Referee #1

We thank the Referee 1 for this review of our manuscript. Below, we address the comments with the comments of Referee 1 in bold and our reply in normal font.

Anema et al. present results from an update to their Solar-Induced Fluorescence of Terrestrial Ecosystems Retrieval (SIFTER) v3 algorithm applied to observations from the GOME-2A sensor, together with extensive comparisons to the previous SIFTER v2. Updates include the use of the latest level 1 GOME-2A radiance product, improvements to better account for instrument degradation, and changes to background ("zero offset") and latitudinal bias corrections.

Space-based measurements of Solar-Induced Fluorescence (SIF) have become established data products and are routinely observed from sensors including GOME, SCIAMACHY, GOME-2, GOSAT, OCO-2, TROPOMI, and OCO-3, with a combined data record that goes back to 1995. SIF is a highly challenging measurement to make from space, and updates and improvements to existing data products are highly welcome to reduce data uncertainty and enhance data consistency and accuracy.

1. Anema et al. demonstrate extensively and convincingly that their v3 SIFTER results present an improvement over the SIFTER v2 product in terms of consistency. However, they do not present any evidence about either version's accuracy. The only comparisons to non-SIFTER data are shown in Figure 13: scaled results from SIFTER are plotted against GPP measurements from FluxSat and FLUXCOM-X products to show that seasonal and inter-annual variations in GPP are reproduced by SIFTER SIF v3 better than v2. This is not evidence for the accuracy of the new version, only for its consistency.
Reply: The quantification of the accuracy of satellite-based SIF product is a well-known limitation and challenge. The difficulty of determining the accuracy of satellite-based SIF products is due to the lack of sufficient independent SIF observations from in-situ and airborne sensors (Rossini et al., 2022). Even with sufficient independent SIF observation, the scale mismatch between satellite observations and in-situ measurements further complicates accuracy assessment. Therefore, the validation of satellite-based SIF products primarily relies on the comparison with other SIF products. However, such cross-product comparisons are also complex as SIF is not measured directly and are further affected by uncertainties emerging from instrumental and retrieval algorithm differences.

2. A wide range of independent satellite-based SIF data products have been publicly released, including that of OCO-2 starting in September 2014 providing about three years of temporal overlap with GOME-2A. While the basic scope of this paper as a "algorithm modifications and product improvement" doesn't have to change, I feel strongly about the need to include, at the least, a comparison plot with independent SIF observations for a perspective on where the SIFTER results fall in relation to data from other instruments. As has been the case for a long time now with minor trace gases like BrO, H2CO, or C2H2O2, SIF is no longer a "first observation" type of measurement, and new data products should be benchmarked against published data records that have been accepted as the current standard.
Reply: We agree with the referee that new data products should be benchmarked against published data records. Our previous dataset, SIFTER v2, was compared to independent SIF data from NASA v28 in van Schaik et al. (2020), showing good consistency, both temporally and spatially. Furthermore, based on tests and fruitful scientific investigations done with SIFTER v2 in Mengistu et al. (2021), Wang et al. (2020), Wang et al. (2022), and Fancourt et al. (2022), we feel that substantial benchmarking has already taken place, and that SIFTER v2, and therefore SIFTER v3, can be considered part of the "current standard". We discuss the comparison of SIFTER v3 against other SIF data further in our reply to comment 5.

3. This is not to suggest that existing records are necessarily correct or that deviating new results are necessarily wrong. SIF in particular is a challenging observation to make, and "dissenting opinions" only help to move the state of these measurements forward. In this particular case, the seasonal peak SIF values shown in Figure 11 appear to be 20-40% higher than those reported from other satellite sensors, for essentially all vegetated regions. That warrants an explanation as to possible sources for these differences and the confidence in the results.

Reply: We thank the referee for this comment and agree that this manuscript will benefit from a comparison against different SIF products, including an explanation regarding differences in SIF. We discuss this further in our reply to comment 5.

