https://doi.org/10.5194/essd-2025-94

Response to Reviewer 2 Comments:

The author provides a systematic work for the development of the Long-term Harmonized multi-satellite SIF (LHSIF) dataset by coordinating SIF satellite observations from GOME, SCIAMACHY, GOME-2, and OCO-2. The TCSIF dataset, also developed by the same author, provides strong support for this study. The overall technical process is complete and feasible, and some interesting results are obtained. This study fits well within the scope of ESSD journal. However, I have some comments about the development process of this dataset. Therefore, my recommendation is major revision.

We sincerely appreciate the reviewer's thoughtful feedback and recognition of our work. In response to the comments, we have implemented the following key improvements to both the dataset and manuscript:

1. **Enhanced harmonization method**
   We replaced the moment matching approach with a CDF-based harmonization strategy, where cumulative distribution functions were constructed separately for different vegetation-climate zone combinations. To preserve seasonal dynamics, the matching was performed independently for each calendar month.
2. **Extended temporal coverage**
   The dataset now includes updated records through 2024.
3. **Strengthened Validation**
   We have strengthened the validation section, incorporated more quantitative evaluations, and revised or removed subjective language to enhance the scientific clarity and objectivity of the manuscript.

Detailed responses to each comment are provided below. Reviewer comments are shown in black, authors' responses in blue, and the corresponding revisions to the manuscript in purple.

Specific concerns and suggestions are outlined as follows:

Major Comment 1： I have concerns regarding the innovation and validation of the dataset. The authors mention two sets of long-term harmonized SIF products (Wang and Wen) and briefly describe their processing steps. However, neither in the Introduction nor in the Results/Discussion sections do I clearly see what specific methodological innovations the authors have introduced or where their data demonstrate superior performance. For instance, while they argue that the CDF method has limitations, there appears to be insufficient validation and justification—such as a comprehensive comparison of overall errors across the three datasets or scenario-specific analyses. I recommend that the authors conduct additional analyses to provide more quantitative evidence supporting the advantages of their dataset.

Response: We sincerely thank the reviewer for this valuable comment. First, we would like to clarify that the main objective of this study is not to propose a new harmonization algorithm, but rather to develop a long-term SIF dataset based solely on satellite observations, with improved consistency and temporal coverage.

In the original submission, we explored moment-matching as an alternative to the CDF method. This consideration was motivated by the inherent limitations of pixel-level CDF normalization, as adopted in previous studies (e.g., Wen et al., 2020; Wang et al., 2022). Specifically, the limited number of overlapping observations at the pixel scale can make it more susceptible to temporal noise, potentially reducing the reliability of the matching.

To address this issue, the revised version adopts an ecological zone-based normalization approach. By grouping pixels into broader ecological regions (e.g., Köppen climate zones), we increased the sample size for each normalization unit, enabling a more reliable implementation of the CDF matching method. As a result, the revised version retains the CDF method. Although the method remains consistent with prior studies, this adjustment reduces the potential instability caused by limited overlapping data in short time series.

We updated the methodology section (Section 2.3) to detail this adjustment and added quantitative assessments in the results section to demonstrate the behavior of the harmonized dataset under this framework.

2.3 CDF matching method

The cross-sensor SIF normalization was implemented using a stratified CDF matching approach to account for environmental variability. Specifically, the stratification was done based on a combination of Köppen climate zones (Beck et al., 2023) and the MODIS land cover types product (MCD12C1; Friedl and Sulla-Menashe, 2022). The degradation-corrected GOME-2A dataset was used as the normalization reference for all other satellite-derived SIF datasets, based on their overlapping periods. The normalization of GOME data was based on the SCIAMACHY dataset, which had been previously normalized with GOME-2 data.

For each sensor pair, the cumulative distribution functions of both the reference and target datasets were calculated across their overlapping temporal coverage. A linear interpolation was used to match the quantiles of the target dataset with those of the reference. Separate CDF transfer functions were derived for each calendar month to account for phenological variations. For SCIAMACHY and GOME, due to the limited temporal overlap, the entire period (from January 2003 to June 2003) was used to construct the CDF function. The complete workflow is shown in Fig. 1.

We also refined the introduction to better situate our study within the existing research landscape, clarifying the methodological distinctions and motivations. The updated paragraph addressing these modifications is provided below:

1 Introduction

****** To address these challenges, **Wen et al. (2020) proposed a harmonization framework** that used the cumulative distribution function (CDF) to integrate SIF datasets from SCIAMACHY and GOME-2 during their overlapping period, resulting in a continuous record from 2002 to 2018.

While Wen's framework laid the foundation for cross-sensor harmonization, it did not explicitly address instrument degradation—a key factor that compromises the long-term consistency of single-sensor records. Such degradation, as observed in GOME-2, poses a significant challenge for long-term consistency and introduces uncertainties in trend analyses (Parazoo et al., 2019).

******

To mitigate this limitation, Wang et al. (2022) attempted to create a temporally corrected long-term SIF product (LT_SIFc*) by correcting the degradation trends in gridded GOME, SCIAMACHY, and GOME-2 SIF products. **However, the method lacks a physically based correction of the actual sensor radiance degradation and instead applies adjustments on the SIF product, which may not accurately reflect the true instrumental change.** This is further complicated by the nonlinear characteristics inherent in the SIF retrieval methodology and subsequent processing procedures (e.g., zero-bias correction and quality filtering), which prevent a direct and linear propagation of sensor degradation into the final SIF retrievals. Recently, the temporally corrected GOME-2A SIF dataset (TCSIF) included a calibration of the radiance measurements of GOME-2A using a pseudo-invariant method (Zou et al., 2024). This correction effectively eliminates the influence of sensor degradation over time, providing a robust benchmark for generating long-term harmonized SIF products.

So far, the cross-sensor consistency of existing long-term SIF records remains to be further evaluated. In this study, we employed the TCSIF dataset as a physically calibrated benchmark to constrain the long-term consistency of GOME, SCIAMACHY, and OCO-2 SIF observations. By harmonizing these multi-sensor datasets against a radiometrically corrected reference, we generated a continuous and temporally

https://doi.org/10.5194/essd-2025-94
consistent SIF product spanning from 1995 to 2024, which is the longest multi-satellite harmonized SIF dataset to date.

To strengthen the justification as suggested, we have conducted an expanded intercomparison of four long-term SIF products (LHSIF, LT_SIFc*, SIF_005, and LCSIF) at both global and regional scales. The results are presented in the new Figure 10 (Section 3.3):

3.3 Validation and comparative analysis

******

The interannual trends of several long-term SIF products—including LHSIF, LT_SIFc*, SIF_005, and LCSIF—were compared. The annual maximum of global monthly SIF was used for comparison. Fig. 10 presents the results for the global scale as well as for several representative climate zones and land cover types.

Among the four SIF products, all except SIF_005 exhibit increasing global trends. LHSIF shows the strongest upward trend at 0.31% yr^{-1}, while LCSIF presents the most stable interannual variation, with a trend standard deviation of only 0.01% yr^{-1}. LHSIF and LCSIF display statistically significant increases on the global scale, whereas the trends for LT_SIFc* and SIF_005 are not statistically significant. The divergent trend between SIF_005 and the other SIF products is further demonstrated on regional scales. For example, in continental cropland regions (Fig. 10d) and temperate deciduous broadleaf forest (DBF) biomes, LHSIF, LT_SIFc*, and LCSIF generally exhibit consistent positive trends, whereas SIF_005 shows a declining trend.

In most cases shown in Fig. 10, LHSIF, LT_SIFc*, and LCSIF exhibit consistent trends. An exception occurs in arid regions, where LCSIF shows an increasing trend while both LHSIF and LT_SIFc* exhibit decreasing trends (Fig. 10j,k). This divergence may be attributed to the machine learning–based nature of LCSIF, which relies heavily on predictor variables and may not fully capture the actual SIF dynamics under stress conditions. In contrast, the observational basis of LHSIF enables it to more directly reflect regional responses to environmental variability.

