Referee 2 (Yann Cohen) report

Referee comment on Froidevaux et al.: Tropical upper tropospheric trends in ozone and carbon monoxide (2005 – 2020): observational and model results

This paper presents climatologies and especially trends in ozone and CO in the tropical upper troposphere, as seen by the MLS satellite instrument and by two models: WACCM and CAM-Chem. The latter has generated two simulations differing by their anthropogenic emission inventories. The time period is quite long (16 years). This paper thus involves important materials, and focuses on a relevant thematic as the model assessment, in the tropical UT that is a complex region.

General comments

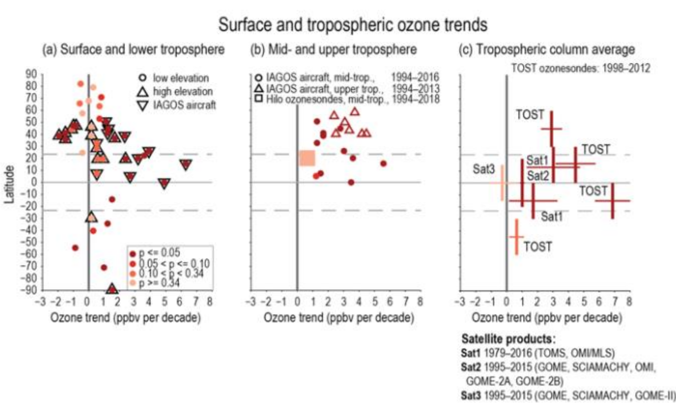
In its ensemble, the paper is unbalanced: it is essentially descriptive, shows lots of details but lacks of interpretation. There is a lot of figures, and several of them are only described without an attempt of explanation, which makes their point hard to get. The paper needs more – and clear – take-home messages to justify the length of the text, and the amount of figures.

Answer: 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 (per referee 1 comments); this change does not modify science results but we hope/think that it has helped, along with all the other changes and adjustments. Per referee 1 comments, 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 2nd step review simpler; there is an added Conclusion and a reference/comparison to a very recent paper on this topic.

Regarding statistical issues of significance, we are caught a bit in the middle of the 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 here 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.

We also have linked this manuscript to the TOAR-II (Tropospheric Ozone Assessment Report, Phase II) special issue. 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



The other numerous but "smallish" changes made in the detailed wording of this manuscript are based on very conscientious reviews that we appreciate. Given that simpler take-home messaging is desired, we will not attempt to (or even pretend we are able to) launch into more complex statistical analyses or language that we imagine very few researchers would find "simpler" or easily understandable, without being a "formal statistician", which none of us are. A few additional useful statistical "details" have been provided (or reworded), nevertheless (see more specific responses below).

Also, some results are interesting but not used. For example, the use of the CEDS emissions with the CAM-Chem model changes the simulated CO trends drastically, but nothing more is said. The very least would be to investigate the question: what are the differences between the two emission inventories that could explain these discrepancies? How do tropical CO emissions behave in both inventories, for example? Or CO precursor emissions? Could the difference be linked to CO primary production, secondary production, or even the CO sink through the whole tropospheric column?
Answer: We have added the following in the text discussion of zonal mean CO trends: "This difference in trends can be explained by significant decreases in Chinese anthropogenic emissions in CEDSv2, despite the increasing anthropogenic tropical CO emissions in both CAMS-GLOB-ANTv5.1 and CEDSv2 (see the new Fig. S4)." We have added the latter Figure to help illustrate the CO anthropogenic emissions issue. Given that this is the only change between the two CAM-chem simulations, we can exclude a change in CO secondary formation or sink between these simulations; we have now mentioned this in Section 2.

