

Reviewer 1

This manuscript by Keppens et al. reports on harmonisation efforts applied to sixteen tropospheric ozone satellite data products which also contributes to CEOS activities and the second TOAR assessment. In contrast to common harmonisation approaches to perform spatiotemporal resampling or bias corrections, here the vertical perspective is harmonised to account for differences in the vertical smoothing and sampling between the different ozone satellite products. The authors first define and then address several harmonisation needs. This is done by harmonising tropospheric top level definitions and a-priori information of several types of ozone data products (nadir profile products, limb/reanalysis-nadir matching products and convective cloud differential products). Large efforts are put in the harmonisation of the nadir profile products for which the CDF framework is used and two methods are eventually selected. The column products are finally harmonised using a fill-in method exploiting CAMSRA data. To investigate the effect of the harmonisation on the ozone products, global maps and time series are compared before and after harmonisation. It is concluded that, despite improvements, the presented harmonisation efforts only partially account for the observed differences between the satellite products and that not only differences in spatiotemporal smoothing and sampling but differences in the measurement uncertainty or other instrument and/or retrieval specific effects are the main challenges in the assessment of different tropospheric ozone satellite products.

Given this comprehensive study demonstrating the challenges of interpreting, using and comparing different (tropospheric ozone) satellite data products the manuscript is well-suited for publication in AMT. However, I suggest to address several (mostly minor and technical) aspects before final publication to further improve the quality of the manuscript.

General Comments:

Sometimes the British English (BE) and sometimes the American English (AE) spelling of individual words is used. As an example, in the abstract both “harmonisation” and “harmonization” can be found separated only by three lines. Although there will be a thorough language editing by AMT, I suggest already now to be consistent throughout the manuscript and to decide on one of both options (probably BE as it is also used in the title).

[Author reply: The authors had opted for BE indeed. Four instances of the AE “harmonization” could be found, where the text literally duplicated project-specific phrasing. This, however, is not necessary, so BE has been adopted throughout.](#)

In general, the consistency of several aspects might be improved throughout the manuscript and the presentation style might be further harmonised. Firstly, this refers to the way things are named (e.g., GOME2 vs. GOME-2). Secondly, there should be a consistent use of abbreviations which are defined. Lastly, also the way information is provided might be addressed. An example here is the consistent representation of geographical coordinates (e.g., 60° N vs. 60°N vs. 60N or 1° x 1° vs. 1° by 1° vs. 1 x 1 degree). See also the specific comments below for other examples.

[Author reply: Thanks for pointing this out. The authors have gone through the entire text to check and correct for inconsistencies. Some are additionally addressed in the technical comments below.](#)

Main comments:

Section 2 (especially Sections 2.1, 2.2 and 2.3) + Table 1: In this section, the different data products relevant for this work are summarised and detailed descriptions of the individual products are given. However, there are some inconsistencies between the information provided in Table 1 and in the text. Also, not for all of the different data products the same information content is provided. For example, for the TROPOMI ozone profile product, no information on the vertical grid resolution and the top level is given, while this information is provided for others, for instance for IASI. So, the

authors may consider harmonising the information content provided for the different products (as far as possible). Further, many abbreviations are used in this section (and in general throughout the manuscript). I suggest to consistently introduce them at their first occurrence and then to mostly use these abbreviations. Another point regarding the naming of the data products: consistent naming should be used for the same product, e.g., GOME2 (in Table 1 and also Fig. 2, 3 + 4) vs. GOME-2 (in the text). The authors might consider revising Section 2 with respect to consistency of naming conventions etc., I think this will be helpful for future readers of this paper. Thereby, some of the specific comments listed below (which is likely not a complete list of small action points) might be considered.

Author reply: The authors are grateful for the reviewer's attentiveness. They agree that Section 2 can be improved in terms of readability in line with the reviewer's comments. Several editorial changes have therefore been made to this section, also in line with the detailed comments by both reviewers. These changes include (1) matching of information between the text and Table 1, (2) information harmonisation between different product descriptions, (3) writing out of abbreviations upon first use and consistent use of the abbreviations afterwards, and (4) consistent product naming throughout the manuscript. Several of these changes are clarified in reply to the technical comments below.

Figures in general: Several different styles are used for the different figures. While this does not pose a major problem, some harmonisation in the figures styles and layouts might be still considered. The authors may also consider using consistent/similar colours for Fig. 3 + 4 and Fig. 5. In the current version, the colours used for the results of the different harmonisation approaches are contra intuitive, e.g., USR is shown in blue in Fig. 3 + 4, while in Fig. 5 blue is used for the WAV results. Several figures lack the units of the shown quantities or even do not contain the latter at all. I suggest adding this information to the axes-labels or colour bars (see also specific comments below). Lastly, some figures might be enlarged which particularly applies to the font sizes of the axes- and colour bar labels (again see also specific comments below).

Author reply: The authors agree with the reviewer's suggestions for figure improvements. They have been individually addressed in the detailed comments below.

Result part: To assess the effects of the harmonisation efforts on the ozone results, a large part focuses on analyses of the spatiotemporal inter-product dispersion. In my understanding the term "dispersion" does not always describe the same "quantity" throughout the manuscript and refers sometimes to slightly different definitions. Although this might be a misunderstanding, I strongly recommend to introduce and define the term "dispersion" somehow. This should include a brief description of how it was determined and what it describes in the respective (sub-)section. This could be done in a similar way as in Section 6.1, where the term "dispersion" for the multi-annual mean maps is explained. I have the impression that this would help future readers of the paper.

