SUMMARY COMMENTS AND SUGGESTIONS FOR REVISION

This Comment follows that of HEGIFTOM Co-Chair R. Van Malderen and amplifies his major points. We refer to the same "Van Malderen et al, in press, 2025" that he does and designate it as "HEGIFTOM-1" because there is a second HEGIFTOM paper in review in the TOAR II collection. We go beyond Van Malderen's comments to 3 summary recommendations:

- The selection and number of ground-based stations for which ozone data are compared to the IASI-CDR product is inadequate in number and 'non-homogenized' datasets should not be used. In contrast to the 7 stations used in this paper, more than 25 stations, homogenized over the 2008-2019 period, need to be used in evaluation of the IASI products and the trends. The 7 sites included no equatorial stations and far too few northern mid-latitude stations, in some cases where more than sonde data are available. The recommended ozonesonde 27 stations appear in Table 1 at the end of this Comment.
- The authors conclude with general, not well-defined speculation on why there is little progress in how computed trends from the new IASI products diverge from the UV-based product trends from the Gaudel et al. (2018) TOAR I paper. More analysis and insights on this issue are needed before the paper is worthy of publication in a quality Copernicus journal. Specific questions are raised for consideration.
- HEGIFTOM-1 is now "THE Reference dataset" for trends comparisons. The mismatch of years (2008-2019 in Boynard et al. vs 2000-2022 in HEGIFTOM-1) is speculated as one reason for why the trends in this paper differ from the ground-based trends in HEGIFTOM-1. We reran the trends in HEGIFTOM-1 for 27 stations for 2008-2019, see Table below, to support a valid comparison. Some of the HEGIFTOM-1 site trend signs changed and uncertainties increased, leading to a clear TOAR-worthy conclusion that trends computed from 12 or 16 years of IASI or ground-based data is inadequate. A revision must include this important result!

Thank you for your kind and constructive feedback. We have carefully considered your comments and made the necessary clarifications and corrections in the revised manuscript. Please find a point-by-point answer below.

SYNOPSIS – This study presents an IASI product over 16 years, consisting of contributions from METOP-A, METOP-B and METOP-C, merged to create the IASI-CDR (Climate Data Record, 2008-2023). Several tropospheric ozone columns (to 450, to tropopause) are presented and compared on a monthly mean basis with (1) the CrIS IR ozone product and with (2) comparable ozonesonde columns from 43 stations over the period. Trends for global mid-latitudes and the tropics are computed using Quantile Regression (QR) for the period 2008-2019 (pre-COVID) and 2008-2023; the latter trends reflect an apparent COVID impact. As in the TOAR I paper in which Gaudel et al. (2018) summarized satellite product trends (2005/2008 – 2016) showing IASI (FORLI version) to be an outlier compared to UV-type satellite records, it appears that the IASI-CDR trends (now 2008-2019, omitting COVID period) is an outlier with the corresponding UV-based time-series. In both cases the greatest discrepancies are in the tropics except for SE Asia where all products display increases of ~1-2 DU/decade. The UV-type satellite products tend to be more variable with regions in the tropical Americas, Africa and Atlantic showing some increases (~2005-2019; Gaudel et al., 2024; Thompson et al., 2024). The IASI-CDR TOAR II-period (Fig. 12 in paper) shows little increase over Europe, a decrease over North America and only modest increases in east Asia, again in disagreement with more variable UV-type satellite trends. IASI 2008-2019 trend comparisons are made with 7 ozonesonde time-series: one subtropical site, 4 northern mid-latitude (majority European), 1 southern mid-latitude (Fig. 16). The authors speculate briefly on causes of the persistent discrepancy between the IASI vs UV trends and what is emerging as the prevailing view of tropospheric changes over the prior ~15 years. Comparisons among comparable IR products eg. IASI-CDR, IASI-KOPRA, and CrIS global ozone products are described with varying degrees of detail or with Figures that are suggestive but not conclusive.