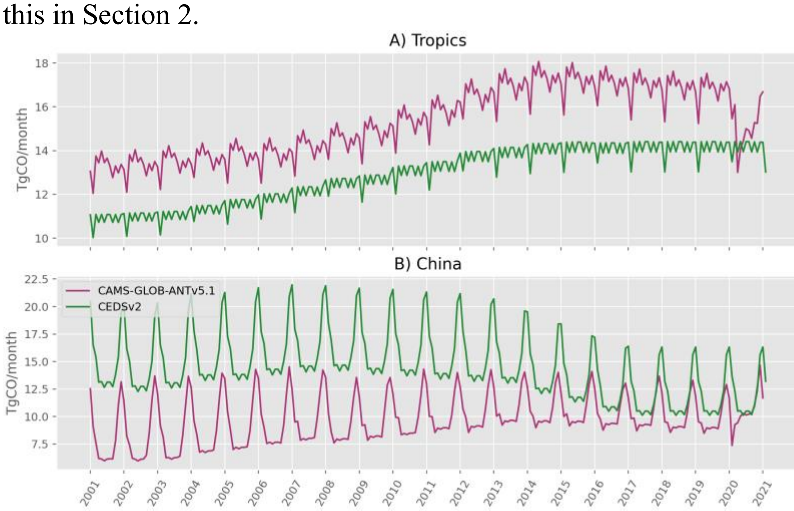


Figure S4. CO emission time series for CAMS-GLOB-ANTv5.1 (Soulié et al., 2024) and CEDSv2 (McDuffie et al., 2020) for (A) the tropics (20°S–20°N) and (B) China.

Another issue is that I found several inconsistencies between the text and the figures, sometimes I even disagree with the result description. Details are given in the "Specific comments" section.
Answer: Thank you; we answer every single comment in the Specifics section below.

Last, the paper is hard to read because of a heavy writing style: phrases are often interrupted by unnecessary brackets, and loaded with successions of adjectives (there are frequent chains of 5 consecutive adjectives, which is too much). The figures are also hard to read, notably the ones with very small characters.
Answer: We have answered every single comment on this front in the Specifics section below.

As the paper needs improvement on several points, I consider my review as major, agreeing with Referee 1. The paper remains promising. I thus hope that my review looks more helpful than heavy, and I am optimistic with the next version of the paper, once it is more organized.

Introduction: *I find the introduction unbalanced, with only one paragraph for ozone versus 3 for CO. The latter is well detailed, with the final focus on the UT, whereas upper tropospheric ozone only has two lines here and there (L52, L134). Reading the introduction, one would not guess that half of the paper focuses on ozone.*
Answer: We do find that there is somewhat more to say about the large CO UT variations, and also regarding the results in our paper, so we give CO somewhat more weight, but the title of the paper speaks for itself regarding including ozone, and the work is fairly balanced overall (even if we feel it does not really have to be measured exactly). With some rewriting and cleaning up, we now have an Introduction that is somewhat more balanced, without loss of information (39% on tropospheric CO, 26% on tropospheric O_3, 20% on the UT part, and 15% on the specific layout and sections of the manuscript itself). We think this is reasonable.

Observations, model simulations, and trend analysis methods:

1/ I know it is already indicated in Table 1, but it should be written explicitly in the text that there are 2 simulations made with CAM-chem, and that the one based on the CEDS emissions is named CAM-chem-CEDS.
Answer: This has now been done in the revised Table and in Section 2.2; throughout the text, we now refer to the simulation names, for more clarity (also per Referee 1 comments).

2/ Could the authors tell about some differences between the CAM-GLOB-ANT and CEDS emission inventories, in order to anticipate the fundamental differences in the CO trends? The only difference mentioned here is that CAM-GLOB-ANT underestimates the CO emissions from South Asia and China, which does not give any clue on why the trends differ so much.
Answer: We added the following explanation and a Figure (see above also) in the Supplement (Fig. S3): "This difference in trends is explain by increasing tropical CO emissions in both CAMS-GLOB-ANTv5.1 and CEDSv2, while there is a decrease in Chinese emissions in CEDSv2 that is not apparent in CAMS-GLOB-ANTv5.1 (see Fig. S3)." The MLS CO trends are negative, in line with Buchholz et al. (2021) for infrared sounders; thus emission inventories with stronger negative trends give improved results.

3/ The CEDS emissions have been found to underestimate the decrease of Chinese emissions since 2013 (e.g. Zhang et al., 2021, https://doi.org/10.5194/acp-21-16051-2021: they mention it in the 3rd paragraph of Sect. 2.1). Could the authors briefly discuss its potential impact, at least on their CO trend calculations?
Answer: The ones used in Zhang et al. (2021) are a previous version (CEDS, v2017-05-18). Here we use the CEDSv2 (McDuffie et al., 2020) with improved negative trends. There is clearly a lack of negative trends in Chinese emissions in CAMS-GLOB-ANTv5.1 (per Figure 12 of Soulié et al. (2024)).

