

S1 SIFTER v3 algorithm implementation on GOME-2A and GOME-2B

Table S1. Shows fit parameters, analyzed periods, etc. for GOME-2A and GOME-2B.

	GOME-2A		GOME-2B
	Nominal swath	Reduced swath	Nominal swath
Level-1b	Release 3 (R3)	Release 3 (R3)	Release 3 (R3)
Cloud parameters	FRESCO+ v2	FRESCO+ v2	FRESCO+ v2
Degradation correction			
- Dependencies	t (day), λ , s	t (day), λ , s	t (day), λ , s
- Reference day ($t=0$)	Jan. 5, 2007	Jan. 5, 2007 *	Nov. 1, 2012
- Polynomial order, p	2	3	5
- Fourier order, q	6	6	6
- Fit period	Jan. 2007–Dec. 2012	Jan. 2007–Dec. 2017 *	Nov. 2012–Nov. 2023
- Applied period	Jan. 2007–15 July 2013	16 July 2013–Dec. 2017	Nov. 2012 - Dec. 2023
Retrieval window			
- Spectral range	734–758 nm	734–758 nm	734–758 nm
- Number of points	118	118	117
Principal components			
- Number of PCs	10	10	10
- Collected data	2007–2012	2007–2012, VZA<35°	2013–2018
- Degradation corrected	Yes	yes	yes
- Pre-treatment method	Autoscaling	Autoscaling	Autoscaling

* values measured in nominal swath interpolated to viewing angles corresponding to reduced swath

We assess the reflectance degradation characteristics of the specific GOME-2 sensors by studying the changes in globally averaged reflectance for each wavelength λ and scan index s . Figure S1 shows that GOME-2A reflectances ($\lambda=740.1$ nm, $s=1$ (most eastern)) show a spurious increase in reflectance between 2007 and 2013, followed by a decrease in reflectance between July 2013 and December 2017. The sharp drop in reflectance from July 2013 is due to the reduction of the GOME-2A swath. For GOME-2B, reflectances are more consistent, particularly between 2012 and 2019, but from 2020, a clear impact of instrument degradation occurs.

Figure S2 shows the corrected reflectance at $\lambda \approx 740.1$ nm and $s=1$ for both GOME-2A (nominal swath) and GOME-2B. The corrected time series of the reflectance shows consistency over the long term and a closer connection in observed global reflectance by both sensors. The small offset (~ 1.1 %) between the average of corrected GOME-2A and GOME-2B reflectances can likely be attributed to small differences in instrumental characteristics, such as the spectral resolution (see Table 1 in the manuscript).

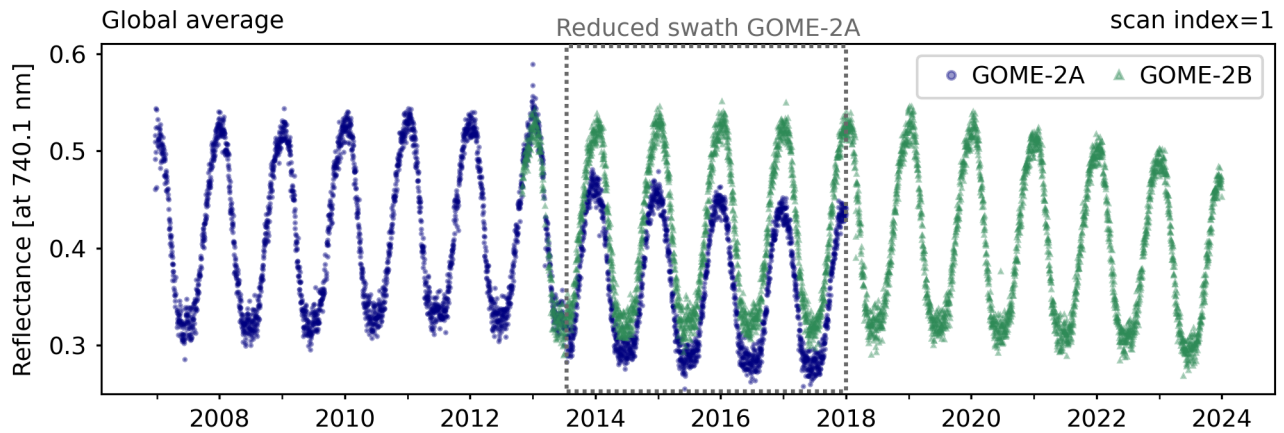


Figure S1. Daily averaged global reflectances (at $\lambda = 740.1$ nm) observed by GOME-2A (in blue) and GOME-2B (in green), and between December 2006–December 2017 and November 2012–December 2023, respectively. The observed reflectances are shown for scan index 1. Due to the swath reduction of GOME-2A from mid-July 2013, the subsequent GOME-2A observations are made under a smaller viewing zenith angle, resulting in decreased reflectance.

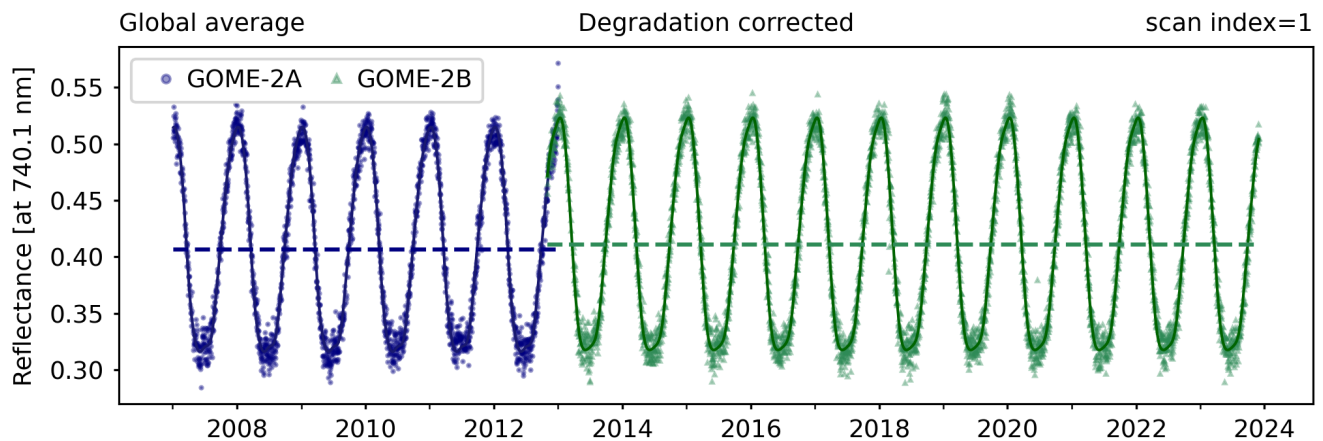


Figure S2. Corrected reflectance at $\lambda \approx 740.1$ nm and scan-index $s=1$ for GOME-2A (under nominal swath, indicated in blue) and GOME-2B (indicated in green) between 2007 and 2023.

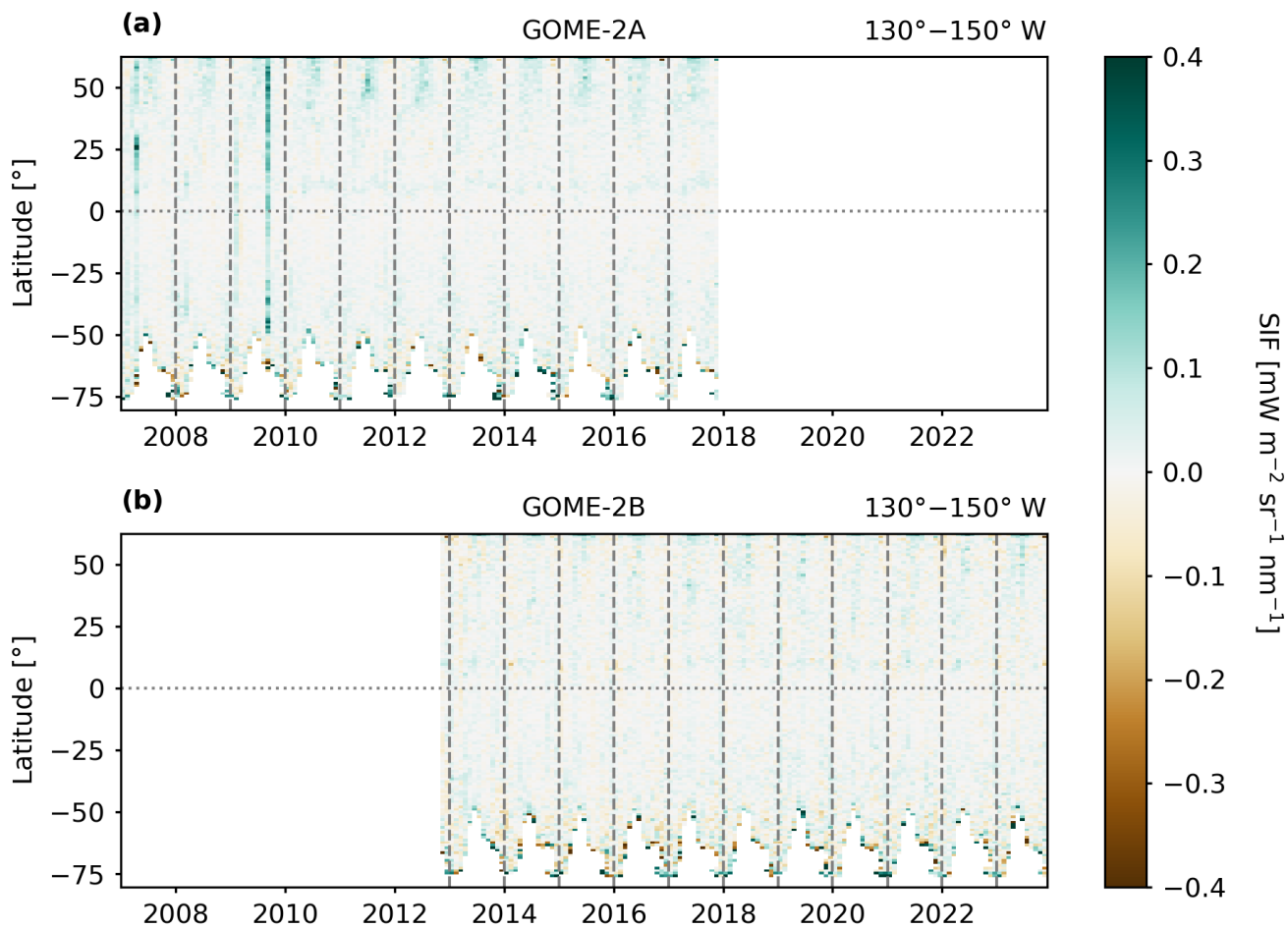


Figure S3. Observed SIF retrieved from (a) GOME-2A and (b) GOME-2B across the Pacific ocean – where SIF is expected to be 0 – per latitude and over time. The SIF values shown *are* adjusted for the latitude bias.

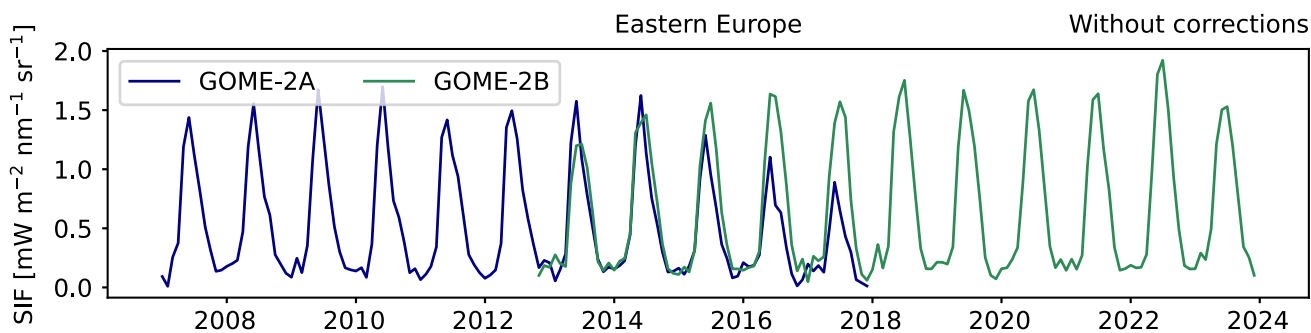


Figure S4. Time series of GOME-2A and GOME-2B SIF uncorrected for degradation and latitude bias correction over Eastern Europe.

