Supporting Information for "Global Retrievals of Solar-Induced Chlorophyll Fluorescence at Red Wavelengths With TROPOMI"

Philipp Köhler[1], Michael J. Behrenfeld[4], Jochen Landgraf[6], Joanna Joiner[3], Troy S. Magney[1,5], Christian Frankenberg[1,2]

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

[4]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA

[5]Department of Plant Sciences, University of California, Davis, CA, USA

[6]SRON Netherlands Institute for Space Research, Utrecht, Netherlands

Contents of this file

1. Text S1 to S6
2. Figures S1 to ~~S8~~ S9
3. Table S1

S1. Retrieval algorithm

As outlined in Sect. 2.2, 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 663 nm – 685.3 nm to include as many solar Fraunhofer lines as possible, while avoiding out-of-band signals and atmospheric absorption lines. The required spectral basis functions (or principal components – PCs) for the data-driven retrieval are derived from TROPOMI measurements over areas assumed to be void of SIF. Specifically, we gridded our far-red SIF retrievals over land to a 0.1° x 0.1° resolution on a monthly basis with biweekly sampling and use only soundings over areas with absolute SIF values lower than 0.05 mW/m^2/sr/nm. This strategy allows us to optimize the number of reference spectra with radiance levels that occur also over vegetated areas. Over the ocean, we defined ocean deserts as regions with chlorophyll concentrations less than 0.03mg/m^3 in the annual average, based on monthly data from 2017 on a 0.1° resolution (downloaded from `https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MY1DMM_CHLORA`). Fig. S1 exemplifies the spatial distribution of potential training areas over land in June 2019 together with the (static) ocean deserts.

In a first step towards selecting the training data, potential training spectra are identified on a weekly basis by screening all TROPOMI soundings with respect to potential training areas, radiance levels within the retrieval window (<50/80 mW/m^2/sr/nm over ocean/land), and cloud fractions (< 0.1). Co-located measurements from the Suomi NPP (National Polar-orbiting Partnership) VIIRS (Visible Infra-red Imaging Radiometer Suite) instrument are used for cloud-screening. The S5P-NPP Cloud product contains the number of VIIRS pixels inside a TROPOMI ground pixel, which are identified as confidently cloudy, probably cloudy, probably clear, and confidently clear. In order to estimate an effective cloud cover, we compute the weighted average of these four values using 1, 0.75, 0.25, and 0 as weights. Soundings that meet all initial criteria are then partitioned into ten radiance bins to select training spectra with a balanced distribution of radiance levels. Typically, there are 100 orbits per week and we sample about ten spectra per orbit and radiance bin, resulting in 10k soundings to perform a singular value decomposition (SVD) and derive the necessary principal components (PCs) for the retrieval. The SVD is done separately over 1) land and ocean as well as for 2) each single spatial row of the detector array (448 in total). We do this for two reasons: 1) reflected radiance levels over land are typically higher and display stronger variations compared to water bodies, and 2) the spectral and radiometric characteristics change slightly across the focal plane. A linear combination of a few PCs can then be used to model all spectra with sufficient accuracy, including sensor specific features. Fig. S2 illustrates the retrieval strategy based on the sample spectrum recorded over the upwelling zone with elevated red SIF values at Peru's coastline (same spectrum as in Fig. 1, location is shown in Fig. 3). The left column of Fig. S2 shows the first ten (ocean) PCs of the spatial row of interest (242/448) 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. PC1 can be interpreted as an average spectrum explaining more than 97% of the variance in the training data, which includes the fractional depth of solar Fraunhofer lines in the absence of any SIF emission. PC2 likely combines typical changes in the spectral reflectance of our reference targets and the slope of the solar irradiance. Another typical instrumental effect can be identified in PC4, which represents a subtle wavelength shift. Additionally, to be able to model variations in the spectral reflectance, we use a set of six Legendre polynomials, each multiplied element-wise with PC1 in order to preserve the fractional depth of the Fraunhofer lines. Variations in the spectral reflectance may originate from the surface or the atmosphere (path radiance), but any elastic scattering in our retrieval window (devoid atmospheric absorption lines) represents a multiplicative effect and does not affect the fractional depth of the Fraunhofer lines. Lastly, two spectral functions are necessary to allow for wavelength shifts of the red SIF peak wavelength as well as varying slopes. For this purpose, we performed a SVD over a set of shifted Gaussians (+/-2 nm with increments of 0.1 nm) with respect to the standard red SIF approximation, a Gaussian peaking at 683 nm with a full width at half maximum of 25 nm (Abbott & Letelier, 1999).