It should be noted that the SIF values presented in this manuscript reflect the instantaneous SIF values, rather than the daily-scaled values that are commonly presented in other work, such as Wen et al. (2020), which analyzed similar geographical regions. For GOME-2A, with a local overpass at 10:30A M, daily-scaled values are approximately 30% lower than the instantaneous values. We acknowledge that this distinction may cause confusion, and we will emphasize in the manuscript that the SIF values presented reflect the instantaneous values.

To enhance the scientific significance of this study, while keeping its focus as an algorithm paper, I recommend the following modifications to the manuscript:

4. Streamline the discussion of the differences between v2 and v3, which can be presented in considerably abbreviated form without sacrificing insights into the modifications.
Reply: We agree with the Referee that the discussion on the differences between v2 and v3 can be more concise. We will revise the "4.2 Processing and improvements of SIFTER v3 retrieval" section. This section will be made clearer by introducing all the changes before discussing and showing the impact of each individual change on the retrieval. Additionally, we will combine Figures 6, 8, and 9 as suggested by the Referee in comment 12.

5. Add a comparison plot to an independent space-based, non-SIFTER SIF data product (e.g., the biomes in Figure 11 could be augmented with data from another satellite instrument or instruments) and a brief discussion of how the data products relate to each other.
Reply: We thank the referee for this comment and agree that the comparison of the SIFTER v3 product against other non-SIFTER SIF data will provide valuable insights into the performance of the new dataset. We will include the comparison of non-SIFTER SIF data products with SIFTER v3 in section 5 and incorporate their time series in Figure 11.

Differences in absolute SIF values across products and instrument arise from instrumental characteristics, such as overpass time, viewing geometry, spectral and spatial resolution, and sampling, as well as differences in retrieval settings, such as the retrieval window. This has been discussed in detail in Parazoo et al. (2020), who compares different SIF products from a variety of sensors (e.g. TROPOMI, GOME-2, SCIAMACHY, GOME, and OCO-2), and discussed their differences.

The below list analyzes which SIF products are potentially suitable and which are less appropriate (or out of scope) to compare against our SIFTER v3.
Appropriate:

- GOME-2A SIF, specifically the GOME-2A NASA product (Joiner et al., 2013, 2016): this comparison is relevant as it limits differences in absolute SIF due to retrieval window setting (both use 734—758 nm) and allows for a "fair" comparison. Furthermore, it provides insight into the impact of our efforts to constrain temporal consistency.
- SCIAMACHY: SCIAMACHY: Comparison against SCIAMACHY SIF data offers a sufficient temporal overlap with GOME-2 of 5 years (2007—2011). The morning overpass of both SCIAMACHY and GOME-2A, as well as similar footprint size, respectively 30x60 km2 and 40x80 km2 over the overlapping period, allow for valuable comparison.

Less appropriate:

- OCO-2 and GOSAT SIF: The restricted global mapping of these sensors limit comparison with GOME-2A SIF data (with much larger spatial resolution) on a monthly and regional scale.
- TROPOMI SIF: Since we only retrieved GOME-2A SIF over the pre-drift period (2007—2017), there is no overlap with TROPOMI SIF data.

Based on the above considerations, we selected the latest version of GOME-2A SIF from NASA (Joiner et al., 2023) as the independent SIF dataset to include in Figure 11.

The following comments are more detailed and editorial in nature and may help the authors during the revision of the manuscript. They are mostly intended as suggestions rather than mandatory points to be addressed, though several issues will benefit from clarifications.

6. Introduction: suggest to include this paper for OCO-2/3 SIF reference
Global GOSAT, OCO-2, and OCO-3 solar-induced chlorophyll fluorescence datasets; R. Doughty et al., Earth Syst. Sci. Data, 14, 1513–1529, 2022 https://doi.org/10.5194/essd-14-1513-2022
Reply: We thank the Referee for this suggestion. Including this reference in the introduction, in lines 20—23, will inform the reader on SIF retrievals from OCO-3.