4/ The tropical lightning NOx emissions is relevant, but adding a sentence about the lightning parameterization would bring relevant complementary information (their frequency, and the vertical distribution of the NOx emissions for each flash).

Answer: We added the following to the text: "The lightning NOx production and its role in ozone formation is reviewed by Verma et al. (2021). This study showed that most lightning activity occurs within deep convective clouds in the tropical and subtropical region. In our study, the emission of NO from lightning is based on the Price parametrization (Price and Rind, 1992; Price et al., 1997). This parameterization is dependent on cloud height, which includes a stronger dependence over land versus ocean (Emmons et al., 2010). The CAM-Chem and WACCM models used here derive tropical (and global) lightning NO_x values of 2.34 (3.23) and 2.79 (4.11) Tg (N) yr^{-1}, respectively (Table 1), with no significant trends over the course of these simulations. These global values are within the generally accepted global range of 3-8 TgN yr^{-1} for lightning NO emission (Schumann and Huntrieser, 2007)."

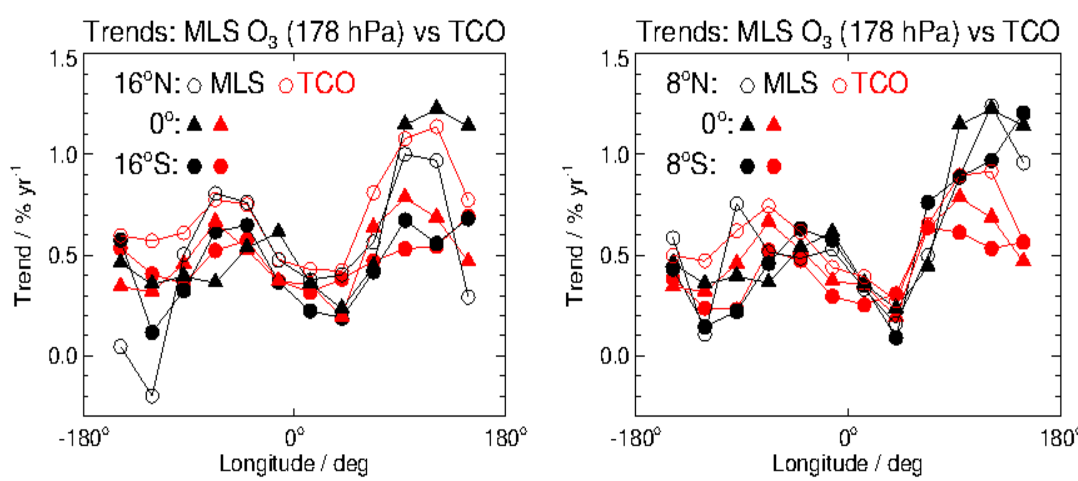
Section 3.2: Zonal mean trends

1/ In Fig. 3, only the 147 hPa level is commented. What about the two levels below?

Answer: We have added some comments pertaining to the three pressure levels, and the text now reads: "The observed average ozone trends at all three pressure levels lie within about 0.3 to 0.5 % yr^{-1}; the peak average trends occur at 178 hPa. There are fairly small latitudinal differences at 178 and 215 hPa. At 147 hPa, the MLS results indicate ~50% larger trends in the NH tropics than in the SH tropics, although this difference is not statistically significant.

2/ No justification is given for focusing on the 12 °S and the 12 °N bins. Is it representative of both tropics? Why not a broader average? I can guess that it is based on the correlation maxima shown in Fig. 13c for the 147 hPa level, but without a clear and explicit justification in the text, the risk is that it looks like a cherry-picking.

Answer: We have added some clarifications regarding the plots displayed in that Figure, to mention that there is some fairly strong correlation in the longitudinal patterns for MLS ozone at 178 hPa and TCO at 12N and 12S. But…we also show that this is not true at all pressure levels, precisely to not just emphasize one pressure level. Regarding "cherry-picking", the adjoining latitude bins for 16S/16N or 8S/8N do not give dramatically different views of the panel below, we add these here for information (likely not needed for manuscript). Thus, we feel that using 12S/12N as a midpoint and the Equator for the deep tropics is a representative "enough" choice. Since tropospheric column (and OMI) sensitivity are centered at lower altitudes than MLS UT O_3 data, we would not necessarily expect the best correlations, but we provided views of some better and some poorer correlations. While this is not the main point of the paper, it is still useful to compare MLS UT trend results to the most relevant satellite-based tropical TCO trends - for the same time period.