https://doi.org/10.5194/essd-2025-94

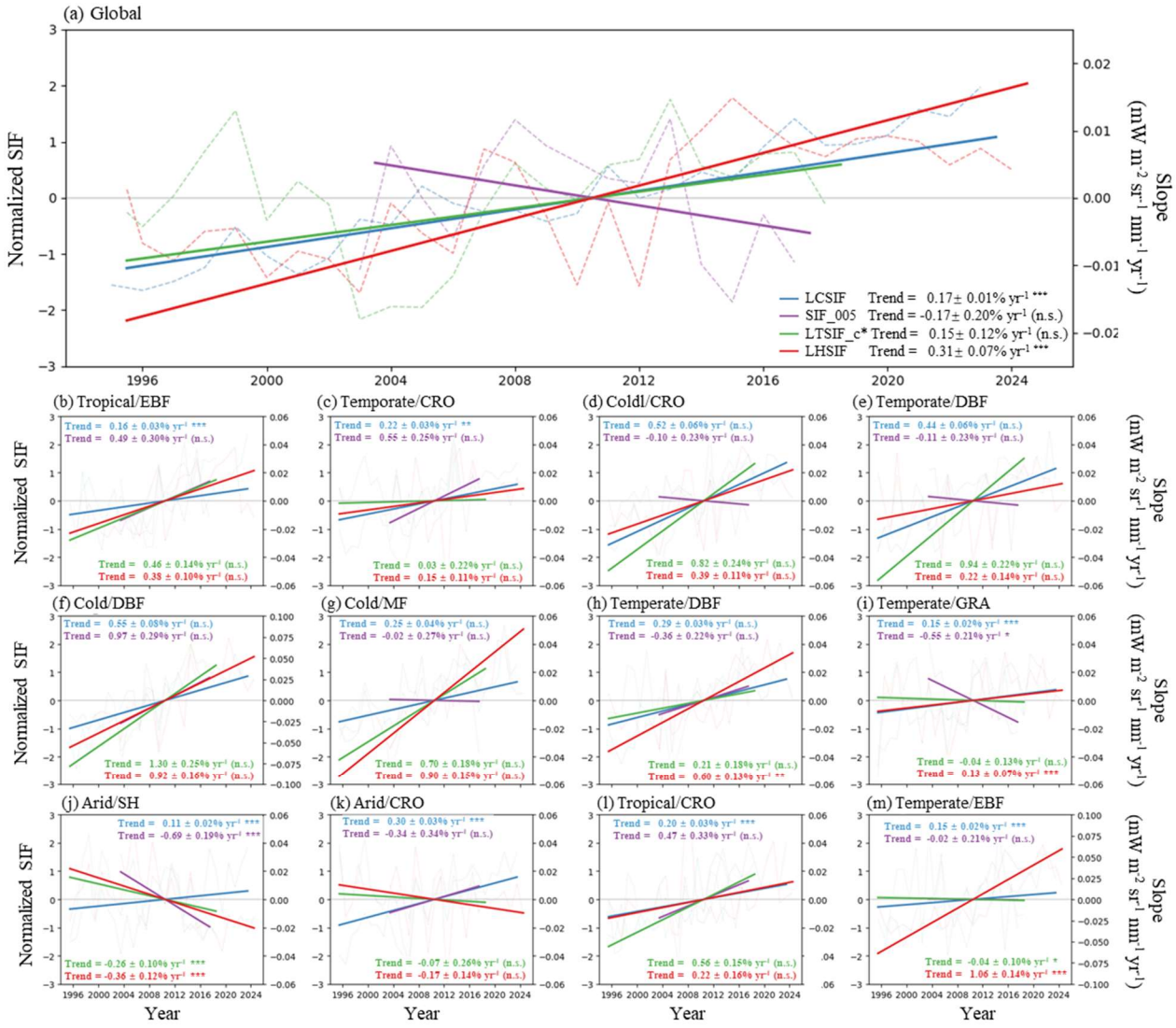


Figure 10. Comparison of interannual variations in long-term SIF products. LHSIF (red), LT_SIFc* (green), SIF_005 (purple), and LCSIF (blue) are compared for (a) the global scale and (b–m) various climatic and vegetation regions. All datasets were normalized using the z-score method. Dashed lines represent yearly maximum values, and solid lines indicate linear trends. To aid visual comparison, trend lines were anchored at the origin (2010, 0). The statistical significance of the trends is indicated as follows: n.s. for not significant ($p > 0.05$), * for significant ($p < 0.05$), and *** for highly significant ($p < 0.001$).

We have added the following paragraph to the revised discussion section to summarize the improvements of our dataset:

4.1 Improvements in cross-sensor harmonization

In this study, we applied a CDF normalization method to harmonize cross-sensor SIF measurements. While the general concept and algorithm are similar to previous studies (Wen et al., 2020; Wang et al., 2022), our data processing framework incorporates several key improvements.

Firstly, the harmonization process is anchored on a temporally corrected GOME-2A SIF (TCSIF) dataset (Zou et al., 2024), which accounts for long-term sensor degradation. The TCSIF product incorporates radiometric correction of GOME-2A sensor degradation using a pseudo-invariant method, and underwent a two-step validation at both radiance and SIF levels. As shown in Zou et al. (2024), the interannual variability of TCSIF shows strong consistency with GPP, providing a more robust reference for long-term harmonization. In contrast, the original GOME-2 SIF used in SIF_005 (Wen et al., 2020) exhibits substantial interannual

fluctuation and a declining trend during 2003–2017, likely due to residual degradation effects (see Fig. 9c and Fig. 10).

Secondly, our harmonization strategy uses GOME-2A as the reference sensor. Its extended data record (2007–2021) provides over five years of overlap with both SCIAMACHY and OCO-2, allowing us to perform a single-step normalization for each. This approach helps reduce the uncertainty propagation associated with multi-step corrections. In contrast, the LT_SIFc* product uses GOME as the benchmark, relying on only a six-month overlap with SCIAMACHY and then sequentially calibrating SCIAMACHY and GOME-2A, which may accumulate uncertainties.

To quantify the impact of overlap duration on harmonization uncertainty, we performed normalization experiments using 6-, 12-, and 24-month overlap periods between GOME-2 and OCO-2. Each experiment was repeated 10 times to assess the variability of the resulting harmonized time series.

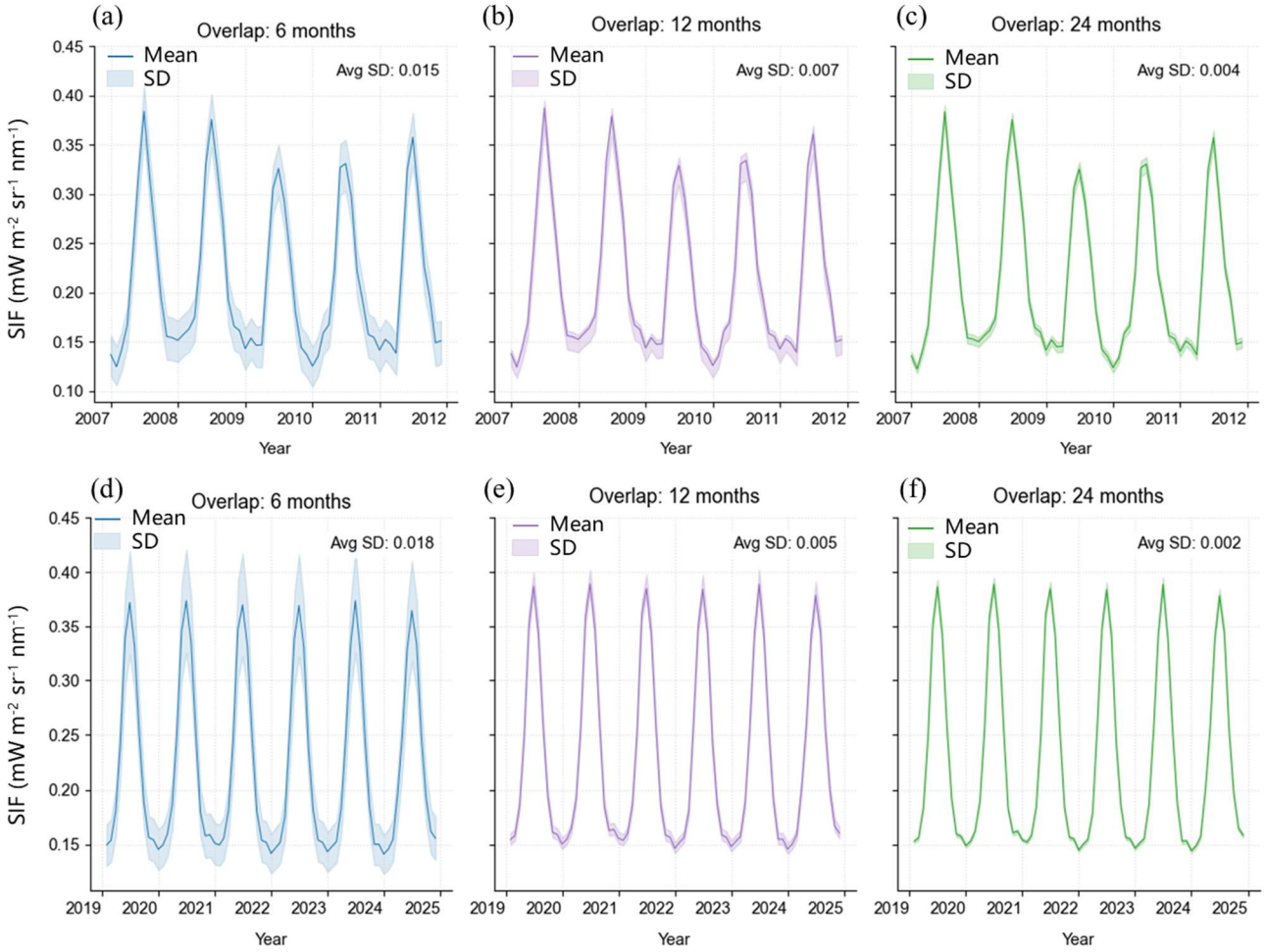


Figure 13. Normalized SIF time series using different overlap durations: 6 months (first column), 12 months (second column), and 24 months (third column) from SCIAMACHY (top row) and OCO-2 (bottom row). The shaded areas represent the standard deviation (SD) across multiple experiments, with units of mW m^{-2} sr^{-1} nm^{-1}.

The results show that a six-month overlap leads to substantially higher variability (7.7% for SCIAMACHY and 8.0% for OCO-2) compared to longer overlaps. As the overlap period increased from 6 to 12 months, the standard deviations of the normalized SIF series decreased from 0.015 to 0.007 mW m^{-2} sr^{-1} nm^{-1} (SCIAMACHY) and from 0.018 to 0.005 mW m^{-2} sr^{-1} nm^{-1} (OCO-2), representing a reduction of over 53.3%. These results confirm that short overlap periods increase normalization uncertainty and highlight the robustness of our chosen strategy, which avoids using GOME directly as the baseline. Nevertheless, users should be aware of the potential uncertainties in the early portion (from 1995 to 2003) of the LHSIF record.

Furthermore, in contrast to the grid-specific matching methods adopted in previous studies, we implemented a region-based and month-specific normalization strategy. The region-based approach allows for a larger sample size within each region, potentially enabling a more robust estimation of the CDF, while the month-specific treatment helps account for seasonal variations in the CDF. The improved stability of interannual trends in the LHSIF product, compared to LT_SIFc* and SIF_005 at both global and regional scales (Fig. 10), may benefit from this normalization strategy.

Once again, we greatly appreciate the reviewer's helpful suggestions, which have led to substantial improvements in both the presentation and validation of our dataset.

Major Comment 2： Several aspects of the validation remain unclear

(1) Fig. 3: Why was only a single year (1998) selected for comparison to evaluate the data performance? This limited temporal scope may not sufficiently represent the overall dataset characteristics.

Response: Thanks for this comment. Although Fig. 3 presents results for a single year (1998), validation at other periods is also provided in the manuscript. For instance, Fig. 2 displays downscaling results for July 1996, and Fig. 4 shows the latitudinal distribution of validation errors across the full time series from 1995 to 2024. In the supplementary materials, Figs. S3–S5 present comparable results to Fig. 3 for the other three sensors in different years.

