**Referee 1 report**

## *Review of Froidevaux et al. about ozone and carbon monoxide trends in the tropical upper troposphere*

*Froidevaux et al. presents ozone ($O_3$) and carbon monoxide (CO) trends in the tropical upper troposphere observed by the Microwave Limb Sounder (MLS) satellite instruments and in comparison with simulations from different models. Although this study should have required an important effort from the authors, I found a number of limitations which would make the review process difficult to complete in a single revision of this manuscript.*

***Major Comments***

*I find this paper is very difficult to read and for different reasons. First, the paper has too many objectives. It discusses the trend of $O_3$ and CO of MLS, but also from different models, which is already broad. Would that make sense to limit the paper to only one species, or only on the MLS trend without including comparison with models? Discussion of the trends are also difficult to read because the authors address many regions in comparison with previous study. I think the paper needs to be restructured by having 2 distinct sections for $O_3$ and CO (if not resubmitting two distinct papers), those being split in (1) presenting the MLS trends, (2) comparing the MLS trends with models, (3) discussing the results w.r.t. existing literature where a table or a figure compiling the different results might support the text.*

**Answer:** Having a new analysis of 15+ years of a satellite-based dataset (and models) regarding changes in tropical UT composition is unprecedented, so it makes sense that one can extract a good amount of new information, and with both $O_3$ and CO, this will lead to a fairly long paper – which has taken years of (part-time) work to produce. There are enough aspects in common at this point for both species that we will not split this up into two submissions, as this would also cause many additional delays for us.

We have re-organized (after step 1, replying to all other comments from both reviews) - by presenting the results for $O_3$ before the CO results; this change does not modify science results but we hope/think that it has helped, along with all the other changes and adjustments. We have also separated the (short) main Conclusions (with shorter cleaner messages, we believe) from the Discussion section (two discussions now, one for $O_3$ and one for CO). In step 1, we prepared one revised Word file, with (quite a few) tracked changes, before any re-organization. We then re-organized in step 2, with fairly minor changes, on top of the step 1 file with all tracked changes "accepted" (by ourselves at least), to make the 2$^{nd}$ step review simpler; there is an added Conclusion and a reference/comparison to a very recent paper on this topic.

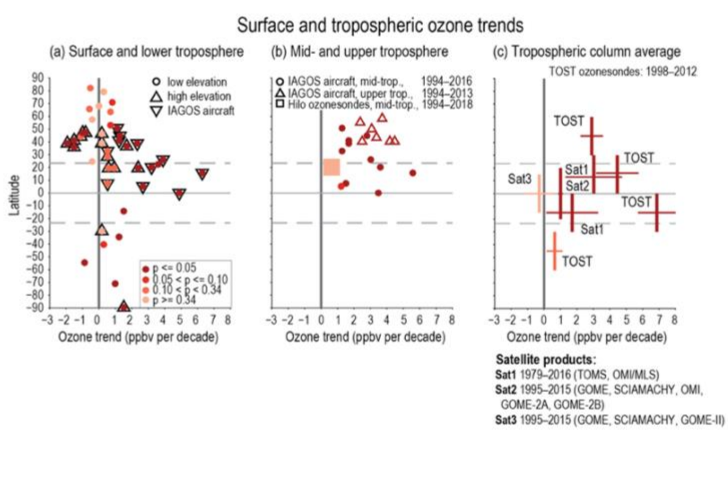
We are not sure that such re-organization really helps the readability, as this is somewhat of a subjective preference, with little impact on the scientific results, but it probably did. Also, we are not claiming we are providing a comprehensive assessment (TOAR-II, for example, will be required for this, and such efforts take a much bigger community yet), but we have made "first-order" comments about past trend results, which often cover somewhat different time periods or regions; the best comparison for $O_3$ is versus TCO for the same period and region, and we made an effort to present that. It is almost impossible to carefully compare results from past work if care is not taken to remove differences between regions or time periods, if this is even possible to assess well (again, not our current goal for this already long paper). We are, however, linking this manuscript to the TOAR-II (ozone-related) assessment, as we were encouraged to do this independently (by O. Cooper and H. Worden), and we obtained permission from the editors to do exactly that, as extra motivation for this manuscript's relevance.

