Quality Assessment of the AC SAF GOME-2 gridded ozone profile data records – author responses

Olaf Tuinder, Peggy Tesche-Achtert and Andy Delcloo

16 February 2026

In this author response, the text in black are the comments from the reviewer, in green are the answers from the authors.

1 Response RC#1, from Arno Keppens

We thank the reviewer for his comments and have attempted to answer them below as best as possible.

1.1 Specific comments

Line 4-5: Add that also the gridded uncertainty or covariance matrix is stored, unless this is not the case. The user community would at least expect it.

In the dataset produced for this paper, the covariance was not stored in the L3 data. We will add this to our operational L3 data generation, and investigate the option to retro-actively add this to the existing dataset.

References are missing for several statements made in the first two paragraphs of the introduction, and for the Metop satellites and GOME-2 instruments in lines 41-44.

Appropriate references have now been added to support the general statements made in the first two paragraphs of the introduction regarding the importance of atmospheric composition monitoring and the role of ozone. In addition, references describing the Metop satellite series and the GOME and GOME-2 instruments have been included in the corresponding section of the manuscript. These additions provide proper context and documentation for the satellite missions and measurements discussed.

The importance of tropospheric ozone (monitoring) is only mentioned in line 40, almost as a side note. The author could better motivate their tropospheric ozone column assessment.

The introduction has been revised to better motivate the importance of tropospheric ozone and its monitoring. Additional text has been added to highlight the role of tropospheric ozone as an air pollutant and greenhouse gas, its relevance for air quality and climate, and the need for continuous and spatially resolved observations. These additions provide clearer context for the tropospheric ozone column assessment presented in this study (Lines 38-44).

If the authors aim at demonstrating the importance of the GOME-2 instrument for atmospheric ozone monitoring in lines 49-62, it seems remarkable that RAL's GOME-2 ozone profile retrieval within ESA's Ozone CCI is missing.

The manuscript has been revised to explicitly acknowledge the retrievals developed within the ESA Ozone Climate Change Initiative, including the RAL retrieval. A short paragraph has been added to place the Ozone CCI products in the context of other long-term GOME-2–based ozone profile records and to clarify their complementary role with respect to the ACSAF gridded ozone profile products investigated in this study (Lines 60–67): The text was changed: The ESA Climate Change Initiative (CCI) and RAL's efforts to construct a homogenized climate data record using multiple sensors and retrieval algorithms are mentioned. Given that the AC SAF GOME-2 gridded ozone profile products are generated within an operational framework with a strong focus on near-real-time processing, long-term stability, and continuity for user-driven applications, both efforts are complementary and jointly contribute to the comprehensive use of GOME-2 measurements for atmospheric ozone monitoring.

In Section 2, references seem to be missing for the offset on line 89, Optimal Estimation on line 92, an assimilated ozone field on line 108 (different from the climatology, which is referred to earlier), and the remarks in Table 1.

References added. One reference to the ATBD for the offset, one to Rodgers (2000) for the OE methodology, one to Eskes 2003 for the Assimilated Total Ozone. We don't see what reference would be missing for Table 1.

Section 2.3 on averaging kernels is too vague in terms of terminology and methodology to be clear. The authors frequently mix (also throughout the text) "averaging kernels" with the "averaging kernel matrix" that combines all retrieval-layer-specific averaging kernels (as the AK matrix rows, uncommonly called "weight(s) curves" with Fig. 1).

An Averaging Kernel is not an easy concept. We see averaging kernel (AK) of an optimal estimation inversion as a matrix containing the weights of the linear effect of a change in the fitted parameters on the retrieved value at a specific position in the state vector. You could make the argument that one individual fitted parameter has an averaging kernel (as a vector), but then how is the full matrix called? Equation 1 treats the AK as a matrix. We've tried to harmonize the terminology throughout the paper.

The authors should be more specific in lines 119-121: What does "the averaging kernel [matrix] cannot be inverted" mean, as one could in theory easily use a pseudo-inverse?