S2 Comparison of GOME-2A and GOME-2B SIF

S2.1 Spatial comparison

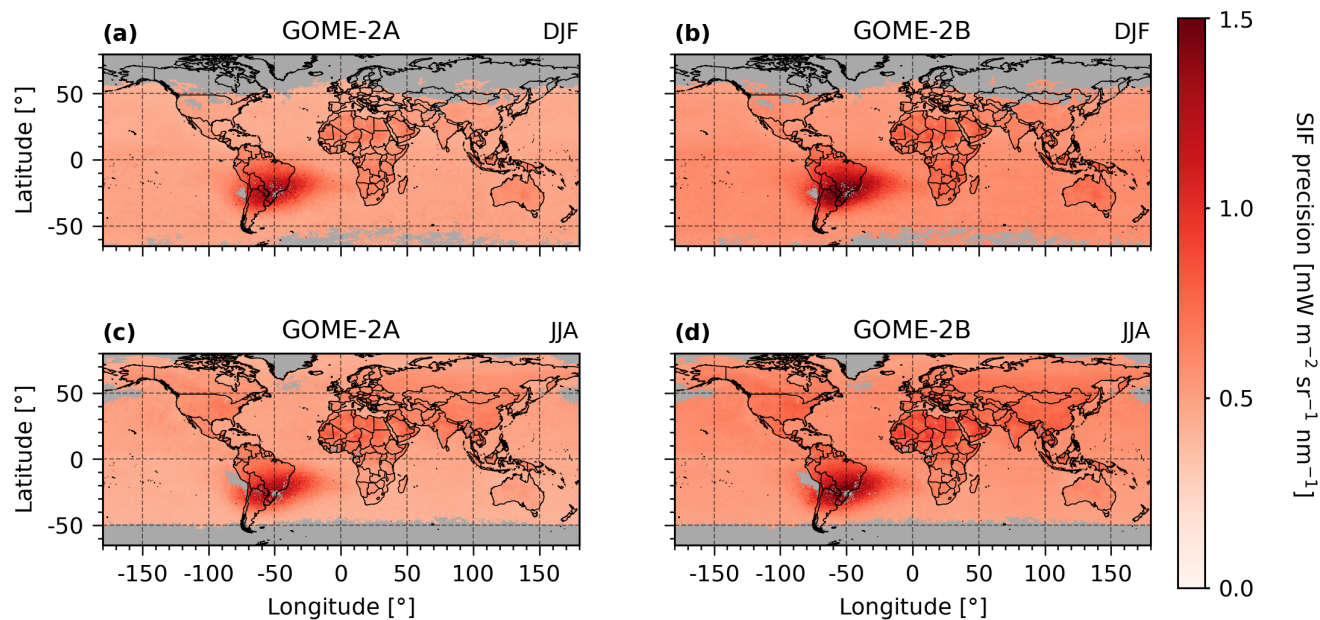


Figure S5. Seasonal averaged SIF precision observed by GOME-2A and GOME-2B over June–August 2016 (a,b), and December–February 2016 (c,d). Larger uncertainty in SIF retrievals are noted over areas influenced by the South Atlantic Anomaly (SAA), including the Amazon region.

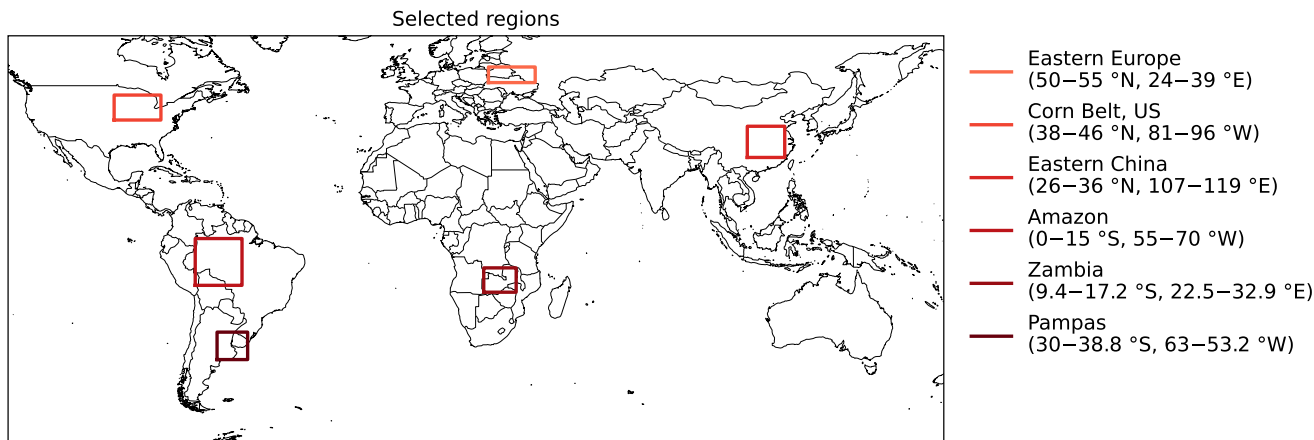


Figure S6. Locations of the studied regions in this study.

Table S2. Seasonally (JJA and DJF 2014–2017) averaged GOME-2A and GOME-2B SIF over different regions and shown over different ranges of the viewing zenith angles. For GOME-2A only the $\pm 35^\circ$ range is shown, as GOME-2A operated under reduced swath in the shown period.

Region	Season 2014-2017	$\pm 35^\circ$		$\pm 52^\circ$
		GOME-2A	GOME-2B	GOME-2B
		[mW m ⁻² sr ⁻¹ nm ⁻¹]		
Eastern EU	DJF	0.09	0.07	0.10
	JJA	1.21	1.22	1.28
US Cornbelt	DJF	0.13	0.10	0.12
	JJA	1.72	1.76	1.81
Eastern China	DJF	0.27	0.26	0.28
	JJA	1.29	1.46	1.51
Amazon	DJF	1.51	1.42	1.57
	JJA	0.88	0.80	0.82
Zambia	DJF	1.01	1.07	1.15
	JJA	0.28	0.29	0.30

Table S3. Regional averages of GOME-2B SIF observations over DJF 2014–2017 and JJA 2014–2017 for the most eastward and westward pixels. Eastward includes ground pixels with viewing zenith angles (VZA)< -35° , whereas Westward includes ground pixels with VZA> $+35^\circ$. This refers to the 6 first and 6 last across-track ground pixels, respectively.

Region	Season 2014–2017	Eastward [mW m ⁻² sr ⁻¹ nm ⁻¹]	Westward [mW m ⁻² sr ⁻¹ nm ⁻¹]	Difference W/E [%]
Eastern EU	DJF	0.10	0.16	-
	JJA	1.18	1.51	+28
US Cornbelt	DJF	0.12	0.17	-
	JJA	1.65	2.08	+26
Eastern China	DJF	0.27	0.33	+22
	JJA	1.42	1.72	+21
Amazon	DJF	1.46	1.97	+35
	JJA	0.75	0.91	+21
Zambia	DJF	1.07	1.40	+31
	JJA	0.29	0.36	+24

S2.3 Collocation

Two co-sampling experiments are performed to assess the comparability of GOME-2A and GOME-2B SIF. Figure S7 illustrates both experiments. Experiment 1 co-samples based on the spatial and temporal overlap, and Experiment 2 co-samples based on the spatial and temporal overlap, and on similarity in viewing geometry. Figure S8 demonstrates examples of accepted

20 spatial overlap between GOME-2A and GOME-2B SIF.

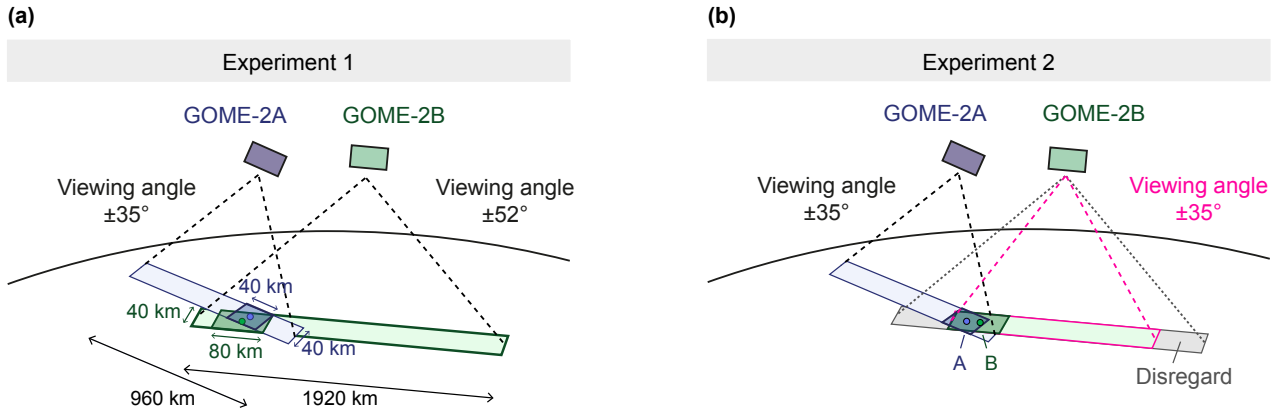


Figure S7. Sketch of the co-sampling of Experiments (a) 1 and (b) 2, in which ground pixels from GOME-2A and GOME-2B are spatially and temporally collocated. In addition to the spatial and temporal collocation, in Experiment 2, the viewing zenith angles of GOME-2B pixels need to align with the range of viewing geometry of GOME-2A. In these experiments, GOME-2A refers to observations done under reduced swath (after July 2013).

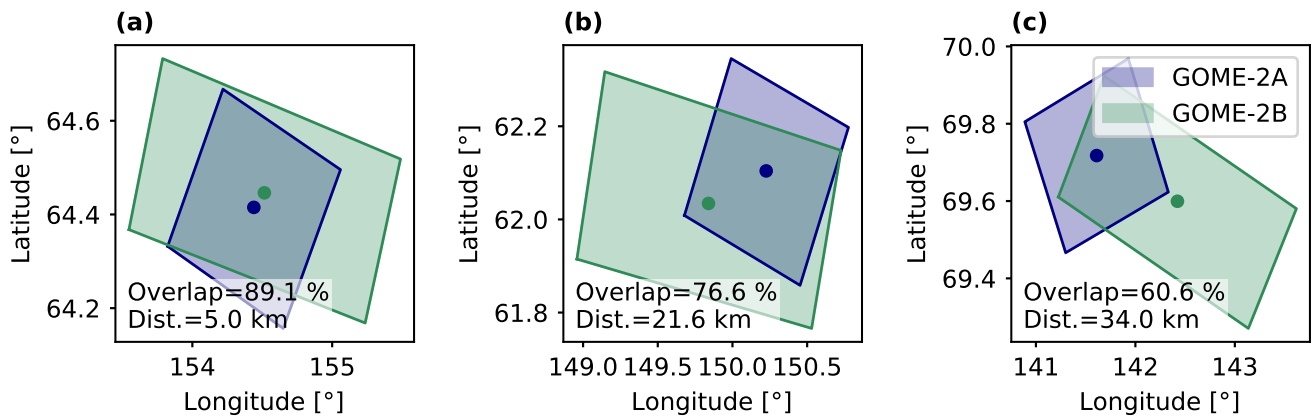


Figure S8. Three examples (a–c) of accepted ground pixels of GOME-2A (in blue) and GOME-2B (in green) when applying collocation. All shown ground pixels are observed over land on 15 June 2014.

