Supporting Information for "Global retrievals of solar induced chlorophyll fluorescence with TROPOMI: first results and inter-sensor comparison to OCO-2"

Philipp Köhler[1], Christian Frankenberg[1,2], Troy S. Magney[2], Luis Guanter[3], Joanna Joiner[4], Jochen Landgraf[5]

[1] Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA.

[2] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA.

[3] Helmholtz Centre Potsdam, German Research Center for Geosciences (GFZ), Potsdam, Germany.

[4] NASA Goddard Space Flight Center, Greenbelt, MD, USA.

[5] SRON Netherlands Institute for Space Research, Utrecht, Netherlands.

Contents of this file

1. Text S1 to S8
2. Table S1
3. Figures S1 to S8

Additional Supporting Information (Files uploaded separately)

1. Captions for Movies S1

S1. Spatial Resolution of TROPOMI NIR Soundings across the Satellite Track

This section addresses why the spatial resolution of TROPOMI ground pixels across the S5P satellite track ranges approximately between 3.5–15 km (Sect. 2.1). TROPOMI's UVN spectrometers integrate the reflected sunlight every second with a two-dimensional (across swath, wavelength) 1024 x 1024 pixel detector array. The 7 km along-track spatial resolution is, therefore, determined by the satellite's velocity. In contrast, the across track spatial resolution depends on the view angle and detector pixel co-adding. The ground pixel size becomes generally larger towards the edges of TROPOMI's wide swath (2600 km), which is due to the Earth's curvature. However, the 1024 spatial detector pixels can be co-added to increase the signal to noise ratio, while the current binning mode of TROPOMI's NIR spectrometer results in 448 across-track ground pixels. The resulting pixel size/area is shown in Fig. S1 as a function of viewing zenith angle/spatial row. The sudden decrease in pixel size towards the edges of the swath means that fewer detector pixels are co-added, resulting in a considerably lower SNR for the affected spatial rows. Using spatial rows below/above 20/427 to retrieve SIF is in principle possible but associated with significantly higher uncertainties, which is why we exclude these spatial rows from our analysis.

S2. Retrieval Algorithm

0094-8276/18/$5.00

As outlined in Sect. 2.3, we exploit the change in the fractional depth of solar Fraunhofer lines, which occurs due to the additive nature of the SIF signal. We chose a retrieval window ranging from 743 nm – 758 nm (encompassing about 130 spectral points), because this wavelength range is devoid of atmospheric absorption features and contains several strong Fraunhofer lines. A sample spectrum of TROPOMI's band 6 and a zoom-in on the retrieval window is shown in Fig. S2. The required spectral basis functions (or principal components – PCs) for the statistically-based retrieval are derived within our retrieval window from TROPOMI data over vegetation free areas (e.g. ocean, ice, and deserts; sampled on a daily basis) through a singular value decomposition (SVD). A linear combination of a few PCs can then be used to model all spectra with a sufficient accuracy, including diverse sensor specific features. The SVD is done separately for each single spatial row of the detector array (448 in total), because spectral and radiometric characteristics change slightly across the focal plane. Fig. S2 (left column) shows the first ten PCs of one spatial row (314/448) together with the percentage of their explained variance. Even though this is a purely statistical approach to reduce the dimensionality of the training data set, a physical meaning can be attached to some PCs. PC_1 can be interpreted as an average spectrum explaining already more than 99% of the variance in the training data, which includes the fractional depth of solar Fraunhofer lines in the absence of any SIF emission. PC_2 likely combines typical changes in the spectral reflectance of our reference targets and the slope of the solar irradiance. Typical instrumental effects can be identified in PC_3 and PC_7, where PC_3 represents a subtle wavelength shift, while PC_7 captures an instrument line shape broadening.

In contrast to the selected vegetation-free spectra (training data), the spectral reflectance of vegetation increases rapidly from the red to NIR part of the spectrum (known as red-edge). Additionally, to be able to model also spectral reflectances with high curvature, we use a set of three Legendre polynomials, each multiplied element-wise with PC_1 in order to preserve the fractional depth of the Fraunhofer lines. The last necessary spectral basis function for our retrieval is a reference SIF emission shape, which is derived from leaf-level measurements conducted by *Magney et al.* [2017] (shown in Fig S2). In sum, the forward model can

now be written as

$$\mathbf{F}_{TOA} = \sum_{i=1}^{n_{PC}} (\gamma_i \cdot \mathbf{PC}_i) + \sum_{j=1}^{3} (\eta_j \cdot \mathbf{P}_j \odot \mathbf{PC}_1) + F_s \cdot \mathbf{h}_f, \quad (1)$$