*During the first reading of the paper, it was not clear how many model simulations were used, two or three? One of the reasons is that two models are used, but one of them is used with two different configurations. It would be clearer to say that the paper is using three model simulations, to label these simulations clearly with label choice different from the name of the model, and to use these labels instead of the model name in the paper.*
**Answer:** This was not clear enough, we agree. This issue has been now fixed as suggested, with model names and simulation names clearly defined throughout the manuscript. More details are also given in the answers to specific questions/comments.

*Throughout the reading of the paper, I did not find a clear motivation for this study. While the introduction discusses the processes affecting $O_3$ and CO in the tropical upper troposphere, why $O_3$ and CO trends matter in this region?*
**Answer:** In answer to this comment, we have added the following paragraph to the Introduction, before introducing the MLS UT measurements: "How do changes in the upper troposphere relate to changes in the lower troposphere, such as changes in emissions? There have not been many such studies in the past, in large part because of the lack of well-sampled long-term data in the upper reaches of the troposphere, where ozone is of radiative significance. While this region is not directly related to surface pollution, fast convection episodes in the tropics would seem to imply that there are some correlations between lower tropospheric and upper tropospheric abundances, and even longer-term trends. Long-range transport of pollution can extend into the UT, and back downward with cross-continental impacts on surface pollution levels. Constraints on chemistry climate models are one important goal for studies of long-term measurements of upper tropospheric composition. Such studies are also expected to contribute to continuing assessments of pollutant trends in the troposphere, such as the Tropospheric Ozone Assessment Report Phase II (TOAR-II), while related model simulations are of interest to continuing assessments of chemistry climate models (e.g., CMIP-7)."

Regarding the TOAR-II relevance, two Tables have been added to the Supplement, with tabulated values and uncertainties from our tropical UT trend results versus latitude, in % per decade and ppbv per decade, at the (separate) request from Owen Cooper, who sent us the Figure below (from IPCC AR6 WG-I, chapter 2) as motivation. We also believe that the simulation results shown here will be of interest to the modeling community, in the context of CMIP-7 (Coupled Model Intercomparison Project-7). A new assessment report would presumably include our MLS trend results versus latitude for the middle panel below, along with any other updates.

**Appendix 1: IPCC AR6 Tropospheric Ozone Trends**



*Moreover, I do not see a clear take home message from the conclusions. At some point, splitting Sect. 4 in two would help in that sense.*
**Answer:** Yes, we decided to split things up into re-organized "Discussion" sections (one for $O_3$ and one for CO), followed by a more focused short Conclusion section with the main messages.

*Also, while there is a discussion of the evaluation of MLS data, I did not find such a discussion about the models especially in the regions surrounding the tropical upper troposphere. In particular, how models represent surface observation of CO? This could help to interpret the difference between models and MLS CO in the tropical upper troposphere. In the case a model does not have a good representation of the regions surrounding the tropical upper troposphere, I would exclude it from this study in order to simplify your message.*
**Answer:** CAM-chem and WACCM have been evaluated in previous studies, and we mention them in the introduction: "Gaubert et al. (2020, 2023) found that this version of CAM-chem tends to overestimate tropospheric oxidants, such as ozone, hydrogen peroxide, nitric acid, and hydroxyl radical, resulting in a shorter lifetime of tropospheric methane and CO, mainly in the northern hemisphere extra-tropics." These 2 papers include comparisons with satellite (MOPITT), aircraft (NSF ARIA, NASA/NIER KORUS-AQ, NASA ATom field study), NDACC, and surface in-situ data and describe the inversion of CO emissions from CAMS-GLOB-ANT.

MOPITT XCO columns are more suited for the evaluation of global (coarse grid) models than surface observations. Thus, we did a comparison with MOPITT in the manuscript and we found indeed that models have a good representation of "the regions surrounding the tropical upper troposphere" (here, we mean the total column, which is weighted towards the lower troposphere). We see no reason to eliminate any of the model simulations, since they were run to try to better understand the MLS data and the comparisons with slightly different simulations. Simulating surface data is not a good enough requirement for a good simulation of the broader troposphere, especially the upper troposphere. We have no plans to try to add such details in this manuscript as we find that studying regional surface data would not necessarily be relevant enough for the upper troposphere; moreover, it would require a huge separate effort, way beyond the scope of this already long manuscript (as implied by this reviewer's comments as well).