Author reply: The authors agree that essentially two types of dispersions appear in this manuscript. Most common is the "inter-product dispersion (IPD)" that has now been defined and abbreviated as such throughout the text and that has been related to the formulas in Section 6.1. The second only appears in the discussion of Figure 5 (last paragraph of Section 4.2), where it serves as an estimator of the satellite data uncertainty. The latter has therefore been added to the beginning of this paragraph: "here as an estimator of uncertainty, not IPD"

Minor/specific and technical comments:

Page 1, line 1: The first Tropospheric Ozone Assessment Report is sometimes referred to as TOAR (mainly in the first part) and sometimes as TOAR-I (mainly in the second part). Please consider a consistent naming throughout the manuscript.

Author reply: "TOAR-I" has been replaced by "TOAR" throughout the text.

Page 3, line 45: It is mentioned that the different ozone products are spanning over 21 years (January 2003 to December 2023). In Table 1, however, some data products start before 2003, e.g., ERS2/GOME v3 starts in 1995. In my understanding these 21 years refer to the final harmonised products. Probably cross-check and comment on that.

Author reply: The 21 years refer to the harmonised products indeed, given the time range of CAMSRA. The indication “spanning 21 years in total (January 2003 to December 2023)” has therefore been moved to the end of the introduction, upon mentioning the harmonisation results section.

Page 3, lines 53-56: Does a reference for this product already exist? If yes, maybe add it here.

Author reply: A reference to Van Malderen et al. (2025) has been added here. This work was already referred to later in the text. Note that the initial reference to the preprint version (Van Malderen et al., 2024) has been replaced by the version published in ACP.

Page 3, lines 59-60: Probably replace “The last section before the conclusions...” by “Section 6....”.

Author reply: An explicit reference to Section 6 has been added.

Page 4, line 76: Already introduce the abbreviation LISA here and leave the long version out in line 122 on page 5.

Author reply: This suggestion has been adopted.

Page 4, line 77 + Table 1: Please consider a consistent naming for GOME-2 (in the text vs. GOME2 in the table) here and also in the rest of the manuscript (e.g., lines 71 or 81).

Page 4, line 81: Please also consider a consistent naming for ERS-2 (vs. ERS2 in the table).

Author reply: For both GOME-2 and ERS-2, we have adopted the official naming with hyphen at all occurrences.

Page 4, line 85: Consider consistency between BE and AE throughout the manuscript.

Author reply: This has been done. See reply to first general comment.

Page 5, line 88: Cross-check the value of 10^{12} molecules of ozone per cubic meter in the troposphere. I feel it should read “molecules per cubic centimeter”.

Author reply: This value is indeed correct, e.g. see Table 1 in Keppens et al. (2018).

Page 5, line 91: Like above for the Hartley band, you might add here the wavelength range for the Huggins bands.

Author reply: The spectral band range relevant to the retrieval has been added: 323-335 nm

Page 5, line 95-96: Can this negative bias and its implications easily be estimated. If so, probably add some information here?

Author reply: Also following a comment by the second reviewer, this sentence has been rephrased as follows: “Therefore, a negative bias in retrieved ozone - proportional to the cloud-covered partial ozone column - is to be expected in the presence of high or optically thick clouds.”

Page 5, lines 100-102: In Table 1, the native resolution is listed to be $28 \times 28 \text{ km}^2$. Probably clarify that.

Author reply: The label “native” has been removed from the column headers in Table 1, to avoid confusion with the native sensing resolution of the satellite instruments.

Page 5, line 110: The time periods given here are not consistent to the ones listed in Table 1.

Author reply: Thanks for having remarked this. The IASI-AB combined dataset only goes until the end of 2022, as correctly indicated in the table. This has been adopted in the text.

Page 5, line 112: Consider using the already introduced abbreviation OE for optimal estimation here.
Author reply: This suggestion has been adopted.

Page 5, line 114: I guess it should be FORLI-O3.
Author reply: FORLI can stand by at itself as it also serves as a basis for the retrieval of other molecules, but this suggestion has nevertheless been adopted, in order to more specific.

Page 6, line 126: The sentence regarding the synergism of IASI and GOME2 is basically the same as the first sentence of this paragraph (page 5, line 121). You may consider skipping one of them or rephrasing a little bit.
Author reply: The authors agree. From the sentence onward, the text has been rephrased as follows: "With both instruments on board the Metop satellite series, this multispectral satellite approach is designed for observing the vertical profile of ozone with enhanced sensitivity, particularly in the lowermost troposphere, as compared to single-band approaches. It offers..."

Page 6, lines 139-140: RAL was already introduced, and the long version might be skipped here.
Author reply: This suggestion has been adopted.

Page 6, lines 134-149: The descriptions in this paragraph give the impression that this product also requires a lot of harmonisation efforts. In case this is not a complete misunderstanding. Could you briefly comment on what was done differently in their harmonisation efforts?
Author reply: Thanks for pointing out this confusion phrasing. The "harmonisation efforts" for the creation of this product are essentially inter-product bias corrections and hence are different from what is done in this work. To avoid confusion, the authors have rephrased this paragraph, omitting the terms homogenisation and harmonisation: "It consists of homogenised and combined ozone profile measurements..." has been replaced by "It combines bias-corrected ozone profile measurements..." and "in order to harmonise the integrated columns" has been replaced by "in order to match the integrated columns"

Page 6, line 154: Probably add "(IUP-UB)" after University Bremen to make a link to Table 1.
Author reply: This suggestion has been adopted.

