, J. A., Zhang, Y., Goulas, Y., Li, Z., Zhang, Q., Kang, M., Li, J., and Moya, I.: Canopy structure explains the relationship between photosynthesis and sun-induced chlorophyll fluorescence in crops, Remote Sensing of Environment, 241, 111 733, https://doi.org/https://doi.org/10.1016/j.rse.2020.111733, 2020.

Dechant, B., Ryu, Y., Badgley, G., Köhler, P., Rascher, U., Migliavacca, M., Zhang, Y., Tagliabue, G., Guan, K., Rossini, M., Goulas, Y., Zeng, Y., Frankenberg, C., and Berry, J. A.: NIRvP: a robust structural proxy for sun-induced chlorophyll fluorescence and photosynthesis across scales, https://doi.org/https://doi.org/10.31223/X5ZS35, 2021.

Doughty, R., Köhler, P., Frankenberg, C., Magney, T. S., Xiao, X., Qin, Y., Wu, X., and Moore, B.: TROPOMI reveals dry-season increase of solar-induced chlorophyll fluorescence in the Amazon forest, Proceedings of the National Academy of Sciences, 116, 22 393–22 398, https://doi.org/10.1073/pnas.1908157116, 2019.

Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., Kraft, S., Middleton, E. M., Miglietta, F., Mohammed, G., Nedbal, L., Rascher, U., Schüttemeyer, D., and Verhoef, W.: The FLuorescence EXplorer Mission Concept—ESA's Earth Explorer 8, IEEE Transactions on Geoscience and Remote Sensing, 55, 1273–1284, https://doi.org/10.1109/TGRS.2016.2621820, 2017.

Du, S., Liu, L., Liu, X., Zhang, X., Zhang, X., Bi, Y., and Zhang, L.: Retrieval of global terrestrial solar-induced chlorophyll fluorescence from TanSat satellite, Science Bulletin, 63, 1502–1512, https://doi.org/https://doi.org/10.1016/j.scib.2018.10.003, 2018.

Frankenberg, C., Butz, A., and Toon, G. C.: Disentangling chlorophyll fluorescence from atmospheric scattering effects in O2A-band spectra of reflected sun-light, Geophysical Research Letters, 38, L03 801, doi:10.1029/2010GL045896, 2011a.

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., Toon, G. C., Butz, A., Jung, M., Kuze, A., and Yokota, T.: New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity, Geophysical Research Letters, 38, L17 706, doi:10.1029/2011GL048738, 2011b.

Frankenberg, C., O'Dell, C., Guanter, L., and McDuffie, J.: Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: implications for its retrieval and interferences with atmospheric CO_2 retrievals, Atmospheric Measurement Techniques, 5, 2081–2094, https://doi.org/10.5194/amt-5-2081-2012, 2012.

Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., Pollock, R., and Taylor, T. E.: Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2, Remote Sensing of Environment, 147, 1–12, 2014.

Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P. E., Gómez-Dans, J., Kuze, A., Suto, H., and Grainger, R. G.: Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements, Remote Sensing of Environment, 121, 236–251, 2012.

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

Guanter, L., Aben, I., Tol, P., Krijger, J. M., Hollstein, A., Köhler, P., Damm, A., Joiner, J., Frankenberg, C., and Landgraf, J.: Potential of the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence, Atmospheric Measurement Techniques, 8, 1337–1352, https://doi.org/10.5194/amt-8-1337-2015, 2015.

Guanter, L., Bacour, C., Schneider, A., Aben, I., van Kempen, T. A., Maignan, F., Retscher, C., Köhler, P., Frankenberg, C., Joiner, J., and Zhang, Y.: The TROPOSIF global sun-induced fluorescence data set from the Sentinel-5P TROPOMI mission. TROPOSIF-L2B [Data set], ESA S5P+ Innovation Datasets. European Space Agency, https://doi.org/10.5270/esa-s5p_innovation-sif-20180501_20210320-v2.1-202104, 2021.

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J. A., Frankenberg, C., Huete, A. R., Zarco-Tejada, P., Lee, J.-E., Moran, M. S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J. M., and Griffis, T. J.: Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence, Proceedings of the National Academy of Sciences, 111, E1327–E1333, 2014.

He, L., Magney, T., Dutta, D., Yin, Y., Köhler, P., Grossmann, K., Stutz, J., Dold, C., Hatfield, J., Guan, K., et al.: From the ground to space: Using solar-induced chlorophyll fluorescence to estimate crop productivity, Geophysical Research Letters, 47, e2020GL087 474, 2020.

