

Referee #2

We thank Referee 2 for this review of our manuscript. Below, we address the comments with the comments of Referee 2 in bold and our reply in normal font.

This paper presents a harmonised GOME-2A and GOME-2B SIF dataset generated using the SIFTER v3 algorithm, along with an in-depth discussion on SIF biases caused by sensor degradation. This work is of significant value for the application and observation of SIF data. However, I have the following concerns:

- 1. The degradation correction is based on the long-term trend of global daily mean reflectance. However, while this trend is attributed primarily to instrument degradation, could it also include contributions from long-term changes in cloud cover, aerosols, or surface albedo?**

Reply: Reported overall trends in global reflectance are much smaller, with MODIS data estimating a decrease in global albedo of 0.0004 between 2002 and 2016 (Li et al., 2018), compared to the impact of throughput loss of up to +10% in terms of reflectance (EUMETSAT, 2022). Therefore the long-term changes from geophysical factors, such as, changes in cloud cover, aerosols, and greening, will be smaller than the accuracy of the degradation correction method used (Tilstra et al., 2012b).

Note, that this point was also raised by Referee #1 in his/her first and second comment. In response to these comments we performed a robustness test over the bright Sahara region. The results indicated similar tendencies. This strongly suggests that the degradation of the sensor led to the observed long-term trends. Furthermore we implement the following textual changes:

- Lines 128-129 will be changed to “Although no major long-term trends are expected, substantial long-term trends are noted (Fig. 2).”
- After line 134 of the manuscript we add: “While geophysical changes, such as variations in cloud fraction, aerosols, and global greening, can affect long-term trends in global mean reflectance, their impact is expected to be minor compared to the observed trends and attributed impact of throughput loss following instrument degradation (Li et al., 2018; EUMETSAT, 2022). Moreover, the pronounced scan-angle dependence of the observed long-term trends strongly indicates an instrumental origin.”

- 2. Figure 3 shows an overall increase in GOME-2A reflectance relative to the reference day across all scan positions during the first six years, followed by a decline. Could you please explain the reason for this initial increase?**

Reply: The reflectance (R) is proportional to the observed radiance (I) to solar irradiance (I_0) ($I:I_0$). Reflectance degradation arises from differences in throughput loss between the observed radiance and solar irradiance over time (Tilstra et al., 2012a). Both follow different optical paths, making them subject to different instrumental effects. The decline in reflectance during the first years of GOME-2A reflects a stronger degradation of the solar irradiance signal compared to the radiance signal. Vice versa, an increase in reflectance, noticed in the later years, represents a stronger degradation of the radiance signal, as described in lines 144—147 of the manuscript.

- 3. Figure 4(a) shows that the spectral features (peaks/troughs) in PC#1 for GOME-2B are less coherent and shallower than those for GOME-2A. The authors attribute this to “higher reflectance uncertainty.” However, could this also indicate a broadening of the Instrument Slit Function in GOME-2B? If the PCs fail to capture the sharp absorption features seen in GOME-2A, this could lead to systematic biases in SIF retrieval.**

Reply: We agree with the referee that the slight difference in the slit functions could, in principle, contribute to differences in the principal components (PCs). While GOME-2A and GOME-2B share an identical instrumental design, small differences in their instrumental properties are unavoidable in practice. GOME-2A has a spectral resolution of 0.48 nm, whereas GOME-2B has a spectral resolution of 0.50 nm; corresponding to a 4% difference in

effective slit function width. Such a difference could indeed lead to minor smoothing of spectral features in GOME-2B. We will include this instrumental difference as a potential source of the observed differences in #PC1 in the revised manuscript.

Overall, GOME-2A and GOME-2B show highly consistent spectral patterns in the PCs, with only minor difference in feature sharpness. The consistency indicates that the dominant transmittance structures are well captured in both sets of PCs. Nonetheless, slight differences in instrumental characteristics may cause systematic biases in the SIF retrieval, as discussed in the manuscript (Lines 297–298). The overall strong agreement between GOME-2A and GOME-2B SIF over the analysed regions suggest that residual differences do not significantly impact the robustness of the retrieved SIF.