In sum, the forward model can now be written as

$$\mathbf{F}_{TOA} = \sum_{i=1}^{10}(\alpha_i \cdot \mathbf{PC}_i) + \sum_{j=1}^{6}(\beta_j \cdot \mathbf{P}_j \odot \mathbf{PC}_1) + \sum_{k=1}^{2}(\gamma_k \cdot \mathbf{h}_k), \tag{1}$$

where α_i, β_j, and γ_k are the state vector elements, $\mathbf{PC_i}$ are the principal components of the SVD, $\mathbf{P}_j$ are the Legendre polynomials, the $\odot$ operator denotes element-wise multiplication, and $\mathbf{h}_k$ are the two functions to model the fluorescence emission spectrum (bold characters indicate variables with a spectral component). In total, we provide 10 PCs, 6 Legendre polynomials, and two functions to model the fluorescence emission spectrum to the retrieval. This means there are initially 18 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 retrieval accuracy and precision as reported by Guanter et al. (2013) and Joiner et al. (2013). However, Köhler, Guanter, and Joiner (2015) proposed to optimize the number of free model parameters by making use of a stepwise model selection, which is also implemented in TROPOMI's far-red SIF retrieval (Köhler et al., 2018). Specifically, we use a backward elimination algorithm to automatize the selection of 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. We find that on average 7 out of 18 state vector elements are automatically selected. $\mathbf{PC}_1$ and $\mathbf{h}_k$ are exceptions from being removed by the backward elimination algorithm to assure that the retrieval estimates the red SIF emission even if its contribution is not significant, in which case $\mathbf{h}_k$ would be dropped by the algorithm. In a final step, the inferred spectrally resolved red SIF estimate is averaged between 680–685 nm (covering the red SIF peak) to report one value.

S2. Sensor Noise

As detailed in Köhler et al. (2018), fewer detector pixels are co-added at the edges of the swath (viewing zenith angles $> 60°$), resulting in a considerably lower Signal-to-Noise Ratio (SNR) for the affected spatial rows. Using spatial rows below/above 20/427 to retrieve SIF is possible in principle but associated with significantly higher uncertainties, which is why we exclude these spatial rows from our analysis.

The SNR within the retrieval window can be estimated by building the ratio between the mean signal level and the standard deviation of the residual. The measurement noise of grating spectrometers is expected to scale with the square root of the signal level. By means of fit residuals for one single day (06/05/2018) we set up spatial row specific SNR models 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). The validity is tested on a different day (07/15/2018) by comparing our SNR model (averaged over spatial rows 20-427) to single retrieval SNRs and the official estimates attached to the L1B data in Fig. S3. The goal is to verify the applicability and performance of our forward model (Eq. 1). Since our SNR estimates agree with the official SNR estimates, we can conclude that there are no over/underfitting issues in the retrieval.

S3. Filtering

Poor retrievals can be identified by the reduced χ^2 (χ^2_{red}), a common statistical metric for the goodness of fit. The χ^2_{red} estimation requires knowledge about the measurement noise/SNR. Here, we use our SNR estimates, which follow the expected scaling with the square root of the signal level. In contrast, visible discontinuities in the official SNR estimates likely originate from stepping through distinct light levels during the pre-flight calibration. In Fig. S4, we compare the retrieved χ^2_{red} to the expected distribution, which can be estimated through the degrees of freedom (166), computed by the number of spectral points in the retrieval window (173) minus the number of state vector elements (7 on average). If we naively use all retrievals by disregarding the trained range of signal levels, the χ^2_{red} distribution is shifted towards higher values with a median of 1.17, pointing to an underfitting of spectra (Fig. S4, left column). However, once we filter for trained radiance levels (Fig. S4, right column), the χ^2_{red} distribution approaches the expected distribution with a median value of 1.03. In order to filter poor retrievals, we accept only retrievals with χ^2_{red} estimates inside the 95% range of expected values, that is $0.8 < \chi^2_{red} < 1.23$.