Page 7, line 164: Probably add "...ranging from 2012 to 2022...".
Author reply: This suggestion has been adopted.

Page 7, line 180: Please introduce the abbreviation EPIC somewhere.
Author reply: The abbreviations in this section have been spelled out upon first use.

Page 7, line 183: Please also introduce the abbreviation LRT here (used in Table 1).
Author reply: This suggestion has been adopted.

Page 8, line 191: The abbreviation CCD was already introduced earlier.
Author reply: Agree. "Convective Cloud Differential" has been replaced by "CCD" here.

Page 8, line 192: 1 x 1 degree might be consistently represented by 1° x 1° as done mainly in the rest of the manuscript.
Author reply: This suggestion has been adopted.

Page 8, line 201: Please change NO2 to NO₂.
Author reply: This suggestion has been adopted.

Page 8, line 203: In Table 1, a 6 h resolution is given. Probably comment on this in the text.

Author reply: The temporal resolution depends on the CAMSRA product. It should indeed have been six hours here. This has been modified in the text.

Page 8, line 208: Please specify SBUV/2.

Author reply: "Solar Backscatter Ultraviolet Radiometer-2" has been added between brackets.

Page 8, lines 207-220: This paragraph gives an overview on effects on the quality of the ozone analysis field caused by changes in the observing system. I wonder how the mentioned effects affect the study presented here. You may comment on that.

Author reply: The authors agree that this overview lacks an indication of the effect of changes in the assimilation data (quality). The following sentence has therefore been added at the end of the paragraph: "The effect of these assimilation data (quality) changes on the reanalysis can be expected to be of the same order as the effect of choosing between a full reanalysis and a yearly climatology that is derived therefrom, as discussed next."

Page 9, lines 224-230: As stated in the text differences are in general below 1 DU. Nevertheless, there seems to be indeed a systematic trend from ca. 2010 onwards which was already mentioned by the authors as a possible risk of this approach. Could you comment on the effect this choice might have on the final harmonised products?

Author reply: The authors had already (partially) addressed this question at the beginning of the paragraph, mentioning that "this choice comes with the risk of introducing a tropospheric ozone trend component from the CAMS reanalysis in the harmonised data that is not in the initial satellite data, it is assumed to provide a better representation of the actual tropospheric (ozone) dynamics." To be fully clear, they have added that "(i.e., the risk of removing actual dynamics using the climatology is larger)" Besides, the less than 1 DU effects mentioned at the end of the paragraph have been expressed in percentage as well: "1 DU (3-4 %) differences"

Figure 1: Please add a space between y-axis label and "[DU]".

Author reply: In reply to a comment by Reviewer 2, the y-axis label has been renamed, including a space before the units.

Page 10, line 266: Consider rephrasing the end of the sentence. Suggestion: "..., including the use of the CAMSRA...."

Author reply: This suggestion has been adopted for readability.

Page 10, line 271: Please consider the consistent use of $1^\circ \times 1^\circ$ instead of 1° by 1° .

Page 10, line 272: Add the unit ($^\circ$) to the numbers.

Page 10, line 273: Compare above ($5^\circ \times 5^\circ$).

Author reply: The authors agree with the reviewer's suggestions. They have been adopted.

Page 11, line 289: I guess it should read: "Such a process..."

Author reply: Both "such process" and "such a process" can be grammatically correct, apparently, but "such a process" is more common. This suggestion has been adopted.

Page 11, line 299: "...use of an a-priori..."

Author reply: A hyphen was forgotten here indeed. It has been added.

Page 11, line 304: You might consider using the above-defined abbreviation OE here.

Author reply: Given the beginning of a new paragraph, we prefer the full expression here.

Page 12, lines 326-335: Please also define the I used in the equations.

Author reply: It has been added just below Eq. (1) that “ I is the unit matrix having the same size as A_j , while...”

Tables 2 + 3:

1. I find it a little bit confusing that CDF is marked as “Not applied in this work.”, whereas on page 12, lines 322-324 it reads “The latest formulation of the CDF framework has therefore been applied in this work...” In my understanding this is no contradiction as all methods rely on the CDF framework. Nevertheless, you might consider commenting on this and make the text/Tables clearer on that point.

2. Further, I find the table captions a little bit confusing but do not have a clear opinion how to improve that, so this remains just a comment.

3. A last point regarding the tables (just a suggestion): You might consider changing the order of the different methods according to the order in the text (basically putting USR to the end).

Author reply: 1. The authors agree that this is ambiguous. The asterisk has therefore been removed for CDF, and an additional horizontal line has been added to the table.

2. The table captions have been adapted in agreement with the first point: “Approaches for harmonising ozone profile observations that include averaging kernels and covariance matrices, based on the general CDF method (first row). Harmonised profiles are obtained from N input profiles and determined by the choice for the new...”

3. The reviewer’s suggestion certainly makes sense, but the authors have decided to stick to the order where the methods not applied in this work are grouped at the end of each table.

Page 13, line 347: You might add “(second row)” to that sentence.

Author reply: This suggestion has been adopted.

Page 14, lines 387-388: After reading this sentence several times, I still have the impression that I do not get it right. You might consider rephrasing it.