To further address the concern of insufficient data verification, we have included additional validation results. We have added histograms for the downscaling residuals (Fig. S2) in the revised version to provide a more quantitative view of the errors. These histograms are based on multiple years and multiple sensors, with the mean and standard deviation of residuals annotated in each panel to enhance the credibility and robustness of the validation. The newly added content is as follows:

Main text：

3.1 Downscaled SIF dataset

The comparison of fine-resolution (0.05°) and coarse-resolution (1°) SIF datasets, derived from GOME, is illustrated in Fig. 2. The top two rows (panels a–f) illustrate the enhanced spatial variability achieved through the downscaling process, revealing finer vegetation patterns and distinct intensity gradients. The downscaled SIF datasets render subtle patterns in SIF more apparent compared to the original coarse-resolution data (panels g–k). Additionally, the downscaling method, which incorporates neighborhood-based pixel searching, effectively fills in data gaps in the original data while preserving spatial continuity. The residual, which was calculated as the difference between the downscaled SIF (which was re-aggregated to original 1° resolution) and the original SIF is shown in panels l–p. It can be observed that in major vegetated regions, the residuals are concentrated within the range of -0.5 to 0.5 mW m^{-2} sr^{-1} nm^{-1}. The histograms of the downscaling residuals across different years and sensors are shown in Fig. S2. Overall, the absolute values of the mean residuals are less than 0.008 mW m^{-2} sr^{-1} nm^{-1}, and the standard deviations are below 0.105 mW m^{-2} sr^{-1} nm^{-1}.

https://doi.org/10.5194/essd-2025-94

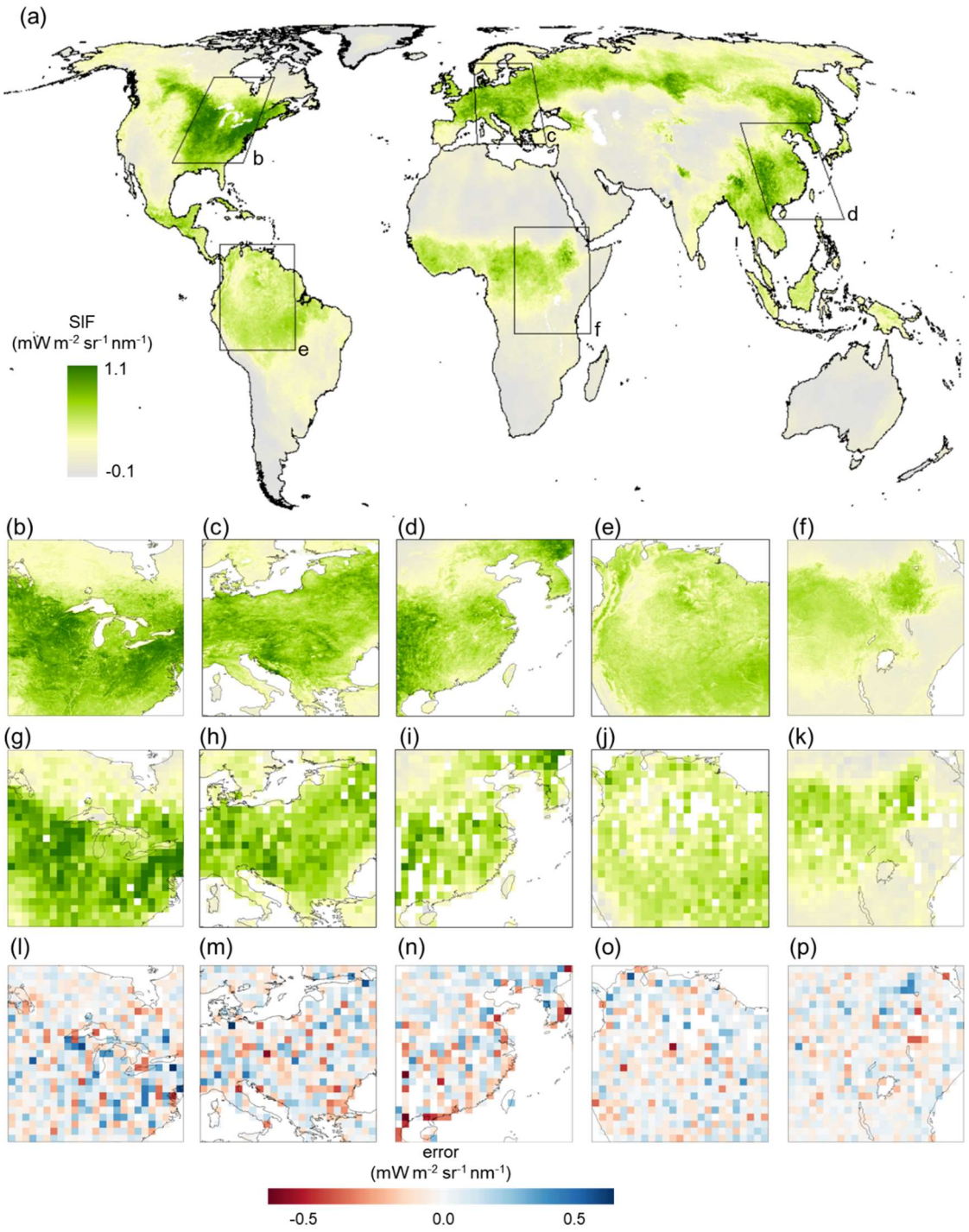


Figure 2. Spatially downscaled SIF maps (a-f) compared to the original SIF maps at a coarser resolution (g–k). SIF data from GOME observations in July 1996 are shown as an example. **The bottom row (l–p) shows the downscaling residuals, which were calculated as the difference between original SIF and downscaled SIF.** Panels b, g, and l depict North America; c, h, and m focus on Europe; d, i, and n depict East Asia (centered on China); e, j, and o represent the Amazon Basin; and f, k, and p show Sub-Saharan Africa.

Supplementary Materials:

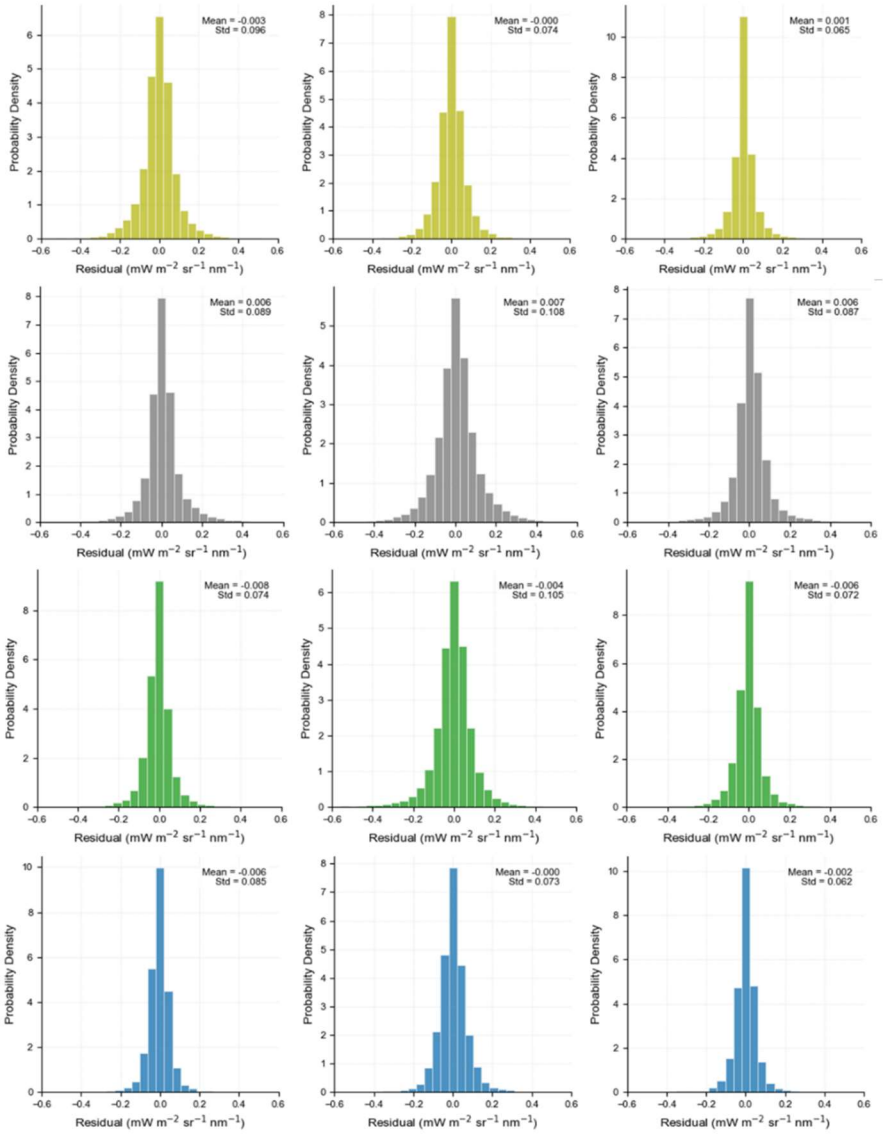


Figure S2. The Histograms of the downscaling residuals for GOME (1996, first row), SCIAMACHY (2008, second row), GOME-2 (2014, third row), and OCO-2 (2018, fourth row). The left, middle, and right columns correspond to results for January, July, and December, respectively.

These additions are intended to better illustrate the temporal stability and generalizability of the dataset across different time windows.

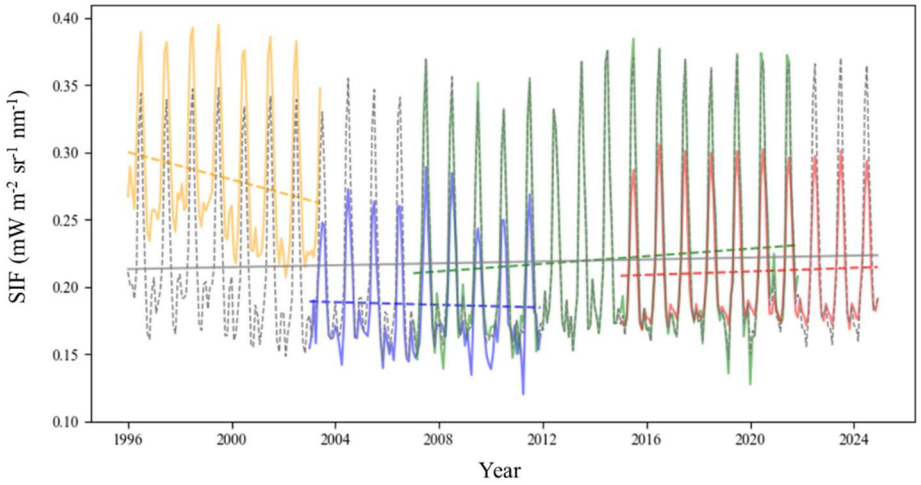
(2) Long-term trend validation: The analysis appears to rely solely on the global average SIF shown in Fig. 5. Furthermore, the description in Lines 275-280 is largely qualitative. More quantitative metrics (e.g., statistical significance tests or error metrics) would strengthen the validation.

Response: Thanks for this comment. In response, we have enriched Fig. 5 with additional quantitative information, including the mean and standard deviation of the global average SIF time series derived from each sensor as well as the harmonized product. The description in Lines 275–280 has also been revised to reflect these quantitative analyses. The updated figure and revised paragraph are provided below for your reference:

3.2 Temporal harmonization

https://doi.org/10.5194/essd-2025-94

The time series of the original SIF datasets from individual satellites and the resulting long-term harmonized SIF dataset (1995–2024) are presented in Fig. 5. Before normalization, substantial inter-sensor discrepancies were observed: mean SIF values ranged from 0.19 mW m^{-2} sr^{-1} nm^{-1} (SCIAMACHY) to 0.28 mW m^{-2} sr^{-1} nm^{-1} (GOME), while interannual trends varied from -0.76% yr^{-1} (GOME) to 0.54% yr^{-1} (GOME-2). Among the original sensor datasets, only the GOME-2 dataset showed a statistically significant trend ($p < 0.05$), whereas other sensors exhibited non-significant variations ($p \geq 0.05$). In contrast, the harmonized LHSIF dataset demonstrated a significant positive trend ($p < 0.001$, Trend = 0.31% yr^{-1}).