OVERALL COMMENT – The paper poses good questions, presents a reasonable approach for its calculations and selection of results and is well-arranged. However, it leaves too many unanswered questions and does not advance the scientific understanding of its ozone trends beyond the first TOAR study. The three most important aspects of the paper that require additional analysis are summarized as follows:

1>> Quality assurance and evaluation of the IASI-CDR. There is reference to a larger set of sonde stations used for comparisons (43 stations; Fig 8 and 10) but Table 2 is inaccurate. The following stations are not available on HEGIFTOM archive because they are not homogenized datasets: Lindenberg, Prague, Tateno, Hong Kong, Broadmeadows, and Macquarie. We recommend using only homogenized datasets for reference datasets and trends. Comparison of trends is restricted to 7 stations, none truly tropical, 15 degrees or less, despite the fact that a number of cited and other TOAR II studies, both satellite and ground-based, focus on the tropics e.g., Froidevaux et al., 2025; Gaudel et al., 2024; Thompson et al., 2025. Time-series comparisons in Fig. 16 show only 7 stations with limited geographical coverage.

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.

Concerning the evaluation of IASI-CDR:

- Extensive discussion of CDR vs earlier FORLI product appears – but no illustrations of why you expect the CDR product to perform better than the earlier one.

While we did not explicitly claim that the IASI-CDR product performs better in all respects compared to the earlier FORLI product, the key improvement lies in its homogeneity and consistency over time, making it more suitable for long-term trend analysis. As shown in Boynard et al. (2018), the IASI-FORLI dataset suffers from inhomogeneities and artificial drifts caused by changes in processing versions. The IASI-CDR product was developed specifically to mitigate these artifacts by reprocessing radiances and auxiliary data consistently over the full record.

To assess the drift of IASI ozone measurements, we examined two sets of criteria:

- Original criteria (criteria 1): at least 3 soundings per month, 70% data coverage (from initial study)
- Updated criteria (criteria 2): at least 2 soundings per month, 50% data coverage (based Van Malderen et al. (2025) selection criteria).

The drift shifts from -1% per decade with criteria 1 to -2.5% per decade with criteria 2. While this change is noticeable, both estimates remain well below the 3% per decade threshold, indicating that the drift is still within an acceptable range.

The increased drift under Criteria 2 is likely due to the inclusion of more stations, which could introduce variability, especially at stations with gaps or incomplete time series. When the stricter Criteria 1 (3 soundings per month and 70% coverage) is applied, the drift returns to -1% per decade, which highlights how the drift estimate is sensitive to the selection criteria.

In contrast, the earlier FORLI product exhibited significant artificial drifts, particularly around 2010, due to changes in processing versions (Boynard et al., 2018).

These results highlight the improved consistency and homogeneity of the IASI-CDR product, making it a more reliable choice for long-term trend analysis compared to the earlier FORLI version.

- IASI-KOPRA product mentioned but why is there no extensive comparison with this product or at least a paragraph comparing results of Dufour et al. (https://egusphere.copernicus.org/preprints/2025/egusphere-2024-4096/) to those shown here, particularly for sonde comparisons?

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 with high certainty and a lack of evidence for trends across Asia and North America.

We have included this discussion in the revised manuscript.

- Vertical discrepancies mentioned in comparisons with sondes (tropical, mid-latitudes, and polar) appear in Fig. 10. Although overall IASI column amounts are compared favorably to the sondes (Line 22, only 2% offset in tropics) can this be due to a cancelling of offsets illustrated in the Figure? Discuss the potential impact on the trends.

The vertical discrepancies observed between IASI and ozonesonde profiles not inherently bias trend estimates, provided these vertical offsets remain stable over time. Our drift analysis of the tropospheric column reveals small temporal drift during the study period, indicating that the magnitude and vertical distribution of these biases have remained reasonably constant. Therefore, while vertical compensation explains the small differences in absolute column amounts, it is unlikely to distort the long-term tropospheric ozone trends derived from IASI.