Fig. S5 illustrates that negative red SIF estimates occur primarily in the vicinity of optically thick clouds, even when the χ^2_{red} filter is applied (Fig. S5c). Since the affected retrievals are classified as satisfactory, we hypothesize that there is an additive spectral signature in the L1B spectra that is unaccounted for, which is modeled sufficiently well by the two spectral functions designed to retrieve the SIF emission. One possibility which might confuse the retrieval algorithm and obtain negative red SIF estimates consists of an added signal that is more pronounced in the shortwave part of the retrieval window and decreases with wavelength. In this context, it should be noted that negative retrieval results are not unphysical *per se* as long as they can be explained by the retrieval noise. However, the retrieval noise leads to positive as well as negative outliers and we observe predominantly negative red SIF values if the stringent radiance filter (Fig. S5d) is not applied. Overall, the following filter criteria are employed to exclude unphysical retrieval results from the analysis:

- $0.8 < \chi^2_{red.} < 1.23$
- Radiance levels <50/80 mW/m^2/sr/nm over ocean/land
- Air-Mass-Factors < 4
- Viewing Zenith Angles $< 60°$

S4. "Zorro"-Experiment

We conducted an experiment in which 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 assess the retrieval accuracy and precision. We used one day (07/15/2018) of TROPOMI orbits (including about 13M single soundings) and added two realistic SIF intensities (SIF@683nm=0.5 / 1 $mW/m^2/sr/nm$) with randomly varying spectral shapes (incl. diverse slopes and peak wavelengths) as shown in Fig. S6. In addition, we degraded the measurements by adding random noise according to the model in Sect. S2, because the original noise level would cancel out when calculating the difference between experiment and reference. For illustration purposes (improved spatial coverage), we applied a relaxed filter of radiance levels ($<$150 $mW/m^2/sr/nm$) before gridding the original and experimental retrieval results to a 0.2° x 0.2° resolution. The difference map exposes the input pattern and illustrates that the retrieval itself performs well, even outside the trained range of radiance levels. However, some areas in the reference map (based on original retrievals) display strongly negative red SIF values pointing to spectral signatures in the L1B spectra, which can interfere with the retrieval in the vicinity or presence of optically thick clouds (see Fig. S5). The comparison between input and ΔSIF@683nm values includes only soundings satisfying all filter criteria (Sect. S7). It can be seen that the retrieval is highly accurate (unbiased), indicating that there is no significant crosstalk between the spectral functions used by the retrieval. The standard deviation of ΔSIF@683nm can be regarded as the *mean* precision error, which amounts to 0.4 $mW/m^2/sr/nm$. However, the estimation of *single* measurement precision errors requires that ocean and land data are analyzed separately with respect to radiance levels as it is done in the following section.

S5. Uncertainty Estimates

Attaching reliable single measurement precision errors is challenging, because red SIF is computed as a superposition of spectral basis functions (multiplied with the two corresponding state vector elements), while the final reported value is an average of the spectrally resolved red SIF in a subset (680-685nm) of the retrieval window (663-685.3nm). In order to bypass an explicit computation, we estimate the precision errors based on the "Zorro" experiment. In particular, we use the difference between the original retrievals and the retrievals with added pattern plus noise (ΔSIF@683nm). Similar to the SNR, it can be assumed that the error is driven by radiance levels. Therefore, we compute signal level dependent error functions for ocean and land separately using the standard deviation of ΔSIF@683nm in distinct radiance bins. To assess the quality of our error estimates, we collected the July 2018 retrievals over potential training areas (surfaces where no SIF emission is expected; ocean deserts and land where our far-red SIF retrievals are near zero) and evaluate the standard deviation in distinct radiance bins. Fig. S7 shows that the predicted single retrieval uncertainties are slightly higher than actually observed over SIF free areas. There is a close agreement over the ocean, resulting in a self consistency and reinforcing confidence in the approach to estimate the uncertainties. Over land, the uncertainty estimates could either be too conservative or simply reflect the challenge to model strong variations in the surface reflectance properties of vegetation. Note that observations over SIF free areas are not available for low radiance levels, the typical radiance range of vegetation. The lack of training data in the relevant radiance levels could explain the higher uncertainties. However, there is a general consistency between the shapes of predicted and observed single measurement precision errors. Over land, we consider the predicted uncertainties to be more realistic than the observed ones, because the prediction is also based on photosynthetically active areas, while the observations are only based on soundings which could have been included in the training data.

S6. Extended MODIS nFLH comparison

In the main manuscript, a quantitative comparison between TROPOMI red SIF and MODIS nFLH is only shown on a monthly basis. For a more detailed comparison, Fig. S7 shows zonal averages of overlapping grid boxes for different aggregation levels in time (daily, weekly, monthly, seasonal) together with the spatial coverage of available grid boxes. The spatial coverage (before co-location) illustrates the potential benefit from TROPOMI red SIF observations, which show an improved spatial coverage on all investigated time scales. Similar to Fig. 2 in the main manuscript, we find remarkably consistent absolute values as well as latitudinal variations across time scales with small discrepancies arising at low latitudes. Given the tendency to retrieve negative values in the vicinity of optically thick clouds, it seems likely that undetected artifacts in the TROPOMI measurements cause a low bias in those regions.

