Dear authors,

I'm writing this comment mainly from my perspective as HEGIFTOM co-chair.

First of all, I want to congratulate you with your nice paper. What I really like about the paper is the honest approach. From the first TOAR activity, and in the paper by Gaudel et al. (2018), the inconsistency in tropospheric ozone time variability (and even in the sign of the trend) among different satellite ozone retrievals (and even between different retrieval methods of the same satellite data) popped up. Your paper tries to contribute to this issue, and you do not hide the overall negative IASI trends, which contrast with UV-VIS satellite tropospheric ozone measurements and ground-based measurements, but really try to find explanations for it (nicely summarized in the conclusions). And the tropospheric ozone decline after 2019 is really a common feature in all those datasets.

However, what I do miss in the introduction (lines 50-60) and in the conclusions (lines 550-554) are more references to results from TOAR-II papers. Please look for more relevant references (also for (regional) trends) in the TOAR-II SI collection of papers: https://acp.copernicus.org/articles/special_issue1256.html.

With my HEGIFTOM hat on, I want to draw your attention to some additional clarifications that are needed when using the HEGIFTOM ozonesonde data. First of all, your Table 2 contains some errors, e.g. Lindenberg, Prague, Tateno, Hongkong, Broadmeadows, and Macquarie are not available as HEGIFTOM time series, because these data have not been homogenized. So, they can be only available as WOUDC data. It might be important to explicitly mention that only the HEGIFTOM (and SHADOZ) data have been homogenized (i.e. corrected for biases), which is very important for use in trend estimation. Also important to note is that the HEGIFTOM and SHADOZ data are identical: the SHADOZ data have been copied on the HEGIFTOM ftp-server.

Thank you very much for your kind and constructive feedback. We appreciate your positive remarks on our work and the helpful suggestions regarding the TOAR-II references and the HEGIFTOM ozonesonde data. We have carefully considered your comments and made the necessary clarifications and corrections in the revised manuscript.

Comment 1: Table 2 contains some errors regarding the availability of HEGIFTOM time series (e.g., Lindenberg, Prague, Tateno, etc.). It might be important to explicitly mention that only the HEGIFTOM (and SHADOZ) data have been homogenized.

Thank you for this important clarification. Indeed, Table 2 includes both homogenized (HEGIFTOM and SHADOZ) and non-homogenized (WOUDC-only) ozonesonde records. Our intention was to use the full set of available ozonesonde data, regardless of homogenization status, to maximize spatial and temporal coverage, particularly for the satellite-sonde comparison, where a wider dataset increases the likelihood of coincidences. We acknowledge, however, that homogenized data are more appropriate for long-term trend analysis. As such:

- For trend estimation, we relied only on homogenized data from HEGIFTOM and SHADOZ.
- For satellite-sonde comparisons, we included all available stations, even if not homogenized (i.e., WOUDC-only).

This has been clarified in the revised manuscript.

Also, I found it very surprising that only 7 stations fulfill your criteria of having at least 3 monthly launches and > 70% sampling. Knowing that most stations launch weekly and overall do not reveal large, consistent, gaps in their time series, this looks rather an underestimation to me. In our HEGIFTOM individual trends paper (see Van Malderen et al., 2025), with at least 2 monthly launches and more or less > 50% sampling in the 2000-2022 time period, we end up with 34 homogenized ozonesonde stations! Are QA/QC criteria driving this small sample? Which ones? Please give more details about the site selection.

Our initial selection criteria may appear overly restrictive, especially when compared to the HEGIFTOM analysis (Van Malderen et al., 2025), which included 34 homogenized stations using a more relaxed threshold of at least two monthly launches and >50% sampling.

In our original analysis, we applied more stringent selection criteria (at least three launches per month and 70% time series completeness between 2008 and 2023). These thresholds, adapted from Lu et al. (2019), were chosen to ensure the reliability of trend estimates by minimizing uncertainty in monthly means and maintaining sufficient temporal overlap with satellite retrievals.

However, in response to your comment, we repeated the trend analysis using the less restrictive criteria of at least two monthly launches and >50% sampling, consistent with the HEGIFTOM approach. This adjustment significantly increased the number of eligible stations from 7 to 27.