- The CrIS-TROPESS a priori looks so much better than the IASI (varies with season and latitude). Although the IASI climatology (Fig. 7) looks reasonable, can the IASI a priori (Fig. B1) – with apparently little seasonal information and only latitude dependence- be a cause of the discrepancies with other products? Does inadequate representation of seasonality (monthly variations in trends are typical and significant in the tropics, for example: Stauffer et al., 2024; Thompson et al., 2021) propagate to trends that disagree with both sondes and UV-products? What additional insights can you derive from scatterplots with ozonesondes (Figs. A1-4)? Comparing to sondes seasonally can help identify discrepancies. There is extensive discussion of similarities and differences with CrIS but it doesn't get to the crux of understanding the large negative ozone trends here that are at odds with other products.

We thank the reviewer for this insightful comment. While the IASI retrievals use a climatological a priori profile (latitude-dependent, non-seasonal), our results show good agreement with CrIS-TROPESS, which uses a spatially and seasonally varying a priori. This suggests that the simplified IASI a priori is not the dominant driver of discrepancies with other products.

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 have revised the manuscript to include this analysis and discussion, emphasizing 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.

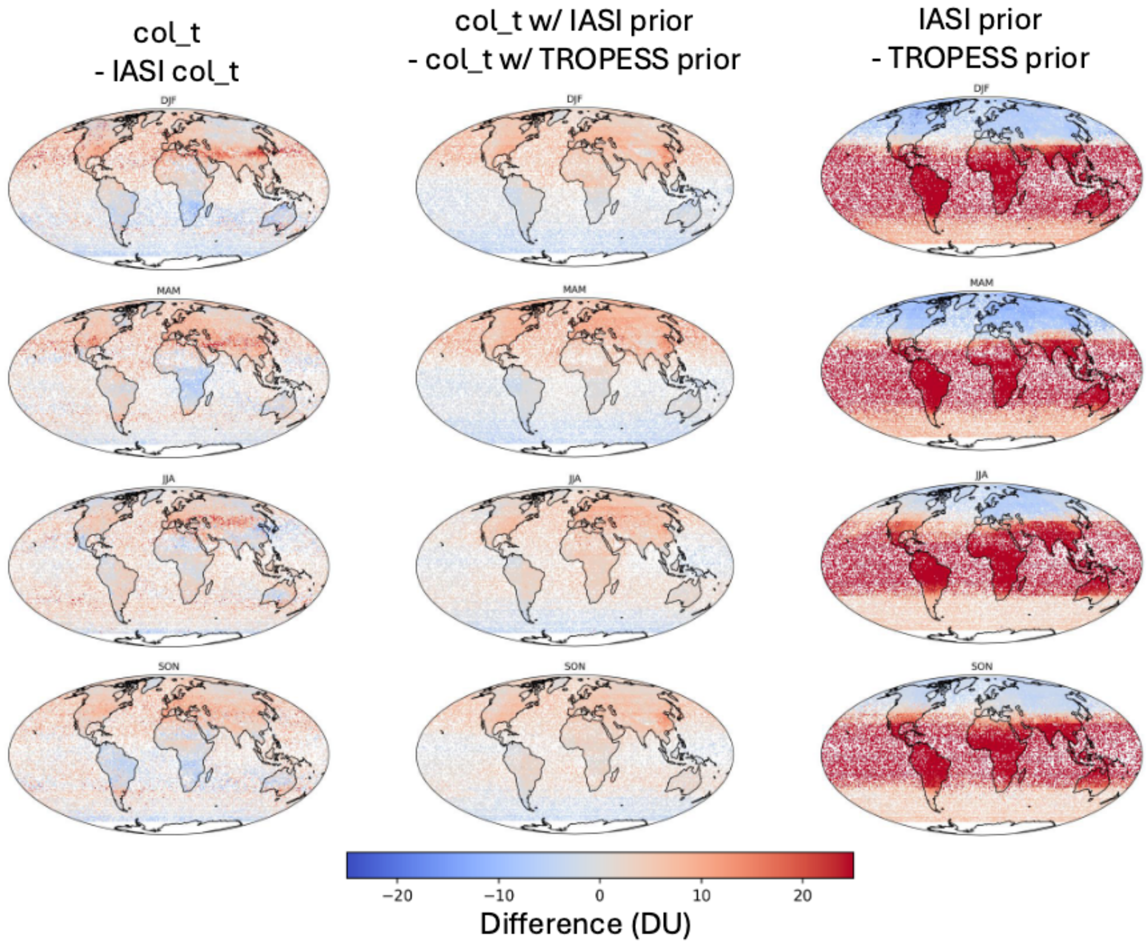

Figure 1. Spatial distribution of tropospheric ozone (Dobson Units) over the period 2016-2022: left: difference between CrIS (TROPESS) and IASI retrievals; middle: difference between CrIS retrievals using the IASI and TROPESS a priori; right: difference between the IASI and TROPESS a priori.