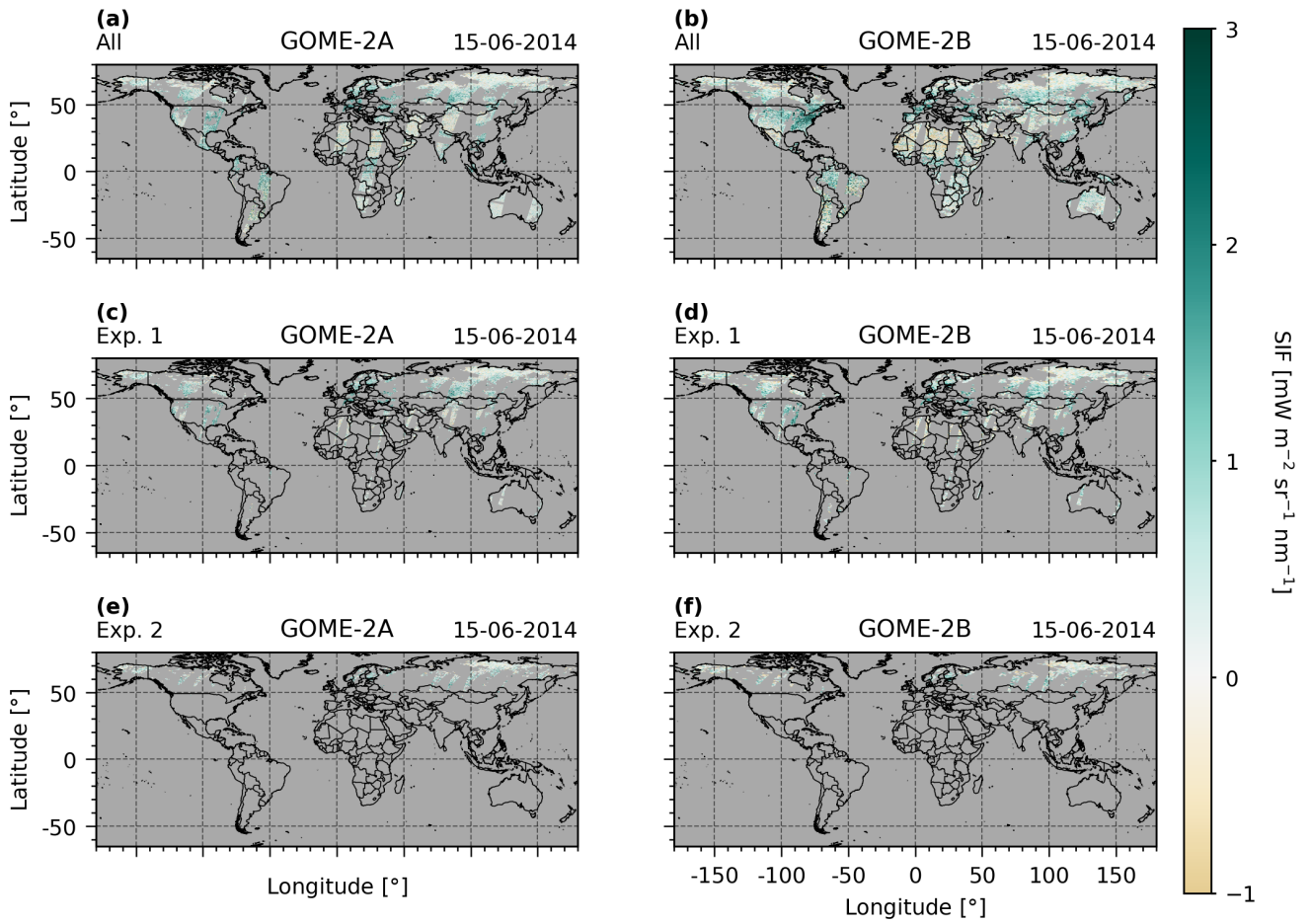


Figure S9. Maps of GOME-2A (on the left) and GOME-2B (on the right) showing the selected ground pixels according to each sampling setting of all data (a,b), Experiment 1 (c,d) and Experiment 2 (e,f). All pixels shown are observed over the land on the 15th of June 2014.

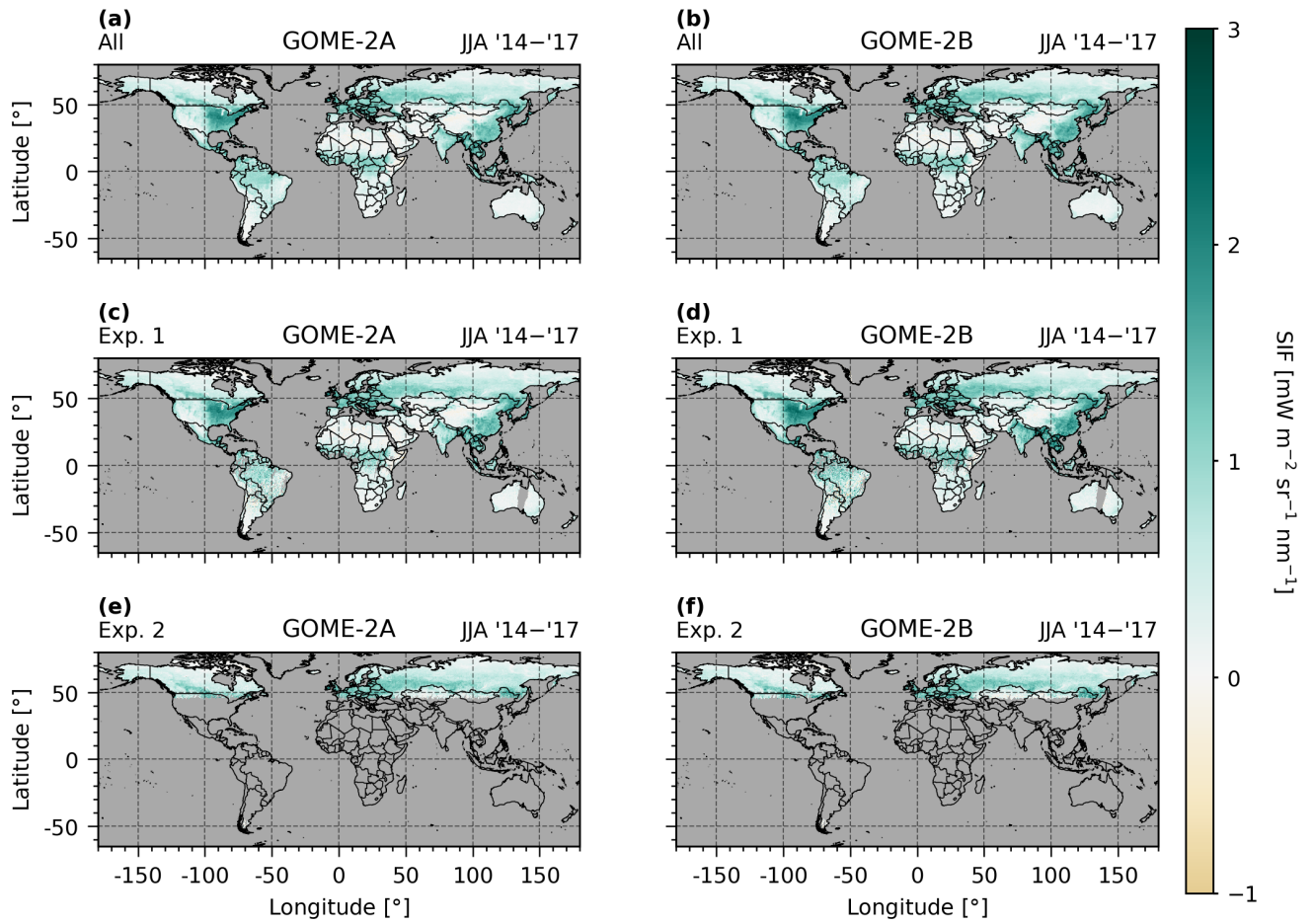


Figure S10. Maps show SIF from GOME-2A (left) and GOME-2B (right), averaged over June–August 2014–2017 and gridded on a $0.5^\circ \times 0.5^\circ$ resolution, using all available data (a,b), data selected according to sampling in Experiment 1 (c,d), and data selected according to sampling in Experiment 2 (e,f), for each experiment. All grid cells that contain valid data are shown. No filtering on counts per grid cell were applied.

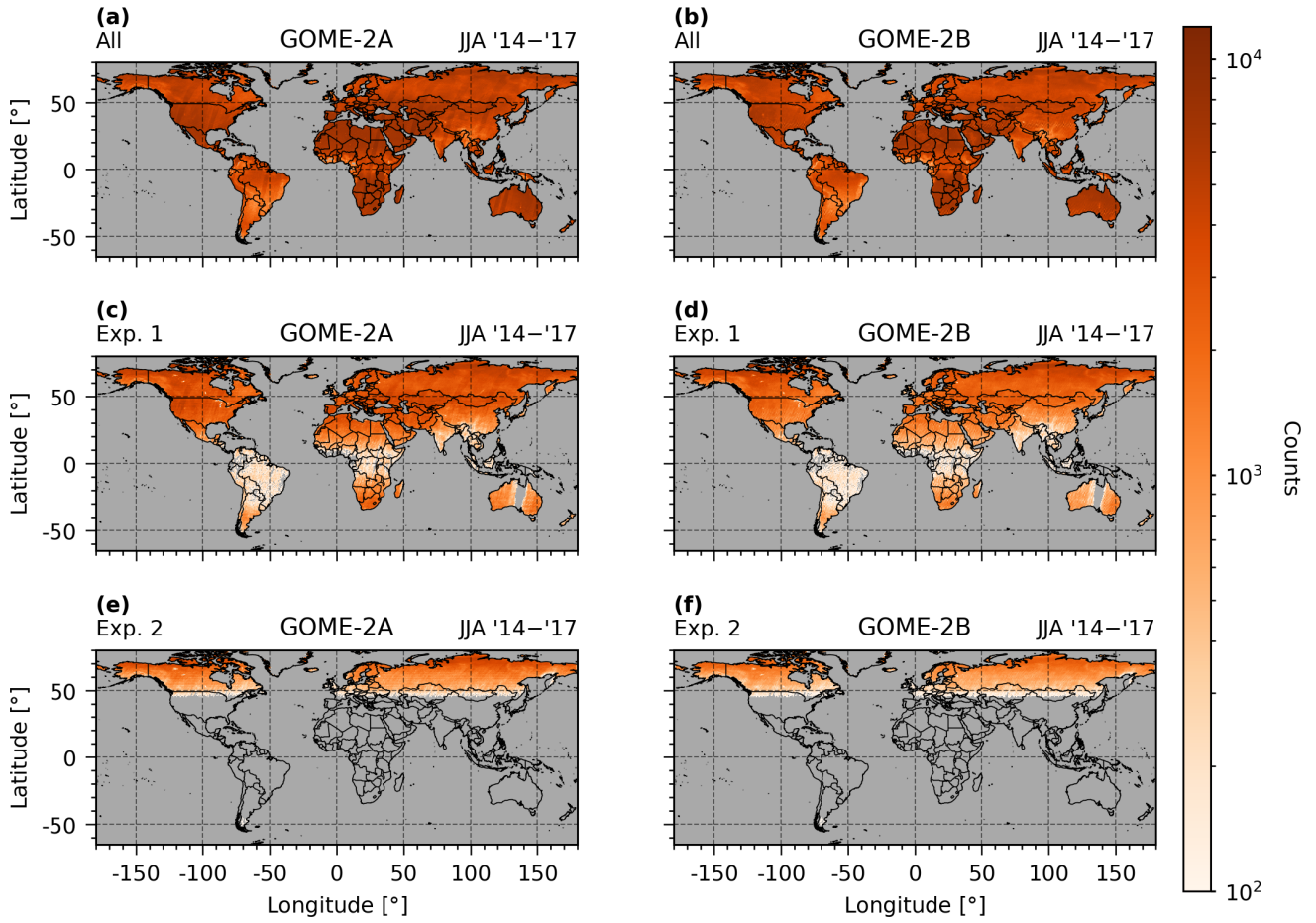


Figure S11. Maps show the counts per grid cell of averaged SIF from GOME-2A (left) and GOME-2B (right) over the JJA 2014–2017, using all available data (a,b), data selected according to sampling in Experiment 1 (c,d), and data selected according to sampling in Experiment 2 (e,f), for each experiment. During the gridding process, each ground pixel was divided into 16 ground pixel segments, whereafter each segment was attributed to a grid cell. The counts represent the number of ground pixel segments within each grid cell.