Importantly, for six of the seven stations included in our initial analysis, trend results remained consistent under the relaxed criteria. For one station, the trend became statistically significant with high certainty, likely due to the more complete time series provided by the less restrictive thresholds. Despite this exception, the overall agreement between the two approaches suggests that relaxing the selection criteria does not substantially affect trend outcomes. We have therefore adopted the relaxed criteria in the revised manuscript to broaden the station sample while ensuring consistent and reliable results.

Following your recommendation, we have also restricted the trend analysis to homogenized ozonesonde records only. Non-homogenized stations have been excluded from trend calculations to ensure consistency. These stations are still utilized for the validation of IASI data but are not included in the long-term trend estimation. This approach ensures a more rigorous and consistent assessment of trends.

Your mostly negative IASI tropospheric ozone trends in the 2008-2023 time period contrasts somewhat with the mixture of insignificant, positive and negative trends that we found for 55 HEGIFTOM sites (IAGOS, FTIR, ozonesonde, Umkehr, Lidar) for the time period 2000-2022 and the sfc - 300 hPa ozone column. To reconcile those trend differences between both studies, trends for the same time period (2008-2019) and for exactly the same metric might be compared (another comment posted by co-authors from our study might follow). W.r.t. this last comment, as you are looking at the time behaviour/trends of different tropospheric ozone metrics (sfc-tropopauze, sfc - 450 hPa, 450 hPa - tropopauze), as we did, although we used different metrices (sfc-300 hPa, sfc-700 hPa, 300-700 hPa), some reference or discussion w.r.t. our results might be appropriate.

We thank the HEGIFTOM co-authors for their valuable comments and for sharing their trend results over the 2008–2019 period. As suggested, we compared our updated trend estimates with the HEGIFTOM results at the 15 stations common to both datasets.

Our analysis shows that, at all 15 stations, the 2σ uncertainties exceed the magnitude of the estimated trends for both IASI and ozonesonde data. This reflects the challenge of detecting statistically significant trends over a relatively short 12-year period, especially when using satellite data.

HEGIFTOM reports statistically significant trends at seven stations, which differs from our findings. We attribute this primarily to differences in data sampling: our analysis uses ozonesonde profiles matched in time and space to IASI overpasses, whereas HEGIFTOM utilizes the full ozonesonde record without this constraint. These methodological differences likely explain much of the variation in trend significance and even direction.

We have included this comparison and interpretation in the revised manuscript to clarify the relationship between our results and those of HEGIFTOM.

In our HEGIFTOM trends paper, we compared and discussed the Payerne tropospheric ozone time series (also used DLM for it) quite extensively, so you might use of this information when discussing the drift between the Payerne ozonesonde and IASI time series

Thank you for the suggestion. The discussion of the Payerne tropospheric ozone time series in the HEGIFTOM paper was very helpful, and we have incorporated this information when addressing the drift between the Payerne ozonesonde and IASI time series.

Recently, as you are aware, Dufour et al. (2025) also looked at the comparison and (regional) trends of another IASI product, IASI-O3 KOPRA, with a limited sample of ozonesondes (Boulder, Payerne, and Uccle are common). You do refer to this study, but I would expect a more detailed comparison of the results of both studies! To my opinion, at least a paragraph might be needed explaining the similarities or the differences between the results of both studies.

Thank you for this suggestion. A more detailed comparison with Dufour et al. (2025) who analyzed IASI-O_3 KOPRA data from 2008 to 2022 using similar regional definitions is indeed relevant to contextualize our results.

For the surface-tropopause column, we find a negative trend with high certainty over Europe (–0.07 ± 0.07 DU/yr, p = 0.03), which aligns well with the trend reported by Dufour et al. (–0.05 ± 0.02 DU/yr, p = 0.03), also classified as high certainty. Over North America and Asia, both studies detect a negative trend, with medium to very low certainty

At the station level, we observe a consistent picture between the two studies. For the stations common to both datasets, trends are negative and statistically significant in all cases except Uccle, where both studies find a non-significant positive trend. The only notable discrepancy is at Boulder: Dufour et al. (2025) report a significant negative trend, while we find a non-significant small positive trend. This divergence may reflect differences in the retrieval sensitivity or sampling characteristics between the two IASI products at this high-altitude site.