where γ_i, η_j, and F_s are the state vector elements, n_{PC} is the number of PCs used in the retrieval, $\mathbf{P}_j$ are the Legendre polynomials, the $\odot$ operator denotes element-wise multiplication, and $\mathbf{h}_f$ is the reference fluorescence emission spectrum (bold characters indicate variables with a spectral component). F_s gives an estimate for the far red SIF peak by normalizing the SIF reference shape ($\mathbf{h}_f$) to one at 740 nm. We provide 10 PCs, 3 Legendre polynomials, and a reference fluorescence emission spectrum to the retrieval. This means there are initially 14 state vector elements to model the top-of-atmosphere (TOA) radiance spectra ($\mathbf{F}_{TOA}$) through an ordinary least squares fit. The number of provided PCs is somewhat arbitrary, but has effects on the retrieval accuracy and precision as reported by *Joiner et al.* [2013]. However, *Köhler et al.* [2015] proposed to optimize the number of free model parameters by making use of a stepwise model selection. The linear forward model (Eq. 1) allows us to follow this approach. Specifically, we use a backward elimination algorithm to automatically select the required model parameters with respect to the goodness of fit balanced by model complexity (number of state vector elements). It has been shown that a potential overfitting (fitting noise) can be avoided, while results remain stable, independent of the number of PCs initially provided to the retrieval. Note that $\mathbf{PC}_1$ and $\mathbf{h}_f$ are excepted from being removed by the backward elimination algorithm. In case of the sample retrieval (Fig. S2), 7 state vector elements have been selected. Consistent to the sample retrieval, we find that on average 7-8 (out of 14) state vector elements are automatically selected. The average Signal-to-Noise Ratio (SNR) within the retrieval window is 2660, estimated by building the ratio between the mean signal level (117 mW/m^2/sr/nm) and the standard deviation of the residual (0.044 mW/m^2/sr/nm). A further assessment of the uncertainty, accuracy, and performance of our retrieval approach is presented in the following.

S3. In-filling of solar Fraunhofer lines by SIF

As described in Sect. 2.3, the SIF retrieval relies on the presence of absorption lines in the solar irradiance spectrum, known as solar Fraunhofer lines. The ratio between intensity in the continuum and intensity in the central wavelength of the Fraunhofer lines will be identical before and after being reflected at any surface as long as there is no source of radiation. However, if a fluorescent target is observed, there is an additive offset to both continuum and solar Fraunhofer lines. This offset affects the Fraunhofer lines, in relative terms, more than the continuum level. By means of averaged fit residuals, the change in the fractional depth (known as 'in-filling') of solar Fraunhofer lines is illustrated in Fig. S3. Specifically, we were running the retrieval for one orbit (passing over the Amazon rainforest) with and without SIF as a state vector element in the forward model (Eq. 1). We then averaged residuals of soundings where the retrieval results (incl. SIF) indicated an amount of $\mathrm{SIF}_{740\mathrm{nm}}$ larger than 1 mW/m^2/sr/nm. In case SIF is neglected as spectral basis function, our forward model is not able to accurately fit the measurements, because the PCs (derived from non-fluorescent targets) intentionally do not capture the in-filling of solar Fraunhofer lines by SIF. In contrast, the averaged residual of fits including SIF as an additive contribution is free from spectral features, underscoring the validity of our forward model and the reference SIF shape.

S4. Uncertainty Estimation

In order to assess the uncertainty of our SIF estimates, the $1-\sigma$ retrieval error is calculated by evaluating the retrieval error covariance matrix given by