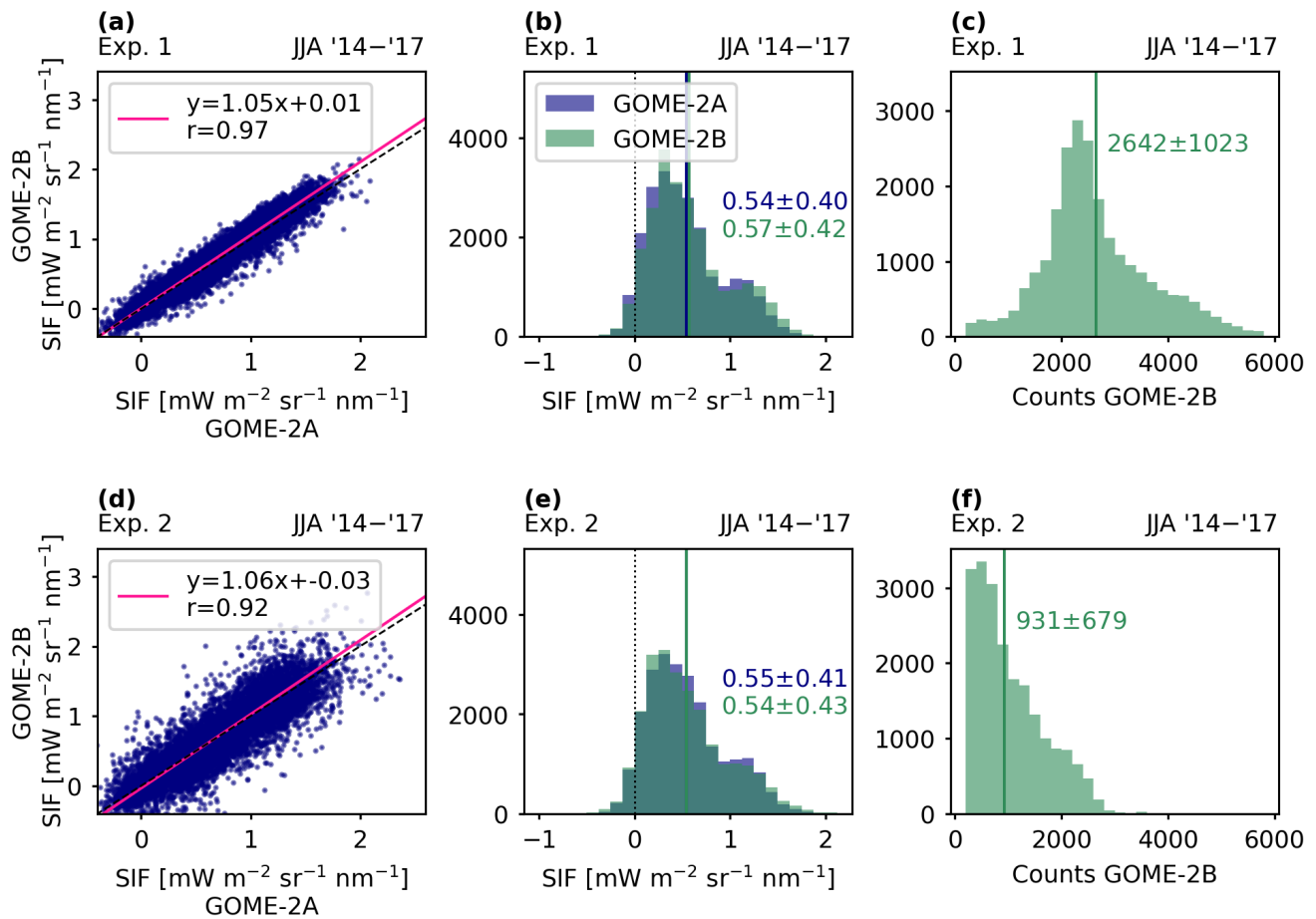


Figure S12. Comparison plots between GOME-2A and GOME-2B SIF observations, following the sampling selection according to Experiment 1 (upper plots) and Experiment 2 (lower plots), using all data (*full*). Subplots (a,d) show the correlation between GOME-2A and GOME-2B SIF using reduced major axis (RMA) regression, and r representing the Pearson correlation. Subplots (b,e) show the histogram plots of GOME-2A and GOME-2B SIF, and subplots (c,f) showing the histograms of the ground pixel segments counts per grid cell.

S2.3.1 Reduced version of Experiment 1

The lower correlation in Experiment 2 may be attributed to the substantially lower number of observations within each grid cell, approximately a 0.35:1 ratio compared to Experiment 1 (Fig. S11). To enable a more balanced statistical comparison between the experiments, we subsampled Experiment 1's observation days to better match the number of observations per grid cell. In this *reduced* setup, Experiment 1 was run using 128 randomly selected days from the JJA 2014–2017 period. The 128 days reflect the approximate 0.35:1 ratio in the average number of GOME-2B pixel segments per grid cell between Experiment 2 and 1 (Fig. S12c,f). Experiment 2 was run using data from all 366 days, but only grid cells with valid observation in both experiments were included in the comparison. Additionally, a stricter requirement for the minimum number of pixel segments per grid cell was set at $\mu - \sigma$, where μ and σ represent the mean and standard deviation of GOME-2B pixel segment counts in the *reduced* Experiment 1.

Table S4 summarises the comparison results between GOME-2A and GOME-2B SIF for the *full* and *reduced* setups in Experiments 1 and 2. The Pearson correlations of 0.95 (Exp. 1) and 0.96 (Exp. 2) indicate strong coherence between the datasets, even without accounting for differences in VZA range. Furthermore, the metrics show a bias of higher GOME-2B

35 SIF values particularly for larger SIF values. However, the comparisons between Experiments 1 and 2 indicate that this bias is largely driven by the inclusion of larger viewing angles in GOME-2B observations. When GOME-2A and GOME-2B SIF observations were matched in VZA range, the bias was reduced to $\sim 2\%$. Our results increase our confidence in the consistency between GOME-2A SIF observations prior to July 2013 and GOME-2B SIF observations (post to July 2013), when both sensors operated with the same 1920 km swath and ground pixel configuration. Note that this reduced experiment has been repeated five times (not shown), yielding similar results.

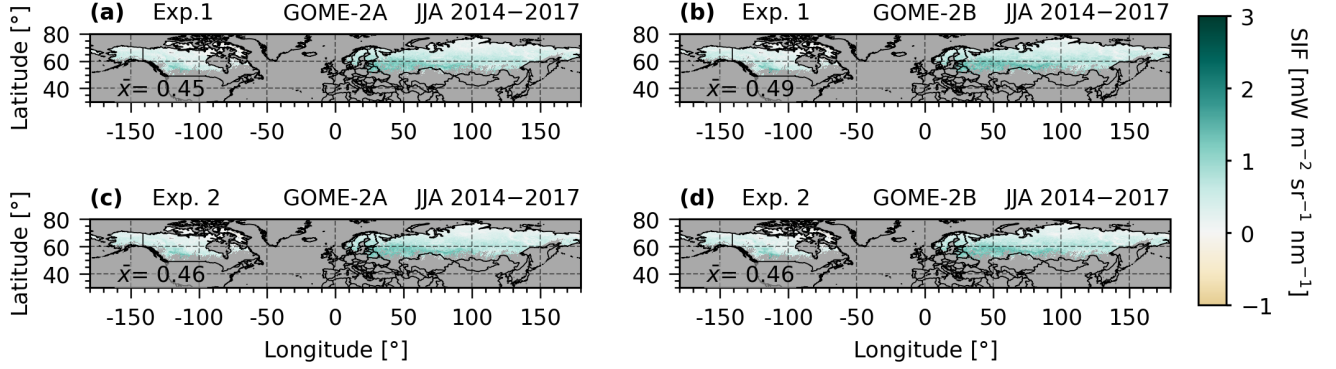


Figure S13. GOME-2A (left subplots) and GOME-2B (right subplots) sampled according to Experiments 1 and 2 and averaged over the June–August (JJA) period between 2014 and 2017 on a $0.5^\circ \times 0.5^\circ$ spatial grid. Grid cells are shown when SIF values were available in both datasets and experiments. Furthermore the number of pixel segments must exceed the μ - σ threshold. With μ and σ representing the average and standard deviation of the average pixel segment counts per grid cell of GOME-2B Experiment 1 data. The mean for each SIF dataset and experiment is shown as $\bar{x}$. The data represents all available data in Experiment 2, but a subsample in Experiment 1 to match the number of ground pixel segment counts in both Experiments.

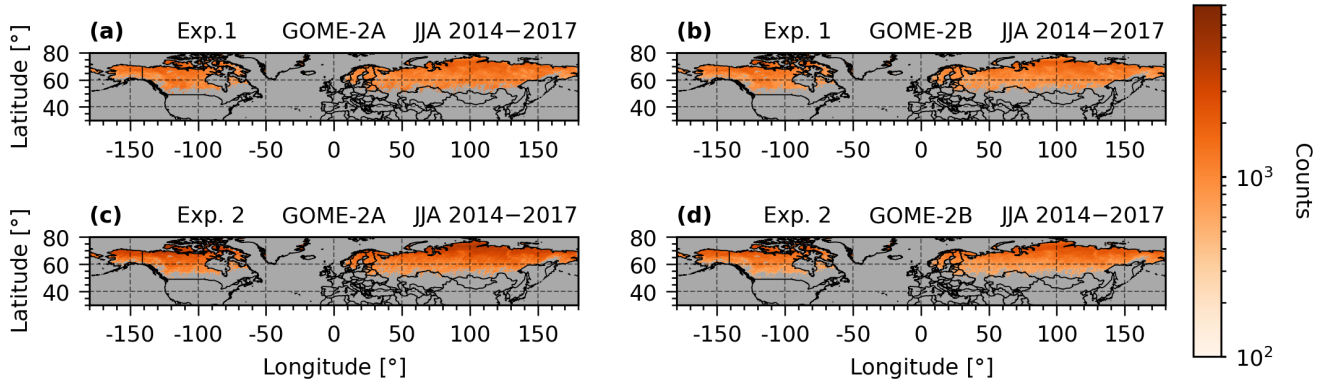


Figure S14. Maps show the counts per grid cell of averaged SIF from GOME-2A (left) and GOME-2B (right) over the JJA 2014–2017, according to sampling in Experiment 1 (a,b), and data selected according to sampling in Experiment 2 (c,d). Experiment 1 uses a subset of the available data (128/366 days) to equalise the counts between both experiments and to allow for a fairer comparison between the two. During the gridding process, each ground pixel was divided into 16 ground pixel segments, whereafter each segment was attributed to a grid cell. The counts represent the number of ground pixel segments within each grid cell.

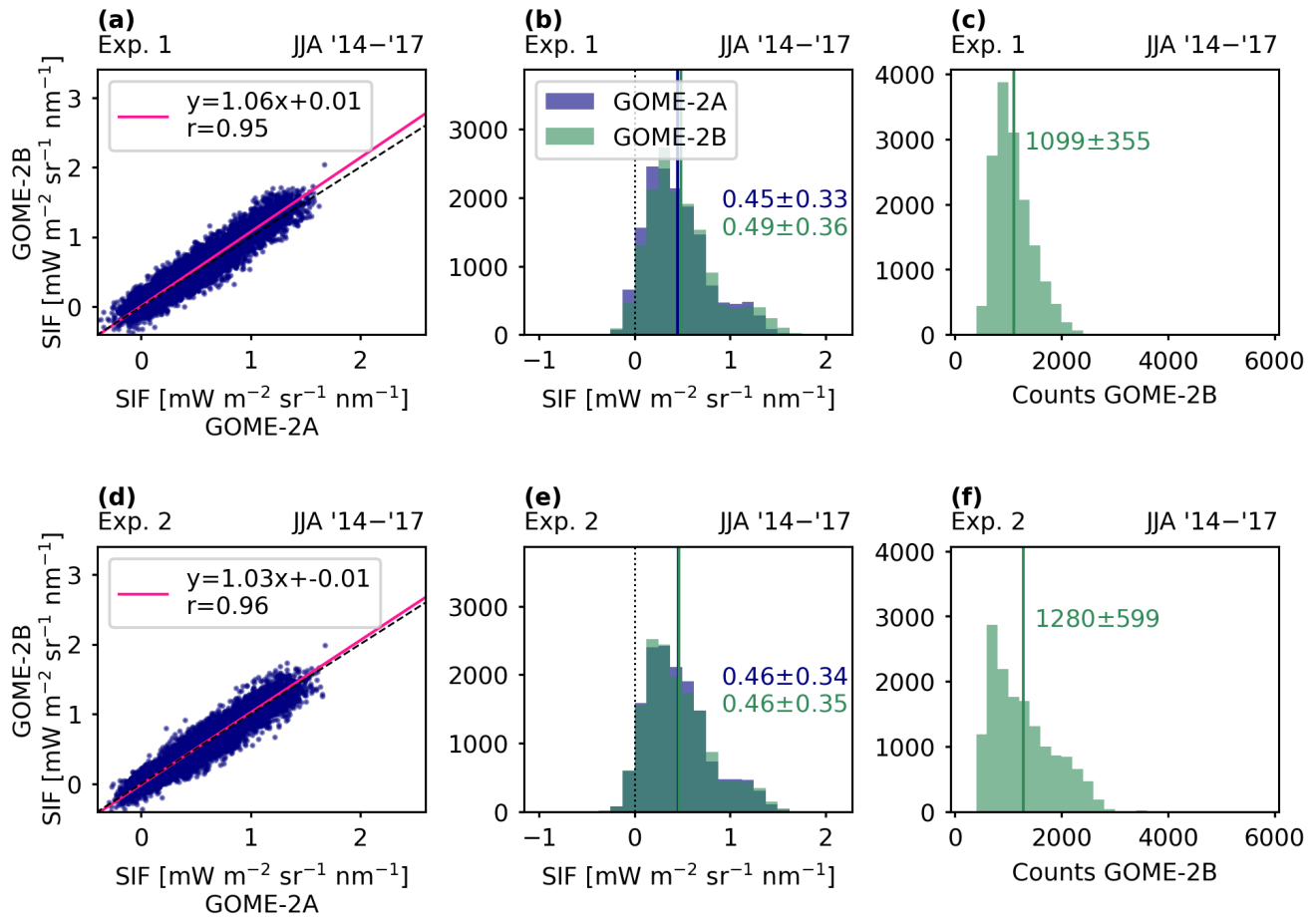


Figure S15. Comparison plots between GOME-2A and GOME-2B SIF observations, following the sampling selection according to Experiment 1 (upper plots) and Experiment 2 (lower plots), using a subset of the data for Experiment 1 to enable an equitable statistical comparison (*reduced*). Subplots (a,d) show the correlation between GOME-2A and GOME-2B SIF using reduced major axis (RMA) regression, and r representing the Pearson correlation. Subplots (b,e) show the histogram plots of GOME-2A and GOME-2B SIF, and subplots (c,f) showing the histograms of the ground pixel segments counts per grid cell.