Jeong, S.-J., Schimel, D., Frankenberg, C., Drewry, D. T., Fisher, J. B., Verma, M., Berry, J. A., Lee, J.-E., and Joiner, J.: Application of satellite solar-induced chlorophyll fluorescence to understanding large-scale variations in vegetation phenology and function over northern high latitude forests, Remote Sensing of Environment, 190, 178 – 187, 2017.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Yoshida, Y., Corp, L. A., and Middleton, E. M.: First observations of global and seasonal terrestrial chlorophyll fluorescence from space, Biogeosciences, 8, 637–651, https://doi.org/10.5194/bg-8-637-2011, 2011.

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., Huemmrich, K. F., Yoshida, Y., and Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2, Atmospheric Measurement Techniques, 6, 2803–2823, 2013.

Joiner, J., Yoshida, Y., Guanter, L., and Middleton, E. M.: New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: simulations and application to GOME-2 and SCIAMACHY, Atmospheric Measurement Techniques, 9, 3939–3967, https://doi.org/10.5194/amt-9-3939-2016, 2016.

Joiner, J., Yoshida, Y., Köhler, P., Campbell, P., Frankenberg, C., van der Tol, C., Yang, P., Parazoo, N., Guanter, L., and Sun, Y.: Systematic Orbital Geometry-Dependent Variations in Satellite Solar-Induced Fluorescence (SIF) Retrievals, Remote Sensing, 12, https://doi.org/10.3390/rs12152346, 2020.

Khosravi, N., Vountas, M., Rozanov, V. V., Bracher, A., Wolanin, A., and Burrows, J. P.: Retrieval of Terrestrial Plant Fluorescence Based on the In-Filling of Far-Red Fraunhofer Lines Using SCIAMACHY Observations, Frontiers in Environmental Science, 3, 78, https://doi.org/10.3389/fenvs.2015.00078, 2015.

Köhler, P., Guanter, L., and Joiner, J.: A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIA-MACHY data, Atmospheric Measurement Techniques Discussions, submitted, 2014.

Köhler, P., Guanter, L., and Frankenberg, C.: Simplified physically based retrieval of sun-induced chlorophyll fluorescence from GOSAT data, IEEE Geoscience and Remote Sensing Letters, 12, 1446–1450, https://doi.org/10.1109/LGRS.2015.2407051, 2015.

515 Köhler, P., Frankenberg, C., Magney, T. S., Guanter, L., Joiner, J., and Landgraf, J.: Global Retrievals of Solar-Induced Chlorophyll Fluorescence With TROPOMI: First Results and Intersensor Comparison to OCO-2, Geophysical Research Letters, 45, 10,456–10,463, https://doi.org/10.1029/2018GL079031, 2018.

Köhler, P., Behrenfeld, M. J., Landgraf, J., Joiner, J., Magney, T. S., and Frankenberg, C.: Global Retrievals of Solar-Induced Chlorophyll Fluorescence at Red Wavelengths With TROPOMI, Geophysical Research Letters, 47, e2020GL087 541,
520 https://doi.org/https://doi.org/10.1029/2020GL087541, e2020GL087541 10.1029/2020GL087541, 2020.

Liu, X., Guanter, L., Liu, L., Damm, A., Malenovský, Z., Rascher, U., Peng, D., Du, S., and Gastellu-Etchegorry, J.-P.: Downscaling of solar-induced chlorophyll fluorescence from canopy level to photosystem level using a random forest model, Remote Sensing of Environment, 231, 110 772, https://doi.org/https://doi.org/10.1016/j.rse.2018.05.035, 2019.

Luus, K. A., Commane, R., Parazoo, N. C., Benmergui, J., Euskirchen, E. S., Frankenberg, C., Joiner, J., Lindaas, J., Miller, C. E., Oechel,
525 W. C., Zona, D., Wofsy, S., and Lin, J. C.: Tundra photosynthesis captured by satellite-observed solar-induced chlorophyll fluorescence, Geophysical Research Letters, 44, 1564–1573, 2017.

Maes, W. H., Pagán, B. R., Martens, B., Gentine, P., Guanter, L., Steppe, K., Verhoest, N. E., Dorigo, W., Li, X., Xiao, J., and Miralles, D. G.: Sun-induced fluorescence closely linked to ecosystem transpiration as evidenced by satellite data and radiative transfer models, Remote Sensing of Environment, 249, 112 030, https://doi.org/https://doi.org/10.1016/j.rse.2020.112030, 2020.

530 Mengistu, A. G., Mengistu Tsidu, G., Koren, G., Kooreman, M. L., Boersma, K. F., Tagesson, T., Ardö, J., Nouvellon, Y., and Peters, W.: Sun-induced Fluorescence and Near Infrared Reflectance of vegetation track the seasonal dynamics of gross primary production over Africa, Biogeosciences Discussions, 2020, 1–23, https://doi.org/10.5194/bg-2020-242, 2020.

Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J.-P., Miller, J. R., Guanter, L., Moreno, J., Moya,
535 I., Berry, J. A., Frankenberg, C., and Zarco-Tejada, P. J.: Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress, Remote Sensing of Environment, 231, 111 177, https://doi.org/https://doi.org/10.1016/j.rse.2019.04.030, 2019.

Pagán, B. R., Maes, W. H., Gentine, P., Martens, B., and Miralles, D. G.: Exploring the Potential of Satellite Solar-Induced Fluorescence to Constrain Global Transpiration Estimates, Remote Sensing, 11, https://doi.org/10.3390/rs11040413, 2019.

Parazoo, N. C., Frankenberg, C., Köhler, P., Joiner, J., Yoshida, Y., Magney, T., Sun, Y., and Yadav, V.: Towards a Harmonized Long-
540 Term Spaceborne Record of Far-Red Solar-Induced Fluorescence, Journal of Geophysical Research: Biogeosciences, 124, 2518–2539, https://doi.org/10.1029/2019JG005289, 2019.

Plascyk, J. A. and Gabriel, F. C.: The Fraunhofer Line Discriminator MKII – An airborne instrument for precise and standardized ecological luminescence measurement, IEEE Transactions on Instrumentation and measurement, IM-24, 306–313, 1975.

Sanders, A. F. J., Verstraeten, W. W., Kooreman, M. L., Van Leth, T. C., Beringer, J., and Joiner, J.: Spaceborne Sun-Induced Veg-
545 etation Fluorescence Time Series from 2007 to 2015 Evaluated with Australian Flux Tower Measurements, Remote Sensing, 8, https://doi.org/10.3390/rs8110895, 2016.

Shan, N., Zhang, Y., Chen, J. M., Ju, W., Migliavacca, M., Peñuelas, J., Yang, X., Zhang, Z., Nelson, J. A., and Goulas, Y.: A model for estimating transpiration from remotely sensed solar-induced chlorophyll fluorescence, Remote Sensing of Environment, 252, 112 134, https://doi.org/https://doi.org/10.1016/j.rse.2020.112134, 2021.

550 Sioris, C. E., Courrèges-Lacoste, G. B., and Stoll, M. P.: Filling in of Fraunhofer lines by plant fluorescence: Simulations for a nadir-viewing satellite-borne instrument, Journal of Geophysical Research – Atmospheres, 108, L4133, doi:10.1029/2001JD001321, 2003.

Smith, W. K., Biederman, J. A., Scott, R. L., Moore, D. J. P., He, M., Kimball, J. S., Yan, D., Hudson, A., Barnes, M. L., MacBean, N., Fox, A. M., and Litvak, M. E.: Chlorophyll Fluorescence Better Captures Seasonal and Interannual Gross Primary Productivity Dynamics Across Dryland Ecosystems of Southwestern North America, Geophysical Research Letters, 45, 748–757, https://doi.org/https://doi.org/10.1002/2017GL075922, 2018.

Sun, Y., Frankenberg, C., Wood, J. D., Schimel, D. S., Jung, M., Guanter, L., Drewry, D. T., Verma, M., Porcar-Castell, A., Griffis, T. J., Gu, L., Magney, T. S., Köhler, P., Evans, B., and Yuen, K.: OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence, Science, 358, https://doi.org/10.1126/science.aam5747, 2017.

Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., and Magney, T.: Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP, Remote Sensing of Environment, 209, 808–823, https://doi.org/https://doi.org/10.1016/j.rse.2018.02.016, 2018.

Turner, A. J., Köhler, P., Magney, T. S., Frankenberg, C., Fung, I., and Cohen, R. C.: A double peak in the seasonality of California's photosynthesis as observed from space, Biogeosciences, 17, 405–422, https://doi.org/10.5194/bg-17-405-2020, 2020.

van Schaik, E., Kooreman, M. L., Stammes, P., Tilstra, L. G., Tuinder, O. N. E., Sanders, A. F. J., Verstraeten, W. W., Lang, R., Cacciari, A., Joiner, J., Peters, W., and Boersma, K. F.: Improved SIFTER v2 algorithm for long-term GOME-2A satellite retrievals of fluorescence with a correction for instrument degradation, Atmospheric Measurement Techniques, 13, 4295–4315, https://doi.org/10.5194/amt-13-4295-2020, 2020.

Walther, S., Voigt, M., Thum, T., Gonsamo, A., Zhang, Y., Köhler, P., Jung, M., Varlagin, A., and Guanter, L.: Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of photosynthesis and greenness dynamics in boreal evergreen forests, Global Change Biology, 22, 2979–2996, 2017.