$$\mathbf{S}_e = (\mathbf{K}^T \mathbf{S}_0^{-1} \mathbf{K})^{-1}, \quad (2)$$

where $\mathbf{K}$ is the Jacobian matrix (formed by model parameters of Eq. 1) and $\mathbf{S}_0$ is the measurement error covariance matrix. $\mathbf{S}_0$ is a diagonal matrix in case of spectrally uncorrelated Gaussian noise, which is an appropriate assumption for grating spectrometers. Here, we use our fit residuals to set-up a retrieval window specific noise model and estimate the retrieval error through Eq. 2 for several orbits. These error estimates can eventually be used to derive the single retrieval error based on the radiance level only. In view of the high data rate of TROPOMI measurements, this is computationally more efficient than propagating the measurement noise attached to the L1B data for every individual sounding. Another benefit arising from this strategy is that fitting-issues would become obvious, because the measurement noise is expected to scale with the square root of the signal level. In other words, the applicability and performance of our forward model (Eq. 1) can additionally be reviewed when comparing the noise model derived from SIF-fit residuals to the noise estimates attached to the L1B data. Specifically, we tested if we can parameterize the measurement noise as $A + B \cdot \sqrt{\text{signal level}}$, where A represents the signal independent noise contribution (read-out noise) and B is the scaling factor of the shot noise (function of signal magnitude). Two different noise estimates were used for this purpose: 1) residual standard error (RSE), and 2) standard deviation of SIF-fit residuals. We find that both noise estimates are in compliance with the theory and lead to very similar noise models without significant differences across the used spatial rows (26th–422th) of the co-added detector pixels. The obtained noise model is then used to evaluate Eq. 2 for several orbits to predict the 1-σ retrieval error. These estimates can in turn be used to derive the retrieval error as function of radiance level only, which increases the computational efficiency of the error estimation dramatically. We acknowledge that this approach needs further inspection. Therefore, we compared the error estimates attached to the L1B data with our noise model derived from SIF-fit residuals and found a very good agreement (not shown). In a second step, we challenged our noise and error model based on real data. Following *Guanter et al.* [2015], we assume that the standard deviation of single retrievals over the Sahara desert (15°N–30°N, 5°W–30°E) can represent the uncertainty. This region is particularly well-suited, because the sparse vegetation cover in combination with miscellaneous surface reflectances allows us to assess the validity of our assumption that the error can be estimated as a function of radiance level. Fig. S4 depicts the standard deviation calculated from real OCO-2 and TROPOMI SIF estimates as well as the predicted 1-σ retrieval error derived from the error model (depending on the signal level only). The consistency between observed and predicted TROPOMI errors reinforces confidence in our retrieval and error estimation, because: 1) high quality fits are achieved with the forward model (Eq. 1), and 2) noise and error estimates are realistic. Note that the OCO-2 curve in Fig. S4 is based on SIF estimates at 757 nm only, whereas an average of SIF at 757 nm and 771 nm would result in lower retrieval errors, while an upscaling to 740 nm would increase the same. Nonetheless, it can be stated that the precision with respect to retrieving SIF is very similar for both instruments.

S5. "Bat Map"-Experiment

We conducted an experiment where we added an artificial SIF signal to real measurements in order to demonstrate the validity of our retrieval approach for various surface types and atmospheric conditions. Additionally, this experiment allows us to review the retrieval accuracy and precision. In particular, we aim to answer following questions:

1. Are there correlations between the reference fluorescence emission shape and other spectral functions used by the retrieval, which lead to interferences and affect the retrieval accuracy, e.g., introduce a location-dependent bias if surface properties and/or atmospheric conditions change?

2. Is the precision on a global scale consistent with the Sahara comparison in Sect. S4?

3. How do slight variations in the 'true' SIF emission shape impact the retrieval results?

We used one day (03/19/2018) of TROPOMI orbits, including about 4.2M single soundings after filtering as described in Sect. 2.4, and added two realistic SIF intensities (SIF_{740nm}=1 / 2 mW/m^2/sr/nm) with randomly varying spectral shapes (while the retrieval still assumes the reference shape). The spectral shapes were computed by linearizing the reference emission shape and altering the slope by +/-0.5° around the central wavelength of our retrieval window (see Fig. S5). In addition, we degraded the measurements by adding random noise according to the model in Sect. S4, because the original noise level would cancel out when calculating the difference between experiment and reference. It might, however, be remarked that we were running this experiment also by adding SIF without altered spectral shapes and without adding noise, whereas all retrieved values matched exactly the input (not shown). Consequently, we can conclude that there is no significant interference between spectral functions used by the retrieval and that there is no impact due to different surface properties and atmospheric conditions. Fig. S5 shows the set-up and results of the more challenging experiment for our retrieval algorithm. The only visible bias is with respect to the higher input intensity (median ΔSIF is 1.96 mW/m^2/sr/nm), which is due to the linearization of the reference shape. A low bias can be expected, because the unaltered linearized SIF shape, which is on average added to the spectra, is greater than one at 743 nm, while the reference SIF shape used by the retrieval is scaled to one at 740 nm. In the ideal case, the standard deviation of the ΔSIF map should correspond to the mean predicted error of the experiment. As can be seen from Table S1, the actual error is slightly higher than predicted. We consider this to be a minor issue, because there is only a difference in the second decimal place (up to 0.04 mW/m^2/sr/nm). Since the discrepancy between predicted and actual error is not significantly increasing for those soundings where we added SIF artificially (with varying spectral shapes), we conclude that slightly steeper or flatter SIF emission shapes do not affect the retrieval precision. However, a spectral shift or a deviation from the assumed SIF emission shape may affect the retrieval accuracy (lead to small biases). Note that this experiment does not propagate the SIF signal through the atmosphere; however, the previous study by *Frankenberg et al.* [2011] has already shown that Fraunhofer line based methods are rather insensitive to atmospheric scattering, even in case of high aerosol optical depths. In sum, the results of the "bat map"-experiment emphasize the validity of our retrieval approach.