By the way, Figs. 7 and 8 are not necessary and I suggest to move them into the appendix, or into the supplementary material.
Answer: We have modified/improved the readability of Figs. 7, 8 (per referee 1 comments), but we also agree with the above and have moved these to the Supplement…

Section 3.3: Mapped trends
L519-540: This paragraph does not propose any comment on the full fit. Still, there is a visible difference between MLS and WACCM in the East hemisphere that would be worth noticing.
Answer: We have added the following text, in answer to this comment. "This somewhat poorer ENSO contribution in the model simulation also exists in parts of the Eastern hemisphere for the semi-annual term; the combination of these differences helps to explain the somewhat poorer overall fits (and variance contributions) for the model than for MLS." We also note that this is also true for the CO case, when this comes up further down in the manuscript. "…but as for O_3, there are somewhat smaller variance contributions in the Eastern hemisphere from ENSO and the semi-annual term than in the MLS case."

L525: Concerning the multiple linear regression model, there is no information on the short-term coefficient, and I did not find more in Froidevaux et al. (2019) on these 3 and 4-month variabilities. An explicit description of this term in Sect. 2 would be welcome then.
Answer: We have done this by adding some clarification to the beginning of Section 2.3 as follows: "Some of the short-term variability can be captured by 3- and 4-month periodic terms (cosine and sine functions with shorter periods than the annual and semi-annual functions); such terms can help to fit intra-seasonal variability and reduce the trend error bars slightly by producing better fits to the time series variability." Incidentally, such regression functional terms have been used previously in atmospheric science literature by others as well.

L541-551: The focus has been given to the 147 hPa level for CO, but nothing is said on the 215 hPa level. However, one could expect the correlation with ENSO to be higher at the latter level. How does it behave there?
Answer: We have added a sentence *(L752–755)*: "At 215 hPa (see Fig. S6), the ENSO variance contribution is slightly larger than at 147 hPa only in a small number of bins, but the overall ENSO-related patterns are not stronger, as seen also in the CO sensitivity coefficients to ENSO in Figure 16[old18] below, which shows only slight differences between the two pressure levels."

Section 4: Discussion and conclusions
L589 - 592: "Potential causes [of the CO discrepancy between MOPITT and the models]..."
→ The authors have also compared the models with MLS through their ozone fields. One could expect both species to help in the understanding of each other's behaviour. Could the ozone overestimation in the UT be representative of the tropospheric column? In this case, could this tropospheric overestimation of ozone be the cause of the CO underestimation through an underestimation of the CO lifetime?
Answer: Indeed, we agree that since ozone photolysis is a significant source of OH, the dominant CO sink, there could be a relationship between ozone overestimation and CO underestimation. However, errors in model CO are also caused by secondary CO formation from a myriad of compounds from vegetation, fires, anthropogenic sources. CO in the UT is also particularly sensitive to transport, with heavily parameterized convective schemes. So, we cannot answer this question with certainty.

L653 - 672: This paragraph starts with discrepancies between satellite-based trends in tropospheric ozone, and ends with a seasonal dependency of these trends from the SHADOZ data set. The message of this paragraph is not clear.

Answer: We have done some slight rewording for the first part, and separated the sonde results out to a 2nd paragraph. *(L891–920)*

Issue about the statistical significance criterion:

L620-622: "We show that the zonal mean O3 tropical UT trend results for different time period choices, with start and end years adjusted by one or two years, do not significantly depart from the 2005-2020 results." → I disagree for CAM-chem-CEDS in the southern tropics, where the difference can even reach 100% of the initial best estimate.

Answer: Yes, we were focusing on the MLS (observed) ozone trend results but did not state this precisely enough. We have now done this, and we note that "there is more sensitivity to the choice of time period in the CAM-chem-CEDS trend results over the southern tropics." This sort of statement is also now mentioned in the main text initial discussion of Fig. 4.

By the way, in the current figure (Fig. 4), some differences here are clearly near the border of the confidence interval, so they are very close to the so-called "significance". Concluding on a "non significant" diagnostic is thus too harsh. I invite the authors to rethink their conclusions about these interpretations. For example in Figs. 4 and 5, even though each latitude bin does not have a 95%-significant signal[1], the spatial coherence[2] is a source of statistical significance.