Wu, X., Xiao, X., Zhang, Y., He, W., Wolf, S., Chen, J., He, M., Gough, C. M., Qin, Y., Zhou, Y., Doughty, R., and Blanken, P. D.: Spatiotemporal Consistency of Four Gross Primary Production Products and Solar-Induced Chlorophyll Fluorescence in Response to Climate Extremes Across CONUS in 2012, Journal of Geophysical Research: Biogeosciences, 123, 3140–3161, https://doi.org/https://doi.org/10.1029/2018JG004484, 2018.

Yang, P. and van der Tol, C.: Linking canopy scattering of far-red sun-induced chlorophyll fluorescence with reflectance, Remote Sensing of Environment, 209, 456 – 467, https://doi.org/https://doi.org/10.1016/j.rse.2018.02.029, 2018.

Yang, P., van der Tol, C., Campbell, P. K., and Middleton, E. M.: Fluorescence Correction Vegetation Index (FCVI): A physically based reflectance index to separate physiological and non-physiological information in far-red sun-induced chlorophyll fluorescence, Remote Sensing of Environment, 240, 111 676, https://doi.org/https://doi.org/10.1016/j.rse.2020.111676, 2020.

Yao, L., Yang, D., Liu, Y., Wang, J., Liu, L., Du, S., Cai, Z., Lu, N., Lyu, D., Wang, M., Yin, Z., and Zheng, Y.: A New Global Solar-induced Chlorophyll Fluorescence (SIF) Data Product from TanSat Measurements, Advances in Atmospheric Sciences, 38, 341–345, https://doi.org/10.1007/s00376-020-0204-6, 2021.

Yin, Y., Byrne, B., Liu, J., Wennberg, P. O., Davis, K. J., Magney, T., Köhler, P., He, L., Jeyaram, R., Humphrey, V., et al.: Cropland carbon uptake delayed and reduced by 2019 Midwest floods, AGU Advances, 1, e2019AV000 140, 2020.

Yoshida, Y., Joiner, J., Tucker, C., Berry, J., Lee, J.-E., Walker, G., Reichle, R., Koster, R., Lyapustin, A., and Wang, Y.: The 2010 Russian drought impact on satellite measurements of solar-induced chlorophyll fluorescence: Insights from modeling and comparisons with parameters derived from satellite reflectances, Remote Sensing of Environment, 166, 163 – 177, https://doi.org/https://doi.org/10.1016/j.rse.2015.06.008, 2015.

Zeng, Y., Badgley, G., Dechant, B., Ryu, Y., Chen, M., and Berry, J.: A practical approach for estimating the es-
cape ratio of near-infrared solar-induced chlorophyll fluorescence, Remote Sensing of Environment, 232, 111 209,
https://doi.org/https://doi.org/10.1016/j.rse.2019.05.028, 2019.

Zhang, Y., Xiao, X., Zhang, Y., Wolf, S., Zhou, S., Joiner, J., Guanter, L., Verma, M., Sun, Y., Yang, X., Paul-Limoges, E., Gough, C. M.,
Wohlfahrt, G., Gioli, B., van der Tol, C., Yann, N., Lund, M., and de Grandcourt, A.: On the relationship between sub-daily instantaneous
and daily total gross primary production: Implications for interpreting satellite-based SIF retrievals, Remote Sensing of Environment, 205,
276 – 289, https://doi.org/https://doi.org/10.1016/j.rse.2017.12.009, 2018.

Zhang, Z., Chen, J. M., Guanter, L., He, L., and Zhang, Y.: From Canopy-Leaving to Total Canopy Far-Red Fluorescence Emis-
sion for Remote Sensing of Photosynthesis: First Results From TROPOMI, Geophysical Research Letters, 46, 12 030–12 040,
https://doi.org/https://doi.org/10.1029/2019GL084832, 2019.

Zhang, Z., Zhang, Y., Porcar-Castell, A., Joiner, J., Guanter, L., Yang, X., Migliavacca, M., Ju, W., Sun, Z., Chen, S., Martini, D., Zhang,
Q., Li, Z., Cleverly, J., Wang, H., and Goulas, Y.: Reduction of structural impacts and distinction of photosynthetic pathways in a
global estimation of GPP from space-borne solar-induced chlorophyll fluorescence, Remote Sensing of Environment, 240, 111 722,
https://doi.org/https://doi.org/10.1016/j.rse.2020.111722, 2020.

Zuromski, L. M., Bowling, D. R., Köhler, P., Frankenberg, C., Goulden, M. L., Blanken, P. D., and Lin, J. C.: Solar-Induced Fluorescence
Detects Interannual Variation in Gross Primary Production of Coniferous Forests in the Western United States, Geophysical Research
Letters, 45, 7184–7193, https://doi.org/https://doi.org/10.1029/2018GL077906, 2018.