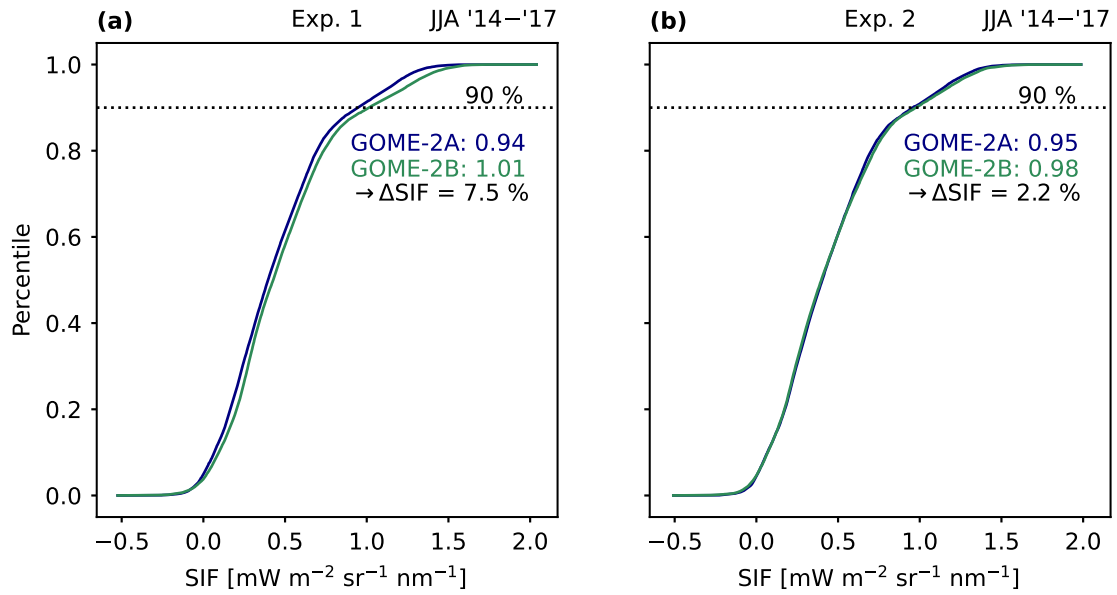


Figure S16. ECDF plots of GOME-2A SIF (in blue) and GOME-2B SIF (in green) in (a) Experiment 1 and (b) Experiment 2. SIF observations are averaged over the JJA 2014–2017 period and cover approximately 50 °N to 80 °N. The dashed black line indicates the 90th percentile line. The value at the 90th percentile for GOME-2A and GOME-2B SIF are shown, as well as the differences between them (in percentages). The data shown here refers to the *reduced* setup, where a selected subset of data is used in Experiment 1 to equalize the pixel counts per grid cell between both Experiments.

Table S4. Comparison of statistical and distributional metrics between GOME-2A and GOME-2B SIF observations averaged over the JJA 2014–2017 period, across two collocation experiments using all available observations within the period (*full*) and a *reduced*-data sensitivity test. The sensitivity test aimed to equalise the number of pixels per grid cell between the two experiments. Observations cover the Northern latitudes between approximately 40° N and 80° N. Data and analysis of the *reduced* sensitivity test are visualised in Figs. S13–S16.

Metric	<i>Full</i>		<i>Reduced</i>	
	Exp. 1	Exp. 2	Exp. 1	Exp. 2
GOME-2A SIF mean	0.54	0.55	0.45	0.46
	mW m ⁻² sr ⁻¹ nm ⁻¹		mW m ⁻² sr ⁻¹ nm ⁻¹	
GOME-2B SIF mean	0.57	0.54	0.49	0.46
	mW m ⁻² sr ⁻¹ nm ⁻¹		mW m ⁻² sr ⁻¹ nm ⁻¹	
Correlation _{A,B}	0.97	0.92	0.95	0.96
Slope	1.05	1.06	1.06	1.03
Intercept	0.01	-0.03	0.01	-0.01
ΔSIF _{90th perc., A,B}	6.9 %	1.7 %	7.5 %	2.2 %
Mean N per point _B	2642	931	1099	1280
Days JJA '14–'17	366	366	128	366

40 S3 Consistency tests of GOME-2A and GOME-2B SIF

Here we provide the full regression results and statistical tests underlying the analysis in Section 5.1 in the manuscript. For clarity, the regression equations used in the analysis are repeated here:

$$y_t = \mu + \alpha t + S_t + \delta U_t + \varepsilon_t \quad (\text{S1})$$

where y_t represents the regional SIF value at month t , μ the monthly mean, αt the linear trend over time, S_t the seasonal component, δU_t the mean level shift term, and ε_t the residual of the observed and fitted SIF.

The seasonal component represents the annual cycle of SIF as a first-order harmonic:

$$S_t = \beta_1 \sin\left(\frac{2\pi(t - \phi)}{12}\right) + \beta_2 \cos\left(\frac{2\pi(t - \phi)}{12}\right) \quad (\text{S2})$$

where β_1 and β_2 control the amplitude, and ϕ represents the phase shift.

Table S5 shows the regression results of the model fit in Eq. S1 for all studied regions. Although the coefficient α is found to be insignificant across all regions, excluding this term substantially affected the step-change coefficient δ , suggesting that an underlying trend is otherwise confounded by δ . Note that the model does not account for interannual variation, limiting detailed analysis of phenological changes or other complex year-to-year dynamics. However, the strong correlation ($R^2 \geq 0.83$) between the modelled and observed SIF indicates a reliable fit for detecting and estimating the step-change coefficient δ . Figure S17 shows the fitted (in pink) and observed SIF (in blue dots) across all six regions.

Table S5 shows the results of the Chow and Likelihood Ratio (LR) tests for all studied regions. Regions with significant p-values are indicated with a star, * for $p < 0.1$ and ** for $p < 0.05$.

Table S5. Regression results of the model fit (Eq. 4 in the manuscript) for all studied regions. Although the coefficient α is found to be insignificant across all regions, excluding this term substantially affected the step-change coefficient δ , suggesting that an underlying trend is otherwise confounded by δ .

Regression results	Eastern Europe	Corn Belt US	Eastern China	Amazon	Zambia	Pampas
R^2	0.85	0.88	0.89	0.88	0.90	0.83
$R^2_{\text{adj.}}$	0.85	0.88	0.88	0.88	0.89	0.83
RMSE [$\text{mW m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$]	0.20	0.22	0.16	0.11	0.12	0.16
ϕ	0.02	0.02	0.60	0.97	0.60	0.60
μ [$\text{mW m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$]	0.53**	0.67**	0.68**	1.22**	0.63**	0.70**
α [$\text{mW m}^{-2} \text{sr}^{-1} \text{nm}^{-1} \text{month}^{-1}$]	3.15e-4	-2.84e-5	4.06e-4	2.02e-4	-4.16e-5	-2.18e-5
β_1	0.10**	-0.07**	0.24**	-0.30**	3.20e-4	-0.25**
β_2	-0.66**	-0.85**	-0.59**	0.30**	0.52**	0.43**
δ [$\text{mW m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$]	9.66e-4	0.06	0.13**	-0.10**	0.03	-0.03

* $p < 0.1$, ** $p < 0.05$

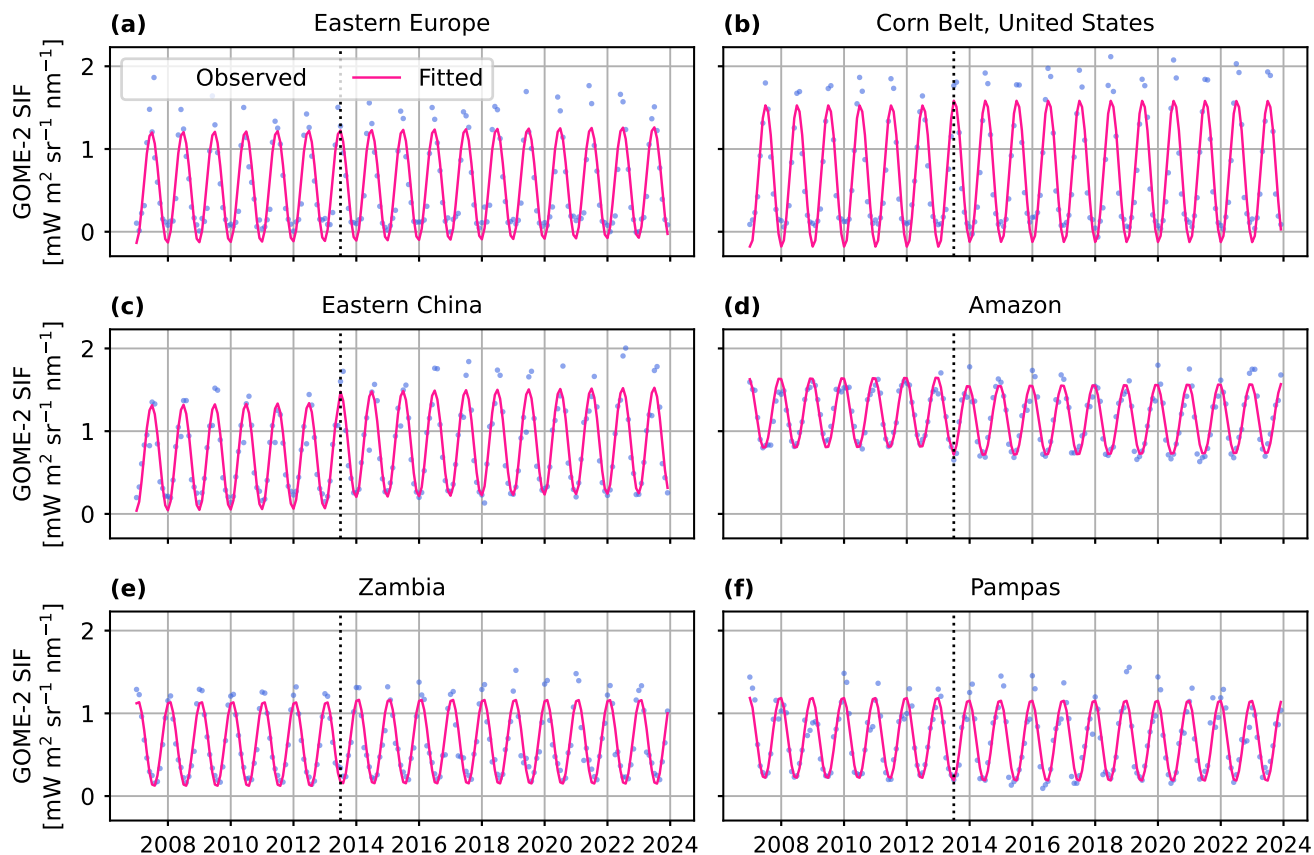


Figure S17. Observed monthly averaged SIF (in blue) and the fitted monthly averaged SIF (in pink).

Table S6. Results of the Chow and Likelihood Ratio (LR) tests for all studied regions.