Overall, both studies consistently depict negative trends in Europe wth high certainty and a lack of evidence for trends across Asia and North America.

We have included this discussion in the revised manuscript.

Also, looking at Fig. B1 (the constant a priori used in the IASI retrievals vs. the latitudinal and seasonal variable a priori used for CrIS), I wonder what the impact of such a time invariant (right?) a priori would have on the calculated trends.... The authors might comment on this.

We have conducted an additional analysis to investigate this point more deeply. Specifically, we applied the IASI a priori (fixed in latitude, not seasonally resolved) to the TROPESS retrieval algorithm, which normally uses a spatially and seasonally varying a priori. The results indicate that the choice of a priori profile clearly influences the retrieved ozone profiles. However, the differences observed between the IASI and TROPESS products are smaller than the differences between their respective a priori profiles. This is further illustrated by Figure 1 showing that the difference between the priors (right maps) is larger than the difference between the retrieved TROPESS and IASI columns (left maps). Additionally, the difference between the priors (right maps) is also larger than the difference observed when swapping the priors (middle maps). While the left and middle maps have comparable magnitudes, their spatial distributions differ to some extent.

These findings suggest that the a priori alone does not fully explain the discrepancies between the datasets. Moreover, the retrieval process is not strictly linear with respect to the a priori assumptions, indicating that a complete retrieval using the IASI a priori would be necessary to accurately quantify its impact (Kulawik et al., 2008).

We will update the manuscript to include this analysis and discussion and emphasize that differences in retrieval methodologies, such as the treatment of prior covariance matrices and the representation of vertical ozone profiles, likely contribute significantly to the observed inconsistencies.