Author reply: Thanks for pointing out this unclarity. This sentence has been rephrased and extended as follows: “This would mean making use of the single-profile Eqs. (4) and (7) for APR and USR, respectively, upon...”

Page 14, line 389: Can you briefly comment on how large the expected discrepancies by this approximation are?

Author reply: These two sentences have been updated as follows: “...appears in the APR method, yielding differences at the percentage level. This simplified approach is hence adopted in this work.”

Page 14, line 391: See above ($1^\circ \times 1^\circ$).

Author reply: This suggestion has been adopted, in order to be consistent with the above.

Page 14 + 15, lines 400-404: Consider slightly rephrasing this sentence (i.e., description of what is depicted in Fig. 2) to make it clearer. This would then also apply to the first sentence of the caption of Fig. 2. Further, I wonder why WAV is chosen as “reference” of the comparison. Probably, I missed it but is the WAV method the common approach? Maybe you could briefly comment on that.

Author reply: The authors agree that this phrasing can be simplified, and have updated the text as follows: “...how much the spatiotemporally averaged relative difference between each product pair changes upon going from WAV to APR (below diagonal) and to USR (above diagonal), expressed in terms of the ratio, in %, between the products’ mean difference before and after harmonisation, for all three top level definitions combined.” The last paragraph of Section 4.1 mentions that APR and USR “are compared with the weighted average (WAV) profile, in order to verify whether the ozone profile data harmonisation reduces the dispersion between the selected products...”

Figure 2, caption: “...closer to each other...”

Author reply: That's clearer indeed. This suggestion has been adopted.

Page 15, line 403: This means...

Author reply: Thanks for having noticed this typo. It has been corrected.

Figures 3 + 4: You might consider adding "Year" as a y-axis-label.

Author reply: This suggestion has been adopted.

Page 16, lines 426-428: Probably discuss (or at least comment) that this increase of ΔSMOD apparently depends also on the top level definition, i.e., there is no clear increase for GOME2 in the middle column of Fig. 3.

Author reply: This suggestion has been adopted by extending the comment between brackets at the end of the paragraph: "yet note the dependence on the tropopause definition and the different vertical scales..."

Page 16, lines 434-436: Probably briefly repeat these reasons. I think keywords might be sufficient.

Author reply: This sentence has been rephrased in order to stress the reasons explained earlier: "It does not seem to hold for the RAL products for reasons explained above (essentially starting from a common prior already)."

Page 16, line 436: Please remove ")".

Author reply: Thanks for having remarked this. The suggestion has been adopted.

Figure 5:

- If possible, consider enlarging the figure or at least the font size of the x- and y-axes-labels. In a printed version, the y-labels are really hard to read. The same applies to the legend of the figure. Like for Fig. 3 + 4 the top-level definition might be used as titles for the two columns.

Author reply: The authors have adopted this suggestion, by reframing Figure 5 with increased font sizes. Top-level definitions have been added as panel titles.

- As mentioned above (see main comments), you may consider using consistent colours for the different methods for Fig. 3 + 4 and Fig. 5.

Author reply: Following both reviewers' suggestions, the colour scheme of Figs. 3, 4, and 5 has been matched.

- You also might specify the abbreviation GND somewhere.

Author reply: The abbreviations SAT and GND that appear in the vertical axis label have been added to the figure caption.

Page 16, line 449: I would argue that the dispersions reach higher than 10 DU for some cases.

Probably increase the range mentioned in the text.

Author reply: The authors agree that if a range is specified, its upper value should be higher than 10 DU. This value has been increased to 15 DU accordingly.

Page 17, line 464: Probably insert the word "only" before +1 % to stress that this is not of great relevance.

Author reply: This suggestion has been adopted.

Page 17, line 467: "spatio-temporal" should be changed to "spatiotemporal" to be consistent to other parts of the manuscript.

Author reply: Thanks for having remarked this. The suggestion has been adopted.

Page 17, line 483: Probably, briefly mention the two selected (out of four) approaches. (In my understanding APR and CAMSRA fill-in.)

Author reply: The authors agree that it would be helpful to mention the two selected methods explicitly. “i.e., APR and the fill-in method” has been added to the end of the sentence.

Figures 6 + 7: Both figures should be enlarged as they are (even on a screen) hard to read. This in particular applies to the font size of the subplot titles. Further, I recommend adding the shown quantity and its unit to the colour bars. This could be done once in the middle for each figure.

Author reply: Following a similar suggestion by both reviewers, Figs. 6 and 7 have been reframed to allow larger panels and font sizes. Only one colour bar has been maintained within each Figure.

Page 18, Title of Sect. 6.1: I suggest introducing the abbreviation MAM here (or somewhere else in the beginning of this section). It is used several times in the following but not introduced before.

Author reply: The authors have adopted this suggestion, by adding “multi-annual mean (MAM)” explicitly at the beginning of the section.

Page 18, line 492: In my understanding it’s an inter-product dispersion. Probably, mention it explicitly. See also main comments above.

Author reply: The inter-product dispersion is intended here indeed. This has been added explicitly.

Page 18, line 495: Probably call it consistently near-global rather than “pseudoglobal” also in the equation.

Author reply: The authors agree and have applied the phrasing “near-global” everywhere in the text, also in agreement with community practice.

Page 18, line 496: Consider adding one or two sentences on what can be learned from Fig. 9. For example, why is the change for GOP-ECV so large or why are there no values for GOME and SCIAMACHY. Although the first aspect is explained below, I think there should be in general a sentence on Fig. 9 (might be a different aspect).