	Eastern Europe	Corn Belt	Eastern China	Amazon	Zambia	Pampas
F-test, Chow	0.12	0.37	0.31	1.44	0.17	1.11
p, Chow	0.99	0.90	0.93	0.20	0.98	0.36
Likelihood Ratio (LR)	1.40	3.14	51.48	26.76	2.57	2.04
p, LR	0.24	7.63e-2*	7.23e-13**	2.30e-7**	0.11	0.15
R ² _{adj.} (full model)	0.849	0.881	0.893	0.888	0.894	0.830
R ² _{adj.} (reduced model)	0.849	0.880	0.863	0.873	0.893	0.829

*p < 0.1, **p < 0.05

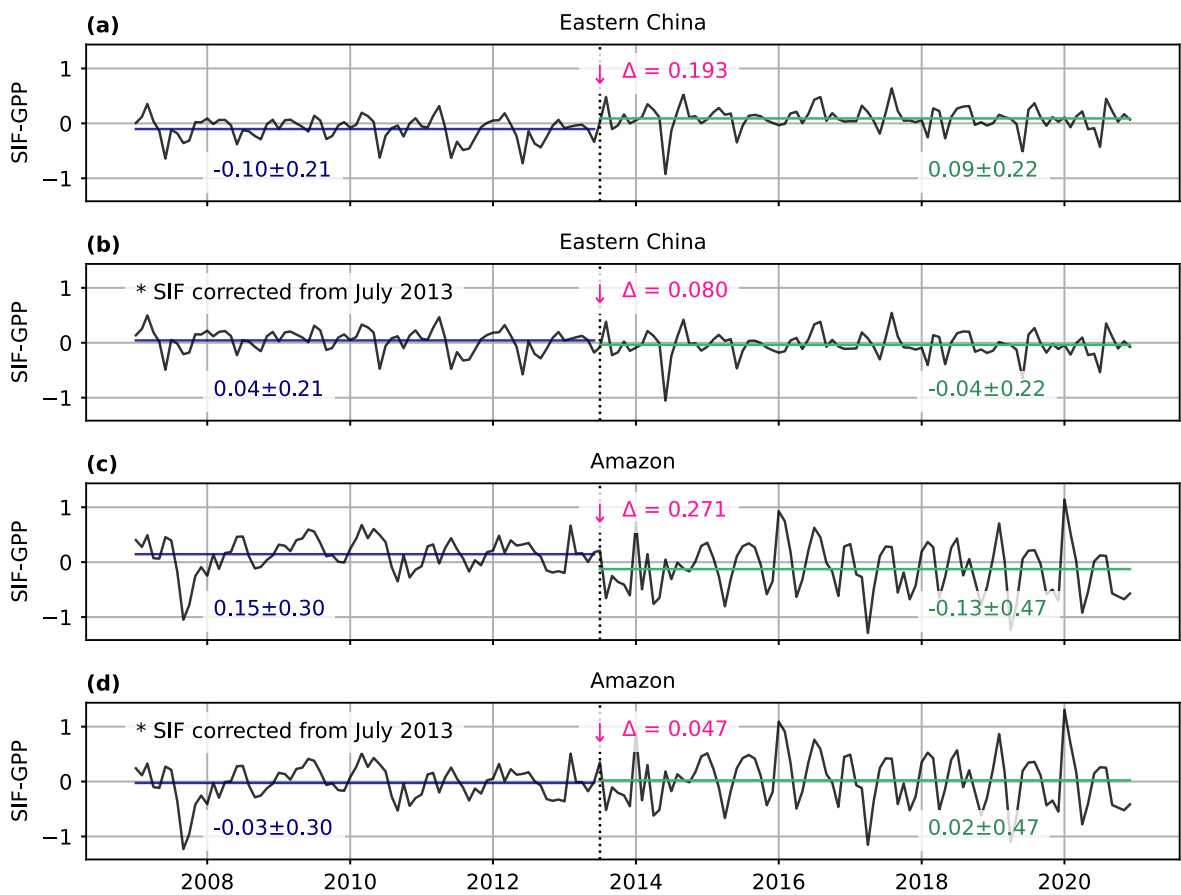


Figure S18. Difference between standardised GOME-2 SIF and standardised FluxSat GPP.

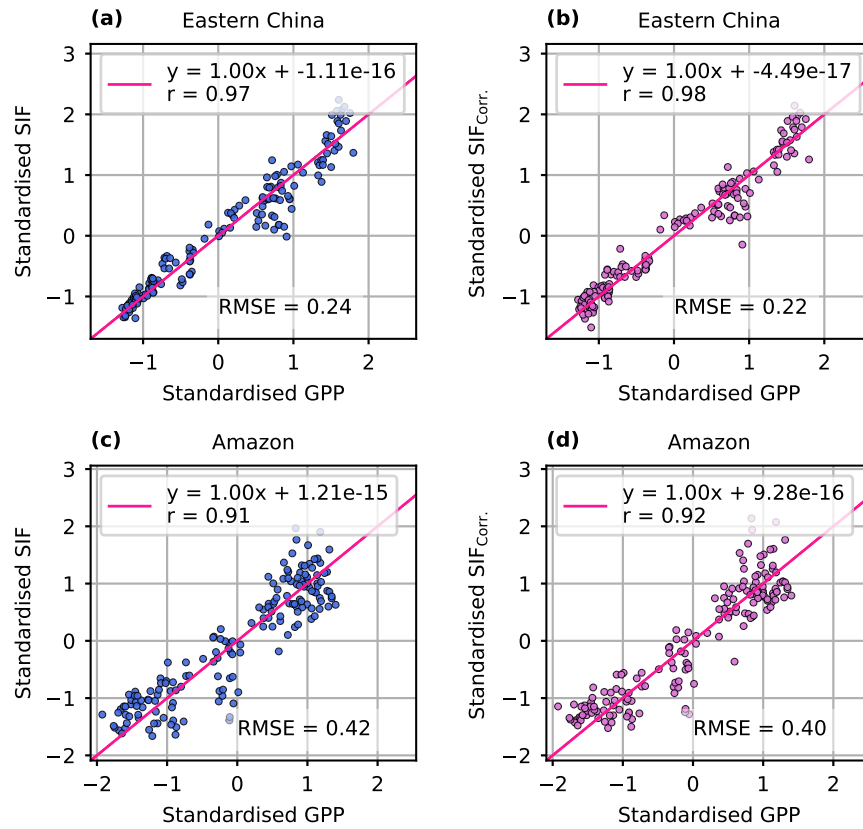


Figure S19. Correlations between standardised GOME-2 SIF and standardised GPP over January 2007–December 2020 for both intersensor offset uncorrected (in blue, plots on the left) and corrected GOME-2 SIF (in pink, plots on the right). Correlations are shown over Eastern China (upper plots) and the Amazon region (bottom plots).

S4 Evaluation of GOME-2 SIF with independent SIF and GPP data

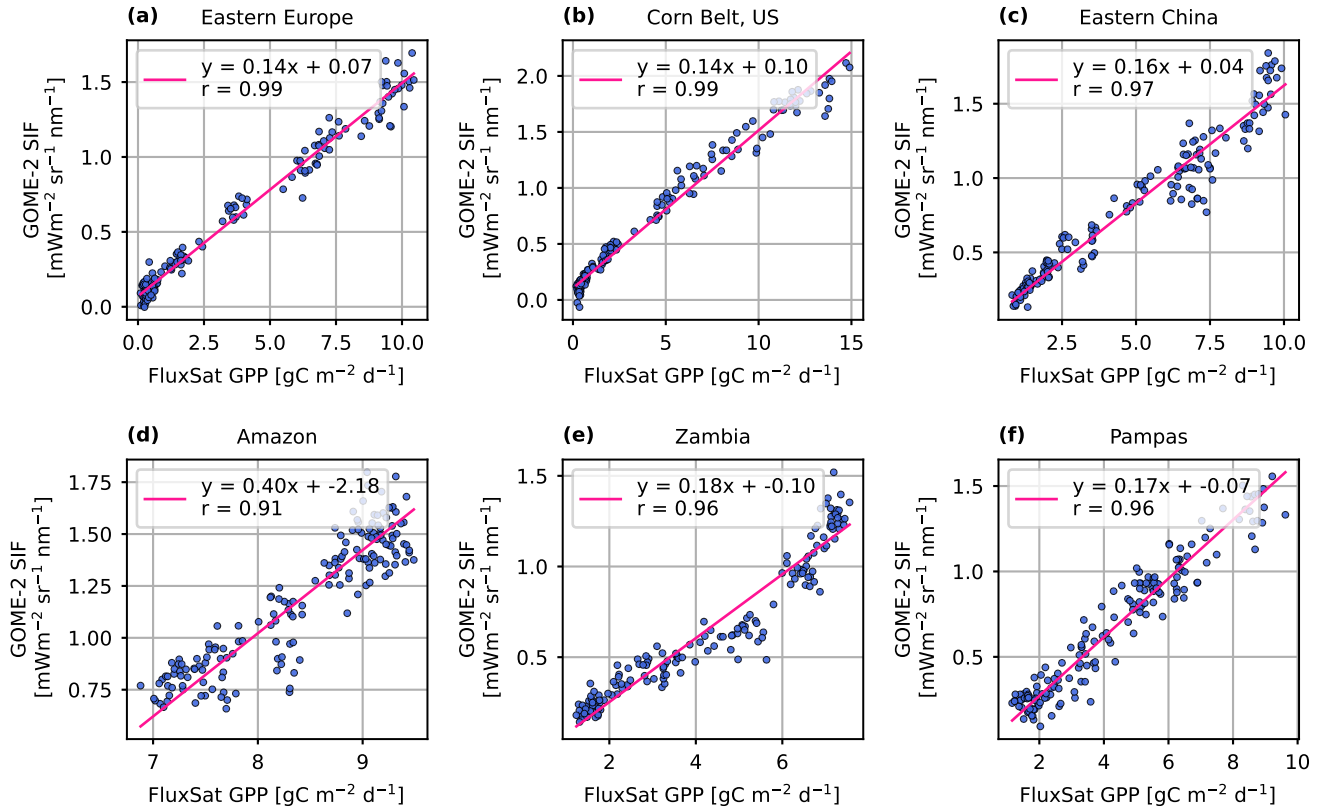


Figure S20. Scatterplots of FluxSat GPP values versus GOME-2 SIF across (a) Eastern Europe, (b) the US Corn Belt, (c) Eastern China, (d) Amazon, (e) Zambia, and (f) Southeastern Australia. Data reflects monthly averages ranging between Jan. 2007 to December 2020. The GOME-2 averages between Jan. 2007 and June 2013 reflect GOME-2A data (under nominal swath configuration), and thereafter reflect the GOME-2B data (also under nominal swath configuration). The regression fit is indicated by the blue dashed line and was obtained using reduced major axis regression. The shown correlation (r) represents the Pearson correlation.

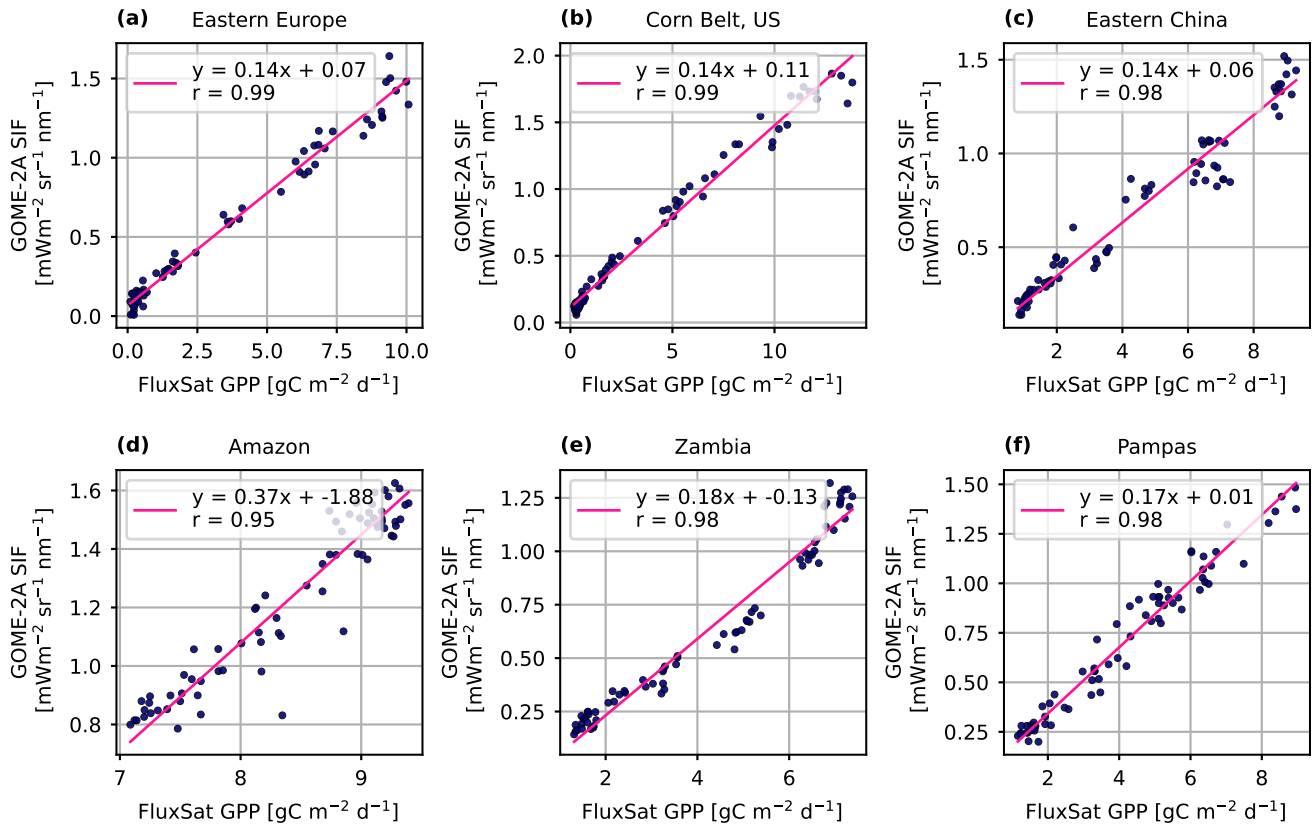


Figure S21. Scatterplots of FluxSat GPP values versus GOME-2A SIF across (a) Eastern Europe, (b) the US Corn Belt, (c) Eastern China, (d) Amazon, (e) Zambia, and (f) Southeastern Australia. Data reflects monthly averages ranging between Jan. 2007 to Dec. 2012. During this period, GOME-2A operated in nominal swath (1920 km) configuration. The regression fit is indicated by the blue dashed line and was obtained using reduced major axis regression. The shown correlation (r) represents the Pearson correlation.