Sensor	mean	std	Trend (yr^{-1} %)
GOME	0.28	0.05	-0.76 (n.s)
SCIAMACHY	0.19	0.04	-0.01 (n.s)
GOME-2	0.22	0.07	0.54 (*)
OCO-2	0.21	0.04	0.13 (n.s.)
LHSIF	0.22	0.06	0.31 (***)

Figure 5. Global-averaged SIF time series derived from GOME (yellow), SCIAMACHY (blue), GOME-2 (green), and OCO-2 (red), along with the long-term harmonized SIF time series (LHSIF, gray dotted line), which aligns the satellite datasets based on overlapping periods. The table on the right summarizes the statistical characteristics of each sensor, including the mean, standard deviation (std), and the annual trend (Trend) averaged over the respective time spans. The statistical significance of the trends is indicated as follows: n.s. for not significant ($p > 0.05$), * for significant ($p < 0.05$), and *** for highly significant ($p < 0.001$).

(3) Fig. 8: What is the rationale for using annual maximum values rather than annual means? Maximum values are generally more susceptible to noise interference. Additionally, how was the claimed 'reduction in uncertainty' quantified?

Response: We appreciate the reviewer's insightful comment. We acknowledge the general concern that maximum values may be more sensitive to noise. We would like to clarify that in our study, the annual maximum SIF is defined as the composite value of the peak month (typically occurring during the June–August growing season) within each year, rather than an instantaneous measurement. This monthly compositing approach minimizes the influence of transient noise or outliers, thereby addressing the reviewer's concern.

We chose to use annual maximum SIF values over annual means for several reasons:

(1) The annual peak SIF better reflects vegetation productivity and ecosystem functioning, as it coincides with the peak growing season when photosynthetic activity is most active and data quality is highest.

(2) Satellite SIF products often suffer from low signal-to-noise ratios, especially in sparsely vegetated regions or during dormant seasons. Using mean values over the full year may be disproportionately affected by these low-quality periods.

Based on these considerations, the use of annual maximum SIF was retained in the revised manuscript.

Regarding the "reduction in uncertainty" mentioned in the original manuscript, we refer to the decrease in the standard deviation of the fitted trend after normalization. This has been clarified in the revised manuscript as below:

The annual maximums of the global-averaged SIF were used to investigate the fluctuation of the worldwide vegetation from 1995 to 2024. Significant interannual fluctuations were found for the SIF time series without normalization, with an overall decline (blue line in Fig. 7a). **The normalized SIF time series reveals a growth rate of 0.31% yr^{-1}. After normalization, the standard deviation of the fitted slope decreases from 0.25% to 0.07%, indicating a reduction in uncertainty.** The boxplot in Fig. 7b further shows a narrower range of SIF values after temporal normalization, suggesting a more concentrated data distribution and improved comparability across sensors.

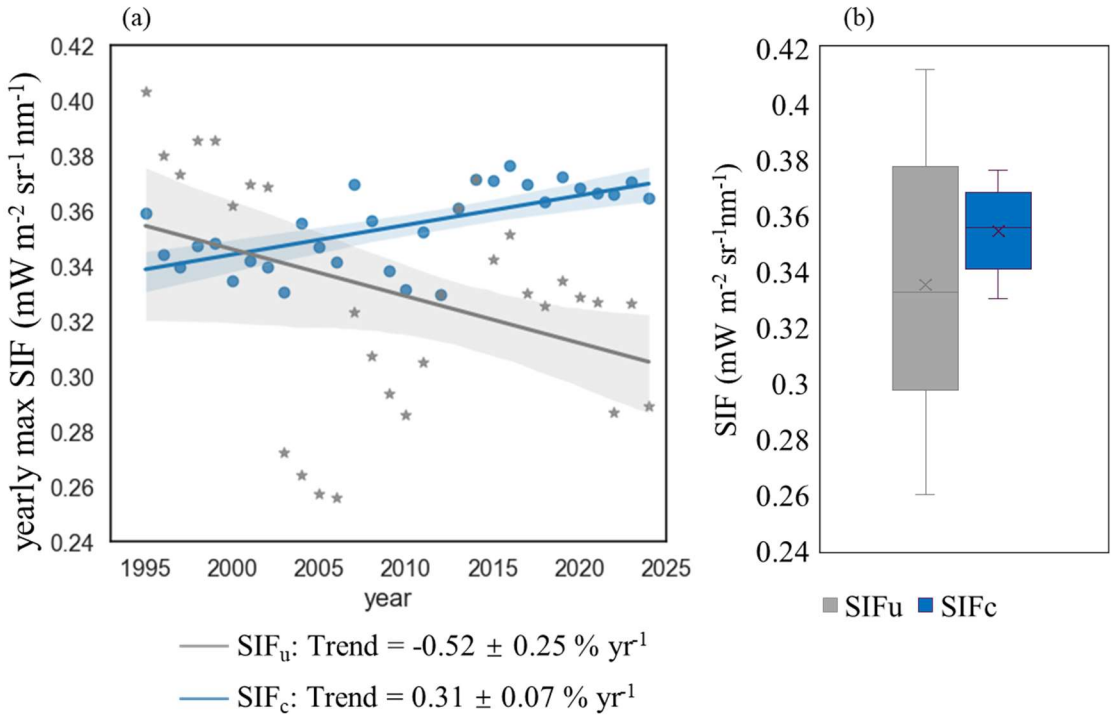


Figure 7. (a) Trend and (b) box plot of the yearly maximum global-averaged SIF of the combined time series before (SIF$_u$) and after (SIF$_c$) normalization during 1995–2024

(4) Temporal inconsistency: Why does Fig. 8 cover 1996–2023 while Fig. 9 starts from 1995? This discrepancy in time ranges should be explicitly justified.

Response: Thanks for pointing out the inconsistency. The difference in time ranges arises from the use of different statistical metrics in Figs. 8 and 9. Specifically, Fig. 8 shows annual averages, while Fig. 9 depicts the trend of annual peak-month values (usually observed in July). Since satellite observations in 1995 only began in July, that year was excluded from Fig. 8 to avoid bias due to incomplete coverage. In contrast, the available data in 1995 were sufficient to determine the peak for that year, allowing its inclusion in the analysis in Fig. 9.

To clarify this point, we have added the following explanation to the caption of Fig. 8 in the revised manuscript:

https://doi.org/10.5194/essd-2025-94

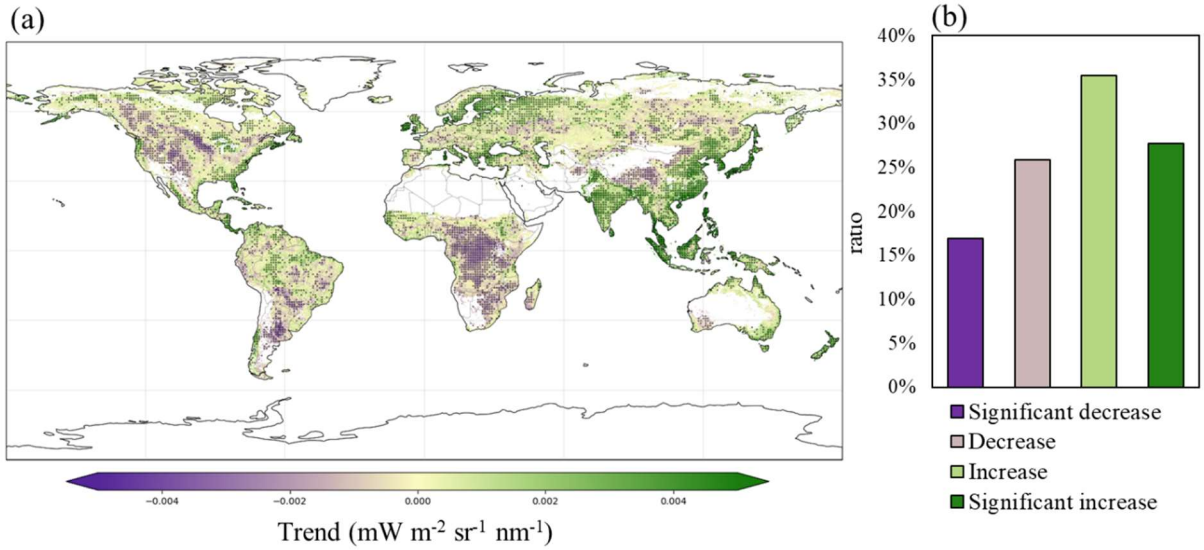


Figure 8. (a) Map of trends in LHSIF for 1996–2024. (b) Percentage of areas in global vegetation covered by four different trend types (significant decrease: negative correlation and $p < 0.05$; decrease: negative correlation and $p \geq 0.05$; increase: positive correlation and $p \geq 0.05$; and significant increase: positive correlation and $p < 0.05$). The black dots in (a) represent statistically significant trends ($p < 0.05$). **The statistics begin in 1996 due to incomplete data coverage in 1995, which only includes the second half of the year (July–December).**

To clarify, these observations are not meant to imply that the authors' approaches are inherently flawed. However, clearer explanations for these methodological choices are necessary to ensure robust interpretation of the results.

Response: We thank the reviewer for the clarification. We agree that clearer explanations can strengthen the interpretation of the results. Accordingly, we have revised the manuscript to provide more quantitative support for the dataset developed in this study.

Major Comment 3： In Section 4.1, the authors show that the overlap between GOME and SCIAMACHY SIF records is limited to only six months. However, the manuscript lacks a detailed explanation for addressing this issue within the current study. In fact, the challenge of short-term overlap is also present in this study. The limited temporal overlap constrains the representativeness of the GOME SIF data and may introduce uncertainty and potential biases in the mean-standard deviation-based matching approach. This issue is evident in Figures 7 and 8. For instance, in Figure 7, noticeable discrepancies between the corrected and uncorrected datasets are observed prior to 2003. It is worth noting that these early-period differences reflect the extent to which the correction improved the quality of this dataset. Therefore, I suggest that the authors further clarify the applicability and effectiveness of this matching method in mitigating biases arising from short-term overlaps, and possibly, supplement the analysis with a quantitative assessment of associated uncertainties.