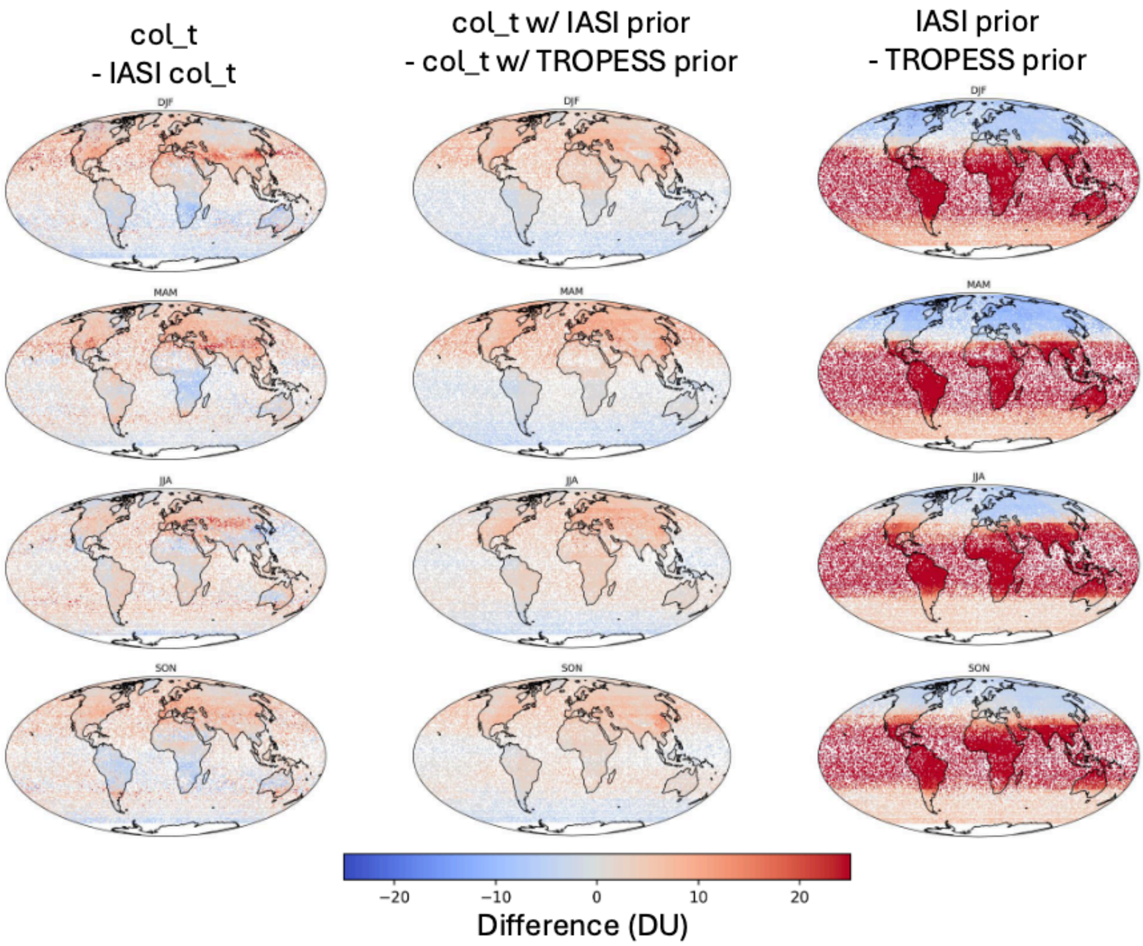

Figure 1. Spatial distribution of tropospheric ozone (Dobson Units) over the period 2016-2022: left: difference between CrIS (TROPESS) and IASI retrievals; middle: difference between CrIS retrievals using the IASI and TROPESS a priori; right: difference between the IASI and TROPESS a priori.

Finally, I am also quite surprised by your small trend uncertainties (e.g. in comparison with ours, which are quite consistent between the QR and MLR trend estimation tools used). You might want to provide some additional information on how exactly those have been calculated.

Trend calculations follow the guidelines established by the TOAR-II Statistics Focus Working Group (Chang et al., 2023b). Specifically, trends are estimated using quantile regression (QR) at the 50th percentile, and uncertainty is assessed using a moving block bootstrap method, which also calculates p-values to evaluate statistical significance. We used the toarstats Python package (https://gitlab.jsc.fz-juelich.de/esde/toar-public/toarstats). The monthly ozone column time series are first deseasonalized by fitting and removing a sine-cosine model with 12- and 6-month periodicities to account for seasonal variations. The relatively small uncertainty ranges reported in the manuscript correspond to ±1σ. To improve clarity and transparency, we now report uncertainties at the 95% confidence level (±2σ) in the revised manuscript.

Thank you for taking my comments into consideration and I wish you good luck in the review of your manuscript!

Thank you very much for your constructive and fair comments. Your balanced and respectful approach was greatly appreciated.

With kind regards,

Roeland Van Malderen

References:
Dufour, G., Eremenko, M., Cuesta, J., Ancellet, G., Gill, M., Maillard Barras, E., and Van Malderen, R.: Performance assessment of the IASI-O3 KOPRA product for observing midlatitude tropospheric ozone evolution

for 15 years: validation with ozone sondes and consistency of the three IASI instruments, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-4096, 2025.
Van Malderen, R., Thompson, A. M., Kollonige, D. E., Stauffer, R. M., Smit, H. G. J., Maillard Barras, E., Vigouroux, C., Petropavlovskikh, I., Leblanc, T., Thouret, V., Wolff, P., Effertz, P., Tarasick, D. W., Poyraz, D., Ancellet, G., De Backer, M.-R., Evan, S., Flood, V., Frey, M. M., Hannigan, J. W., Hernandez, J. L., Iarlori, M., Johnson, B. J., Jones, N., Kivi, R., Mahieu, E., McConville, G., Müller, K., Nagahama, T., Notholt, J., Piters, A., Prats, N., Querel, R., Smale, D., Steinbrecht, W., Strong, K., and Sussmann, R.: Global Ground-based Tropospheric Ozone Measurements: Reference Data and Individual Site Trends (2000–2022) from the TOAR-II/HEGIFTOM Project, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-3736, 2025.

Kulawik, S. S., Bowman, K. W., Luo, M., Rodgers, C. D., and Jourdain, L.: Impact of nonlinearity on changing the a priori of trace gas profile estimates from the Tropospheric Emission Spectrometer (TES), Atmos. Chem. Phys., 8, 3081–3092, https://doi.org/10.5194/acp-8-3081-2008, 2008.