Figures A1–A4 present scatterplots between IASI and ozonesonde data for four vertically integrated subcolumns (surface–300 hPa, 300–150 hPa, 150–25 hPa, and 25–3 hPa), chosen to reflect the vertical sensitivity of the retrieval. These comparisons, which aim to validate the four subcolumns defined in Boynard et al. (2018) with the maximum sensitivity, show overall consistency (with biases typically <20%). However, these comparisons are not broken down by season. We agree that a seasonal comparison could provide further insight into discrepancies and trend behavior, and this will be explored in future work.

Finally, we note that our trend results align well with those from Dufour et al. (2023), reinforcing confidence in the robustness of our approach.

2>> Inadequate number of ground-based (GB) reference sites. There are two aspects of this issue.

First, the authors show 43 potential stations (Table 2, Fig. 8) but make trend comparisons with only 7 ozonesonde data sets. Perhaps they don't understand the fine points of which data are appropriate to use (see Van Malderen comment). There is no need to speculate, as the authors have done, on why their results do not resemble those from mostly UV sensors. All satellites now have the HEGIFTOM-1 trends at individual stations (some with

multiple instruments, not only sondes) as the gold standard independent reference at 55 sites total for 2000-2022. The HEGIFTOM reprocessing includes references of the data for each instrument type (sonde, FTIR, UV Umkehr) to a global absolute standard. Your trends analysis should add at least 20 stations (exclude polar sites where IASI-CDR struggles with DOFS) to have a more representative picture of IASI performance. Note that of the 7 reference sites (Boulder, Hilo, Lauder), there is more than one GB record for comparison. The stations in the Table below have sufficiently temporally dense records for comparison.
Second, Line 330 states that months with only 1-2 sondes/month give 'inadequate' results for 50%-ile trends. In the accepted version of HEGIFTOM-1, it is shown that these trends (computed with QR or MLR) are unaffected by cutting from 4-5 sondes/ month to 2; only the uncertainty changes (increases). That is a second justification for using more sonde locations for the GB comparisons - candidates in the Table.

Please see answer above, which indicates that in the revised manuscript, 27 stations are used for the assessment of the trends.

3>> Brevity of the IASI record is a concern for 2008-2019 trends. For comparison, we re-ran the QR trends for the authors' selected 7 stations as well as 20 other HEGIFTOM reference ozonesonde stations (excluding polar sites). The results are listed in the Table below, which also includes results from the HEGIFTOM-1 2000-2022 trends for surface-300 hPa (in ppbv/decade) as well as XO3 (ppbv/decade) and DU (DU/decade) for surface to tropopause tropospheric columns as a reference.

Note for the recalculated 2008-2019 ozonesonde trends, that 4 out of the 7 author selected stations change sign of trend from the longer time series to the shorter time series. For example, Boulder, a station with a high certainty ($p<0.05$) associated with a negative trend for the longer time series (2000-2022), has a slightly positive trend with large uncertainties for the 2008-2019 period. Of the 27 ozonesonde stations listed in the table, 9 sites have trend sign changes. Discussion on the point of reduced reliability and value of the shorter time series (12 or 16 years vs 23 years in HEGIFTOM and sonde studies*) is needed and is now a view that can be made with confidence. In a sort of "reversal" of your paper's message, a significant advance and outcome of your paper, with the contracted (2008-2019) HEGIFTOM calculation, is that TOAR II needs to recognize the limitations of datasets that cover fewer than ~20 years!
The uncertainties in the trends also increase with the shorter time series (ie. double those for the 2000-2022 time period – see Table below).