The averaging kernel is a singular matrix. Based on your suggestion we investigated the use of the pseudo-inverse on the averaging kernel using the linalg.pinv function that is available in numpy. This resulted in a checkerboard pattern with non-realistic values. Also the verification of the inverse operator (the matrix multiplication with the original averaging kernel) did not produce the diagonal unity matrix. Our conclusion is that, at this moment, the AK can not be inverted. We updated the text to reflect this.

In addition, why can a reference data set with a high vertical resolution easily be compared to a satellite data set, and not the other way around?

We should distinguish high vertical sampling from high vertical resolution. We define the sampling (rate) as number of measurements per vertical unit or time. The vertical resolution is the local representativity of the measurement. If the reference data set has a very limited local vertical representativity, i.e. the measurement responds only to local changes to the concentration of the measured parameter and is not dependent on concentrations elsewhere in the atmosphere, then the parameter has a high resolution.

In order to compare two data sets they need to be brought on the same sampling and resolution. It is easier to go from measurements with a high vertical sampling with a high vertical resolution (like balloon ozone sondes) to a 'measurement' from satellite by calculating an average balloon sonde value that matches the lower vertical sampling of the satellite and applying the (broader) Averaging Kernel from the satellite. The other way, from the satellite to the balloon sonde does not work because there broader satellite sampling and resolution can not be compared to single high resolution balloon sonde values.

The averaging kernel smoothing approach that is applied later in the text should be explained with reference to Eq. (1).

The description of the averaging kernel smoothing has been revised to explicitly reference Eq. (1) and to clarify how the satellite averaging kernels are applied to the ground-based ozone profiles. Additional text has been added to explain the smoothing operation and its role in adjusting the ground-based profiles to the vertical sensitivity of the satellite retrievals.

The authors should add to that what happens to the lower-resolution FTIR data, which are the result of a sort of retrieval themselves and hence come with averaging kernels as well. Is AK smoothing applied on the satellite data or bi-directionally then, as suggested in the literature?

FTIR ozone profiles are retrieval products with their own averaging kernels and a priori information. In the revised version of the manuscript, we explicitly account for this by applying a bidirectional averaging-kernel treatment following Rodgers and Connor (2003). Specifically, the satellite averaging kernels are applied in the Dobson Unit (DU) domain to the FTIR profiles, while accounting for the FTIR a priori and vertical sensitivity. This approach ensures that differences between satellite and FTIR data are not artificially introduced by inconsistent treatment of vertical resolution. For high-vertical-resolution reference data (lidar and ozonesondes), only the satellite averaging kernels are applied, as these measurements can be considered close to the true atmospheric state. The manuscript has been revised accordingly to clarify the treatment of FTIR data and to ensure methodological consistency with established practices in intercomparison studies (e.g., Calisesi et al., 2005).

Lines 128-129: Do you mean that each retrieved pixel is divided into 10 by 10 km sub-pixels who's centres need to be within a grid cell to be included? This is not fully clear.

That is correct. For a nominal pixel size (80 × 40km, the sub-divion in 8 × 4 leads to sub-cells of 10 × 10 km. The centers of these sub-cells are projected into the grid and the data is handled within the cell in which they fall. This is updated in the text.

Line 147 refers to the retrieval (co)variance (matrix?) and error (profile?), while these are not introduced in Section 2.

The discussion of weighted mean was removed, as the only values we used in this study in the validation of the ozone profiles are the Arithmetic Mean.

Lines 185-186 and Figure 2: Only the AK matrix dispersion (in terms of interpercentiles) at a specific location is shown, while upon constructing averaged averaging kernels for entire latitude bands, one would rather be interested in the AK matrix dispersion within each entire band. Showing this for all 18 bands might be too much for the main text, but should be considered as a supplement, especially given the known non-triviality of applying mean averaging kernels, as discussed in the work of von Clarmann and Glatthor (https://doi.org/10.5194/amt-12-5155-2019), which the authors should introduce in their analysis.