Author reply: What is requested by the reviewer at this point appears in the text later. The authors have nevertheless extended this sentence to cover more information about Figure 9: “Near-global mean columns before (red) and after (blue) harmonisation are shown in Figure 9 for the same fourteen satellite data products and CAMSRA (see discussion below).”

Figure 8:

1. Again, consider adding the shown quantity and its unit to the colour bars (once should be sufficient).

Author reply: This suggestion has been adopted.

2. It would be also helpful for the reader to add titles to the columns indicating the difference between the columns (like mentioned in the caption).

Author reply: This suggestion has been adopted.

3. Probably, change “...after minus before harmonisation” to something like “upon harmonisation” (just a suggestion).

Author reply: This suggestion has been adopted for readability.

Figure 9: Consider adjusting the y-label to make the representation of geophysical coordinates consistent to the caption and the rest of the manuscript (see also general comment above).

Author reply: This suggestion has been adopted.

Page 18, lines 506-507: Probably, add in brackets where this can be seen. In my understanding Fig. 8, left column.

Author reply: The authors have adopted this suggestion by adding “(Figure 8 left)” to this sentence.

Page 18, lines 509-512: Also here, please add where (which plot) this can be seen.

Author reply: This sentence still refers to the mentioning of Figures 6 and 7 in the sentence before.

Page 19, line 518: I think it should read: left and right in the brackets instead of blue and red. Please cross-check.

Author reply: Thanks for having noticed this error. This has been corrected.

Figure 10:

1. Please consider enlarging the font size of the axes-labels and the legend (at least the size of the ones in Fig. 11).

Author reply: The authors have reframed Figure 10 to allow for larger font sizes.

2. Further, I suggest adjusting the y-axis-label to a more practical name for tropospheric column.

Author reply: The y-axis label has been changed to “tropospheric ozone column [DU]”.

Figure 11:

1. Same as Fig. 10 for the y-axis-label.

Author reply: The y-axis label has been changed to “IPD [DU]”.

2. In my understanding not the tropospheric column but the dispersion is shown.

Author reply: Thanks for having remarked this. The IPD is shown indeed. This has now been made clear (see previous).

3. Further, please add the unit DU to the y-axis-label.

Author reply: This suggestion has been adopted.

4. Why does it say “L3” in the legend although L2 is in the title and should be correct (similar for the titles in Fig. 10)?

Author reply: “L3” in the legend refers to the harmonisation being assessed for L3-type data. The title in the first column, on the other hand, refers to the Level-2 nadir profile input data, as mentioned in the figure caption. The authors believe that this is now sufficiently clear.

Section 6.2. In general: Consider also adding some statements on Fig. 10 in the text or include the figure in the considerations in the paragraph on page 19, lines 524-532.

Author reply: The authors agree that the discussion can be more explicitly linked to Figure 10 (and 11). The first sentence of the second paragraph has therefore been extended as follows: “The overall conclusion from Figs. 10 and 11...”

Page 19, lines 524-532: Consider adding where the different conclusions can be seen, e.g., compare Fig. 11, upper panel or so.

Author reply: This suggestion has been adopted, by making reference to the panels in Figure 11 throughout the text.

Figure 12:

1. Please add the unit to the y-axis-label and enlarge the font sizes of labels and legend.

Author reply: The y-axis label has been changed to “IPD [DU]”. The authors have reframed Figure 12 to allow for larger font sizes.

2. I think the dispersion is shown, please add this information to the caption.

Author reply: Thanks for having remarked this. The caption has been updated with “of the IPD [DU]” to make this clear.

3. Further, consider using consistent representations of the geographical coordinates in the legend (compare above).

Author reply: This suggestion has been adopted by updating the legend entries.

4. I am again little bit confused about the L3 in the title of the panels.

Author reply: “L3” in the title refers to the harmonisation being assessed for L3-type data. The authors believe that this is now sufficiently clear, but have nevertheless updated Figure 12 accordingly.

Page 19, lines 533-536: For some regions (e.g., 40° S – 60° S) there seem to be systematic seasonal cycles of the inter-product dispersion. I wonder if these seasonal cycles are understandable. If so, you might comment on that.

Author reply: The authors agree that this seasonality requires some clarification: The end of this paragraph has therefore been updated as follows: “It is hence clear that harmonisation of the profile and column datasets does not lead to better constraints of the seasonal cycle at the zonal to near-global level, although the seasonality of the IPD changes within some latitude bands, like 40-60° north and south. This seasonality, which is due to differences between sounders in their sensitivity to seasonal tropospheric ozone changes, may indeed be affected by changes in the vertical sensitivity upon harmonisation.”

Page 20, lines 553-554: You might briefly repeat why these two methods were chosen. While this was done in the main part, I think it would be helpful for the reader to get a short reminder here in the conclusions section.

Author reply: The authors have adopted this suggestion. This sentence has been updated as follows: “In order to avoid the known difficulties of the maximum-likelihood and information-centred representations, only the a-priori profile replacement (APR) and unit-sensitivity representation (USR) methods have been selected in this work...”

Page 20, lines 555-559: Like the previous comment: consider briefly mention which of the two was chosen and why.

Author reply: The authors have adopted this suggestion by adding the following sentence: “Eventually, our harmonisation assessment focused on those methods that are (re)constrained by the CAMS reanalysis, i.e., APR for the profile products and the fill-in method for the column data.”