References

Abbott, M. R., & Letelier, R. M. (1999). Algorithm theoretical basis document chlorophyll fluorescence (MODIS product number 20). *NASA (http://www. modis. gsfc. nasa. gov/data/atbd).*

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

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A., Middleton, E., ... Frankenberg, C. (2013). 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*(10), 2803–2823.

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

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

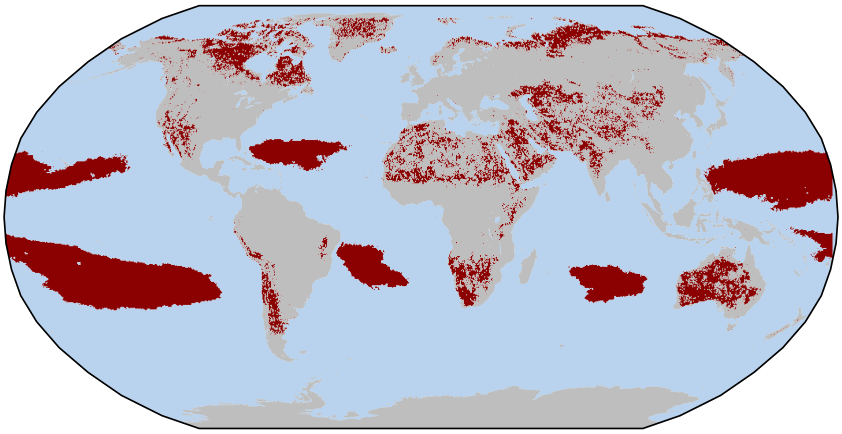

Figure S1. Potential training areas in June 2019, assumed to be void of SIF.

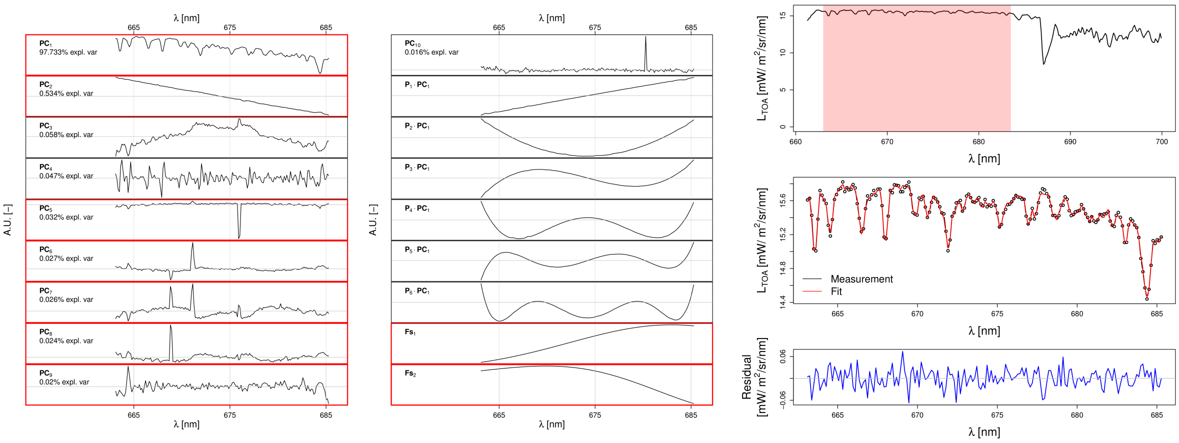

Figure S2. Sample retrieval based on one sounding recorded in the vicinity of Peru's coastline (location is shown in Fig. 3). The left columns show all spectral functions (10 PCs, 6 Legendre Polynomials element-wise multiplied with PC_1, and two functions to model the red SIF emission), which were provided to the retrieval algorithm, while the red boxes indicate the automatically chosen ones. The upper panel on the right shows the measured TROPOMI spectrum in band 5 together with the retrieval window (shaded area in red). The second panel is a zoom-in on the retrieval window and shows the measured (black) and modeled (red) spectrum. The residual is shown in the bottom panel.