In appendix of the updated paper, the 10, 50 and 90 percentile for the averaged Averaging Kernels for all 18 latitude belts are shown. These values are based on the AK's of all converged ozone profile retrievals of GOME-2A in the a particular latitude belt in 2009. In Figure 2 we now show the subset of the 10, 50 and 90 percentile for the $40° - 50°N$ latitude belt based on the re-calculation above. These new plots are very similar to the previous ones obtained from only the Uccle station colocation.

Fig. 3: Plot titles, legends, and axis labels should be made more brief and clear.

Plot titles, legends, and axis labels have been modified. They are now shorter and more clear.

The plotting colours do not seem to correspond to what is indicated in the legend, assuming that M01 to M03 corresponds to Metop-A to Metop-C, respectively.

Yes, M01 is GOME-2B and M03 is GOME-2C. We checked the colors, and they seem to correspond to the correct instruments in the legend.

It should be clarified in the text how the retrieval degrees of freedom are obtained and why two modes appear in the DFS of GOME-2A before 2013.

The Degrees of Freedom is calculated as the trace of the AK matrix. This is now mentioned in the text. The DFS of GOME-2A has a discontinuity around September 2009, when an instrument de-contamination attempt caused a permanent degradation of the transmission and sensing of light through the instrument and detector. This is now mentioned in the text.

Given the coarse (effective) resolution of FTIR observations, one would expect Table 3 to include FTIR observation specifics.

Table 3 has been revised to include typical characteristics of FTIR ozone profile observations. The effective vertical resolution and typical precision of FTIR retrievals, reflecting their limited number of degrees of freedom, are now documented and referenced. This addition provides a more complete overview of the different ground-based measurement techniques used in the validation and better reflects the differing vertical sensitivities of the instruments (Table 3, Lines 255–258).

Section 4.1: Only 20 ozonesonde sites are used for validation, while the EVDC (also hosted at NILU, including WOUDC and NADIR data) covers 60 ozonesonde stations providing data in the period under study. Moreover, within the TOAR-II HEGIFTOM working group, the data of 40 ozonesonde stations has been homogenised for long-term analyses. The authors should hence justify or correct for their (limited) ozonesonde data selection.

We are using ground based datasets which need to be available for the whole time series (2007-2022) and we wanted the L3 reprocessed dataset to be validated against a consistent set of reference stations, we consider including the full TOAR-II HEGIFTOM out of scope for the current study. In the future when all ground based reprocessed datasets are available from TOAR-II HEGIFTOM, we can consider this.

Lines 268-269: The authors' phrasing is too brief to be clear here. What does it mean to apply Eq. 1 if that does not contain a reference profile? The term (also throughout the text) "(AK) retrieved ground-based" data is very confusing. Some ground-based reference data are sort of retrievals by themselves. What the authors refer to are "ground-based data corrected for vertical smoothing difference errors" which is commonly abbreviated as "smoothed ground-based data" but not "retrieved".

We agree that the term "(AK) retrieved ground-based" was misleading. Equation 1 is applied to the original ground-based ozone profile (lidar, MWR, and ozonesonde) using the satellite a priori profile and averaging kernels in order to account for differences in vertical smoothing between the satellite and ground-based observations. The resulting profiles therefore represent ground-based data smoothed to the vertical resolution of the satellite retrieval. Throughout the revised manuscript, the terminology has been corrected accordingly, and we now consistently refer to these data as "smoothed ground-based profiles". The corresponding description around Lines 268–269 has been expanded to explicitly clarify the application of Eq. 1 and the meaning of the resulting profiles.

Sections 5 and 6: It is unclear why the authors leave a gap in the vertical extent of their analysis, as seen from the unmatched altitude ranges in the second and third columns of Table 4. Moreover, Sections 5 and 6 do no longer refer to FTIR data. Where are they in the analysis?

FTIR observations are included in the analysis of Sections 5 and 6 as part of the ground-based station network and contribute to the band-integrated partial columns and time series in the upper stratosphere.

Is this only as part of Table 5, based on all reference stations shown in Fig. 4? The authors should provide an indication of the consistency of the validation results between the different reference data sources.