Response: Thank you for this valuable comment. We agree that the limited temporal overlap between sensors can introduce potential uncertainties in the harmonization process. In the revised manuscript, we have addressed this issue by explicitly testing how the length of the overlap period affects the performance of the CDF-based matching method. Specifically, we conducted a set of quantitative experiments to evaluate the uncertainty associated with short overlap durations. The relevant additions can be found in Section 4.1. The results underline the potential uncertainty associated with the early part of the SIF time series (1995–2003). We have added a cautionary note to alert readers to the greater uncertainty in this early period due to the short overlap between GOME and SCIAMACHY, summarized as follows:

4.1 Improvements in cross-sensor harmonization

******

Secondly, our harmonization strategy uses GOME-2A as the reference sensor. Its extended data record (2007–2021) provides over five years of overlap with both SCIAMACHY and OCO-2, allowing us to perform a single-step normalization for each. This approach helps reduce the uncertainty propagation associated with multi-step corrections. In contrast, the LT_SIFc* product uses GOME as the benchmark, relying on only a six-month overlap with SCIAMACHY and then sequentially calibrating SCIAMACHY and GOME-2A, which may accumulate uncertainties.

To quantify the impact of overlap duration on harmonization uncertainty, we performed normalization experiments using 6-, 12-, and 24-month overlap periods between GOME-2 and SCIAMACHY/OCO-2 (Fig. 13). Each experiment was repeated 10 times to assess the variability of the resulting harmonized time series.

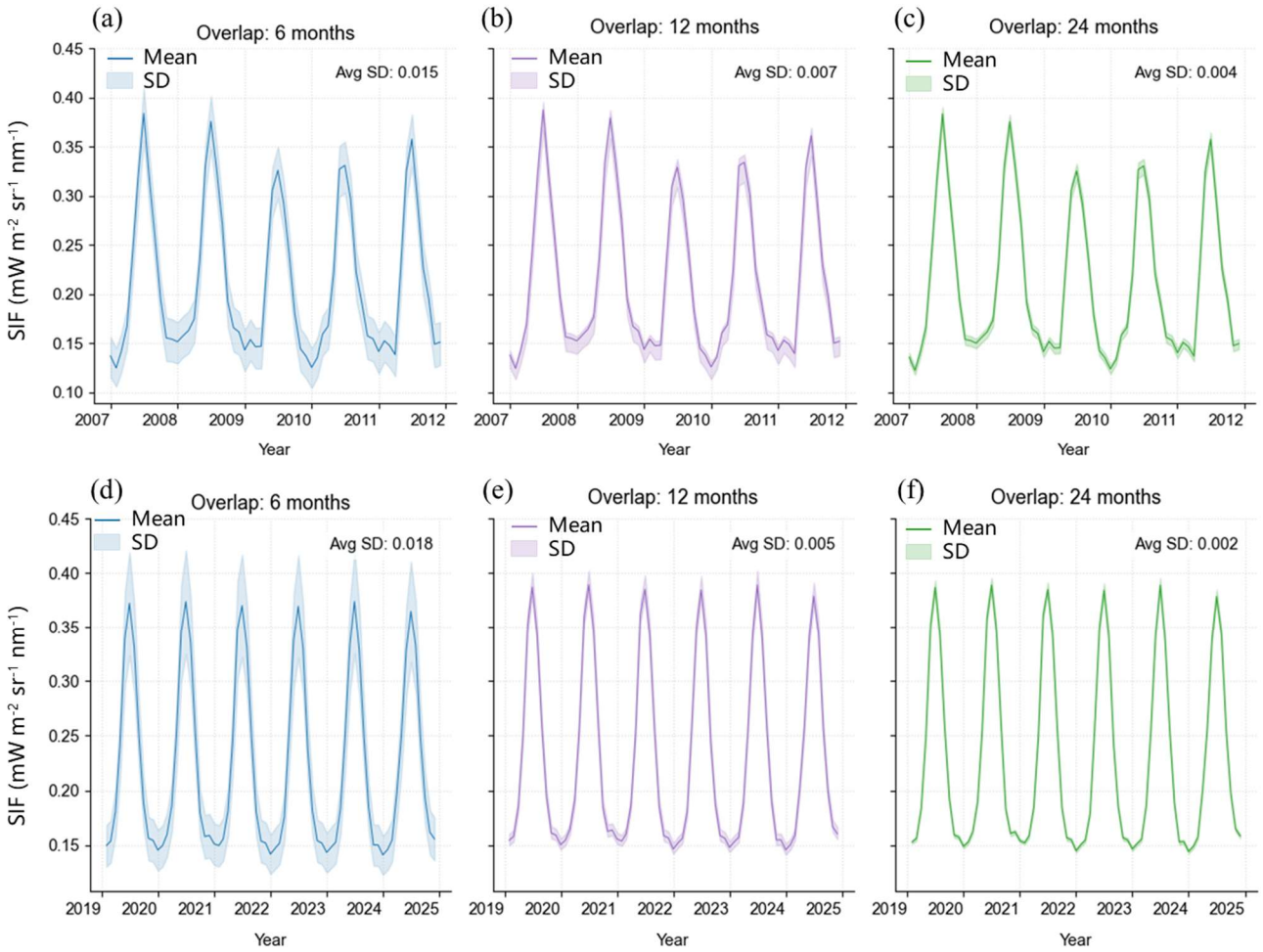


Figure 13. Normalized SIF time series using different overlap durations: 6 months (first column), 12 months (second column), and 24 months (third column) from SCIAMACHY (top row) and OCO-2 (bottom row). The shaded areas represent the standard deviation (SD) across multiple experiments, with units of mW m^{-2} sr^{-1} nm^{-1}.

The results show that a six-month overlap leads to a higher standard deviation in SIF time series compared to longer overlaps. As the overlap period was extended from 6 to 12 months, the standard deviations of the normalized SIF series decreased from 0.015 to 0.007 mW m^{-2} sr^{-1} nm^{-1} (SCIAMACHY) and from 0.018 to 0.005 mW m^{-2} sr^{-1} nm^{-1} (OCO-2), representing a reduction of over 53.3%. **These results confirm that short overlap periods increase normalization uncertainty and highlight the robustness of our chosen strategy,**

which avoids using GOME as the baseline. Nevertheless, users should be aware of the potential uncertainties in the early portion (from 1995 to 2003) of the LHSIF record.

Major Comment 4： The authors' descriptions in multiple sections of the manuscript are overly qualitative and subjective, lacking sufficient experimental support. It is recommended that the authors revise the relevant language to provide more detailed explanations or supplement with additional experiments.
Line 86; 268-269; 391; 394; 398;
Response: We appreciate the reviewer's suggestion. We have revised the manuscript to reduce subjective or qualitative expressions and incorporated additional quantitative analyses to support our conclusions.

First, we introduced quantitative indicators for time series trend analysis in the methods section, as detailed below:

3.4 Temporal Trend Analysis Metrics

To assess long-term trends in vegetation dynamics, we employed the Mann-Kendall (MK) test, a non-parametric method suitable for detecting monotonic trends in time series data, using the Python package pyMannKendall (Hussain and Mahmud, 2019). Trend estimation uncertainty was quantified by the standard deviation and 95% confidence intervals of the estimated temporal trend.

The detailed modifications are as follows:
(1) Line 86
Original:
"A precise, reliable, harmonized, and global high-resolution SIF dataset is not yet available for long-term vegetation monitoring."
Revision:
To avoid emphasizing subjective terms such as "precise," "reliable," and "harmonized" in the introduction, we removed this sentence. Instead, we revised the text as follows:

So far, the cross-sensor consistency of existing long-term SIF records remains to be further evaluated. In this study, we employed the TCSIF dataset as a physically calibrated benchmark to constrain the long-term consistency of GOME, SCIAMACHY, and OCO-2 SIF observations.

We also added more quantitative accuracy assessments and comparisons with other products in Section 3.3 of the revised manuscript to support the reliability of the LHSIF dataset (please refer to the response to Major comment 1).
(2) Line 268-269
Original:
"As shown by the temporally averaged residuals (Fig. 4 (b)), the absolute mean residuals for most locations are less than 0.05 mW m^{-2} sr^{-1} nm^{-1} for regions below 70° N. This result highlights the global applicability of the LUE-based downscaling method."
Revision:
We added quantitative descriptions and replaced the subjective statement with a more informative explanation:

The temporal and spatial distributions of the spatial downscaling residuals were analyzed (Fig. 4). **The monthly mean residuals across different latitudes and months were generally below 0.2 mW m^{-2} sr^{-1} nm^{-1} (Fig. 4 a). In addition, the regions with relatively larger residuals (e.g., > 0.1 mW m^{-2} sr^{-1} nm^{-1}) were mainly located in high-latitude areas.** As shown by the temporally averaged residuals (Fig. 4 b), for most areas below 70°N, the absolute mean residuals are less than 0.05 mW m^{-2} sr^{-1} nm^{-1} for regions below 70° N.

These results indicate that the downscaling method maintains high consistency with the original data across a broad range of temporal and spatial conditions.

(3) Line 391-394

Original:

"The approach proposed adopted in this study is robust compared to CDF matching. By modifying only the mean and standard deviation of the SIF data distribution, the method avoids overfitting and enhances stability. This less stringent approach ensures harmonization while preserving the inherent variability of the datasets. Therefore, the LHSIF dataset provides an unprecedented long-term harmonized SIF dataset that is theoretically reliable and demonstrates a reasonable temporal trend."

Revision:

We have revised the manuscript to remove the overstatement about the reliability of the method. Rather than emphasizing the superiority of our approach, we present additional analytical results to objectively demonstrate its advantages. In fact, the revised version adopts a CDF-based harmonization method similar to that used by Wang et al., and therefore, we have removed the original subjective claims regarding methodological robustness.

Instead, we have added a comparative analysis of the seasonal variations derived from LHSIF and LT_SIFc* products to highlight the improved consistency and continuity of our results. The relevant revisions are detailed below:

The interannual trends of several long-term SIF products—including LHSIF, LT_SIFc*, SIF_005, and LCSIF—were compared. The annual maximum of global monthly SIF was used for comparison. Fig. 10 presents the results for the global scale as well as for several representative climate zones and land cover types.

Among the four SIF products, all except SIF_005 show increasing global trends. LHSIF exhibits the strongest upward trend at 0.31% yr^{-1}, while LCSIF presents the most stable interannual variation, with a trend standard deviation of only 0.01% yr^{-1}. LHSIF and LCSIF display statistically significant increases on the global scale, whereas the trends for LT_SIFc* and SIF_005 are not statistically significant. The divergent trend between SIF_005 and the other SIF products is further demonstrated on regional scales. For example, in continental cropland regions (Fig. 10d) and temperate deciduous broadleaf forest (DBF) biomes, LHSIF, LT_SIFc*, and LCSIF generally exhibit consistent positive trends, whereas SIF_005 shows a declining trend.