S6. Directional Effects

Sect. 2.2 illustrated important TROPOMI specific measurement features related to the sun-surface-sensor geometry. Here, we aim to quantify the impact of the directionality of the SIF emission [e.g., *Joiner et al.*, 2012; *Guanter et al.*, 2012; *Köhler et al.*, 2018; *Yang and Van Der Tol*, 2018] on TROPOMI SIF measurements. We selected 81,205 cloud free soundings (acquired between February and July 2018) over homogeneous evergreen broadleaf forest (EBF) areas in the tropics (latitudes between -23° and 23°). This biome is photosynthetically active all year round, while changes in canopy structure remain limited compared to biomes with a strong seasonal cycle. Based on the quadratic fit in Fig. S6, the SIF intensity can be up to 25% higher when the phase angle approaches zero degrees (hot spot - all leaves within the field of view are directly illuminated). However, most measurements were taken at phase angles between 20°-60°, where the intensity change amounts only up to 10%. A BRDF correction for space borne SIF observations is challenging, because the relationship may vary from biome to biome and potentially depends on many confounding factors such as the growing stage and vegetation density. Yet, for certain applications such as the assessment of seasonal cycles in the tropical rainforest, it might be essential to account for directional effects. One possibility to mitigate the directionality is to normalize SIF retrievals by the continuum radiance level in the NIR (rel. SIF). Fig. S6 demonstrates that this approach works reasonably well, even though it results in a slight overcompensation: the intensity around the hot spot is now about 15% lower compared to extreme phase angles (70°). This observation is in line with the results presented by *Köhler et al.* [2018], where the intensity change in NIR reflectance was higher than for SIF observations. It should further be considered that the normalization will increase the sensitivity to cloud cover (NIR reflectance increases, while SIF decreases).

S7. Data Availability and Cloud Cover

We gridded the one week of TROPOMI SIF between 04/02/2018 – 04/08/2018 to a spatial resolution of 0.2° x 0.2° using cloud filters of increasing strictness (0.8 to 0) to evaluate how the data availability is affected. The time frame has been selected with respect to the solar irradiance at this time of year, so that a SIF value can potentially be attached to most grid cells over land. Fig. S7 illustrates that a stringent cloud filter would primarily affect the spatial coverage in tropical regions (coverage drops down to 20%), which can be explained by the frequent cloud cover in these latitudes. A less dramatic decrease occurs at higher latitudes, where the coverage still reaches about 80% when the data is filtered for cloud free soundings only. It might be noted that the VIIRS cloud fraction contains no information on the cloud optical thickness. This information would, however, be necessary to evaluate the attenuation of the SIF signal due to cloud cover. For example, a substantial fraction of the emitted SIF may pass the atmosphere in case of optical thin clouds, while the same cloud fraction with optical thick clouds might have a shielding effect. Another aspect to consider is that photosynthesis is driven by instantaneous illumination. Evaluating only cloud free measurements will, therefore, lead to a clear-sky bias. In order to evaluate the order of magnitude of such effects in detail, comparisons to ground based measurements are needed.

S8. Latitudinal Comparison between TROPOMI and OCO-2 SIF

The co-located and rigorously filtered OCO-2 and TROPOMI SIF data from June 2018 have been used in Sect. 3.2 to evaluate the consistency on the global scale. Here, we relaxed the filtering (only the cloud fraction of co-located TROPOMI soundings has been limited to 0.1) to maintain a sound basis for a comparison of latitudinal gradients. In total, 54,793 TROPOMI measurements and 541,140 OCO-2 measurements serve as a basis for this analysis. The zonal SIF averages are displayed in Fig. S8, showing a very strong agreement (R^2 value of 0.95). Especially the tropics and mid latitudes on the northern hemisphere show elevated SIF values as can be anticipated during this time of year. However, absolute OCO-2 SIF values within the photosynthetically active latitudes are somewhat higher. The observed bias might be caused by residual cloud cover, differences in acquisition time and viewing-illumination geometry, inhomogeneous land cover types within single TROPOMI footprints, or absolute radiometric calibration differences. In order to exclude any systematic biome dependent bias, further studies will need to inspect this discrepancy more in detail.

S9. Time-lapse of TROPOMI SIF (Weekly Composites)

Movie S1. Weekly composites of length-of-day corrected SIF from 03/05/2018 (week 10) to 07/29/2018 (week 30), animation is available in the HTML version.

Table S1: Comparison between predicted and actual 1-σ retrieval errors. The unit is given in mW/m^2/sr/nm.

Input SIF	Retrieved SIF (median)	Predicted 1-σ error	Actual 1-σ error	# of soundings
0	0.0008	0.36	0.39	2.937.664
1	0.98	0.39	0.43	160.700
2	1.96	0.38	0.41	1.087.211