We will mention the difference in spectral resolution as a potential source of the observed difference in #PC1 by changing lines 182-183 from “The less defined pattern in GOME-2B may reflect its generally higher reflectance uncertainty in the NIR compared to GOME-2A. ” to “The less defined pattern in GOME-2B may reflect its generally higher reflectance uncertainty in the NIR and/or its slightly courser spectral resolution (0.50 vs 0.48 nm, Table 1) compared to GOME-2A”

- 4. Line 175: The PCs are derived from different time periods (2007–2012 vs. 2013–2018). Given potential interannual variability in atmospheric conditions (e.g., water vapour) over the Sahara, this temporal mismatch might introduce non-instrumental differences into the PCs. I suggest verifying the PCs using the overlapping period (2013–2017) for both sensors to isolate purely instrumental differences.**

Reply: We appreciate the referee’s suggestion to use the overlapping period for verification of the PCs to isolate instrumental differences. However, in the case of GOME-2A and GOME-2B, the overlapping period is affected by both instrumental degradation (most pronounced period for GOME-2A) and differences in sampling due to the reduced swath configuration of GOME-2A. This will complicate the comparison and isolation of instrumental differences. In addition, the reduced swath of GOME-2A during this period limits the sampling of viewing zenith angles to $[-34^\circ, +34^\circ]$, compared to $[-54^\circ, +54^\circ]$ during the observations made during the 2007–2012 period.

We computed both sets of PCs from a large selection of spectra, spanning 6 years of data, to ensure robust sampling of a wide range of atmospheric conditions. Furthermore, the Sahara region is considered to be radiometrically stable. Therefore, we do not expect substantial differences in the derived PCs related to atmospheric variability.

- 5. Line 225: The observation that GOME-2B is more negative over barren areas but more positive over high-vegetation regions implies a multiplicative bias (gain difference) rather than a simple offset. If so, is the "simple additive correction" proposed later sufficient to address this amplitude-dependent discrepancy?**

Reply: We thank the referee for this comment. While the observed pattern of more negative values over barren area and more positive values of SIF may suggest a multiplicative component, this pattern is not uniquely indicative of a multiplicative gain difference. In theory, the SIF signal over barren areas should be 0. However, several factors, including instrumental and atmospheric, can impact the apparent relative in-filling of the Fraunhofer lines, and thus the retrieved SIF. For example, recent work (Sanghavi et al., 2025), shows that atmospheric inelastic scattering causes in-filling of Fraunhofer lines which can introduce significant biases in SIF over vegetation-free regions (up to $0.35 \text{ mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$), which may contribute to the slight difference in SIF between GOME-2A and GOME-2B over barren regions. Additionally, when accounting for the different viewing geometry sampling, the noted difference in GOME-2A and GOME-2B is not observed consistently across all regions. A true gain difference would be expected to produce a systematic scaling between the sensors. Hence, this would suggest that the discrepancies cannot be attributed to a simple multiplicative bias. Given this, the additive correction proposed in this study (Section 5.1) is designed based on the observed differences between GOME-2A and GOME-2B SIF over vegetative regions, where the SIF signal is most relevant.

To further assess whether the discrepancy between GOME-2A and GOME-2B includes a multiplicative bias, we extend the tested model (Eq. 4 of the manuscript) by including an additional term allowing the seasonal amplitude to change after the transition at T_0 , resulting in the following equation:

$$y_t = \mu + \alpha t + S_t^* + \delta U_t + \epsilon_t \quad (\text{Eq. 1})$$

With S_t^* as:

$$S_t^* = \beta_1 \sin\left(\frac{2\pi(t - \phi)}{12}\right) + \beta_2 \cos\left(\frac{2\pi(t - \phi)}{12}\right) + \zeta_1 U_t \sin\left(\frac{2\pi(t - \phi)}{12}\right) + \zeta_2 U_t \cos\left(\frac{2\pi(t - \phi)}{12}\right) \quad (\text{Eq. 2})$$

We fit this model (Eq. 1) to the regions of Eastern China and the Amazon, where we noticed small discrepancies between GOME-2A and GOME-2B in the manuscript. Table 1 shows the regression results of this model fit. To allow easy comparison against the model with solely an additive factor, we've included the regression results from Eq. 4 of the manuscript within brackets in blue.

Table 1. Regression results of the model (Eq. 1) for Eastern China and the Amazon region. The results of the model used in the manuscript (Eq. 4 of the manuscript) are shown within brackets and in blue for comparison.

Regression results	Eastern China	Amazon
R^2	0.90 (0.89)	0.89 (0.88)
$R^2_{\text{adj.}}$	0.89 (0.88)	0.89 (0.88)
RMSE [$\text{mW m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$]	0.16 (0.16)	0.11 (0.11)
ϕ	0.02 (0.02)	0.97 (0.97)
μ [$\text{mW m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$]	0.67** (0.68**)	1.21** (1.22**)
α [$\text{mW m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$]	4.11e-4 (4.06e-4)	1.99e-4 (2.02e-4)
β_1	0.23** (0.24**)	-0.24** (-0.30**)
β_2	-0.51** (-0.59**)	0.30** (0.30**)
ζ_1	0.02	-0.01**
ζ_2	-0.13**	0.01
δ [$\text{mW m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$]	0.13** (0.13**)	-0.10** (-0.10**)

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

The results in Table 1 show that the inclusion of the multiplicative terms (ζ_1, ζ_2) only leads to a minor improvement in model performance by a slightly increase in R^2 , but with no change in RMSE. Although in both regions, one of the interaction terms (ζ_1 or ζ_2) is statistically significant, the magnitude of these coefficients is smaller than the corresponding seasonal coefficients (β_1 or β_2). Furthermore, the other model parameters do not change substantially by the inclusion of the multiplicative terms, indicating that the model's structure is not fundamentally altered by these terms. While a multiplicative component cannot be fully excluded, the dominant discrepancy is captured by an additive offset. Finally, we advise caution in using the corresponding fit coefficients ζ_1 and ζ_2 for bias correction between GOME-2A and GOME-2B SIF, as this could distort genuine seasonal patterns. We will include this additional analysis in the supplementary S3 of the manuscript.