*When discussing differences between model and MLS and/or their trends, it is very important to make sure that differences and/or trends are significant. When citing trends from other papers, make sure they are significant at the 2-sigma level (and avoid citing non-significant trends). There are many discussions where this is not clear that it is the case (in particular around the hatched regions in Fig. 12 et 14). If these cases need to be discussed, then it should be justified, e.g. because other studies based on other datasets show differences and/or trends which are significant. Otherwise, I would not discuss these cases because they do not help to simplify the whole message of the paper.*
**Answer:** While this largely makes sense for the take-home short conclusions, where we now have the main points more cleanly, we are also sensitive to Referee 2's comments on statistics, namely that just ignoring all differences outside the 2-sigma error bars is not really justified or the best (categorical) approach…trend patterns and tendencies matter to some extent, especially when/if there is consistency in terms of patterns or over broad regions. Indeed, recommendations

to discuss significance in more nuanced ways have also been adopted by TOAR-II assessment guidelines, which refer to Wasserstein et al. (2019) (mentioned by referee 2).

Thus, we keep some of the maybe secondary but worthwhile comments in the discussion sub-sections, but we have tried to minimize this in the “take-home” fairly short set of conclusions. We are somewhat caught in the middle of these somewhat different viewpoints from both referees, and probably cannot satisfy each referee 100%, but we have tried our best (and we will aim for 95%...:)). We see some value in such a “compromise approach”, but we do lean more towards the guidance from referee 2 (and TOAR). We hope that the revised version, with many changes, will be essentially good enough (without reaching “perfection”) to now mostly satisfy both referees, who made constructive (and often also quite detailed) comments.

Regarding more specific comments in this review here, we have indicated in the revised text that several of the trends that are referenced from the literature are indeed significant results, given the quoted error bars (see the specific comments/answers below).

### *Other General Comments*

1. *The multivariate linear regression method is quickly introduced in Sect. 2.3 where the reader is pointing to the Appendix 3 of Froidevaux et al. (2019) which is fine. However, the choice of the different proxies that are used in this study, and the way they are connected to the trend analysis carried out in the paper should be reminded to the reader (e.g. the connection between CO from biomass burning and ENSO).*
**Answer:** Alright, we have added some proxy variable explanations in Sect. 2.3, as follows *(L349–355)*: “…annual and semi-annual periodicities, to account for these known variabilities in atmospheric composition, with 3- and 4-month periodic components to better fit shorter-term (intra-seasonal) variations, which also helps to reduce the trend error bars. In addition, we include functions describing multi-year variations caused by the QBO (which mostly affects the stratosphere) and by ENSO, which has been tied, for example, to regional droughts and biomass burning events, with related increases in convection and transport of surface pollution into the upper troposphere.”

2. *MLS $O_3$ profiles display vertical oscillations in the tropical UTLS as stated in the MLS Data Quality Document (Livesey et al., 2022). This should be mentioned in Sect. 2.1 where it should be justified that trend would not be affected by these oscillations. It is also necessary to give an estimation of the bias introduced by these oscillations that would be useful to remind when discussing the comparison between MLS and the model simulations. The justification of not using the averaging kernels should also be mentioned in that section.*
**Answer:** We have made some changes along the lines suggested. Please see also our comments regarding this in the specific comments section below.

3. *I suggest discussing the time series plots (Figs. 7 and 8) before addressing the trend analyses. By using this order, the authors will visually introduce the evolution of $O_3$ and CO in the tropical upper troposphere before addressing their trends.*
**Answer:** While we have agreed to mention these examples of time series and regression fits earlier in the manuscript, we also agree with Referee 2, who would much prefer to see these Figures in the Supplement. We have thus done that (with some readability improvements in the

plots), in part also to simplify and lighten the paper. While we (and authors of other papers mentioning regression fits) do like this sort of plot, we can also agree that showing these Figures (as examples only) will not really change the main UT trend and variability results.

4. *There is a long discussion about the CO climatologies at 12°S and N (Fig. 9), mainly on the disagreement between MLS and models at 12°N. What is less discussed is that MLS CO at 12°N displays very different climatology than at 12°S, which is not the case for the models whose climatologies are rather similar at 12°S and N. What would be the reason of this difference between 12°S and 12°N in MLS CO?*
**Answer:** We believe that we have answered this question to the best of our ability, based on the following text in the discussion regarding the Figure above *(L550–571)*:
"The MLS CO curves show the two maxima previously observed in seasonal analyses of biomass burning events, with related upward injections of CO and their subsequent transport to the UT being implicated. Based on fire counts from satellite data (see e.g., Duncan et al., 2003, 2007), a March biomass burning maximum has been associated with the northern hemisphere (mainly from Southeast Asia, but also from northern Africa); outflow from the Asian monsoon contributes to the August NH maximum. The September/October maximum arises from the southern hemisphere (Indonesia, Malaysia, Southern Africa, Brazil)."