7. Figure 1: "different hues of grey" isn't working well; would suggest "orange", "light red", "dark red" or something similar, to show pre-6.3.3 processor version. Alternatively, time frames of each processor version should be included either in the figure description or indicated in the plot (shading, lines, etc.) to give the reader an idea which processor version was used when.
Reply: We appreciate the Referee's comment on the clarity of Figure 1. We agree that the different hues of grey that represent the v5.3, v6.0, and v6.1 versions should be changed into different shades of red.

8. The mid-2013 drop must be the change in throughput related to the switch to narrow swath. But why exactly does the reflectance drop? Should that not be taken care of by updates to the radiometric calibration? The atmosphere doesn't change with the switch to a reduced swath, and vicarious calibration or cross-sensor radiometric comparisons (MODIS, etc.) should provide information on the actual radiance levels.
Reply: The mid-2013 drop is indeed caused by the switch to the reduced 960-km swath (see lines 114-116). The reduction in swath led to observations being made at a smaller range of viewing zenith angles (from between ±55° to ±35°). These more nadir-like observations result in decreased reflectance due the bidirectional reflectance distribution function (BRDF) of the Earth's surface, which causes reflectance to vary with viewing geometry (Tilstra et al., 2021).

9. How do the 740 nm R3 reflectances in this figure relate to the equivalent 747 nm R3 reflectances in Figure S2 that show a smooth transition across the switch to reduced swath?
Reply: We acknowledge that the inconsistency in respective wavelength of the reflectance shown in Figure 1 and Figure S2 is confusing. Therefore, we will change Figure S2 such that it shows the degradation corrected reflectance at 740 nm. The effect of changed viewing geometry, due to the swath reduction, on the reflectance differs with wavelength.

10. Figure 2: Why are post-2013 reflectances not shown? Would it be instructive to limit pre- and post swath reduction reflectances to the extend of the reduced swath?
Reply: To obtain correction factors for the reflectances for the entire 2007-2017 record, we used two fit periods: 2007-2012 and 2007-2017 (see Table 1) that account for the change in viewing geometry due to the swath reduction in 2012. Figure 2 shows the fits across the 2007-2012 period. It serves as demonstration of the applied fit on the global averaged reflectance data. We will include the 2007-2017 (under reduced swath geometry) fits in the appendix.

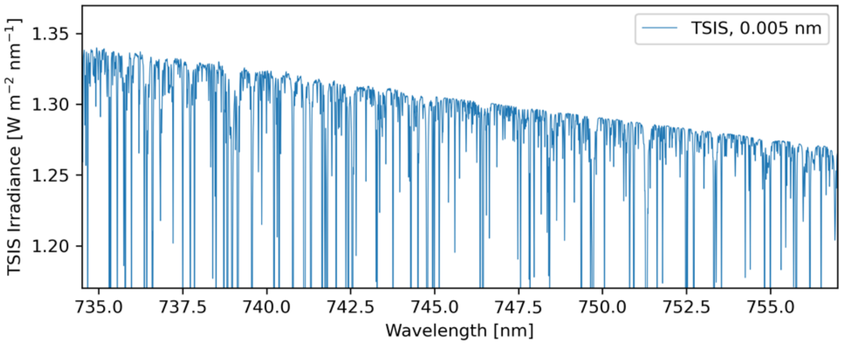
11. Figures 3&4: These could be combined, since they are principally showing the same thing; as for visual cosmetics, discrete color levels (12?) might introduce some structure into the monotone Figure 4.
Reply: Although Figures 3 and 4 both show the correction factors over time, they both do show different elements. Figure 3 shows the scan-angle dependency, whereas Figure 4 shows the wavelength dependency. Therefore, they could not be combined.

12. Figures 6,8,9: Those panels could be combined into a single figure (with shared x-axes to save vertical space).
Reply: we thank the referee for this comment. We agree that Figures 6, 8, and 9 should be combined with shared x-axes. This will also help with the streamlining of the discussion on the differences between SITER v2 and v3.