Table 5 is based on all available ground-based reference stations shown in Fig. 4 and includes contributions from lidar, microwave radiometer (MWR), FTIR, and ozonesonde observations, depending on data availability in the respective altitude ranges. This has now been clarified in the manuscript. In addition, we have now performed a station-to-station consistency analysis for the different reference sites, comparing satellite ozone partial columns with lidar, FTIR and microwave radiometer (MWR) observations. The manuscript has been updated accordingly to clarify these results. (Line 320 to 335).

Finally, the authors should discuss their findings with respect to existing GOME-2 ozone profile retrieval products and uncertainties, both their own Level-2 input as third-party Level-2/3 data.

The aim of this study is to specifically verify and validate the ACSAF level-3 ozone profiles against ground based reference data. We do not include validation of third party ozone data (like the RAL and C3S L2 and L3) this would widen the scope of the study too much. Also the validation if the C3S products is done inside the C3S project itself.

Fig. 5: The plot layout in Figs. 6 and 7 is much more clear than in 5, which seems to have pairwise identical plots in the first column, and the L3 a priori profiles (green lines) missing.

Figure 5 has been regenerated with additional lines.

Fig. 8: Why is the tropospheric column not included? Please provide an indication of the corresponding relative differences. Given that L3 data are considered, global maps of differences over the entire time series could be very insightful too.

Figure 9 has been created showing the tropospheric time series of the difference.

Data references: The data availability statement should also include the ground-based reference data. FTIR data are not mentioned in the author contributions. Only the NDACC network is acknowledged, while other network sources appear in the text.

The data availability statement has been revised to explicitly include the ground-based reference data used in this study, namely ozone lidar, microwave radiometer (MWR), FTIR, and ozonesonde observations obtained from the NDACC archive. In addition, the acknowledgements have been updated to explicitly mention FTIR data and to clarify that all ground-based reference measurements were provided through the NDACC network. The contributions of the NDACC station Principal Investigators are now more clearly acknowledged.

1.2 Technical corrections

Throughout the text, acronyms should be spelled out upon first usage, e.g., SAF, GOME, FTIR, DU, and STDEV in the abstract.

The authors feel that the abstract should be short and concise, and is not the place where all acronyms should be spelled out. In the introduction and elsewhere in the paper we try to explain these terms and acronyms as much as possible the first time it is used.

Abstract: Repetition of lines 13 and 18.

The duplication in the text is removed.

Line 15: “In the lower…”

Corrected.

Line 36: "ter-molecular interactions" looks like a typo.

We rephrased the description of the ozone production mechanism to avoid potential ambiguity in the terminology, while preserving the physical meaning. It is now described as the production of ozone through three-body molecular recombination.

Adding references to the subsequent manuscript sections in the last paragraph of the introduction would improve readability.

We have added a number of references to the introduction. The last paragraph describes the contents of the paper and the authors do not see what references would be needed in this section.

Line 87: "ozone partial column density values (in DU)" contains a contradiction between the quantity and the units.

Corrected, 'density' removed.

Use appropriate quantities and symbols in Eqs. 2, 6, 7, 8, 9, 10, 11, and 13.

The equations in the gridding section are intentionally left unit-less, as they are also apply to other parameters in the Level-3 files. The equation of the relative difference is adapted to better indicate the ozone from the gridded data and the ozone from the reference instruments, smoothed with the averaging kernel.

"seen" in Table 2 probably refers to the number of sub-pixels being included?

That is correct. However, we changed it to 'number of sub-pixels used', to prevent anthropomorphizing an algorithm.

Where only layer numbers are mentioned, like in line 174, please add the corresponding altitude or pressure.

Pressures were added to the layers mentioned in this section.

The coloured legend in the central panel of Fig. 1 is insufficiently clear to leave it to that.

We understand that the central legend is not readable printed on paper, but for readers that use the PDF version of the paper can zoom in all the way and see the relation of the individual layer numbers associated with their nominal pressures. We could remove the legend, but we prefer to keep it this way to provide this detail to the reader of the digital version of the paper.

Quantify the "relatively small differences" of line 182.