## *Specific Comments*

*L13: Replace "…chemistry climate models. The models…" by "chemistry climate model simulations. The simulations are from…"*
**Answer:** Yes, this has been done.

*L26: "…CAM-chem and WACCM…". This is confusing, is there two or three model simulations?*
**Answer:** "the WACCM simulation (WACCM-CEDS) and both CAM-chem simulations have similar trends…" This should make it clear that there are three simulations (one from WACCM and two from CAM-chem model, the CAM-chem-CAMS and CAM-chem-CEDS simulations).

*L139-142: Again, it looks only two model simulations are used in the paper.*
**Answer:** We have clarified as follows *(new L163)*: "Altogether, we use one WACCM simulation as well as two separate CAM-chem simulations (the latter two having different anthropogenic emission inputs for CO), as described in Sect. 2, where we provide more details about the MLS data and these model simulations."

*L57: Replace "carbon monoxide (CO),…" by "carbon monoxide – CO,…"*
**Answer:** Yes, done *(new L72).*

*L68: CO produced by biomass burning are also from incomplete combustion so it is redundant with the first part of the sentence. Please, update.*
**Answer:** Yes, done *(L85).*

*L83: I am not sure to understand the word "priorihydrocarbons", could you define it?*
**Answer:** This was a glitch/typo, the corrected word is just "hydrocarbons".

*L92: Is the trend from 2000-2010 significant?*
**Answer:** Yes it is, according to the reference(s) quoted in this paragraph; we have also slightly expanded upon the descriptions of significant trends in these sentences (and corrected the above time period to be 2000–2012). *(L105)*

*L113: Same question here, is the trend significant?*
**Answer:** Yes, the trend results we refer to here are also significant, and we have mentioned this specifically now *(L130)*. Of course, the actual reference(s) provide more nuanced and regional details.

*L123-135: What is the point here?*
**Answer:** We have added motivating sentences in this reorganized paragraph *(L122–140)* regarding the complexity of the UT region. We see no overwhelming reason to change much else in this paragraph, which provides useful references regarding past studies of this region.

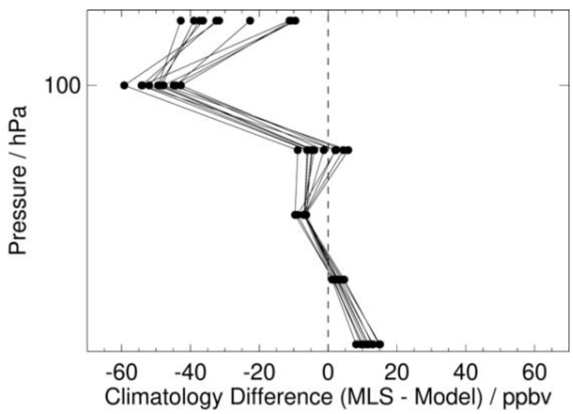
*L138-141: It is not clear how many model simulations are used in the paper? Two or three?*
**Answer:** This is mentioned more clearly in the Table and here (now) in the text *(L156–167)*: We have also made this clearer elsewhere throughout the revised manuscript, including the relevant (revised) Figures.

*L186-192: Could you provide, roughly the precision of $O_3$ and CO in % as well?*
**Answer:** Certainly - the approximate percent uncertainty estimates have been added in parentheses for all the cases given in this section.

*L212: "of70" => "of 70"*
**Answer:** Yes, this has been fixed.

*L305-320: The use of the averaging kernels and introducing the vertical oscillation in $O_3$ profiles must be moved to Sect. 2.1.*
**Answer:** We have agreed to move most discussion of MLS profile oscillations, as well as biases, in the introduction about MLS data (Sect. 2.1). Our longer response (to referee 2) about these issues is copied here for completeness.

From past MLS ozone validation results with respect to ozonesonde profiles, in particular the work by Hubert et al. (2016), we have known that systematic vertical oscillations about mean correlative profiles exist in the MLS UTLS ozone retrievals (see also Livesey et al., 2022, and below). Some unaccounted for low-level systematics in the measurement system (including the retrieval system) seem to lead to these small oscillations; a retrieval vertical grid width that is slightly too narrow might also lead to such oscillations, and the sharp gradient near the tropopause may play a role. We can also illustrate such oscillations by using a (smooth) model ozone long-term climatology as a fixed profile reference, see the plot below, showing MLS climatological (2005–2020) average difference profiles (from avg. model) for a number of 4-degree-wide latitude bins between 20S and 20N.