13. Figure 7: The SIFTER 3 ILS is lost in the line width, to the point that the visual effect is somewhat strange; suggest to reduce line thickness (or switch v2 and v3 thickness), and/or include a zoom of, e.g., the 741-742 nm region.

Reply: We agree with the Referee that the current infilling of the differences between the irradiances of SIFTER v2 and SIFTER v3 (in blue, Figure 7) appears visually somewhat strange. Including a zoom at a specific region indeed provides clarity.

For general information: The TSIS solar reference spectrum is becoming more widely adopted as the standard irradiance reference; absolute radiometric levels differ slightly from Chance/Kurucz (see image below). https://lasp.colorado.edu/lisird/data/tsis1_hsrs_p1nm



Reply: Thanks for pointing us to this improved solar reference spectrum. We intend to study the impact of this improvement on our SIF-retrievals in future studies.

Zero-Level Offset Adjustments:
First, for reference: practically every existing SIF retrieval approach neglects the effect of inelastic Raman scattering on the Fraunhofer lines. This introduces an error in SIF retrievals that, while negligible over high-SIF biomes, disproportionally affects low-SIF regimes and, with that, necessarily zero-level offset corrections. A study to quantify this effect is currently under review (and thus not available to the authors of this manuscript).
Reply: Thanks for making us aware of this development.

14. SIFTER v3 switched to including fully cloudy pixels for background correction, which means more implicit variability of rotational Raman scattering in the background references. More clouds will mean less atmospheric Raman scattering, hence less reduction in Fraunhofer line depth and thus less "erroneous SIF" over non-fluorescing surfaces. Does this conform with the change in background correction values shown in Figure S7? That figure is a little hard to interpret (and it may also benefit from a tightening of the plot range to ±0.5 or ±0.4).
Reply: Considering cloudy pixels to obtain the zero-level adjustments indeed lead to fewer "false" positive SIF over non-vegetative surfaces (lines 321—325). We agree that the plot range could be tightened to ±0.5.

15. Figure 10: Is the the latitude-dependent ILS is known, or can it be derived from in-flight spectra? Would that help with the latitude-dependent offset correction?
Reply: The exact latitude-dependent varying slit function is not exactly known as this effect is thought to arrange from multiple sources (Joiner et al., 2012), with the thermal instability across orbit as one of the prominent causes. Understanding the latitude-dependence of the ILS will certainly help, but by itself is not sufficient for a comprehensive correction. This is because the measurements of GOME-2A with its ILS over our reference area (the Sahara desert, 5°-25° N) are not necessarily representative for other latitudes. We are considering testing whether additional reference areas (at other latitude zones) are helping to reduce the latitude-dependent offsets.

General:
16. Ever so often, use "allow" instead of "enable"
Reply: we appreciate this comment regarding the over-use of "enable". We agree that this should be modified.

17. Line 73: Check the font – is it "oh cee ell oh" or "oh cee eye oh"? (OClO or OCIO)
Reply: We thank the Referee for his attentiveness. OClO is chlorine dioxide and is written as OClO (with 'ell', not 'eye').

18. Line 88: proceeded → performed
Reply: we will change "proceeded" to "performed" in line 88.

19. Line 90: Are there any details on this "drop of throughput"? Specifically why does it affect the reflectances?
The throughput tests concerned temporary substantial increases of the instrument's temperature. The changed detector temperatures during the second throughput test unexpectedly resulted in a loss of signal of the solar measurements (EUMETSAT, 2009, 2022). The reflectance is affected when the throughput for the earthshine optical path and solar optical path changes in a different way (EUMETSAT, 2022). Hypothetically, when they both change in the same way, the reflectance would not be affected.

20. Line 133: "narrow swath" and "nadir static" supposedly are a special observation modes?
Reply: Observations under narrow swath are done under a swath of 320 km, instead of the default 1920 km swath (Munro et al., 2016). For one day in each 29-day observation cycle, GOME-2A observes under narrow swath. Nadir static is an observation mode where the scan mirror points nadir without scanning, and which is used for monthly calibration of the instrument (Munro et al., 2016). The different viewing geometries of these two special observation modes result in deviating reflectance in comparison to the observations done under nominal swath (that are used for the SIF retrieval), therefore we excluded observations under narrow swath and nadir static from our analysis.