In most cases shown in Fig. 10, LHSIF, LT_SIFc*, and LCSIF display consistent trends. An exception occurs in arid regions, where LCSIF shows an increasing trend while both LHSIF and LT_SIFc* exhibit decreasing trends (Fig. 10j,k). This divergence may be attributed to the machine learning–based nature of LCSIF, which relies heavily on predictor variables and may not fully capture the actual SIF dynamics under stress conditions. In contrast, the observational basis of LHSIF enables it to more directly reflect regional responses to environmental variability.

https://doi.org/10.5194/essd-2025-94

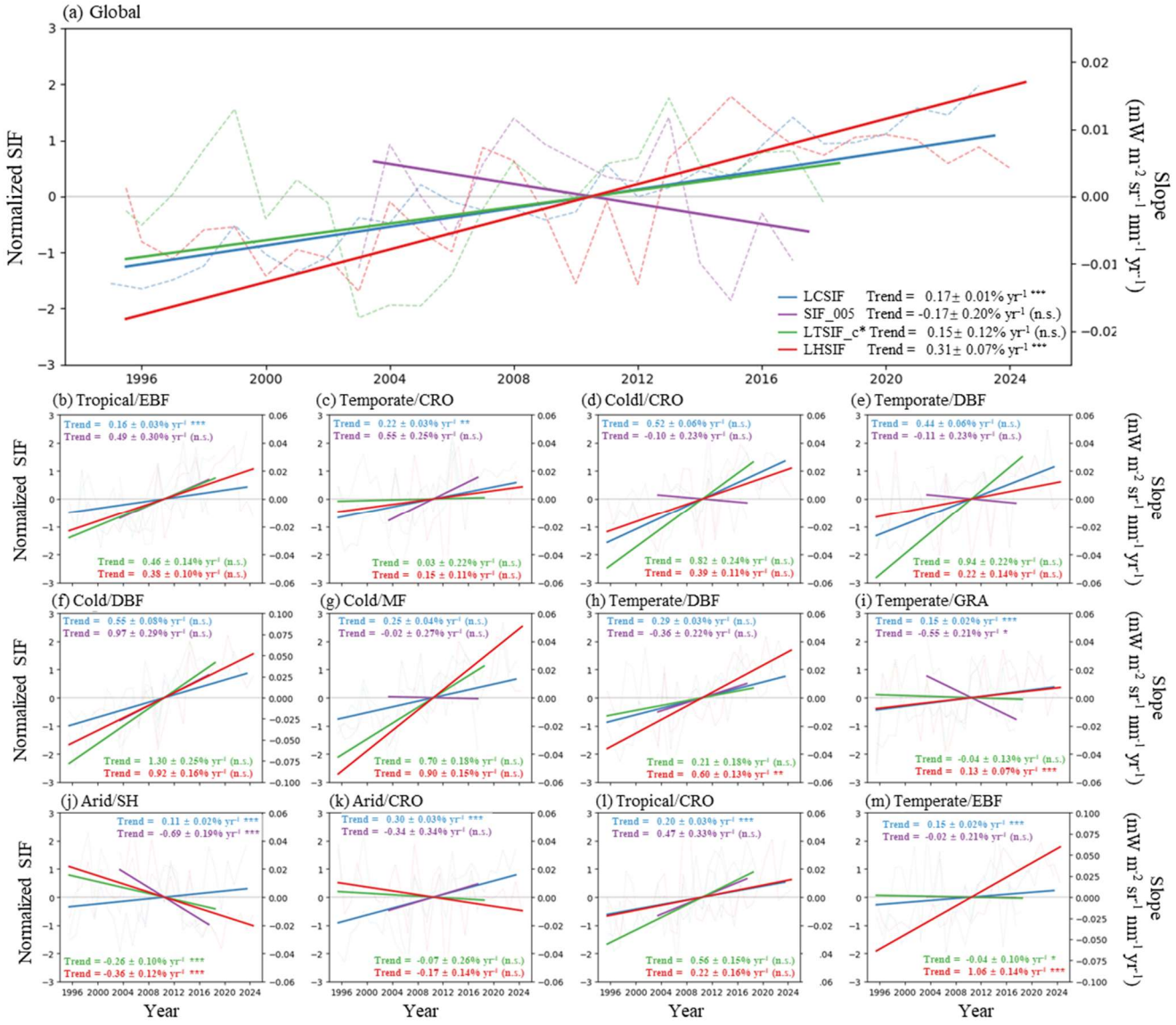


Figure 10. Comparison of interannual variations in long-term SIF products. LHSIF (red), LT_SIFc* (green), SIF_005 (purple), and LCSIF (blue) are compared for (a) the global scale and (b–m) various climatic and vegetation regions. All datasets were normalized using the z-score method. Dashed lines represent yearly maximum values, and solid lines indicate linear trends. To aid visual comparison, trend lines were anchored at the origin (2010, 0). The statistical significance of the trends is indicated as follows: n.s. for not significant ($p > 0.05$), * for significant ($p < 0.05$), and *** for highly significant ($p < 0.001$).

(4) Line 398

Original:

Despite its high performance, it is crucial to remember that this statistical methodology is not a rigorous calibration procedure.

Revision:

We added a more detailed explanation, as shown below:

Although the normalized series yields consistent seasonal and interannual trends across sensors, it is important to note that this normalization approach does not involve a physically based calibration (e.g., radiometric calibration based on pseudo-invariant targets). As a result, residual sensor-specific biases may still exist.

Major Comment 5: To more comprehensively evaluate the quality of the LHSIF dataset, it is recommended that the authors incorporate the LCSIF dataset (Fang et al., 2025) for comparative experiments in their study. Comparing these two datasets can reveal the potential strengths and limitations of LHSIF, particularly in terms of long-term trend analysis and responses across different ecosystems. Furthermore, the long-term record of LCSIF can serve as a reference to validate the consistency and accuracy of LHSIF data during the earlier years.

Response: Thanks for this valuable suggestion. In response, we have incorporated a comparative analysis between the LHSIF and LCSIF datasets, including assessments of long-term trends and ecosystem-specific responses. The corresponding results have been added to the revised manuscript as below:

The interannual trends of several long-term SIF products—including LHSIF, LT_SIFc*, SIF_005, and LCSIF—were compared. The annual maximum of global monthly SIF was used for comparison. Fig. 10 presents the results for the global scale as well as for several representative climate zones and land cover types.

Among the four SIF products, all except SIF_005 exhibit increasing global trends. LHSIF shows the strongest upward trend at 0.31% yr^{-1}, **while LCSIF presents the most stable interannual variation, with a trend standard deviation of only 0.01% yr^{-1}. LHSIF and LCSIF display statistically significant increases on the global scale, whereas the trends for LT_SIFc* and SIF_005 are not statistically significant.** The divergent trend between SIF_005 and the other SIF products is further demonstrated on regional scales. For example, in continental cropland regions (Fig. 10d) and temperate deciduous broadleaf forest (DBF) biomes, LHSIF, LT_SIFc*, and LCSIF generally exhibit consistent positive trends, whereas SIF_005 shows a declining trend.

In most cases shown in Fig. 10, LHSIF, LT_SIFc*, and LCSIF exhibit consistent trends. An exception occurs in arid regions, where LCSIF shows an increasing trend while both LHSIF and LT_SIFc* exhibit decreasing trends (Fig. 10j,k). This divergence may be attributed to the machine learning–based nature of LCSIF, which relies heavily on predictor variables and may not fully capture the actual SIF dynamics under stress conditions. In contrast, the observational basis of LHSIF enables it to more directly reflect regional responses to environmental variability.

https://doi.org/10.5194/essd-2025-94

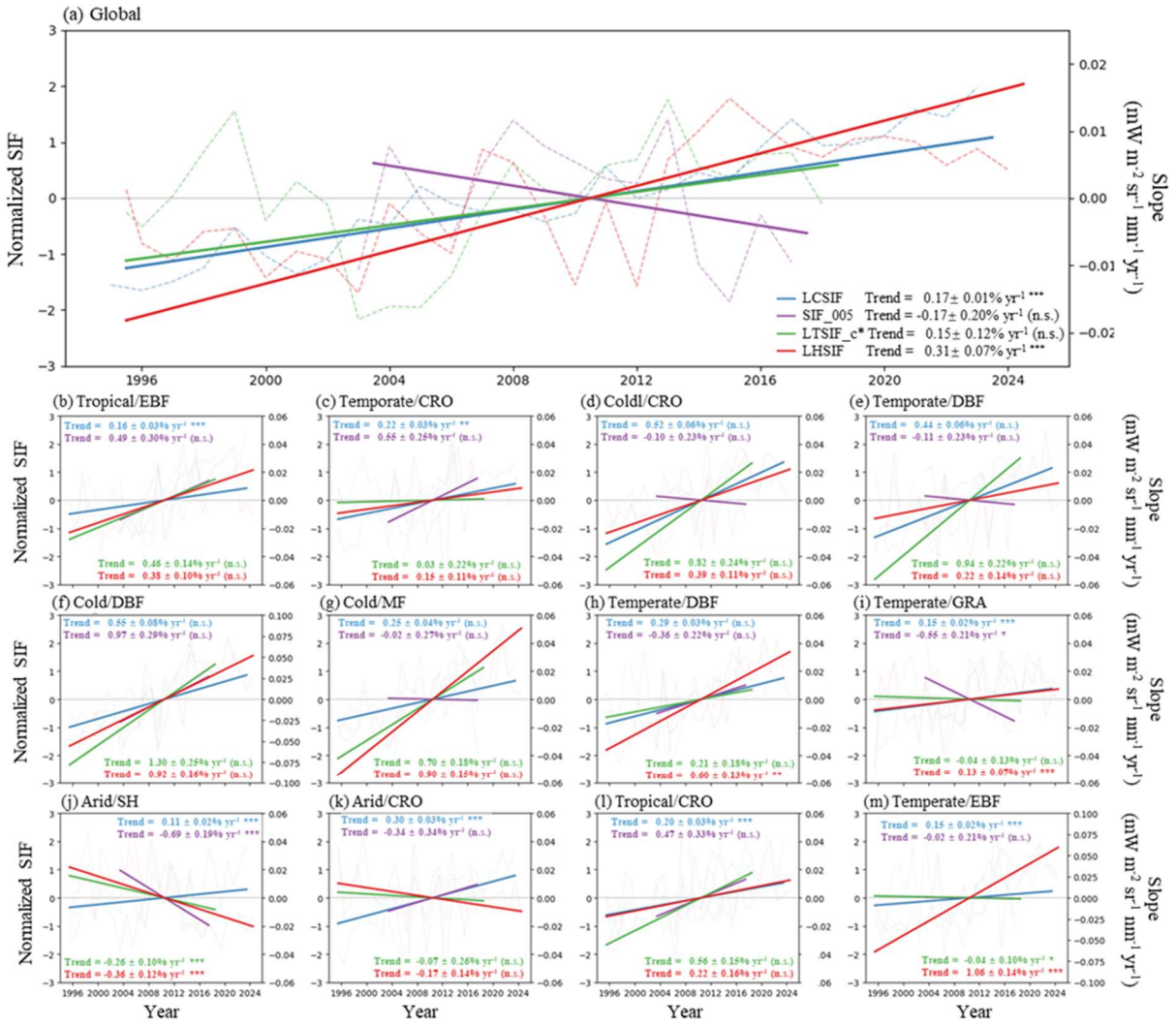


Figure 10. Comparison of interannual variations in long-term SIF products. LHSIF (red), LT_SIFc* (green), SIF_005 (purple), and LCSIF (blue) are compared for (a) the global scale and (b–m) various climatic and vegetation regions. All datasets were normalized using the z-score method. Dashed lines represent yearly maximum values, and solid lines indicate linear trends. To aid visual comparison, trend lines were anchored at the origin (2010, 0). The statistical significance of the trends is indicated as follows: n.s. for not significant ($p > 0.05$), * for significant ($p < 0.05$), and *** for highly significant ($p < 0.001$).