Johnson et al. (2018) have investigated the effect of pre-computed representative Aveaging Kernels in a wide variety of cases. They find a maximum difference $< 5 - 10\%$ across the whole profile for all cases they tested. This is now mentioned in the paper.

Caption of Fig. 1: Does "at 45° N" mean for the band going from 40° to 50° N?

That is correct. We added the following text: '(i.e. the $40°N - 50°N$ band)'.

It would be helpful to add the diagonals in the colour plots of Figs. 1 and 2 to guide the eye.

Dashed lines on the diagonals have been added in Figs 1 and 2 to guide the eye of the user.

Line 223: Wrong usage of tildes.

The tildes are actually the symbol for 'similar', indicating a value around the given number.

Line 240: "profile" instead of "spectrum"?

Thanks for spotting that, spectrum was changed to profile

Lines 254-255: Do you mean comparing mean reference data with the Level-3 grid cell that overlaps with the station location?

We compare not just the overlap of the L3 grid cell but all in a 100 km radius. The comma's in the text have been moved. The text now reads: "For each station, we computed a monthly mean ground based ozone profile, and compared it to the monthly mean level-3 ozone data obtained with a horizontal distance smaller that 100 km around the location of the ground-based stations."

Captions of Figs. 6 and 7: Explain the meaning of r in the second column.

We clarified that r denotes to the correlation coefficient between the level-3 satellite and smoothed ground-based time series indicated

Line 357: In "use of merged Metop GOME-2 datasets" it is unclear whether these already exist to the community.

A merged (and homogenized) data sets where no distinction can be made any more between original instruments does not exist yet. But given that the GOME-2 ozone profiles are good quality across the study time period, this is an option to consider as a Climate Data Record.

References somehow do not appear in alphabetical order.

Thanks for spotting that. References are now alphabetically, and with authors with multiple papers they are ordered by year.

2 Response to RC#2, from Juseon Bak

We thank the reviewer for her comments and have attempted to answer them below as best as possible.

2.1 Specific comments

1. L298 : Please better specify about "a seasonal dependency present in the dataset" and this was also the case in the level 2 product".

We describe the behaviour of the level-2 ozone profiles in many validation reports of the AC SAF concerning the ozone profile product (https://acsaf.org/valreps.php). There is a significant seasonal behaviour present in the dataset, which makes it necessary to do the validation over a full year, in order to exclude a bias because of this behaviour.

2. L313 said "30-50 km range, where the application of the averaging kernel is essential due to the limited vertical sensitivity of nadir-viewing instrument". In general, the nadir-UV observations better perform for retrieving stratospheric profiles than tropospheric profiles (OMI/Liu et al 2010, TROPOMI/Keppens et al. 2024), due to the Rayleigh scattering-induced vertical sensitively. Liu et al. (2010) also show the less importance of applying averaging kernels for evaluating the stratospheric column ozone.

-Keppens et al.: 5 years of Sentinel-5P TROPOMI operational ozone profiling and geophysical validation using ozonesonde and lidar ground-based networks, Atmos. Meas. Tech., 17, 3969–3993, https://doi.org/10.5194/amt-17-3969-2024, 2024.

-Liu, X et al. Validation of Ozone Monitoring Instrument (OMI) ozone profiles and stratospheric ozone columns with Microwave Limb Sounder (MLS) measurements, Atmos. Chem. Phys., 10, 2539–2549, https://doi.org/10.5194/acp-10-2539-2010, 2010.

We agree that nadir-viewing UV instruments generally provide good sensitivity to stratospheric ozone due to Rayleigh scattering, and that averaging-kernel effects are less critical for evaluating broad stratospheric column ozone, as shown by Liu et al. (2010) and Keppens et al. (2024). In our study, the application of averaging kernels is motivated by the comparison of vertically resolved ozone profiles and partial columns in relatively narrow altitude ranges (30–40 km and 40–50 km) with high-vertical-resolution ground-based observations. In this context, differences in vertical smoothing between satellite and ground-based measurements can significantly affect the comparison, even in the stratosphere. The manuscript has been revised accordingly to clarify this point and to better reflect the role of averaging kernels in profile-based validation (Line 351 to 358).

