The manuscript '*An inter-comparison of tropospheric ozone reanalysis products from CAMS, CAMS-Interim, TCR-1 and TCR-2*' presents a description and extensive evaluation of tropospheric ozone from four recent global chemical reanalyses: CAMS-iRean, CAMS-Rean, TCR1 and TCR2. The study performs very detailed comparisons between the reanalyses and independent observations of surface, profile and column ozone and assesses the relative performance of the reanalyses. This includes some very nice analyses of specific aspects of the reanalyses such as the representation of the diurnal ozone cycle. I am really impressed by the amount of work that went into this study and I applaud the authors for their very thorough and well-organized analysis. This type of paper is not easy to write but it is very important for the scientific community, especially as the understanding of the importance of chemical reanalyses is growing.

The paper is well thought out, well-written and well-organized. It is certainly worth of prompt publication. I have only several minor suggestions for edits and some technical corrections.

Minor general comments

1. Tables and the discussion of reanalyses' performance: it would be really good to have the RMSE values shown in percent in addition to absolute values (ppbv, DU). Having absolute values alone makes it difficult to judge if the RMSE is large or not. I appreciate that sometimes, particularly when the mean ozone is low, large percent values may be misleading but that shouldn't be an issue if both, absolute and relative RMSE is shown. In my specific comments I point to some places where having percentages would be particularly useful but it would really be best to have them in all the tables.
2. It would be helpful to have a schematic figure, similar to Davis et al. 2017 Fig. 1, showing the ozone observations assimilated by each reanalysis, indicating whether a bias correction was applied or not, and, as an added benefit, showing time periods of the reanalyses.
3. The authors often use the word 'model' as synonymous with 'reanalysis', e.g. L273, L307, L316, L318, and in many other places. I suggest limiting the use of this term to the instances where you are really talking about a model (e.g. 'model levels' or 'chemistry model', etc.)
4. Section 5 contains detailed discussions of reanalyses comparisons against multiple data sets. It's easy to lose the big picture in all these details. It would be really helpful to include 2-3 sentence summary highlighting the key results at the end of each subsection as it is already done in Subsection 5.3.

Specific comments and technical corrections

L35 climate-change → climate change

L34-36. This sentence conflates two different things: (1) the importance of ozone forcing for climate and (2) a lack of impact of improved ozone representation on long-term weather forecasts. I suggest splitting it into two sentences.

L38. This deserves more references than just the two that are provided.

L120-121 'to evaluate their fitness for purpose for the various types of application described above'. This sounds a little awkward. Please, consider rephrasing.

L158. Was there any kind of bias correction applied to these ozone data, as in CAMS-REAN? Maybe I missed that information. As I stated in my general comments a figure summarizing all these data types and how they're used in each reanalysis would be useful. This information could also be added to tables 2, 3 and 4.

Table 2. For the profile data types it would be helpful to include the vertical ranges or at least the lowest levels assimilated.

L200. Why couldn't they be filtered out?

L231 & L247.Livesey et al., 2011 is not in the reference list. If this is the MLS version 4.2 data quality and description document then its latest version is from 2018 (https://mls.jpl.nasa.gov/data/v4-2_data_quality_document.pdf). Is the MLS data quality screening based on some earlier guidelines? Note that v4.2 didn't exist in 2011.

L245-251. TES should also be mentioned here for completeness.

Table 4. According to the table TCR-1 uses MLS v3.3. It's version 4.2 in the text (L230).

L273. Data has been collocated → data **have** been collocated

L299 'any of the reanalysis model resolutions is considered too coarse' please correct the grammar

L310. What's the frequency of EMEP data?

L325. I think 'multiannual' is one word. At least, please be consistent; 'multi-annual' it's hyphenated a few lines below.

Figure 2 appears to be repeated or at least I can't discern any difference between the top two and bottom two rows. In addition, please, explain in the caption what 'Season: AYR' means ('all year'?) or remove it from the legend.

L325-342. What about the large discrepancy between the sondes and all the reanalysis near the surface at NH subtropics and, to a lesser extent, the tropics?

L341-346. The comparison with ACCMIP would be easier to see if the biases shown here were given as percentages in addition to absolute values.

L370. Could you briefly justify the use of an 'ozonopause' rather than more commonly used lapse rate or dynamical definitions of the tropopause? In addition, because of the high vertical resolution of ozonesondes they're likely to attain 150-ppbv threshold at very different (and somewhat random) altitudes than the reanalyses. How does that impact these comparisons?

L373. SH midlatitudes also look messy, especially TCR reanalyses. The absolute RMSE may be less than at high latitudes but relative to the mean column it looks quite high. Here and elsewhere it might be helpful to provide percent values for the mean biases and RMSE.

L390. 'These figures'. It's **one** figure (multiple panels)!

Figure 4. The caption says that ozonesondes are shown in black but since what's shown is biases w.r.t. the sondes the latter are not really shown at all, are they? I suggest deleting that sentence. Also, please state that numbers of observations are shown as gray dashed lines, even if it's obvious from the previous figure. As a side note, I'm not against multi-panel figures but I don't think I've seen one with 21 panels before.

L401. Why is MIPAS relevant to the troposphere? Is that an indirect impact of assimilating total ozone with stratospheric ozone constrained by MIPAS data? The same question applies to line 451-453 (Antarctic ozone).

L430-433. Any idea what happens around 2010-2011 that causes this improvement over Japan?

L445-446. The CAMS reanalyses show some large departures before 2005, especially at 382 hPa. Can you comment on that?

L507. But it's not exactly the same period, is it? Figure 6 shows aggregated data from 2005 to 2012 and S3 is extended through 2014.

Figure 8. The caption says 'left' and 'right'. It should be 'top' and 'bottom'. Alternatively, the panels could be labelled.

L652. 'Figure 11'. I think it should be '12'.

L727. Here and elsewhere, please provide percentages in addition to absolute values. How large (small) is 6 ppbv in this case?

L806. Could you expand on this? It would be very helpful to include a paragraph with specific recommendations for the users: What kind of studies are these reanalyses good for? Which reanalyses are recommended for a particular type of study and which ones are less reliable? Are there any types of problems for which these reanalyses are not useful?

This is partially addressed in the second to last paragraph where the authors delineate some issues related to trend and long-term variability studies using reanalyses but I think this type of discussion could be expanded to other areas.

L 810. Do you really mean 'any' models or is it 'many' models?

References

Davis, S. M., Hegglin, M. I., Fujiwara, M., Dragani, R., Harada, Y., Kobayashi, C. et al. (2017). Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP,. *Atmos. Chem. Phys.*, 17, 12743-12778, https://doi.org/10.5194/acp-17-12743-2017