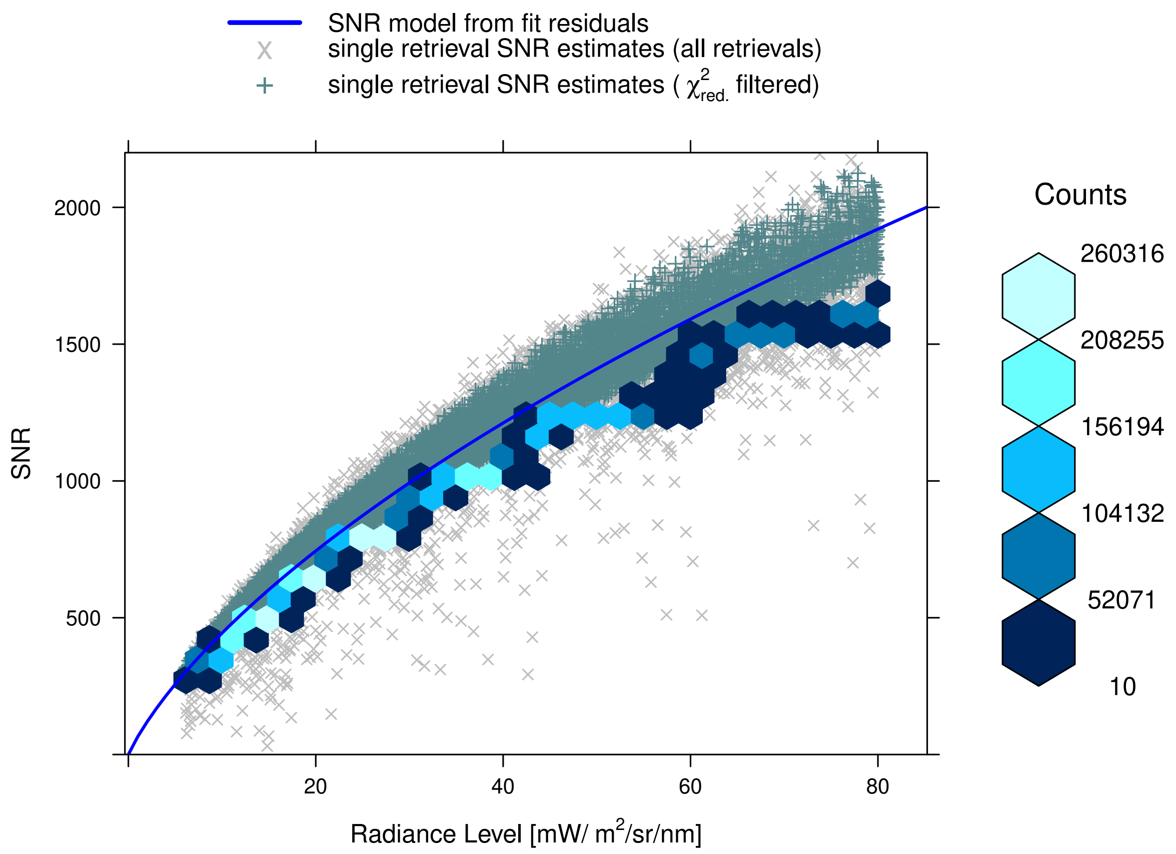

Figure S3. The SNR model (averaged over spatial rows 20-427) derived from 06/05/2018 fit residuals is displayed as blue line, gray crosses denote 10k randomly sampled single retrieval SNRs ~~and~~ from 07/15/2018, while plus signs denote 10k randomly sampled SNRs after applying the χ^2_{red} filter. All 07/15/2018 official ~~L1B~~ SNR estimates (attached to the L1B data) are displayed on top as hexagonal bin plot with the color table to the right indicating the number of soundings per bin.