The oscillations are larger in the region close to 100 hPa, which is above where we are analyzing UT tropical data (we focus on the 3 bottom pressure levels, 215, 178, and 147 hPa). So the above oscillations (not the mean biases) are small, of order 1–3 ppbv, or 2-6%, for the UT levels we use in the manuscript. We do not need to worry much about this in terms of the model average difference from MLS profiles. More importantly, regarding potential impacts on trends, a truly systematic effect will just not change a trend result. Also, if there was an issue of long-term stability/drift for the MLS data, the whole profiles, including the small oscillations, should be affected; the MLS ozone data have been one of the most stable of all remote sensing measurements systems (as we point out in the manuscript as well). Moreover, any random effects affecting differences from correlative data would not show up as a systematic, and would average out over the long-term. One can always wonder, or speculate, about unknown unknowns or systematics that might change with time, but uncovering such an actual issue outside the combined errors of MLS and sondes (for example) would be quite challenging. Sonde profiles are not perfect either, and some studies have in fact used MLS profiles to help identify poorer quality sonde results (or sonde sites), in the more recent years. We did not want to pretend everything is perfect in the MLS $O_3$ data, despite their generally praised high quality overall. We also could not embark on more detailed studies of such a potential issue at this time, with no expectation of a quantitative characterization, let alone a fix, if one could actually pinpoint something that might affect trend results.

Thus, we cannot realistically change much or analyze this topic in more detail for this paper, but we did add to and modify the text (now in Section 2.1) to state: "There are also some biases in MLS tropical UT ozone values, which tend to be on the high side (by 10–20%) with respect to ozonesonde data (see Hubert et al., 2016, Fig. 6), but the above issues are systematic in nature. While we think that neither these biases nor the small vertical oscillations (a few % in magnitude in the region of interest here) would play a major role in changing our MLS UT trend results, given the trend uncertainties, any time-dependent effect, if it exists, would be quite difficult to characterize, or provide a fix for."

Also, the relevant general model profile smoothing comments have now been placed in the model introductory section (2.2). However, the portion of the original lines dealing with specific comparisons of the MLS and model climatologies has been left in the main text where they were, as we believe that this is the more relevant section to discuss that sort of thing (and we believe that it flows better that way). (*L386–391*)

*L311: Could you mention the figure of Hubert et al. you are referring?*
**Answer:** Certainly, we have added "(see their Figs. 6 and 8)" to the text.

*L325-326: replace "model/MLS differences" by "differences between models and MLS", and later, "model/MLS bias" by "bias between models and MLS".*
**Answer:** Done, this has been changed as suggested.

*L348: "Figure 3… and the two models…". I see three model curves on the figure, not two. See also the Major Comments section above.*
**Answer:** Yes, we have modified all the related statements, including here ("three simulations").

*L351: "2-sigma level" of what?*
**Answer:** We have clarified this as follows: "positive and significant (meaning that a zero trend lies outside the 2σ estimate of trend uncertainty)".

L352-353: "The average…". Is the average also for the 3 levels 147, 178, 215? Please clarify.
**Answer:** Yes, we have added "(averaging all three pressure levels)".

*L353-355: I don't understand the meaning of "(we have used the rms of these from the three pressure levels in Fig. 3)"?*
**Answer:** We have clarified this uncertainty estimate by the following: "… 2σ trend uncertainty (calculated here as the root mean square of the 2σ trend uncertainties at all three pressure levels in Fig. 3)." *(L475–476)*

*L359-362: "If a larger…". I don't understand what you mean here, please, rephrase.*
**Answer:** We can explain this more thoroughly, and the point is that the results will not be changed drastically if one uses broader latitude bins, in terms of the significance of the different results; this is because the variability in the time series for various latitude bins is correlated enough (in the tropics at least) that broader latitude averages do not help to reduce the uncertainties very much. We have now written, more specifically *(L500–507*): "Indeed, if broader latitude regions were analyzed for trends, the corresponding trend uncertainties would be reduced, which could make some of the compared trends differ by more than their 2σ error bar variability. However, the trend error reduction in our testing with a 20°-wide latitude bin instead of a 4° bin is only 5–10%, meaning that the uncertainties get divided by much less than the square root of the number of small latitude bins used (an error reduction result corresponding to zero correlation in the temporal variability between bins, e.g., if random noise alone was present). Thus, we do not readily obtain more significant differences in these trend comparisons by just averaging over broader regions." We do not believe that this is often appreciated (or studied) enough.