21. Equation 1: by itself, this doesn't provide much information that couldn't be conveyed by text alone. Can the full equation be provided?
Reply: The full equation is given in Tilstra et al. (2012). We agree with the Referee that it's useful to repeat the full equation in this manuscript. We model the global mean reflectance (R^*) by:

$$R^*_{\lambda,s}(t) = P^p_{\lambda,s}(t)\left[1 + F^q_{\lambda,s}(t)\right]$$

$P^p_{\lambda,s}$ is the polynomial with degree *p*, describing the long-term trend in R^*, and $F^q_{\lambda,s}$ is a finite Fourier series with order *q*, describing the seasonal variation. λ is the wavelength, *s* the scan angle and *t* is time.

With:

$$P^p_{\lambda,s}(t) = \sum_{m=0}^{p} u^m_{\lambda,s} t^m$$

And

$$F^p_{\lambda,s}(t) = \sum_{n=1}^{q} [v^n_{\lambda,s}\cos(2\pi nt) + w^n_{\lambda,s}\sin(2\pi nt)]$$

22. Line 162: "Scanner Angle" (and "Scanning Angle" or "Scan-Angle") → "Scan Angle"
Reply: We thank the referee for this comment. We indeed use different variations throughout this manuscript. We will modify this and consistency use "scan angle".

23. Line 165: "sensor-switch" supposedly is the change to the reduced swath? That term is a bit confusing, "swath reduction" would be better.
Reply: The "sensor switch" indeed refers to the reduction of the swath. We agree that it's clearer to refer to the reduction of the swath.

24. Line 171: post-sensor switch→post sensor-switch
Reply: See our comment on point 23.

25. Line 182: "relative filling-in of solar Fraunhofer absorption lines" → "reduction of solar Fraunhofer line depth in the radiance spectra" that makes it clear in which spectra this is happening, and it also makes it intuitive that the overall effect is SFLs showing up as enhancements (peak)s in the I/I0 reflectances, from a combination of SIF and inelastic Raman scattering.
Reply: The word "relative" is of importance in line 182. With the re-emission of chlorophyll fluorescence by vegetation the radiance spectra are enhanced both inside and outside the solar Fraunhofer lines. However, due to the low radiance value within the absorption lines, the addition due to fluorescence leads to a noticeable higher relative enhancement within the solar Fraunhofer lines than outside these lines. We do agree to add "in the radiance spectra" in line 182 to enhances the clarity.

26. Figure 8a: Absolute uncertainties remain the same, thus relative uncertainties increase by 15% - does that hold true in general?

Reply: Due to changed slit function in SIFTER v3, the absolute values of SIF decreased with respect to SIFTER v2. Figure 12 shows the correlation between SIF uncertainty in SIFTER v2 vs SIFTER v3. The correlation between absolute SIF values between SIFTER v2 and SIFTER v3 is shown in Figure S12. The decrease in absolute SIF values is not higher than the decrease in the SIF uncertainties, therefore we don't expect a significant change in the relative SIF uncertainties.

27. Line 282: Can a reduction from 0.068% to 0.063% in RMSE really be considered "substantial" or "significant"? While that is indeed a reduction of ~10% in relative RMSE, what are the corresponding values in terms of absolute radiance (~?x10-4)?

Reply: We agree with the referee that the word "substantial" is slightly excessive. However, considering the sensitivity of the SIF retrieval an improvement of 0.063% to 0.068% should not be overlooked. The effect of this improvement is also noted as a decrease on the order of 10% in absolute SIF uncertainty (Figure 12).

References

EUMETSAT (2009). GOME-2 FM3 Long-Term In-Orbit Degradation - Status After 1st Throughput Test, EUM/OPS-EPS/TEN/08/0588.

EUMETSAT (2022). GOME-2 Metop-A and -B FDR Product Validation Report Reprocessing R3, EUM/OPS/DOC/21/1237264.