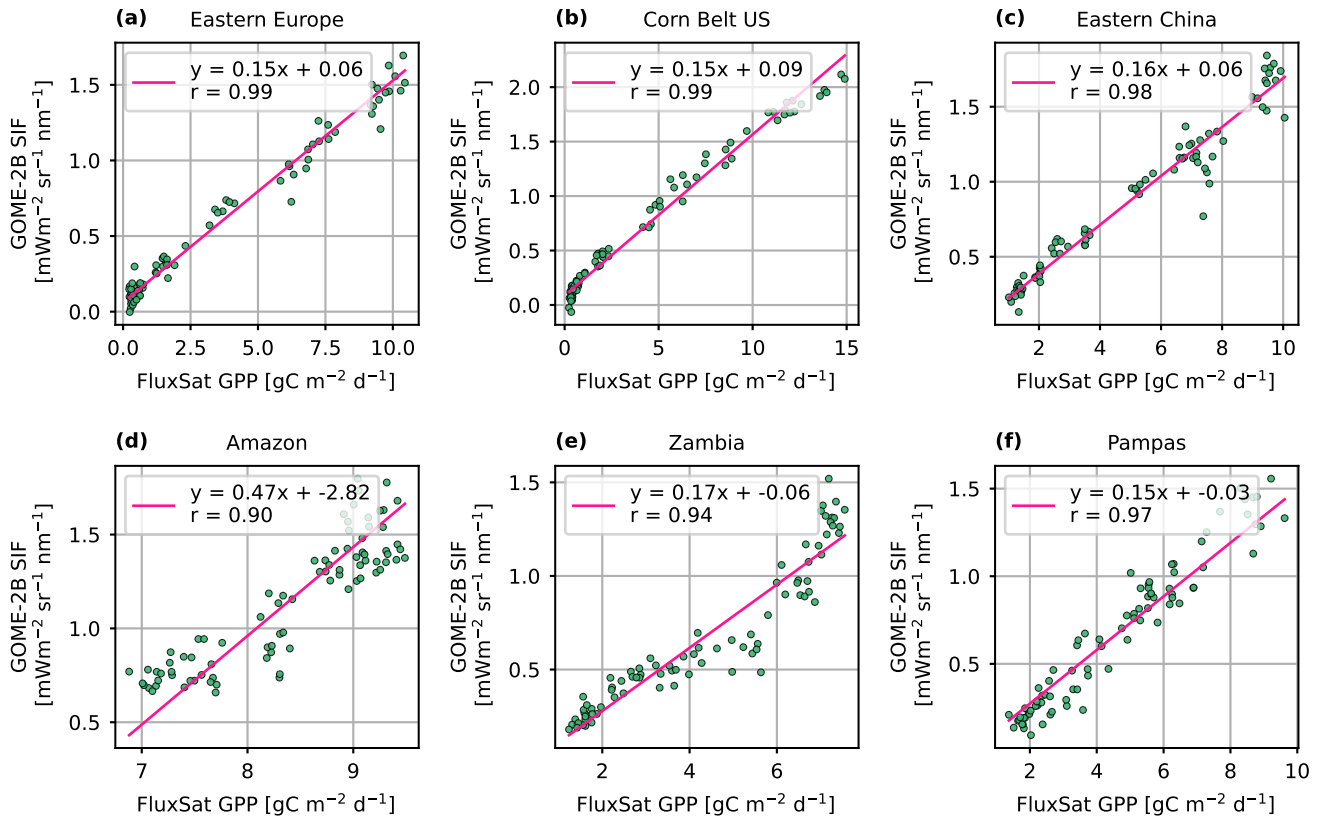


Figure S22. Scatterplots of FluxSat GPP values versus GOME-2B SIF across (a) Eastern Europe, (b) the US Corn Belt, (c) Eastern China, (d) Amazon, (e) Zambia, and (f) Southeastern Australia. Data reflects monthly averages ranging between Jan. 2014 to December 2020. The regression fit is indicated by the blue dashed line and was obtained using reduced major axis regression. The shown correlation (r) represents the Pearson correlation.

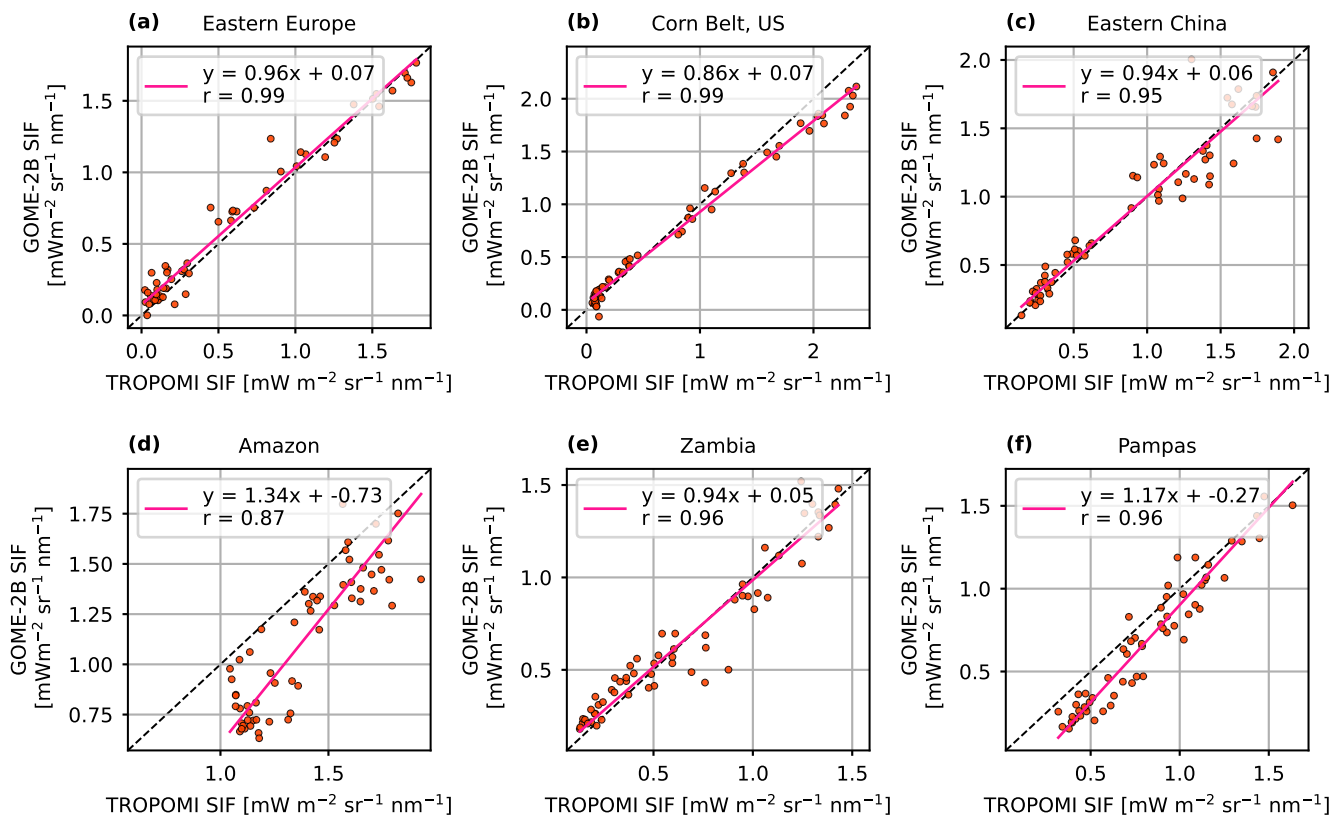


Figure S23. Scatterplots of TROPOMI SIF values versus GOME-2B SIF across (a) Eastern Europe, (b) the US Corn Belt, (c) Eastern China, (d) Amazon, (e) Zambia, and (f) Southeastern Australia. Data reflects monthly averages ranging between February 2018 and December 2022. The regression fit is indicated by the blue dashed line and was obtained using reduced major axis regression. The shown correlation (r) represents the Pearson correlation.

Table S7. Summary of obtained correlations between FluxSat GPP and GOME-2 records (GOME-2A, GOME-2B, combined GOME-2) across all six studied regions. The correlations are obtained from standardised data.

Regions	FluxSat GPP vs GOME-2 SIF		
	GOME-2A SIF	GOME-2B SIF	GOME-2A,B SIF
	Jan. 2007–Dec. 2012	Jan. 2014–Dec. 2020	Jan. 2007–Dec. 2020
Eastern Europe	$y = 1.00x + 2.16e-17$ $r = 0.99$	$y = 1.02x + 7.34e-19$ $r = 0.99$	$y = 1.00x + 1.48e-16$ $r = 0.99$
Corn Belt, US	$y = 1.00x - 2.16e-17$ $r = 0.99$	$y = 1.02x + 1.23e-16$ $r = 0.99$	$y = 1.00 + 6.34e-17$ $r = 0.99$
Eastern China	$y = 1.00x - 3.08e-17$ $r = 0.98$	$y = 0.99x + 3.50e-16$ $r = 0.98$	$y = 1.00x - 5.82e-17$ $r = 0.97$
Amazon	$y = 1.00x - 6.46e-16$ $r = 0.95$	$y = 1.13x + 7.00e-16$ $r = 0.90$	$y = 1.00 + 1.59e-15$ $r = 0.91$
Zambia	$y = 1.00x - 1.83e-16$ $r = 0.98$	$y = 0.97x - 1.66e-16$ $r = 0.94$	$y = 1.00 - 7.14e-17$ $r = 0.96$
Pampas	$y = 1.00x + 1.14e-16$ $r = 0.98$	$y = 0.86x + 3.48e-16$ $r = 0.97$	$y = 1.00x - 2.22e-16$ $r = 0.96$

S5 Sensitivity tests of GOME-2 SIF

To obtain a first-order estimate of the systematic errors within the GOME-2 SIF retrieval, we perturbed several distinct retrieval steps: (i) the degradation correction applied to the reflectance data, (ii) the obtained principal components (PCs) to model the atmospheric transmission, (iii) the interpolation across the slit function, and (iv) the latitude bias correction. For each step, settings were altered within realistic ranges, and the resulting divergence in SIF was evaluated over the Congo Basin area (13° S–6° N, 14–31° W). The maximum difference in average SIF between default (SIFTER v3) and perturbed settings was used as a proxy for the systematic uncertainty contribution of the specific retrieval step. For steps (i), (ii), and (iv), results are shown for GOME-2B SIF on 14 January 2017. For step (iii), slit function interpolation, sensitivity results are shown for GOME-2A SIF on 5 January 2008 and were obtained by Anema et al. (2025).