*L366-380: Is the MLS trend significant or not? It seems not regarding the number given.*
**Answer:** We have clarified that the MLS CO decreasing trend at 147 hPa is different from zero, while the smaller MLS CO trend at 215 hPa is not *(L526–527)*: "based on the error bars, the CO trend from MLS at 147 hPa is different from zero, while the corresponding MLS trend at 215 hPa is not."

*L370. I must say that the Fig. 5 does not allow a proper reading of the error bars while, on the other hand, Fig. 6 do where it looks like the trend is significant. This must be clarified. If it turns out that the trend is not significant, it is difficult to credit sentence "However, there is not as negative a tendency in the latter two model UT CO trends as in the MLS CO trends…"*
**Answer:** Agreed, and we have removed the imprecise nature of the above statements; we mainly point out that the CAM-chem-CAMS trend results for CO depart significantly from the MLS CO

trend results at the northernmost tropical latitudes. We removed the following portion of text: "However, there is not as negative a tendency in the latter two model UT CO trends as in the MLS CO trends, especially if one considers the aggregate values from different latitude bins; thus, there is some room for further improvements in the modeled tropical CO UT trends, although the trends being compared have fairly large error bars." Also, the readability of Figure 5 (as well as the readability of several other Figures) has now been improved.

*L384: Which CO time series? Those shown in Fig. 8?*
**Answer:** Correct, and we have now specified which sample plots we refer to (the plots which are now in the Supplement).

*L422-421: "The fits from the models to the CO behavior at 12°S are quite good." Do you mean "The fits from the models to the MLS CO…"?*
**Answer:** Correct, this is what we mean, so we have added "MLS" in this sentence.

*L422: "These curves…" Do you mean for MLS CO because I do not see double peaks in models at 12°N (Fig. 9a). Also, replace "peak" by "maxima" (and also later).*
**Answer:** Yes, this is primarily seen in the MLS CO curves, so we changed the wording accordingly, as well as wording changes *(L550–571)* to "maxima" or "maximum" instead of "peak(s)".

*L485-486: "with an overall better/good agreement between the CAM and MLS mapped $O_3$ trends." I do not agree here, I don't find that CAM agrees better than WACCM with MLS at 215 hPa. Please, comment.*
**Answer:** We only partially concur; we have deleted the "overall better/good agreement" portion and made this comment more specific, as follows *(L653–656)*: "At 215 hPa, the more strongly positive trends in CAM-chem-CEDS than in WACCM-CEDS over the Australian region (bottom right quadrant, south of the equator) contribute to the better correspondence between the zonal mean $O_3$ trend results (in Fig. 3c) between CAM-chem-CEDS and MLS over the southern tropics." As implied by the above sentence change in the manuscript, something in the mapped simulated $O_3$ trends at 215 hPa (over and north of Australia) has to account for the better general correspondence in the zonal mean trends versus MLS for CAM-chem-CEDS.

*L501-506: I don't see the point here, could you clarify?*
**Answer:** The meaning/investigation regarding the bottom right panel (d) in this (improved) Figure has been explained better now. Please see the sentences regarding the MLS versus TCO trends Figure (old Figure 13, Figure 11 in step 1 revisions) *(L673–683)*: "However, the agreement between MLS UT $O_3$ and TCO trends is often worse for other MLS pressure level choices; this can be deduced from panel (d), where R (correlation coefficient) values relating to the longitudinal variations obtained from MLS at different pressures versus the longitudinal variations in TCO are displayed as a function of latitude (y-axis). In fact, one might not expect the MLS ozone UT trends to track the TCO trends very well, given that TCO measures the entire column whereas MLS measures trends in a vertical region about 5 km wide in the upper troposphere, but this was worth looking into. Regional variability and horizontal sampling differences between MLS and OMI will also play a role (see Thompson et al., 2021, for variability aspects of sonde-derived tropospheric trends). Our comparisons imply that the

correlation between lower and upper tropospheric ozone trends is not a strict "one-to-one mapping", but there are nevertheless some similarities between these regions."