Fancourt, M., Ziv, G., Boersma, K. F., Tavares, J., Wang, Y., & Galbraith, D. (2022). Background climate conditions regulated the photosynthetic response of Amazon forests to the 2015/2016 El Nino-Southern Oscillation event. Communications Earth & Environment, 3(1), 209.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Middleton, E. M., Campbell, P. K. E., Yoshida, Y., & Huze, A. (2012). Filling-in of near-infrared solar lines by terrestrial fluorescence and other geophysical effects: simulations and space-based observations from SCIAMACHY and GOSAT. Atmospheric Measurement Techniques, 5(GSFC-E-DAA-TN9416).

Joiner, J., Y. Yoshida, P. Koehler, C. Frankenberg, and N.C. Parazoo. (2023). L2 Daily Solar-Induced Fluorescence (SIF) from MetOp-A GOME-2, 2007-2018, V2. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/2292

Mengistu, A. G., Mengistu Tsidu, G., Koren, G., Kooreman, M. L., Boersma, K. F., Tagesson, T., ... & Peters, W. (2021). Sun-induced fluorescence and near-infrared reflectance of vegetation track the seasonal dynamics of gross primary production over Africa. Biogeosciences, 18(9), 2843-2857.

Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., ... & Eisinger, M. (2016). The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing–an overview. Atmospheric Measurement Techniques, 9(3), 1279-1301.

Parazoo, N. C., Frankenberg, C., Köhler, P., Joiner, J., Yoshida, Y., Magney, T., ... & Yadav, V. (2019). Towards a harmonized long-term spaceborne record of far-red solar-induced fluorescence. Journal of Geophysical Research: Biogeosciences, 124(8), 2518-2539.

Rossini, M., Celesti, M., Bramati, G., Migliavacca, M., Cogliati, S., Rascher, U., & Colombo, R. (2022). Evaluation of the spatial representativeness of in situ SIF observations for the validation of medium-resolution satellite SIF products. *Remote Sensing*, *14*(20), 5107.

Van Schaik, E., Kooreman, M. L., Stammes, P., Tilstra, L. G., Tuinder, O. N., Sanders, A. F., ... & Boersma, K. F. (2020). Improved SIFTER v2 algorithm for long-term GOME-2A satellite retrievals of fluorescence with a correction for instrument degradation. Atmospheric Measurement Techniques, 13(8), 4295-4315.

Tilstra, L. G., De Graaf, M., Aben, I., & Stammes, P. (2012). In-flight degradation correction of SCIAMACHY UV reflectances and Absorbing Aerosol Index. Journal of Geophysical Research: Atmospheres, 117(D6).

Tilstra, L. G., Tuinder, O. N., Wang, P., & Stammes, P. (2021). Directionally dependent Lambertian-equivalent reflectivity (DLER) of the Earth's surface measured by the GOME-2 satellite instruments. Atmospheric Measurement Techniques, 14(6), 4219-4238.

Wang, S., Zhang, Y., Ju, W., Porcar-Castell, A., Ye, S., Zhang, Z., ... & Boersma, K. F. (2020). Warmer spring alleviated the impacts of 2018 European summer heatwave and drought on vegetation photosynthesis. *Agricultural and Forest Meteorology*, *295*, 108195.

Wang, S., Zhang, Y., Ju, W., Wu, M., Liu, L., He, W., & Penuelas, J. (2022). Temporally corrected long-term satellite solar-induced fluorescence leads to improved estimation of global trends in vegetation photosynthesis during 1995–2018. ISPRS Journal of Photogrammetry and Remote Sensing, 194, 222-234.

Wen, J., Köhler, P., Duveiller, G., Parazoo, N. C., Magney, T. S., Hooker, G., ... & Sun, Y. (2020). A framework for harmonizing multiple satellite instruments to generate a long-term global high spatial-resolution solar-induced chlorophyll fluorescence (SIF). Remote Sensing of Environment, 239, 111644.