Reviewer 2

Keppens et al. applied consistent corrections to the O₃ retrievals from 16 satellite products in order to harmonize the respective data records. The paper is generally well written and matches the scope of the ACP/AMT special issue on Tropospheric Ozone. I recommend publication after dealing with the following issues:

General comments

1. The authors discuss 6 methods for the harmonization of O₃ profiles in section 4.1. However, results are only presented for two of them. It remains unclear if the limited improvement of the harmonization efforts is due to this choice, and if the other methods like WAV or CDF would yield better results. Please check the results at least for CDF as well. I would propose to start the analysis with the comparison to the independent "truth", i.e. the ozonesondes. From this comparison, the two best/most promising methods could be selected and discussed in more detail afterwards. And, actually, from Fig. 5, I would consider WAV to be overall "better" than both APR and USR, as it works quite ok for all cases, while both APR and USR show quite large values for some cases. In any case, it should be clearly motivated which of the methods are selected for what reason, and the poorer performance of the methods that have been discarded needs to be demonstrated.

Author reply: The authors agree that throughout this long section, the motivation for the eventual selection of APR and USR may have been somewhat hidden. Based on editorial suggestions by Reviewer 1, several changes have already been made in this section. It should now be clear that only five data fusion and harmonisation methods are discussed, which can all be derived from the complete data fusion (CDF) framework (cf. updates in text and Tables 2 and 3). Besides, the methods with zero prior constraints (MLR and ICR) had been thoroughly examined in previous work (Keppens et al., 2022) and concluded not to be suitable for our nadir ozone profile harmonisation purposes. This has now been explicated in the text. Finally, the beginning of the last paragraph of Section 4.1 has been rephrased, in order to make reference to all methods discussed, and clearly motivate the selection of APR and USR: "Eventually, because of the difficulties that come with the MLR and ICR methods as mentioned above, two CDF-based vertical profile harmonisation methods are considered in this work, being the direct a-priori profile replacement (APR) and the unit-sensitivity representation (USR). Both are compared with the weighted average (WAV) profile, in order to verify whether the ozone profile data harmonisation reduces the inter-product dispersion (IPD) between the selected products (also see next section)." As such, the authors believe it is clear that the weighted average (WAV) is merely a data fusion technique, which serves as a reference for the harmonisation approaches. On the other hand, the authors also agree that the WAV method is performing more consistently stable (which should not surprise). Therefore, in the TOAR-II tropospheric ozone assessment paper from satellite data, the WAV method is considered (i.e. omitting prior information harmonisation eventually).

2. The authors conclude that there must be other instrument/retrieval specific reasons for the large discrepancies between the products. Please add a bit more discussion about what instrument and/or retrieval issues this could be, and how to deal with (and hopefully resolve) the remaining differences in future studies.

Author reply: The authors agree that the text could have been more specific on the reasons for inter-product discrepancies. The conclusions section has been updated with the following sentence: "This implies that a substantial part of the inter-product differences are instrument and/or retrieval-specific (mainly including measurement uncertainties, auxiliary data, and retrieval implementation) and hence need to be addressed at Level-1 and Level-2 data processing, preceding harmonisation." Providing details on how exactly this can be done, is considered to be outside of the scope of this work.

3. Please check the Copernicus figure guidelines: <https://www.atmospheric-measurement-techniques.net/submission.html#figurestable> In particular the plots with red and green lines are problematic.

Author reply: In reply to detailed comments by all reviewers (also see below), the authors have updated the formatting of most figures, also to be in better agreement with Copernicus figure guidelines.

Detailed comments

Line 39: What does "lack of harmonisation of ... their units" mean? The result of a particular approach/method should not depend on the units used!?

Author reply: It was meant here that even upon having identical quantities, products sometimes still differ in their unit representation, e.g. having VMR in ppmv versus ppbv. We have rephrased this sentence as follows: "Additional confounding factors are time-varying biases and the lack of harmonisation of geophysical quantities (including differences in units)."

Line 53: Two approaches are selected, but it is not clear which one was used for the finally applied harmonization.

Author reply: At this point in the introduction, the authors prefer to stick to mentioning how many harmonisation approaches are considered, without having to go into detail on a specific selection, which would require quite extensive methodological descriptions this early in the text. Following a similar comment from Reviewer 1, however, the authors have stressed the eventually selected methods later in the text (beginning of Section 6).

Section 2: For the applied harmonization, datasets of lv2 O3 profiles and their AKs are needed. I think this should be clearly stated in the beginning of section 2 or 2.1.

Author reply: The authors agree that this should be clear from the beginning. The following has therefore been added to the first paragraph of Section 2.1: "Vertical smoothing harmonisation can be applied to these Level-2 datasets only if the data include initial prior profiles, averaging kernels and covariance matrices (see Section 4)."

Line 71/72: Please provide a reference.

Author reply: Unfortunately, it is not fully clear to the authors what exactly the requested reference should refer to. The first paragraph of Section 2.1 provides a general overview of the nadir ozone profile products covered in this work, while details about the products, including references, appear in the subsequent paragraphs.

Table 1: Please add references and DOIs in another column.

Author reply: Given the current size of the table (already full paper width), the authors believe that adding another column will not improve its readability. References to the data are provided in the text, DOIs are added where applicable (not commonly available).