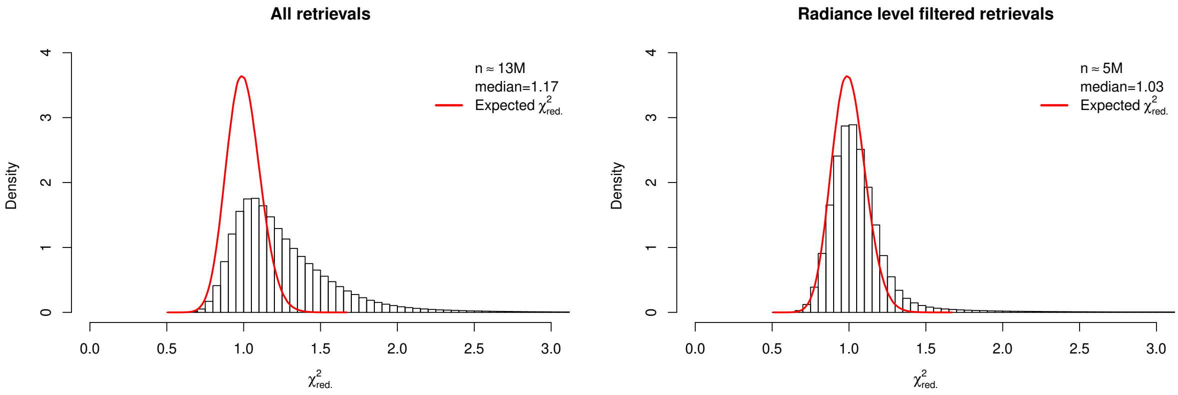

Figure S4. Expected vs observed χ^2_{red} distribution.

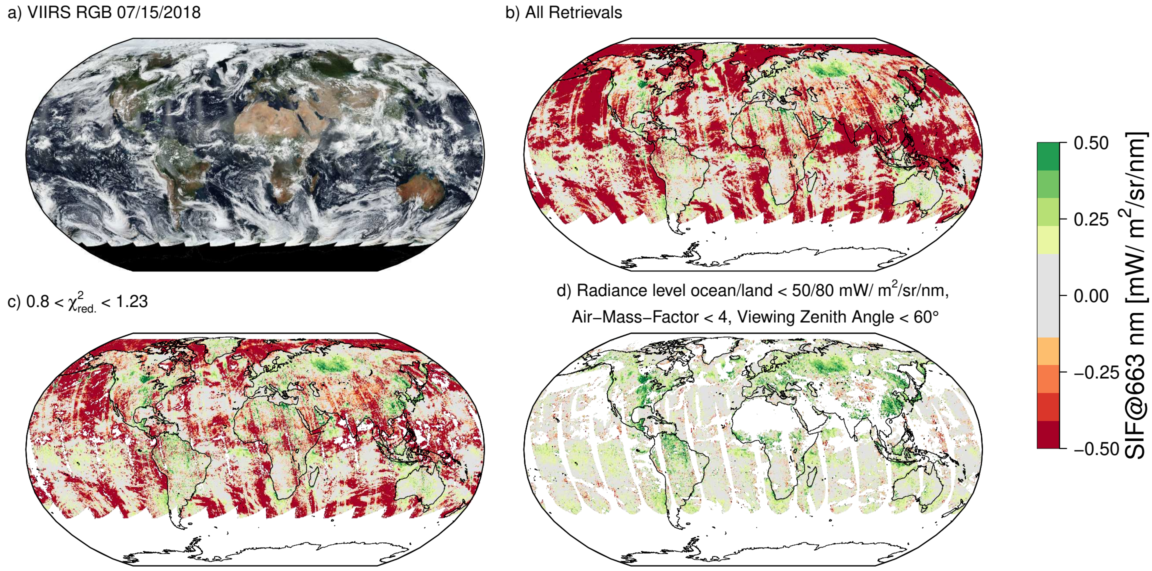

Figure S5. Impact of the filter criteria. The VIIRS RGB image (a) is shown together with the gridded retrievals (one day, 07/15/2018) after applying no filter (b), the χ^2_{red} filter only (c), and additionally the radiance, air-mass-factor, and viewing zenith filter (d).