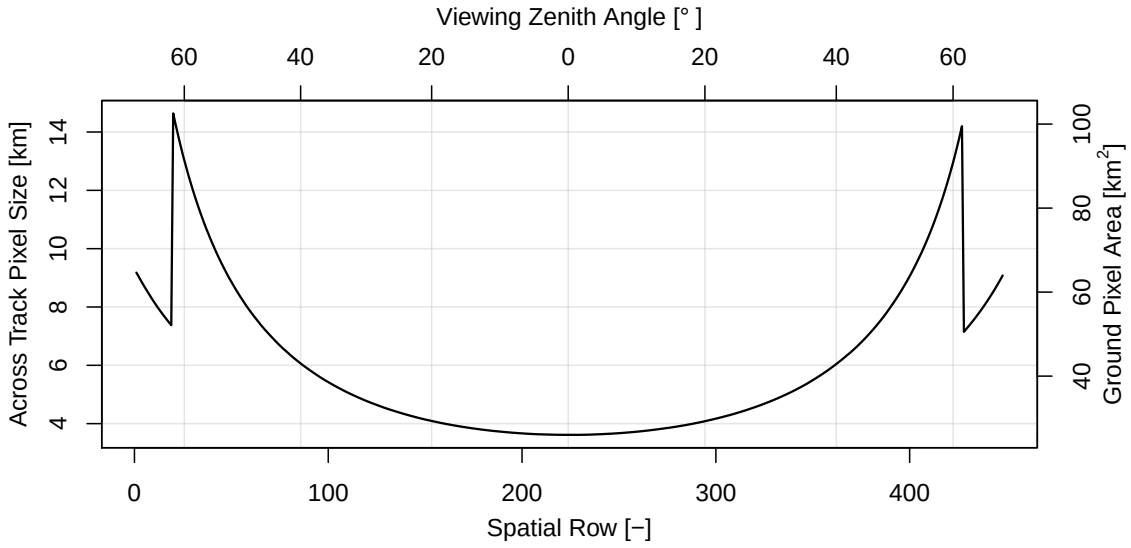

Figure S1: Across track pixel size/ground pixel area as function of spatial row/viewing zenith angle computed from soundings at the equator (same orbit as in Fig. 1).

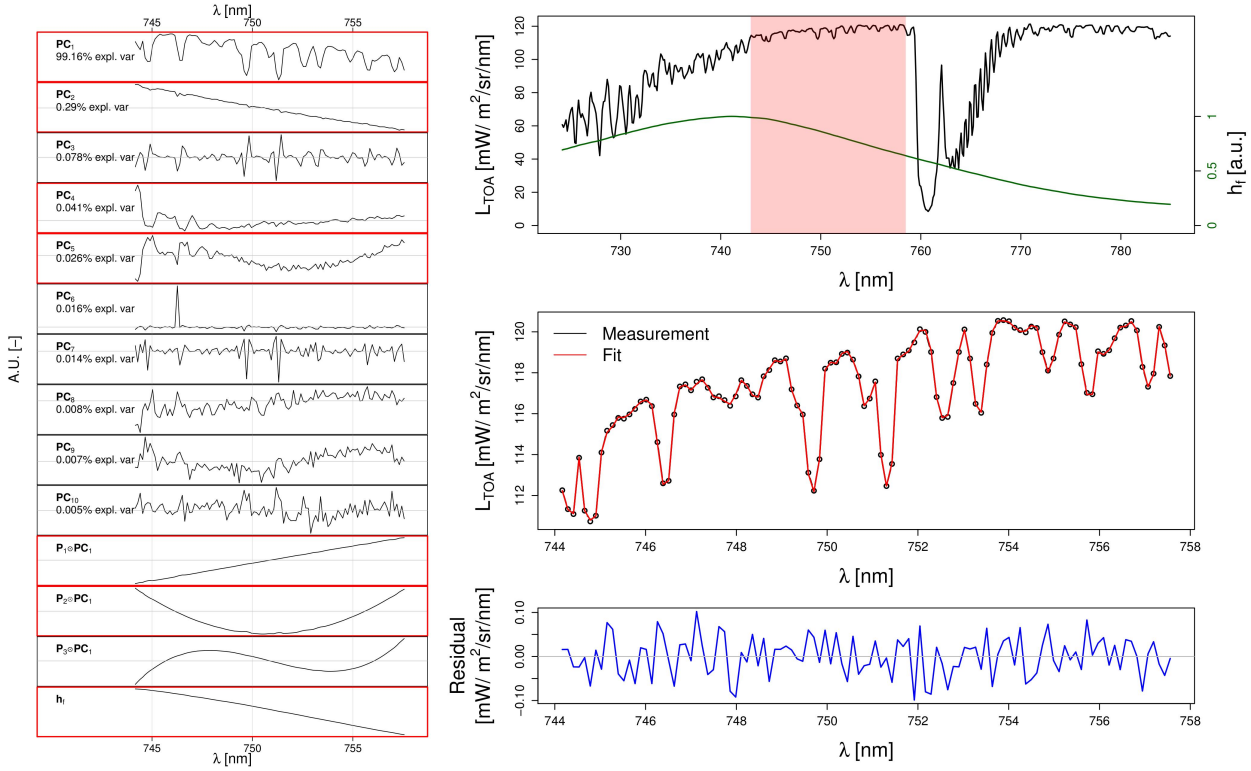

Figure S2: Sample retrieval over the Amazon rainforest (11/28/2017 at 1.48658°N, 57.19547°W, SIF@740nm=1.9 mW/m^2/sr/nm). The left column shows all spectral functions, which were provided to the retrieval, while red boxes indicate the automatically chosen ones. The upper panel on the right shows the measured TROPOMI spectrum in band 6 together with the reference SIF emission shape (green) and the retrieval window (red). The second panel shows the measured (black) and modeled (red) spectrum. The residual is shown in the bottom panel.