Thank you for highlighting the critical role of time series length in trend analysis. We fully agree that longer records, such as the 23-year HEGIFTOM dataset (2000–2022), offer more robust and stable trend estimates with reduced sensitivity to interannual variability.

Our primary trend analysis is based on the 16-year period from 2008–2023, which we consider sufficiently long to yield meaningful and statistically robust results. The shorter 2008–2019 window was included specifically to examine potential short-term impacts of the COVID-19 pandemic. As expected, trends over this shorter period show larger uncertainties and greater sensitivity to variability, including sign changes, as also observed in your HEGIFTOM recalculations.

To assess sensitivity to time series length, we expanded our analysis to 27 ozonesonde stations and compared trends for both 2008–2019 and 2008–2023. Results show that extending the record length significantly reduces trend uncertainties across most stations.

While we agree that datasets shorter than ~20 years must be interpreted with caution, especially in isolation, we believe that 16 years strikes a reasonable balance between capturing recent variability and ensuring trend reliability. Moreover, the general consistency between IASI trends and those from ozonesondes and other studies (e.g., Dufour et al. 2025) lends confidence to our findings.

We appreciate your observation that our results underscore the limitations of shorter records, a conclusion we fully support and that further motivates the need for sustained long-term observations.

On your paper (also noted by R. Van Malderen) the reported uncertainties in Fig. 16 seem small for only 12 years of data. Can you check those and discuss your bootstrap method in more detail?

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 a moving block bootstrap method is used to assess the uncertainty of the derived trends and to calculate p-values for evaluating their 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 eliminate seasonal variations. The relatively small uncertainty ranges reported in the manuscript correspond to uncertainties expressed as ±1σ. To improve clarity and consistency, we have revised the manuscript accordingly and now report uncertainties as ±2σ.

*In addition to the sondes studies you have referenced, Stauffer et al., 2024; Thompson et al., 2021; Thompson et al., 2024, there is an excellent new sonde trends paper submitted on Réunion SHADOZ and SAOZ time-series (1998-2021) submitted to Earth and Space Science: https://essopenarchive.org/doi/full/10.22541/essoar.174594999.98715985/v1

'HEGIFTOM-1" below is posted. Final version is in press.
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.

Also referred to is Thompson et al., submitted, 2024, Posted in 2025:
Thompson, A. M., Stauffer, R. M., Kollonige, D. E., Ziemke, J. R., Cazorla, M., Wolff, P., and Sauvage, B.: Tropical Ozone Trends (1998 to 2023): A Synthesis from SHADOZ, IAGOS and OMI/MLS Observations, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-3761, 2025.

TABLE. This is essentially an update of Table 1 in HEGIFTOM-1, Van Malderen et al., in press, 2025, run again with the QR method, same as employed in Boynard et al. Ozonesonde stations with homogenized data and sufficient sample size are listed and exclude near-polar regions. Yellow-coded lines represent the 7 author-selected IASI-sonde comparison stations. Orange-coded lines indicate where the sign of the trend changes based on the different periods of trends calculations (23 years vs. 12 years).

TABLE IS POSTED IN SUPPLEMENT

Citation: https://doi.org/10.5194/egusphere-2025-1054-CC3IASI-Boynard Paper Comment – 13 May 2025
Anne Thompson, Debra Kollonige, Ryan Stauffer

References

Boynard, A., Hurtmans, D., Garane, K., Goutail, F., Hadji-Lazaro, J., Koukouli, M. E., Wespes, C., Vigouroux, C., Keppens, A., Pommereau, J.-P., Pazmino, A., Balis, D., Loyola, D., Valks, P., Sussmann, R., Smale, D., Coheur, P.-F., and Clerbaux, C.: Validation of the IASI FORLI/EUMETSAT ozone products using satellite (GOME-2), ground-based (Brewer–Dobson, SAOZ, FTIR) and ozonesonde measurements, Atmos. Meas. Tech., 11, 5125–5152, https://doi.org/10.5194/amt-11-5125-2018, 2018.

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