[1] Here, the signal has to be understood as the difference between two lines.

[2] i.e. the autocorrelation with respect to latitude

More generally, it is not because the confidence intervals are large that nothing can be said about the differences: if I understand well the discussions on this topic, the 2-sigma interval is considered as a validity threshold only for a simplicity purpose (and also because "it is the way we always do"), but it is just arbitrary and cannot be interpreted as "inside this interval, nothing happens, but outside, something happens". The purpose here is not to blame any user of this approach (I did it in my first paper, before I heard about the "recent" discussions on this topic), but to remind that some care must be taken on the use of the "not statistically significant" statement, as suggested by some editors in the American Statistician Association (Wasserstein et al., 2019: https://doi.org/10.1080/00031305.2019.1583913). More specifically in the field of atmospheric chemistry, see Chang et al. (2021: https://doi.org/10.1525/elementa.2021.00035).

As Reviewer 1 suggests to remove the citations of "non-significant" trends in order to make the discussion lighter, the objective is sound but I would disagree with a classification that would oppose the "p = 0.051" and "p = 0.049" cases, because they are, on the contrary, quasi equal in terms of significance.

Answer: We view this as more of an interesting general comment than a specific criticism, as mentioned by the referee himself, but we do not want to downplay this completely, and we respect real statisticians. These advances in statistical thinking and guidance can be either somewhat vague or hard to grasp for specifics to apply to a given topic, or without many months of further studies, which we do feel goes beyond the scope of our current work, and at least regarding the main conclusions here (e.g., measured UT O_3 is trending upward, and measured UT CO is trending slightly downward). We will not completely remove some of the much used language of "statistical significance", because we have to replace it by something better, and it is not really obvious how to do this, even the statisticians seem to agree to that.

Having said this, we have at least already tried to address some aspects of this issue, and some of the specifics from this referee. Of relevance as well, we do mention that if different latitude bins show a strong temporal correlation in the time series' atmospheric variability, there will not be as large a reduction in the error bars as if there was almost no correlation. We have tested this specifically and we have reworded some of our text. See the sentences following the description of Fig. 3 zonal mean trend results *(L500–507)*: "However, the trend error reduction in our testing with a 20°-wide latitude bin instead of a 4° bin only reaches 5–10%, meaning that the uncertainties get divided by much less than the square root of the number of small latitude bins used (an idealized error reduction corresponding to zero correlation in the temporal variability between bins, e.g. if random noise alone was present). Thus, we do not obtain much more significant differences in these trend comparisons by just averaging over broader regions."

Therefore, although latitudinal auto-correlation can play a role, we do not believe there is a need to rethink (or at least greatly reword) most of our text regarding the levels of significance (agreement or disagreement), given this lack of appreciable reduction in trend error bars for significantly broader latitude regions. We agree to point out a few patterns that are nevertheless worth mentioning in terms of positive aspects in some comparisons even if the "simple" analysis of error bar ranges does not strictly "allow for" a strong statement of "significant agreement".

We can also appreciate the comments by referee 1 regarding the need to try to focus more on "highly significant" results and "simpler messaging", which hints at the fact that this is a fairly long manuscript already… (see more in our specific answers to referee 1 and our efforts at re-organization, a somewhat painful task that hopefully helps improve the readability, among other changes).

Specific comments

L39: The abstract ends with a mention of the scientific context, instead of combining the paper results with the bibliography, or showing a perspective: what is the piece of information added by this study? What should be done next?

Answer: At the risk of making the Abstract too long, we have adjusted its last sentence and added a few sentences at the end, to try to answer this comment and the perspective issue. See the following last paragraph of the Abstract:

"The MLS-derived upper tropospheric tropical trends in these species arise from a well-sampled multi-year data set, with the results showing a first-order correlation to large-scale changes in lower tropospheric composition (O_3 increases and CO decreases). We find that there are broad similarities (and a few differences) between the measured UT trends and corresponding results from model simulations, which incorporate state-of-the-art representations of the complex interplay between emissions, photochemistry, convection, and transport in the upper troposphere and lower stratosphere. These results will contribute to the continuing assessments of tropospheric evolution, in particular the large community efforts regarding TOAR-II and CMIP-7." If this becomes too long for the Abstract, an alternative would be to just include the last sentence above; the Conclusion section might be a better place for keeping the longer version of the above text.