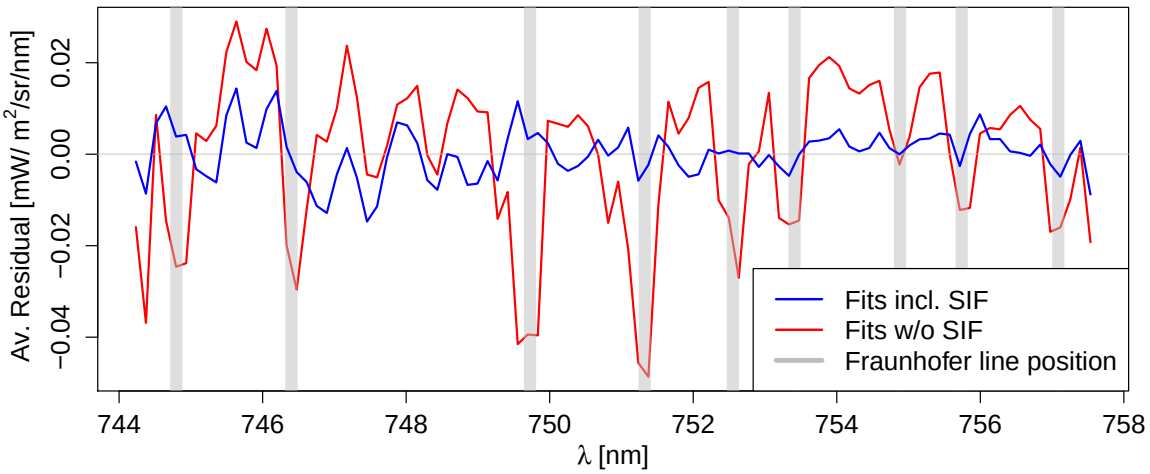

Figure S3: Averaged residuals of one arbitrary TROPOMI orbit where retrieval results (incl. SIF) indicated an amount of SIF_{740nm} larger than 1 mW/m^2/sr/nm. Please see main text for detailed description.

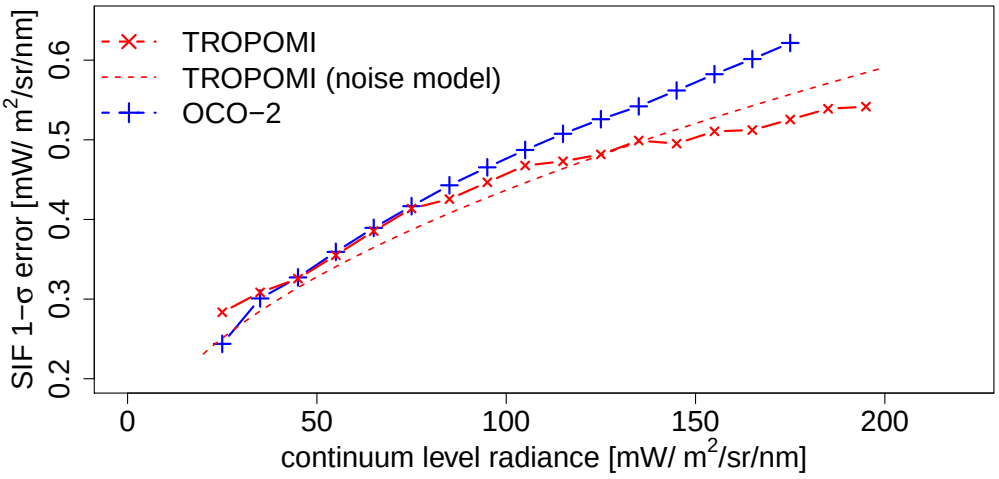

Figure S4: 1-σ retrieval error as function of radiance level derived from real OCO-2 and TROPOMI measurements over the Sahara desert (15°N–30°N, 5°W–30°E) together with the predicted error using the TROPOMI noise model as described in Sect. S4.