6. **Line 247 : suggests that eastern pixels might have higher SIF due to lower SZA. However, Figure 8 clearly shows that western pixels ($VZA > 0$) exhibit significantly higher SIF than eastern pixels during peak seasons (e.g., US Corn Belt in JJA). This indicates that the azimuthal anisotropy (sunlit/hot-spot effect on the west) dominates over the SZA timing effect.**

Reply: We thank the referee for this observation. The intention of Section 4.2 of the manuscript is to discuss the various different mechanisms contributing to the viewing geometry dependence of SIF. The relative importance of each factor can vary depending on vegetation type, canopy structure, region, and time of observation. We first describe the effect that each mechanism can have individually. Figure 8 of the manuscript shows the combined influence of multiple mechanisms. The SZA timing effect, discussed in line 247, refers to the increase in illumination for eastern pixels (relative to nadir and western pixels), which can lead to higher SIF. Additionally, the hot-spot effect can lead to even larger enhancements towards the western side of the swath, as clearly shown in Figure 8. The referee is right to observe the dominance of the hot-spot effect in the shown regions during the peak seasons. We will enhance the clarity of the text by clearly stating that we first discuss the various mechanisms affecting SIF, rather than the combined effect of all mechanisms.

Specifically, we will add the following after line 239 “In the following, we describe each individual mechanism contributing to these dependencies on SIF. The relative importance of these mechanisms depend on the conditions, such as canopy structure and illumination geometry.”

7. **Fig 10 shows that restricting viewing angles reduces the bias to 2.1%. However, as noted, the sampling is geographically limited to Northern latitudes. Given that tropical rainforests (e.g., Amazon) have distinct canopy structures and much higher SIF intensities—and showed the largest biases in Fig. 6—can validation based primarily on high-latitude data sufficiently demonstrate global consistency? Is there a risk that the sensors agree well at low/moderate SIF levels (high latitudes) but still diverge at high SIF levels (tropics)**

Reply: The reduced swath configuration of GOME-2A limits the ability to perform strict co-sampled comparisons with GOME-2B to Northern latitudes. We acknowledge that this restricts a full validation of the consistency over tropical regions.

However, our analysis presented in the manuscripts indicates that the dominant driver of inter-sensor differences is the viewing geometry. In Fig. 8 of the manuscript, we show the relationship between GOME-2A and GOME-2B SIF as a function of VZA. This relationship is highly consistent across different regions, including tropical regions, such as the Amazon. This indicates that the VZA-dependence of the bias is not fundamentally different in the tropics compared to higher latitudes.

Furthermore, confidence in the consistency of the sensors over high SIF levels is provided by the similar relationships observed between GOME-2A and FluxSat and between GOME-2B and FluxSat across several high-SIF regions, such as Eastern China and the Corn Belt (Fig. S2.21 and Fig. S2.22 of the manuscript). These results do not indicate systematic divergence at higher SIF levels (or lower latitudes). While we can't fully exclude region-specific effects, the available evidence does not suggest a systematic divergence between GOME-2A and GOME-2B at higher SIF levels.

References

- EUMETSAT (2022). GOME-2 Metop-A and -B FDR Product Validation Report Reprocessing R3, EUM/OPS/DOC/21/1237264.
- Li, Q., Ma, M., Wu, X., & Yang, H. (2018). Snow cover and vegetation-induced decrease in global albedo from 2002 to 2016. *Journal of Geophysical Research: Atmospheres*, 123(1), 124-138.
- Tilstra, L. G., De Graaf, M., Aben, I., & Stammes, P. (2012a). 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., & Stammes, P. (2012b). A new method for in-flight degradation correction of GOME-2 Earth reflectance measurements, with application to the Absorbing Aerosol Index. In Proceedings of the 2012 EUMETSAT Meteorological Satellite Conference (Vol. 570).

Sanghavi, S., Frankenberg, C., Nelson, R. R., O'Dell, C. W., Rosenberg, R., & Joiner, J. (2025). Impact of Raman scattering on SIF retrievals from hyperspectral satellite observations. *Geophysical Research Letters*, 52(15), e2024GL112777.