Table 1: "Native sampling" is misleading, as e.g. OMI sampling is better than 48 km along track, or TROPOMI is better than 28 km along track. I assume that multiple pixels were co-added in the RAL retrieval. This information should be added to the table (for instance as footnote).

Author reply: Following a related comment by Reviewer 1, the label "native" has been removed from the column headers in Table 1, to avoid confusion with the native sensing resolution of the satellite instruments.

Line 88: "per cubic meter" -> " m^{-3} ".

Author reply: This suggestion has been adopted.

Line 96: "where high or thick cloud is extensive" -> "in presence of high or optically thick clouds"

Author reply: Agree. The text has been rephrased as follows: “Therefore, a negative bias in retrieved ozone - proportional to the cloud-covered partial ozone column - is to be expected in the presence of high or optically thick clouds.”

Line 201: CAMS is used as transfer standard. However, this is not an independent dataset, as the satellite measurements are used during assimilation. This aspect has to be discussed. Please also state whether all or only a subset of the presented data sets are included in CAMS assimilation.

Author reply: None of the ozone profile or tropospheric ozone column products discussed in this work are assimilated in CAMS. Therefore “not involving any of the retrieval products discussed in this work” has been added to Section 2.4.

Line 233: Please add a general O3 sonde reference here.

Author reply: The authors agree that a reference would be helpful here. A reference to Tarasick et al. (2021) has been added here.

Section 4.1: This is a long section providing many details of the different harmonization methods. I think it would be good to enhance structure and add subsubsections for each method.

Author reply: This suggestion has been adopted. Subsubsections have been added to Sections 2.1 and 2.2.

Table 2 last line: What is W_i^{-1} ?

Author reply: This is the generalised pseudo-inverse of a regridding matrix W . This is explained in the text between Eqs. (6) and (7).

Tables 2/3: Please merge those tables into one (first three columns are identical anyhow).

Author reply: The authors have considered this possibility indeed, but for now stick to two separate tables because of the length of the equations. A decision on one or two tables will be made with the editors upon typesetting the final manuscript.

Line 451: USR is far worse for lapse rate tropopause and all satellites except OMI. And for OMI, APR is performing poorly. This should be mentioned as well.

Author reply: The authors agree that the current text may appear incomplete. This text part has been rephrased as follows: “As before, the USR harmonisation effectively works for the lower-resolution (both horizontal and vertical) RAL retrieval products with an order of 10 % difference dispersion reduction, while it does not for the other products. The APR method slightly benefits (by about the same percentage) the higher-resolution TROPOMI and (combined) IR retrievals yet performs poorly for OMI.”

Fig. 1: Please consider a more helpful label of the y-axis.

Author reply: This suggestion has been adopted. The y-axis label has been changed to “deseasonalised anomaly of ozone partial column between 450 hPa – LRT [DU]”

What is the reason for the drop from ~2 DU to ~0 DU in 2004?

Author reply: As indicated in the figure caption, “Vertical lines (yellow) indicate changes in the CAMS assimilation system as detailed in Inness et al. (2019).” The drop occurs at the second yellow line, corresponding to the introduction of several new ozone retrievals to the operational assimilation system (Table 2).

Fig. 3:

- The first row should also be discussed first in the caption.

Author reply: The captions of Figs. 3 and 4 have been rephrased to follow the order of the panel rows.

- Actually, I do not understand what DeltaSMOD means - just a difference of the CAMSRA model data for harmonized vs original profiles? What does that tell me? Is a Delta SMOD of 10 DU good or bad? Please consider if you need to show this quantity, and if yes, please explain it more clearly in paper and figure caption and motivate why this is a helpful information.

Author reply: The third paragraph on the performance of the nadir ozone profile harmonisation mentions (after minor edits based on comments by Reviewer 1) that “The first row (Δ SMOD) shows the difference between the CAMSRA model data (black dashed lines) before and after averaging kernel smoothing. The latter is given by [formula] (Keppens et al., 2019) and the difference hence captures to what extent temporal changes in the averaging kernels (and thus retrieval sensitivity) affect the product time series (Pope et al., 2024). Δ SMOD will increase when the retrieval overall loses sensitivity in time, as is e.g. the case for GOME-2A and TROPOMI (yet note the dependence on the tropopause definition).” The authors believe that this is now sufficiently clear.

- The meaning of colors changes in the last row. This is confusing. Please use completely different colors for the last row.

Author reply: Following both reviewers’ suggestions, the colour scheme of Figs. 3, 4, and 5 has been matched.

Fig. 4:

- I am confused by the second row, which shows the prior columns. For p450, the prior is just flat, whereas the respective subplot in Fig. 3 shows a clear seasonal cycle. Why are the *priors* so different for Fig. 3 and Fig. 4?

Author reply: Figures 3 and 4 show results for different ozone profile retrievals using different prior profiles and constraints. This is indeed reflected in the second rows of the figures, highlighting the differences between retrieval setups, and providing an indication of what the effect of the prior profile choice could possibly be.

- The meaning of colors changes in the last row. This is confusing. Please use completely different colors for the last row.

Author reply: Following both reviewers’ suggestions, the colour scheme of Figs. 3, 4, and 5 has been matched.

Fig. 5: The individual stations are hard to read, and they are not discussed anyhow. Thus I would propose to show histograms of the dispersion across all stations instead.

Author reply: Following a comment by both reviewers, this figure has been reframed and relabelled to increase font sizes. On the other hand, the authors prefer keeping the dispersion information for all stations individually, as this provides a view on meridian features and is of use to the TOAR-II HEGIFTOM working group.