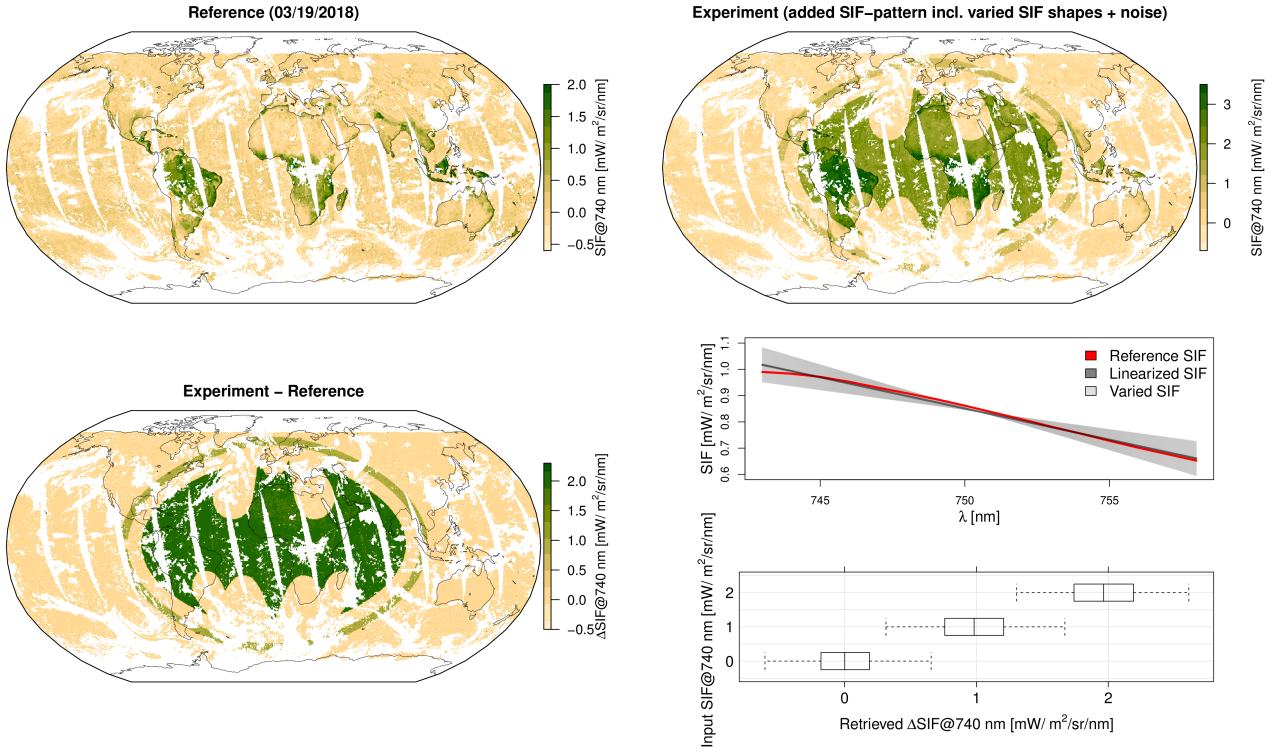

Figure S5: Summary of the "Bat map" experiment. The global maps are based 4.2M soundings.

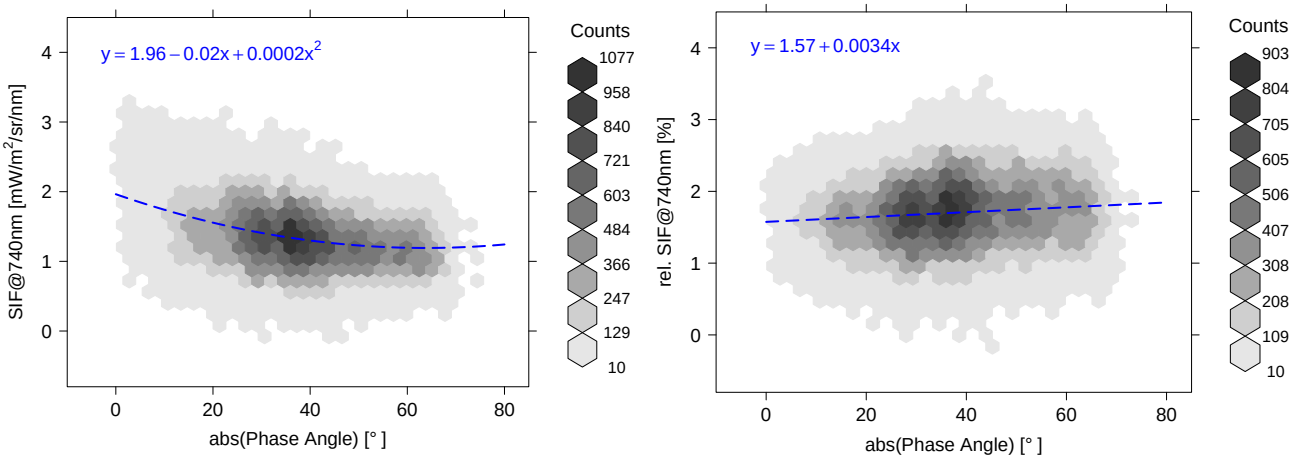

Figure S6: SIF and rel. SIF (normalized by NIR reflectance) in dependence of the observational phase angle for cloud free soundings over EBF regions in the tropics.