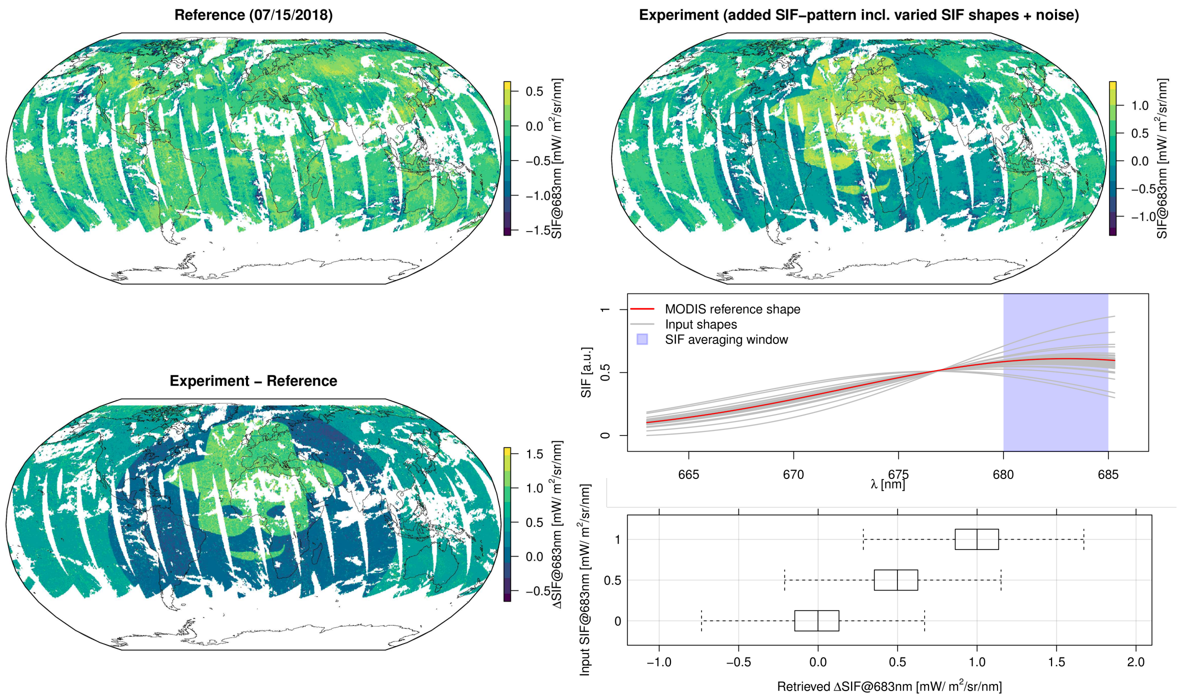

Figure S6. Summary of the "Zorro"-Experiment

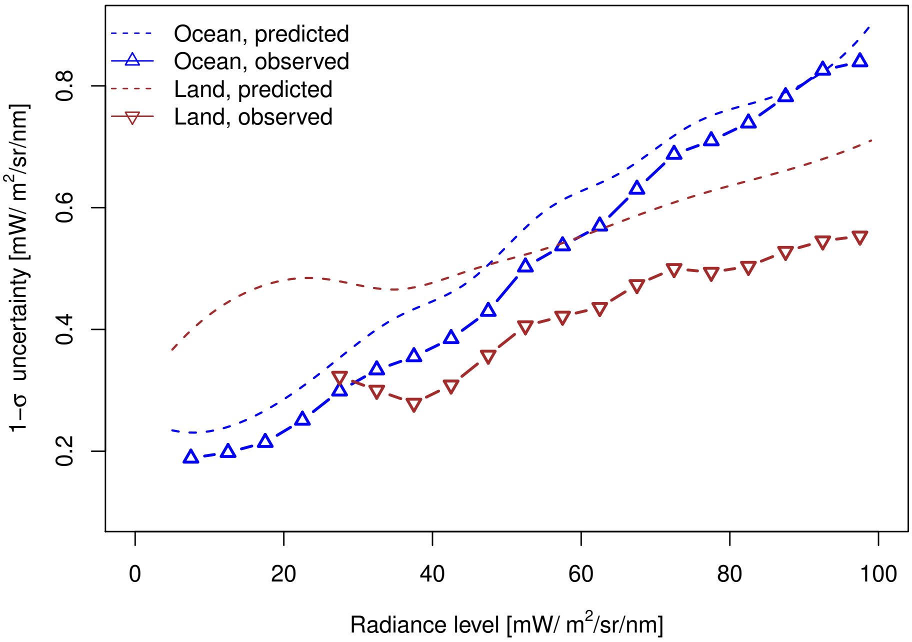

Figure S7. Predicted and observed uncertainties over the ocean and land. Predictions are based on the standard deviation of ΔSIF@683nm from the "Zorro"-experiment in distinct radiance bins. Observations are comprised of the July 2018 retrievals over potential training areas.