*L540: The $R^2$ figures at 215 hPa should be shown in the supplement.*
**Answer:** These Figures for 215 hPa (for both $O_3$ and CO) are now provided in the Supplement.

*L547: "over the South Atlantic region, which is likely linked to biomass burning periods in this region". I guess there is no biomass burning above the Atlantic ocean, so I would revise this sentence.*
**Answer:** Yes, we have changed this to specify "biomass burning in Africa and related CO transport…"

*L657-660: "The TCO…" I am completely lost with this sentence, please, clarify.*
**Answer:** Please see below for the revised wording of these 2–3 sentences regarding referenced TCO analyses and trend results. We hope/expect that this is now clearer *(L895–901)*.
"The TCO analyses by Ziemke et al. (2019) using combined OMI and MLS ozone columns showed that the TCO trends are larger in the 2005-2016 time period than in the two decades before 2005; for the 2005–2016 period, the derived TCO trends in the tropics are about 0.4–0.7 % $yr^{-1}$ (see also Gaudel et al., 2020). These two investigations found regional differences in the TCO trends, with maxima over India, Southeast Asia, the eastern Pacific region, and the tropical Atlantic, while they obtained near zero or slightly negative TCO trends over the Western Pacific."

*L695-695: "Therefore, …" I do not agree about the room for improvement for the models since the trends of MLS and models (CAM-chem-CEDS and WACCM) agree within their uncertainty.*
**Answer:** We agree that this is somewhat overstated, but now this is replaced by: "these average trend results are statistically in agreement, although the MLS CO trends are generally more negative than the simulation results."

*L698-L699: "… clearly not matching the MLS derived negative… trends" As long as the MLS trend is not significant, I would not say that MLS has a negative trend.*
**Answer:** Agreed, for the most part; we have also made these sentences (see the answer above as well) more specific regarding where exactly the CAM-chem-CAMS CO trends disagree significantly with the MLS CO trends *(L942–959)*: "For the CO trends, the average tropical MLS UT trend is $-0.25 \pm 0.30$ %$yr^{-1}$, whereas the corresponding trends from CAM-chem-CEDS and WACCM-CEDS are close to zero ($0.0 \pm 0.14$ %$yr^{-1}$) for this region; these average trend results are statistically in agreement, although the MLS CO trends are generally more negative than the simulation results. However, the CAM-chem-CAMS simulations (which use CAMS anthropogenic CO emissions, see sect. 2.2), yield positive average tropical UT CO trends ($+0.22 \pm 0.19$ %$yr^{-1}$). More specifically, these simulated trends are significantly different from the MLS CO trends in the 12°N-24°N latitude bins."

*L709-710: Is the negative trend between -0.5 and -1.5%/yr significant? If not, I would exclude this comment.*
**Answer:** Yes, the quoted decreasing CO trend results from (mainly) MOPITT column data are most often significant, especially in the northern hemisphere. Thus, we see no need to really change the main thrust of this related text.

### *Table and Figures*

*Table 1: Add a new column for the Model Name and make sure the Model Designation is clearly different from the Model Name.*
**Answer:** This is a worthwhile change, thank you - please see the new Table 1, which should address this comment; we have also used the simulation names throughout the paper and in the Figures, so this is now clearer everywhere.

*Fig. 3 and 5 to redo with better choice of colours and thicker error bars to improve readability.*
**Answer:** Mainly, we have added enough thickness to the lines and error bars; changing colors is subjective, and here, we are planning to keep the cyan and blue to represent the two flavors/simulations of the same model (CAM-chem). The overall readability has been improved.

*Fig. 7 and 8: Around the linear trend lines, add the uncertainty of the trend in order to see if the trend is significant or not, and if the difference between model and MLS are also significant or not. Also, use different colours. Magenta is difficult to see, Orange and red are not easy to distinguish. And increase the thickness of the lines.*
**Answer:** We did add the trend ranges (two sigma) to these plots as dashed lines around the solid average trend results; however, to truly see if something is significant, we feel that one still needs to study the actual numbers and error bars (and the overall patterns, even if individual trends seem to overlap, as mentioned to some extent by referee 2). Also, the line thicknesses were increased. We changed one colors (now has purple rather than orange and blue rather than cyan) and readability has been improved enough, in our view. However, as suggested by referee 2, and to simplify the main text somewhat, these detailed plots have now been moved to the Supplement (with related discussion remaining in the main text).