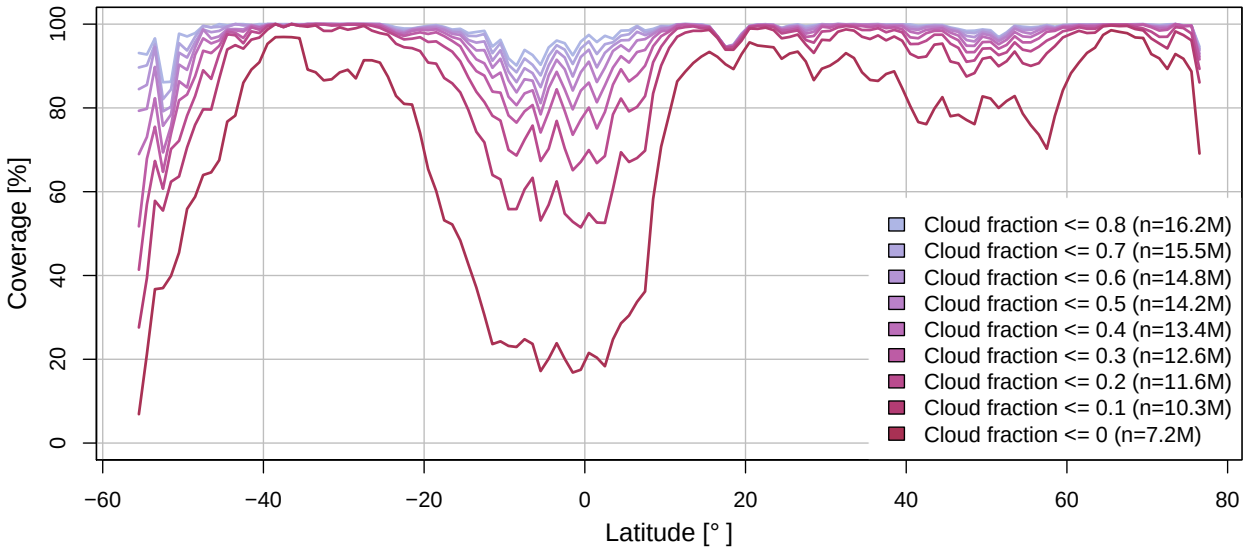

Figure S7: Zonal coverage of land grid cells (0.2° x 0.2°) of TROPOMI SIF for filters of increasing strictness (0.8 to 0). The number of underlying measurements (n) is denoted in the legend.

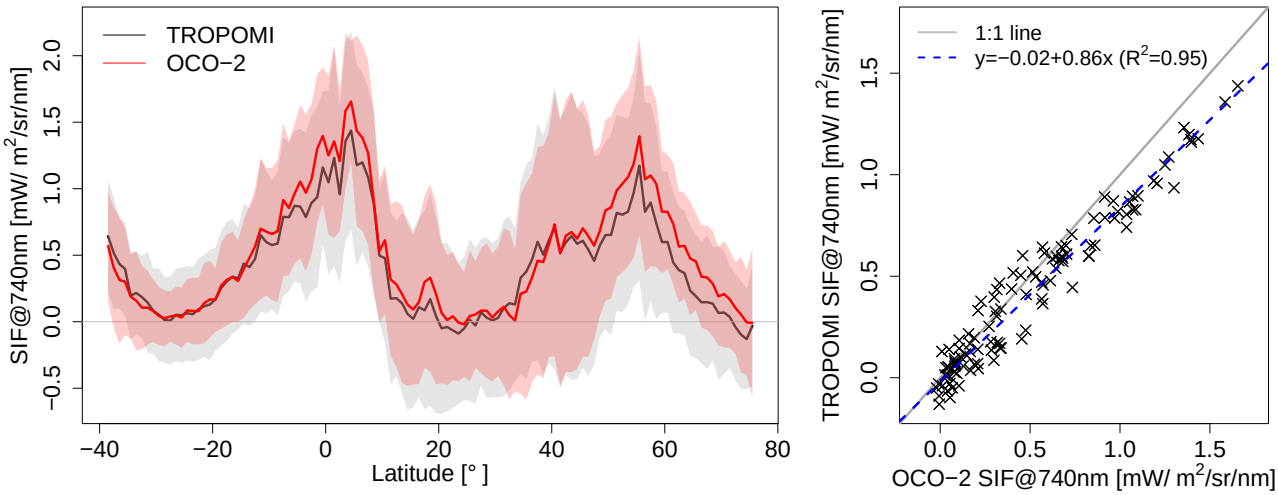

Figure S8: Zonal averages (1° latitude steps) of co-located OCO-2 and TROPOMI SIF retrievals. Shaded areas indicate the longitudinal standard deviation.