Figs. 6&7:

- The color bar needs only to be shown once per figure.

- Please do not rotate the maps.

Author reply: Following a similar suggestion by both reviewers, Figs. 6 and 7 have been reframed to allow larger panels and font sizes. Only one colour bar has been maintained within each Figure.

- Please show all 16 products.

Author reply: GOME and SCIAMACHY are not included as their time range does not overlap with the selected period (2018-2022), as indicated in the second footnote.

Fig. 10: Some products show considerable drifts; this should be stated and discussed (also how far this affects the comparisons made), and exclusion of these "drifting" products might be considered.

Author reply: the authors agree that the observation of clear drifts must be mentioned in the main text. In Section 6.2, a note has been added: “Note that the considerable drifts observed for some UV-vis products are hardly affected by their APR harmonisation.” The discussion of how these drifts affect tropospheric ozone comparisons and assessments, however, is foreseen in the TOAR-II satellite

data assessment (end of 2025), as mentioned in the Conclusions section. Drifts have now nevertheless been mentioned explicitly with the conclusion that a substantial part of the inter-product differences are instrument and/or retrieval-specific.

Comments by Owen R. Cooper (TOAR Scientific Coordinator of the Community Special Issue) on: This review is by Owen Cooper, TOAR Scientific Coordinator of the TOAR-II Community Special Issue. I, or a member of the TOAR-II Steering Committee, will post comments on all papers submitted to the TOAR-II Community Special Issue, which is an inter-journal special issue accommodating submissions to six Copernicus journals: ACP (lead journal), AMT, GMD, ESSD, ASCMO and BG. The primary purpose of these reviews is to identify any discrepancies across the TOAR-II submissions, and to allow the author teams time to address the discrepancies. Additional comments may be included with the reviews. While O. Cooper and members of the TOAR Steering Committee may post open comments on papers submitted to the TOAR-II Community Special Issue, they are not involved with the decision to accept or reject a paper for publication, which is entirely handled by the journal's editorial team. Comments regarding TOAR-II guidelines:

TOAR-II has produced two guidance documents to help authors develop their manuscripts so that results can be consistently compared across the wide range of studies that will be written for the TOAR-II Community Special Issue. Both guidance documents can be found on the TOAR-II webpage: <https://igacproject.org/activities/TOAR/TOAR-II>

The TOAR-II Community Special Issue Guidelines: In the spirit of collaboration and to allow TOAR-II findings to be directly comparable across publications, the TOAR-II Steering Committee has issued this set of guidelines regarding style, units, plotting scales, regional and tropospheric column comparisons, and tropopause definitions.

The TOAR-II Recommendations for Statistical Analyses: The aim of this guidance note is to provide recommendations on best statistical practices and to ensure consistent communication of statistical analysis and associated uncertainty across TOAR publications. The scope includes approaches for reporting trends, a discussion of strengths and weaknesses of commonly used techniques, and calibrated language for the communication of uncertainty. Table 3 of the TOAR-II statistical guidelines provides calibrated language for describing trends and uncertainty, similar to the approach of IPCC, which allows trends to be discussed without having to use the problematic expression, "statistically significant".

General comments:

Since this paper was submitted to AMT in late 2024, the OMI/MLS product has been updated to take advantage of the new OMI Collection 4 L1b retrievals (Kleipool et al., 2022), which correct for instrument drift through the end of 2024. Will this updated OMI/MLS product be incorporated into your analysis? Does this new version of the OMI dataset impact any of the other products in your analysis?

Author reply: The authors are indeed aware that in the meantime Col3 OMI/MLS products have been replaced with Col4 OMI/MLS products on the NASA website. The Col3 and Col4 data do not much differ, however, especially for zonal means. Accordingly, the following clarification has been added to the NASA data description: "Note that the OMI total ozone column product used here is based on the Collection 3 level-1b data, which in the meantime has been replaced by Collection 4 (Kleipool et al., 2022), but the latter has not yet been thoroughly evaluated."

Two of the panels in Figure 6 and Figure 7 have missing data due to the South Atlantic Anomaly (SAA) (Finlay et al., 2020). While this phenomenon is well known to satellite experts, many general readers are not familiar with the SAA. Please consider adding a brief explanation for the missing data.

Author reply: The authors agree that the spatial gaps in Figures 6 and 7 require an explanation. The following text has therefore been added to Section 6.1: "The gaps that appear in the maps of GOME-2A/B are due to screening of the south-Atlantic anomaly (Finley et al., 2020)."

References:

Finlay, C.C., Kloss, C., Olsen, N. et al. The CHAOS-7 geomagnetic field model and observed changes in the South Atlantic Anomaly. *Earth Planets Space* 72, 156 (2020). <https://doi.org/10.1186/s40623-020-01252-9>

Kleipool, Q., N. Rozemeijer, M. van Hoek, J. Leloux, E. Loots, A. Ludewig, E. van der Plas, D. Adrichem, R. Harel, S. Spronk, M. ter Linden, G. Jaross, D. Haffner, P. Veefkind, and P. F. Levelt, Ozone Monitoring Instrument (OMI) collection 4: establishing a 17-year-long series of detrended level-1b data, *Atmos. Meas. Tech.*, 15, 3527–3553, <https://doi.org/10.5194/amt-15-3527-2022>, 2022.

Author reply: These references have been added to the references list, given that they are referred to in the manuscript update.