1. Figure 6. The correlation between L3 and ozonesonde is negative for column amount of 40-50 km, consistently from GOME A/B/C. The time-series clearly show some shift. This result should be deeper addressed. I believe, it could be rooted from satellite a priori data.

We indeed observe a systematic phase shift between the satellite and ozonesonde time series in the 40–50 km altitude range, which leads to a negative correlation despite a generally good agreement of the mean vertical profiles. This behavior can be explained by the combination of (i) narrow altitude-band integration, (ii) occasional crossings of the satellite and ground-based profiles in this region, and (iii) the increased influence of the apriori profile in the upper stratosphere. A seasonal phase mismatch in the apriori representation can therefore result in an apparent anti-correlation when compared to high-vertical-resolution ozonesonde observations. We have updated the test on Line 361 to 372 to better explain the observed negative correlation.

2. We believe that a L3 product intended for long-term applications should primarily be evaluated in terms of its capability to reproduce robust trends, rather than focusing on systematic mean biases. Please include a quantitative assessment of trend consistency between the ozonesonde observations and the L3 product, to see how well the L3 product reproduces ozonesonde-derived trends.

In this paper we are mainly focus on assessing the quality of this product to verify if the product is within the committed error bounds (conform the ACSAF ozone profile ATBD). A trend analysis requires much more detail and investigation in different ozone-relevant regions, and this can be a topic for a follow-up paper. For this study we feel that it is out of scope. Having said that, we observe no significant trends in relative difference between the L3 ozone data and the reference data, so this gives confidence that the L3 product can be used for trends in a future study.

3. This work reports validation statistics based on high-resolution reference data that been convolved to the GOME vertical resolution. While this approach is useful for assessing other retrieval errors excluding the smoothing error, the resulting statistics do not represent the actual accuracy and precision experienced by end users. The authors should therefore primarily provide validation metrics without vertical smoothing, and additionally (as supplementary information) report validation metrics with vertical smoothing.

The application of averaging-kernel smoothing is essential when comparing measurements with substantially different vertical resolutions. The L3 satellite profiles represent retrieval products with limited vertical sensitivity and are influenced by their a priori profiles. Direct comparison with high-vertical-resolution ground-based observations therefore introduces representativeness errors related to vertical smoothing differences. In our analysis, this effect is clearly demonstrated quantitatively. The correlation coefficients (r) shown in the time-series panels (Figure 5, 6 and 7) are consistently higher when the satellite data are compared with averaging-kernel-smoothed ground-based observations than when compared with the original unsmoothed data. This behavior is observed across stations and altitude ranges and is further confirmed by the station-to-station consistency analysis, which shows reduced bias spread and improved agreement after smoothing. These results demonstrate that the smoothing procedure is a necessary step to ensure physically meaningful and vertically consistent comparisons between satellite and ground-based ozone profile data.

2.2 Technical comments

L216: please revise “4 % Stubi et al. (2008)”

Fixed, the references are now in brackets

L223: Please revise 13 hPa, 30-32 km.

Fixed.

L241: Please revise (Garcia et al. (2022))

Fixed the double brackets.

L294 Corresponding levels derived from ozonesonde data, following (Table 4) ==> “ozonesonde data” for improving readability.

What was meant is that the ozone at the corresponding pressure levels are compared. We added the word ’pressure’ to the text to make this clear.

L294 The averaging kernels are applied *to high-resolution ozone profiles* for this evaluation.

Your text suggestion was adopted in the text.

L295. Table 5 summaries the validation statistics as function of latitude and vertical regimes specified in Table 5 for GOME-2 A/B/C, respectively. In addition, Figure 5 supplements the validation results from the mid-latitude stations in terms of time-series approach.

Your text suggestion was adopted in the text.

L319. Inconsistence of terminology between “retrieved ground-based” (L319) and sat adj (Figure 6-7). And, “retrieved” is not proper.

We have changed the terminology througout the paper to smoothed ground-based profile, groundbased and L3