It is important to note that the LHSIF product is derived solely from observed satellite SIF measurements, aiming to preserve the original signal characteristics to the greatest extent possible. In contrast, machine learning-based SIF products, such as LCSIF, are generated through data-driven modeling approaches and represent simulations of SIF signals, rather than direct satellite observations. While machine learning-based reconstructions often yield smoother time series, this smoothness does not necessarily imply higher fidelity. In some cases, it may result from overfitting seasonal patterns, potentially suppressing high-frequency variations that carry meaningful biophysical signals(Ma et al., 2022; Chen et al., 2025).

Detailed Comment 1: Line 17: How did you define the reduction in overall error? Compared to what?

Response: Thanks for this comment. We have added the following supplementary explanation in the abstract.

The resulting harmonized dataset shows a 49% reduction in inter-sensor differences compared to the uncorrected data and exhibits a stable interannual increase of 0.31 ± 0.07% yr^{-1}.

A detailed explanation of the result can be found in Sec. 3.2 as follows:

In all tested scenarios, the normalization process substantially reduced the differences between the two sensors. Overall, the MSD decreased by more than 49% following normalization.

Detailed Comment 2: Line 27: garnered significant attention(s)
Response: Thanks, it has been corrected.

Detailed Comment 3: The second and third paragraphs can be merged to one paragraph after reducing some sentences about SIF-GPP relationship which is not the focus in this study.
Response: Thanks for this comment. These two paragraphs have been revised as below:

Following the publication of the initial global SIF map from the Greenhouse Gases Observing Satellite (GOSAT), interest in the SIF-GPP association greatly increased (Frankenberg et al., 2011; Guanter et al., 2012; Joiner et al., 2011). Subsequent satellite-based analyses have consistently revealed strong spatial and temporal correlations between SIF and GPP, showcasing remarkable alignment between SIF and GPP in terms of spatial distribution and seasonal variability (Anav et al., 2015; Li et al., 2018; Verma et al., 2017; Yang et al., 2015; Guanter et al., 2014; Zheng et al., 2024). However, these results are mostly based on coarse-resolution SIF datasets such as the Global Ozone Monitoring Experiment (GOME)-2, leading to potential spatial mismatch issues. Additionally, the SIF-GPP link varies by vegetation type, emphasizing the critical need for SIF datasets with higher spatial resolution and spatiotemporal consistency to better support ecosystem monitoring and interpretation.

Detailed Comment 4: Line 65: delete "itself"
Response: Thanks, it has been deleted.

Detailed Comment 5: Line 70-74: Could you provide a direct explanation of where Wang's research may have fallen short or failed to consider comprehensively?
Response: Thanks for this comment. We strengthened the limitation of Wang's research in the revised manuscript as below:

To mitigate this limitation, Wang et al. (2022) attempted to create a temporally corrected long-term SIF product (LT_SIFc*) by correcting the degradation trends in gridded GOME, SCIAMACHY, and GOME-2 SIF products. **However, the method lacks a physically based correction of the actual sensor radiance degradation and instead applies adjustments at the SIF product level, which may not accurately reflect the true instrumental change. This is further complicated by** the nonlinear characteristics inherent in the retrieval methodology and subsequent processing procedures (e.g., zero-bias correction and quality filtering), which prevent a direct and linear propagation of sensor degradation into the final SIF values. Recently, the temporally corrected GOME-2A SIF dataset (TCSIF) included a calibration of the radiance measurements of GOME-2A using a pseudo-invariant method (Zou et al., 2024). This correction effectively eliminates the influence of sensor degradation over time, providing a robust benchmark for generating long-term harmonized SIF products.

Detailed Comment 6: Line 84-85: Wen's research has significant overlap with yours, so I'm unclear why it was only mentioned so briefly and at such a late stage.
Response: Thanks for this comment. We acknowledge the importance of Wen's research and have revised the introduction as below:

Despite the availability of multiple satellite SIF products, most have a temporal coverage shorter than 10 years, and large discrepancies have been observed between different SIF products (Parazoo et al., 2019). These

temporal inconsistencies may stem from differences in retrieval algorithms, absolute radiometric calibration errors, instrumental artifacts, directional effects, and variations in satellite overpass times and footprint sizes (Zhang et al., 2018b; Bacour et al., 2019). **To address these challenges, Wen et al. (2020) proposed a harmonization framework that used cumulative distribution function (CDF) to integrate SIF datasets from SCIAMACHY and GOME-2 during their overlapping period, resulting in a continuous record from 2002 to 2018.**

While Wen's framework laid the foundation for cross-sensor harmonization, it primarily focused on aligning overlapping periods and did not explicitly address instrument degradation, a key factor that compromises the long-term consistency of single-sensor records. Such degradation, as observed in GOME-2, poses a significant challenge for long-term consistency and introduces uncertainties in trend analyses (Parazoo et al., 2019). For instance, Yang et al. (2018) reported diverging trends between EVI and SIF, attributing the latter's decline to reduced photosynthetic activity. However, Zhang et al. (2018a) argued that this conclusion was impacted by the deterioration of the GOME-2A instrument. Further research by Koren et al. (2018) showed that the decline in SIF persisted even after correcting for sensor degradation. The SIFTERv2 product (Van Schaik et al., 2020) employed in Koren's study was simply corrected using linear models; the reliability of SIFTERv2 decreased significantly after 2016, limiting its application for long-term trend analysis.

To mitigate this limitation, Wang et al. (2022) attempted to create a temporally corrected long-term SIF product (LT_SIFc*) by correcting the degradation trends in gridded GOME, SCIAMACHY, and GOME-2 SIF products. However, the nonlinear characteristics inherent in the retrieval methodology and subsequent processing procedures (e.g., zero-bias correction and quality filtering) presented a challenge. The bias in sensor observations was not linearly transferred to the SIF product. This approach risks introducing inaccuracies, because the trend corrected in the SIF product may not correspond to the true sensor degradation. Recently, the temporally corrected GOME-2A SIF dataset (TCSIF) included a calibration of the radiance measurements of GOME-2A using a pseudo-invariant method (Zou et al., 2024). This correction effectively eliminates the influence of sensor degradation over time, providing a robust benchmark for generating long-term harmonized SIF products.

Detailed Comment 7: Line 87: 'A precise, reliable, harmonized, and global high-resolution SIF dataset is not yet available for long-term vegetation monitoring.' Based on the preceding sections of the introduction, the authors do not appear to have clearly explained why the prior data were neither precise nor reliable.

Response: Thanks for this comment. We have revised the introduction to more clearly articulate the limitations of previous datasets, thereby clarifying the motivation for developing the LHSIF product. The revised text is as follows:

While Wen's framework laid the foundation for cross-sensor harmonization, it did not explicitly address instrument degradation—a key factor that compromises the long-term consistency of single-sensor records. Such degradation, as observed in GOME-2, poses a significant challenge for long-term consistency and introduces uncertainties in trend analyses (Parazoo et al., 2019). For instance, Yang et al. (2018) reported diverging trends between EVI and SIF, attributing the latter's decline to reduced photosynthetic activity. However, Zhang et al. (2018a) argued that this conclusion was impacted by the deterioration of the GOME-2A instrument. Further research by Koren et al. (2018) showed that the decline in SIF persisted even after correcting for sensor degradation. The SIFTERv2 product (Van Schaik et al., 2020) employed in Koren's study was simply corrected using linear models; the reliability of SIFTERv2 decreased significantly after 2016, limiting its application for long-term trend analysis.

To mitigate this limitation, Wang et al. (2022) attempted to create a temporally corrected long-term SIF product (LT_SIFc*) by correcting the degradation trends in gridded GOME, SCIAMACHY, and GOME-2

SIF products. **However, the method lacks a physically based correction of the actual sensor radiance degradation and instead applies adjustments at the SIF product level, which may not accurately reflect the true instrumental change. This is further complicated by the nonlinear characteristics inherent in the retrieval methodology and subsequent processing procedures (e.g., zero-bias correction and quality filtering), which prevent a direct and linear propagation of sensor degradation into the final SIF values.** Recently, the temporally corrected GOME-2A SIF dataset (TCSIF) included a calibration of the radiance measurements of GOME-2A using a pseudo-invariant method (Zou et al., 2024). This correction effectively eliminates the influence of sensor degradation over time, providing a robust benchmark for generating long-term harmonized SIF products.

So far, the cross-sensor consistency of existing long-term SIF records remains to be further evaluated.

Detailed Comment 8: Line 131: Oco-2/Oco-3 → OCO-2/OCO-3

Response: Thanks, it has been corrected.

Detailed Comment 9: Line 207: Is this spatial resolution (0.072727° × 0.072727°) commonly referred to?

Response: Thanks for this comment. Yes, this spatial resolution is commonly used in related studies. Please refer to (Chen et al., 2019)and (He et al., 2021) for more details.

Detailed Comment 10: Line 221: There appear to be numerous NDVI products based on AVHRR data. Is this a newly developed set? Could you please provide some additional introduction?