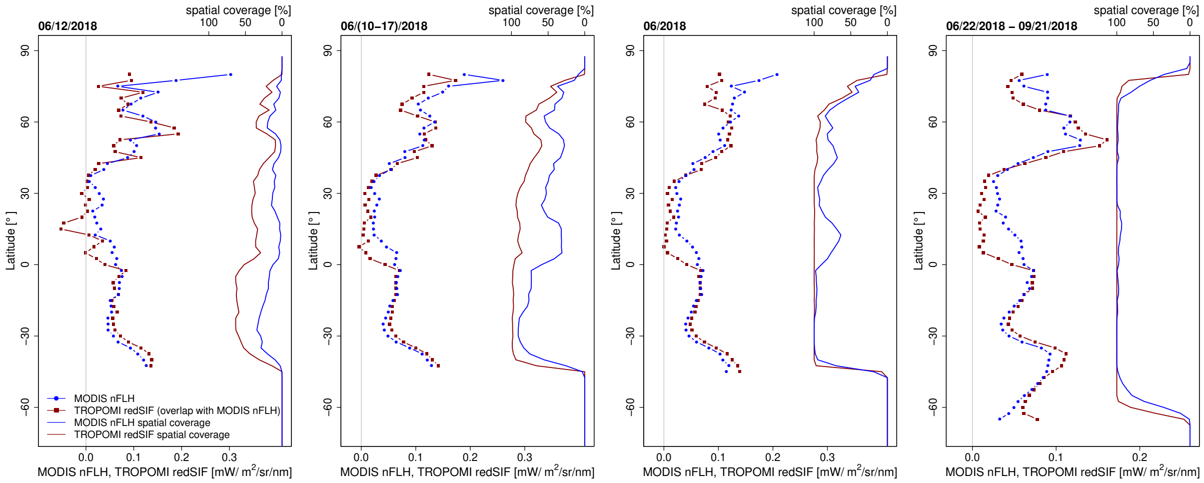

Figure S8. Overlapping zonal averages of TROPOMI red SIF and Aqua/MODIS nFLH together their spatial coverage (before co-location) on a daily, weekly, monthly, and seasonal basis.

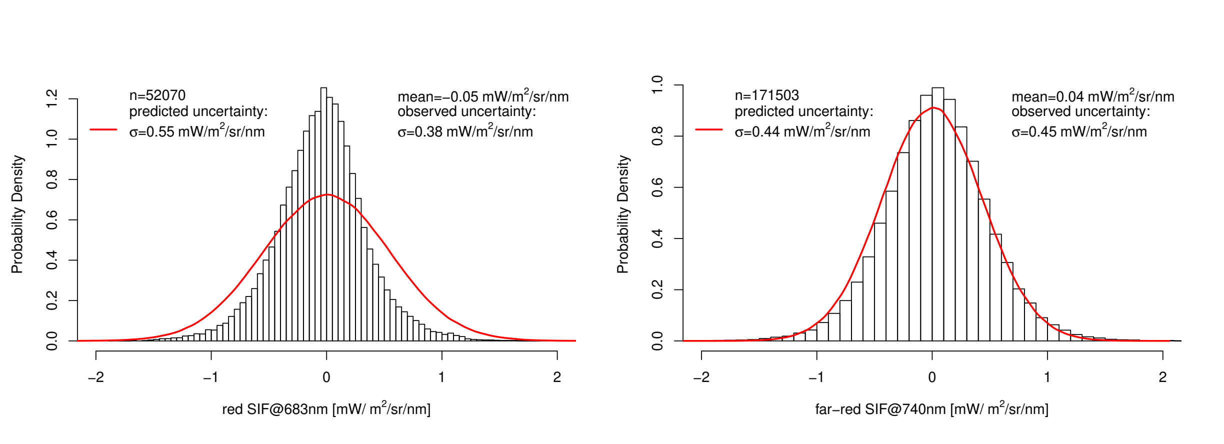

Figure S9. Single retrieval results (05/2018-12/2019) over barren surfaces (Fig. 4 in the main manuscript) are used to test our uncertainty estimates on a regional scale. The red line represents the predicted probability density (assuming the area is void of SIF), while the histograms show the observed probability density. For the red SIF retrievals, the average uncertainty (standard deviation of retrieval results) appears to be slightly lower than predicted (predicted/observed: 0.55/0.38 mW/m^2/sr/nm), while the prediction and observations match remarkably well for the far-red SIF retrievals (0.44/0.45 mW/m^2/sr/nm).

Table S1. Global radiant power of SIF per wavelength unit in [TW/μm] derived from gridded monthly averages (sum of integrals in Fig. 2 of the main manuscript).

	July 2018	October 2018	January 2019	April 2019
red SIF@683nm (ocean)	42.11	43.97	37.47	49.95
MODIS nFLH (ocean)	49.05	59.83	52.16	54.22
red SIF@683nm (land)	39.73	22.52	26.97	26.87
far-red SIF@740nm (land)	279.57	164.41	150.38	166.77