Degradation correction

In SIFTER v3, GOME-2 reflectances are corrected for degradation trends by fitting a combination of a polynomial and a Fourier series to global reflectance over time (Eq. 1 in the manuscript). The default settings for GOME-2B use a polynomial order $p=5$ and Fourier order $q=6$. For the sensitivity test, we reduced the polynomial order to 2, as used for GOME-2A over January 2007–December 2012 (Anema et al., 2025). The lower polynomial order is expected to underfit the strong decrease GOME-2B reflectance from around 2020. This change increased averaged SIF by $0.06 \text{ mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$, shown in Table S8.

Table S8. Summary of sensitivity tests on the degradation correction, comparing polynomial order $p=5$ (default in SIFTER v3), and $p=2$. Results are shown for 615 pixels in the Congo Basin (13° S–6° N, 14–31° W) on 14 January 2017, retrieved from GOME-2B observations. Pixels of both tests were co-sampled and had to meet the requirements of autocorrelation < 0.2 and cloud fraction < 0.3 . All other retrieval settings were kept at the default setting for GOME-2B. The zero-level offset correction is not applied to these results.

Tests	SIF value	SIF uncertainty
Degr. correction settings	[$\text{mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$]	[$\text{mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$]
$q=6, p=5$	1.24 ± 0.68	0.69 ± 0.08
$q=6, p=2$	1.30 ± 0.67	0.68 ± 0.08
$\text{Max}(\Delta \text{SIF value}) = 0.06 \text{ mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$		

Principal components (PCs)

75 The SIFTER v3 retrieval fits a modeled reflectance to the observed reflectance (Eq. 3 in the manuscript). To model the re-
flectance, a set of 10 principal components (PCs) is used to represent the atmospheric transmission. These 10 PCs are obtained
from a large collection of spectra over the Sahara Desert. We tested two perturbations: the pre-processing of the collected
spectra and the number of PCs used. Using variance scaling (as in SIFTER v2) instead of standard deviation scaling (as in
SIFTER v3), increased the averaged SIF by 0.04 mW m⁻² sr⁻¹ nm⁻¹ and increased the SIF uncertainty by 0.06 mW m⁻²
80 sr⁻¹ nm⁻¹ (Table S9), consistent with earlier GOME-2A results obtained by Anema et al. (2025).

The number of PCs used in the fit retrieval is a known source of uncertainty (van Schaik et al., 2020; Köhler et al., 2015).
Too many PCs could result in overfitting, while too few PCs may fail to capture all spectral information. SIFTER v1 originally
used 35 PCs, which were reduced to 10 PCs in SIFTER v2 (van Schaik et al., 2020). The number of PCs in comparable SIF
retrievals also varies, for example, Köhler et al. (2015) uses 5 PCs for GOME-2 and SCIAMACHY SIF, whereas Guanter et al.
85 (2021) uses 4 or 8 PCs to retrieve TROPOMI SIF, depending on the retrieval window. We therefore progressively reduced the
number of PCs from 10 (baseline in SIFTER v3) to 4 and repeated the retrieval (results in Table S9). Average SIF was highest
with 10 PCs and lowest with 6 PCs, giving a maximum difference of 0.47 mW m⁻² sr⁻¹ nm⁻¹. In absolute terms, the reported
SIF uncertainty values were lower for nine or fewer PCs than for 10 PCs; however, the relative uncertainty was lowest for 10
PCs. Nonetheless, these results suggest that using fewer PCs than 10 could potentially enhance the retrieval fit, and reducing
90 the number of PCs should be explored further in future studies.

Taken together, changes in PC pre-processing and the number of PCs, showed a maximum spread of 0.51 mW m⁻² sr⁻¹
nm⁻¹, with the number of PCs being the dominant contributor.

Table S9. Summary of sensitivity tests on the PCs, comparing different pre-processing settings and numbers of PCs. The SIFTER v3 default
setting is standard deviation (SD) scaling and the use of 10 PCs. Results are shown over 587 pixels in the Congo Basin (13° S–6° N, 14–31°
W) on 14 January 2017, retrieved from GOME-2B observations. Pixels of each test were co-sampled and had to meet the requirements of
autocorrelation < 0.2 and cloud fraction < 0.3. All other retrieval settings, such as the degradation correction, were kept at the default setting
for GOME-2B. The zero-level offset correction is not applied to these results.

Tests		SIF value	SIF uncertainty
Num. of PCs	PC scaling	[mW m ⁻² sr ⁻¹ nm ⁻¹]	[mW m ⁻² sr ⁻¹ nm ⁻¹]
10	PC _{var}	1.27±0.70	0.75±0.08
10	PC _{SD}	1.23±0.67	0.69±0.08
9	PC _{SD}	1.00±0.59	0.57±0.06
8	PC _{SD}	0.91±0.61	0.53±0.06
7	PC _{SD}	0.88±0.63	0.52±0.06
6	PC _{SD}	0.76±0.59	0.53±0.06
5	PC _{SD}	0.79±0.59	0.54±0.07
4	PC _{SD}	0.80±0.62	0.54±0.08
Max(ΔSIF value) = 0.51 mW m ⁻² sr ⁻¹ nm ⁻¹			

Latitude bias correction

This post hoc adjustment accounts for observed biases in SIF, which are most likely caused by varying slit functions across
95 latitude. To correct for this bias, regression coefficients between reflectance (at 744 nm) and SIF are obtained from collected
data over ocean regions, up to 14 days back within each 1° latitude band (van Schaik et al., 2020; Anema et al., 2025). In
SIFTER v3, two changes were introduced compared to v2: the cloud fraction filter (<0.4) was removed, allowing data to be
used, and the correction was based on subsets of both the Pacific and Atlantic Oceans (Anema et al., 2025).

For the sensitivity test, we applied the latitude bias correction using both the SIFTER v2 and SIFTER v3 settings. The
100 averaged SIF over the Congo Basin differed by 0.05 mW m⁻² sr⁻¹ nm⁻¹ between the two tests (Table S10).

Table S10. Summary of sensitivity tests on the latitude bias correction, comparing settings as done in SIFTER v3 and SIFTER v2. Results are shown over 624 pixels in the Congo Basin (13° S–6° N, 14–31° W) on 14 January 2017, retrieved from GOME-2B observations. Pixels of both tests were co-sampled and had to meet the requirements of autocorrelation < 0.2 and cloud fraction < 0.3. All other retrieval settings were kept at the default setting for GOME-2B.

Tests	SIF value	SIF uncertainty
Lat. bias correction settings	[mW m ⁻² sr ⁻¹ nm ⁻¹]	[mW m ⁻² sr ⁻¹ nm ⁻¹]
SIFTER v3	0.88±0.69	0.70±0.08
SIFTER v2	0.93±0.69	0.70±0.08
Max(Δ SIF value) = 0.05 mW m ⁻² sr ⁻¹ nm ⁻¹		

Slit function interpolation

One of the algorithm improvements in SIFTER v3 was the enhanced interpolation of the slit function. The slit functions are used in the range of around 612 nm to 770 nm. In SIFTER v2, only slit functions from fully sampled detector pixels ($n=10$) were used, whereas in SIFTER v3, all slit functions from the key data are used ($n=765$), including both fully and not fully sampled pixels, interpolated by EUMETSAT (Anema et al., 2025). These slit functions are convolved with the solar irradiance reference spectrum of Chance and Kurucz (2010), which is then used both for high sampling interpolation of the measured solar irradiance spectrum to the radiance wavelength grid of the radiance and for modeling the reflectance in the SIF retrieval ($\bar{E}_0$ in Eq. 6 of Anema et al. (2025)).

For our sensitivity study, we use the GOME-2A results reported in Anema et al. (2025) for the Congo Basin on 5 January 2008. Average SIF decreased from 1.25 to 1.07 mW m⁻² sr⁻¹ nm⁻¹ when using the slit functions as in v2 and v3, respectively, a difference of 0.18 mW m⁻² sr⁻¹ nm⁻¹ (Table S11).

Table S11. Summary of sensitivity tests on the slit function interpolation, comparing the used slit functions as done in SIFTER v2 and v3. Results shown over 633 pixels in the Congo Basin (13° S–6° N, 14–31° W) on 5 January 2008, retrieved from GOME-2A observations. Pixels of both tests were co-sampled and had to meet the requirements of autocorrelation < 0.2 and cloud fraction < 0.3. All other retrieval settings were kept at the default setting for GOME-2B. The zero-level offset correction is not applied to these results. The results from this sensitivity test were run by Anema et al. (2025).

Tests	SIF value	SIF uncertainty
Slit function interpolation	[mW m ⁻² sr ⁻¹ nm ⁻¹]	[mW m ⁻² sr ⁻¹ nm ⁻¹]
SIFTER v3	1.07±0.62	0.54±0.07
SIFTER v2	1.25±0.63	0.57±0.08
Max(Δ SIF value) = 0.18 mW m ⁻² sr ⁻¹ nm ⁻¹		

Error propagation

Across all sensitivity tests, the PC settings contributed to the largest systematic difference of 0.51 mW m⁻² sr⁻¹ nm⁻¹, followed by the slit function interpolation of 0.18 mW m⁻² sr⁻¹ nm⁻¹, while degradation and latitude bias corrections contributed less than 0.1 mW m⁻² sr⁻¹ nm⁻¹. The PCs used to model the atmospheric transmission represent the dominant source of systematic uncertainty in the SIFTER retrieval. Assuming the contributions from the four retrieval steps are independent, we add them in quadrature:

$$\sigma = \sqrt{\sigma_{\text{degr. corr.}}^2 + \sigma_{\text{PCs}}^2 + \sigma_{\text{lat. bias corr.}}^2 + \sigma_{\text{slit function}}^2} \quad (\text{S3})$$

This gives a combined systematic uncertainty of 0.55 mW m⁻² sr⁻¹ nm⁻¹.

These results are illustrative, but provide a valuable indication of the order of magnitude of systematic errors in the SIF retrieval. In addition to quantifying systematic errors, properly accounting for spatiotemporal error correlations is crucial

for accurate uncertainty quantification (Glissenaar et al., 2025; Woolliams et al., 2018). This is especially of importance in harmonisation efforts, as errors within products retrieved from similar sensors and algorithms are likely correlated (Giering et al., 2019). Robust error quantification prevents misinterpretation of instrumental artefacts or intersensor divergence as real trends; it could also help unravel the origin of noted intersensor offsets in certain regions.

References

- Anema, J. C. S., Boersma, K. F., Tilstra, L. G., Tuinder, O. N. E., and Verstraeten, W. W.: Improved Consistency in Solar-Induced Fluorescence Retrievals from GOME-2A with the SIFTER v3 Algorithm, *Atmospheric Measurement Techniques*, 18, 1961–1979, <https://doi.org/10.5194/amt-18-1961-2025>, 2025.
- 130 Chance, K. and Kurucz, R. L.: An Improved High-Resolution Solar Reference Spectrum for Earth’s Atmosphere Measurements in the Ultraviolet, Visible, and near Infrared, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 111, 1289–1295, <https://doi.org/10.1016/j.jqsrt.2010.01.036>, 2010.
- Giering, R., Quast, R., Mittaz, J. P. D., Hunt, S. E., Harris, P. M., Woolliams, E. R., and Merchant, C. J.: A Novel Framework to Harmonise Satellite Data Series for Climate Applications, *Remote Sensing*, 11, 1002, <https://doi.org/10.3390/rs11091002>, 2019.
- 135 Glissenaar, I., Boersma, K. F., Angloul, I., Rijdsdijk, P., Verhoelst, T., Compernelle, S., Pinardi, G., Lambert, J.-C., Van Roozendaal, M., and Eskes, H.: TROPOMI Level 3 Tropospheric NO₂ Dataset with Advanced Uncertainty Analysis from the ESA CCI+ ECV Precursor Project, <https://doi.org/10.5194/essd-2024-616>, 2025.
- 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 Dataset from the Sentinel-5P TROPOMI Mission, <https://doi.org/10.5194/essd-2021-199>, 2021.
- 140 Köhler, P., Guanter, L., and Joiner, J.: A Linear Method for the Retrieval of Sun-Induced Chlorophyll Fluorescence from GOME-2 and SCIAMACHY Data, *Atmospheric Measurement Techniques*, 8, 2589–2608, <https://doi.org/10.5194/amt-8-2589-2015>, 2015.
- 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.
- 145 Woolliams, E. R., Mittaz, J. P., Merchant, C. J., Hunt, S. E., and Harris, P. M.: Applying Metrological Techniques to Satellite Fundamental Climate Data Records, *Journal of Physics: Conference Series*, 972, 012 003, <https://doi.org/10.1088/1742-6596/972/1/012003>, 2018.