Response: Thanks for this comment. We have added more information in Sec. 2.5.3 as below:

2.5.3 AVHRR vegetation indices

Global NDVI and NIRv datasets from 1995 to 2021, derived from the AVHRR sensors, were utilized in this study. These datasets were developed by Jeong et al. (2024) based on the AVHRR Long-Term Data Record version 5 (LTDR V5) surface reflectance product. **To address temporal inconsistency in long-term AVHRR records, a three-step correction was applied, including cross-sensor calibration, orbital drift correction, and machine learning-based harmonization with MODIS vegetation indices. This post-processing significantly improved the temporal consistency of NDVI and NIRv from 1982 to 2021, as verified using detrended anomalies and trends at calibration sites. The final product enables more robust analyses of long-term vegetation dynamics and reduces spurious trends due to sensor artifacts.**

Detailed Comment 11: Line 238-243: Could be more **quantitative**

Response: Thanks for this comment. To improve the quantitative evaluation of the downscaling results, we have revised Figure 3 and added additional statistical indicators, including RMSE and the fitted regression line. The updated figure and its description are as follows:

The distribution of monthly SIF before and after spatial downscaling is shown using GOME as an example (Fig. 3), while results for the other three satellites are provided in Figs. S3–S5. The spatially downscaled SIF (0.05° × 0.05°) was re-aggregated to 1° × 1° or 0.5° × 0.5° resolution for comparison with the original coarse-resolution SIF. **The results demonstrate that the SIF values from the re-aggregated pixels are generally consistent with the original SIF values, closely clustering along the 1:1 line and showing strong agreement ($R^2 > 0.73$, RMSE < 0.11 mW m^{-2} sr^{-1} nm^{-1}),** indicating that the LUE-based downscaling model effectively captures the relationship between SIF and its driving variables.

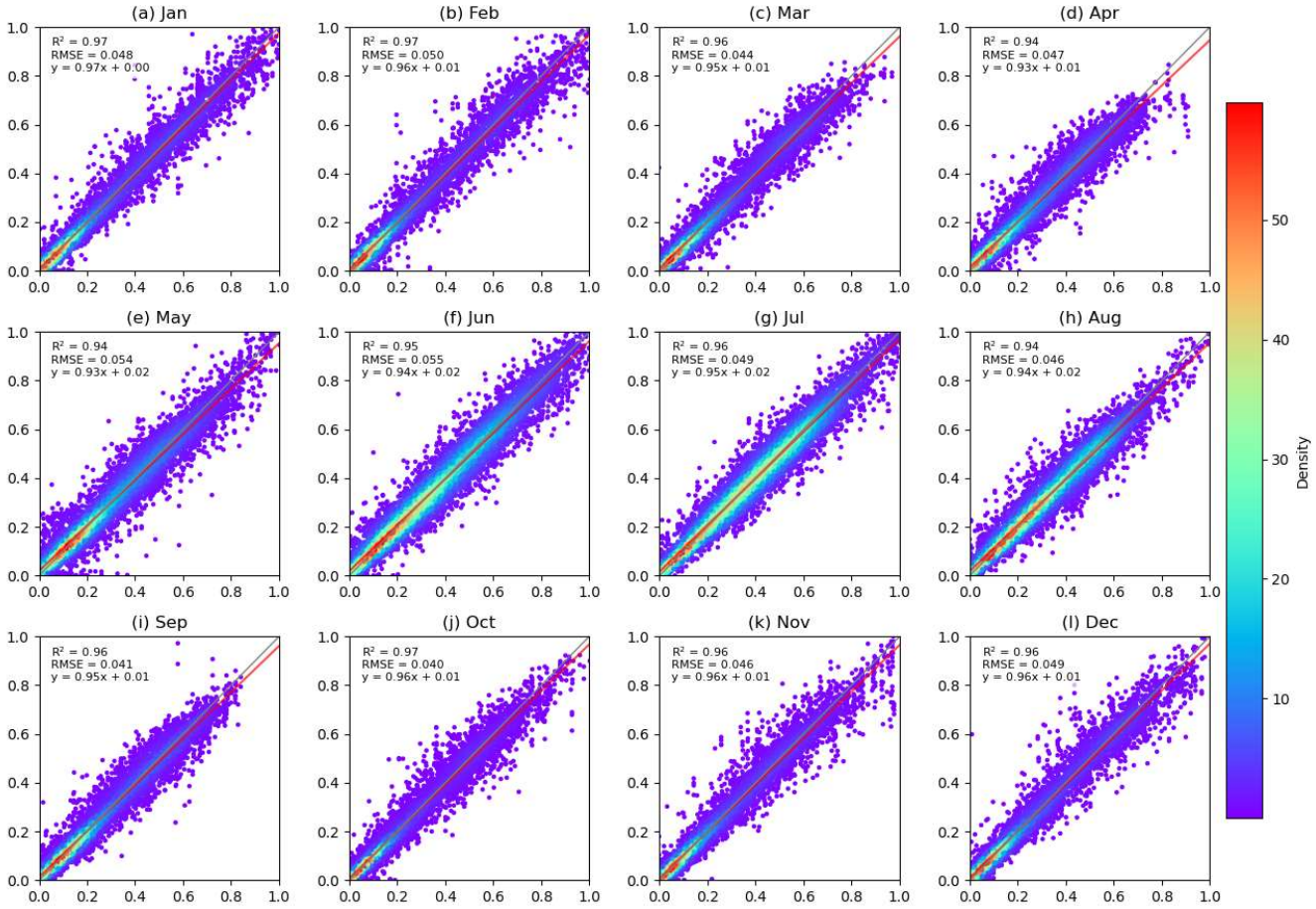


Figure 3. The relationship between the reaggregated GOME SIF (SIF_reagg) and the original GOME SIF (SIF_original) for 1998 (by month).

In addition, we sincerely apologize that the previous version of Figure 3 was generated using an outdated version of the GOME dataset due to a data management oversight. This has now been corrected: the figure has been re-generated using the updated and consistent dataset. Importantly, this correction does not affect the main conclusions of the study, but it improves the accuracy and consistency of the results.

Detailed Comment 12:. Figure 5: the order of GOME and SCIAMACHY in the legend is opposite.
Response: Thanks for this comment. We have checked and confirmed that the legend labels are now correct in the revised version.

Detailed Comment 13:.Line 267: as the difference between ? some words missed
Response: Thanks for this comment. The missing explanation has been added as follows:
The residuals are calculated as the difference between the reaggregated SIF (SIF_reagg) and the original SIF (SIF_original).

Detailed Comment 14: Line 340-345: These belong to Methods. Please check throughout the manuscript.
Response: Thanks for this comment. Corresponding sentences have been moved to Sec. 2.5.4 as below:

In addition, tower-based SIF observations from the ChinaSpec network, including sites such as DM, GC, HL, XTS, and AR (Zhang et al., 2021), were used to validate the accuracy and spatiotemporal consistency of the long-term SIF dataset generated in this study. The locations and cover types of the ChinaSpec sites used

are listed in Table S1. **To ensure consistent comparisons, the tower-based SIF at 760.6 nm was converted to 740 nm using an empirical correction factor of 1.48 (Du et al., 2023). Additionally, the original half-hourly tower-based SIF data were temporally upscaled to daily and monthly values with the aid of PAR and NDVI, following the method described by Hu et al. (2018).**

Detailed Comment 15: Line 343: please add the citation for this factor
Response: Thanks for this comment. We have added the citation below:
To ensure consistent comparisons, the tower-based SIF at 760.6 nm was converted to 740 nm using an empirical correction factor of 1.48 **(Du et al. 2024).**

Detailed Comment 16: Lines 26, 144, 158, etc.: please add the space between the text and the brackets
Response: Thanks for this comment. Spaces have been inserted before the brackets throughout the manuscript.

Detailed Comment 17: Section 4.1: Could you provide more concrete examples demonstrating the advantages of using TCSIF compared to the uncorrected GOME-2A data employed in the other two studies?

Response: Thanks for this comment. As demonstrated in Zou et al. (2024), the TCSIF product shows a more stable interannual trend during the GOME-2A period (2007–2021): 0.705 ± 0.152% yr^{-1}, compared to 1.247 ± 0.234% yr^{-1} in LT_SIFc* and –0.078 ± 0.356% yr^{-1} in SIF_005. The advantages of TCSIF are also observed in the LHSIF dataset, which displays more consistent trends (see Fig. 10). In contrast, SIF_005 shows a declining trend in several regions, likely due to uncorrected sensor degradation.

As the temporal correction is not the focus of this study, we respectfully refer the reviewer to Zou et al. (2024) for a more detailed comparison and methodology.

Detailed Comment 18: Line 374: Missing word.
Response: Thank you for pointing this out. The sentence at Line 374 was removed during revisions made in response to other suggestions.

Detailed Comment 19: Line 385-387: Providing concrete examples rather than textual descriptions would enable readers to better understand.
Detailed Comment 20: Line 391: Please see above.

Response: Thanks for these comments. To address these concerns, we revised the relevant sentences to avoid subjective descriptions and instead provided clearer quantitative comparisons. Additional supporting analyses have been incorporated into Section 3.3 and the discussion part. For more details, please refer to our responses to Major Comment 1 and Major Comment 5.

The references cited in the response are listed below:
Chen, J. M., Ju, W., Ciais, P., Viovy, N., Liu, R., Liu, Y., and Lu, X.: Vegetation structural change since 1981 significantly enhanced the terrestrial carbon sink, Nature Communications, 10, 4259, 10.1038/s41467-019-12257-8, 2019.
Chen, S., Liu, L., Sui, L., Liu, X., and Ma, Y.: An improved spatially downscaled solar-induced chlorophyll fluorescence dataset from the TROPOMI product, Scientific Data, 12, 135, 2025.
Didan, K.: MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006. [dataset], https://doi.org/10.5067/MODIS/MOD13A2.006, 2015.

https://doi.org/10.5194/essd-2025-94
He, Q., Ju, W., Dai, S., He, W., Song, L., Wang, S., Li, X., and Mao, G.: Drought Risk of Global Terrestrial Gross Primary Productivity Over the Last 40 Years Detected by a Remote Sensing-Driven Process Model, Journal of Geophysical Research: Biogeosciences, 126, e2020JG005944, https://doi.org/10.1029/2020JG005944, 2021.
Hu, J., Liu, L., Guo, J., Du, S., and Liu, X.: Upscaling Solar-Induced Chlorophyll Fluorescence from an Instantaneous to Daily Scale Gives an Improved Estimation of the Gross Primary Productivity, Remote Sensing, 10, 1663, 2018.
Ma, Y., Liu, L., Liu, X., and Chen, J.: An improved downscaled sun-induced chlorophyll fluorescence (DSIF) product of GOME-2 dataset, European journal of remote sensing, 55, 168-180, 2022.
Zou, C., Du, S., Liu, X., and Liu, L.: TCSIF: a temporally consistent global Global Ozone Monitoring Experiment-2A (GOME-2A) solar-induced chlorophyll fluorescence dataset with the correction of sensor degradation, Earth Syst. Sci. Data, 16, 2789-2809, 10.5194/essd-16-2789-2024, 2024.