L49: "Global anthropogenic emissions dominate the natural NOx sources"

→ Given that the study focuses on the UT, it would be worth mentioning the particular case of the free troposphere, where lightning is the major source of NOx (especially in the tropics). See for example the review from Verma et al. (2021):

Verma, S., Yadava, P.K., Lal, D.M. et al. Role of Lightning NOx in Ozone Formation: A Review. Pure Appl. Geophys. 178, 1425–1443 (2021).

https://doi.org/10.1007/s00024-021-02710-5

Answer: The model estimates of lightning NOx do not show a significant trend. We added information about the model lightning NOx in section 2 (including a reference to Verma et al., 2021).

L60: "decreases in NOx emissions over some parts of the world" → Which parts of the world?

Answer: We have specified this further as follows *(L71 onward)*: "Changes in tropospheric ozone precursor emissions (e.g., from NO_x, carbon monoxide – CO, and volatile organic compounds) have been implicated as causes for global tropospheric ozone change over the past few decades (Zhang et al., 2016; Zheng et al., 2018; Liu et al., 2022; Wang et al., 2022). Souri et al. (2017) and Zhang et al. (2106), for example, discussed the existence of decreases in NOx emissions over developed countries following emission regulations after the turn of the century."

L62: Concerning the "significant reductions in northern hemisphere tropospheric ozone values" during the COVID pandemic, it is not as simple. It would be worth mentioning some cases (at least western Europe) where nighttime ozone has increased because of the reduction of its titration by NO, and that the particular weather conditions during this spring (more sunlight, caused by a anomalously strong polar vortex) tended to enhance daytime ozone.
Answer: We have not stated that it was simple and the article we cited considers emissions, chemistry (including diurnal variations in photochemistry), as well as weather patterns. From these references, we do not see a problem stating that "significant reductions in northern hemisphere (NH) tropospheric ozone values were observed in 2020 and 2021, although the tropical decreases are much smaller."

L88: "[...] model simulations, which generally showed slight underestimates of [...] CO abundances"
→ Not that slight, seeing that Park et al. (2021) find an underestimation of CO even while doubling the Indonesian fire emissions in October. Another point concerning this citation: it is worth mentioning that the model used by Park et al. (2021) is CAM-chem, as for the current paper.
Answer: Yes, we have removed "slight"; we have also added the mention of CAM-chem as the model used in that study.

L110: The paragraph starts with in situ data, and thus gives the wrong impression that it will focus on it. I kindly ask the authors to start the paragraphs with the main idea (here: upper-tropospheric CO) before entering into detail.
Answer: Yes, we have done this at the start of the paragraph *(L122)*.

L113: "Decreasing CO emissions from anthropogenic and biomass burning sources" → Is it really decreasing from biomass burning sources?
Answer: Yes, there is evidence regarding decreases in fire emissions (see Jiang et al., 2017; Andela et al., 2017). These references were added in the text at this location.

L130 - 135: Is it still linked to ENSO, as the paragraph seems to be focused on it? Does the complexity of the UT need another paragraph instead? Or should it start the paragraph, as it seems that it is dedicated to the UTLS? As before, the first sentence gives the wrong impression that the paragraph focuses on the effects from ENSO on the UTLS, whereas it rather seems to discuss about the UTLS complexity.
Answer: There has been some rewording and additional text here *(L122-153)*.

L132: Wang et al. (2022) is relevant but presents a single model analysis. It would be worth giving a more representative picture of the aircraft NOx research by citing one or some multimodel analyses. Some examples: Hoor et al. (2009), Sovde et al. (2012), Brasseur et al. (2016), or there is also the review from Lee et al. (2021).
Respectively, the DOIs are:
- 10.5194/acp-9-3113-2009
- 10.5194/gmd-5-1441-2012
- 10.1175/BAMS-D-13-00089.1
- 10.1016/j.atmosenv.2020.117834

Answer: This seems quite appropriate, thank you for the specific suggestions. We have added the two more general studies (Brasseur et al., Lee et al.) and the earlier study by Hoor et al. to the mentioned text, where we refer to aircraft NO_x.

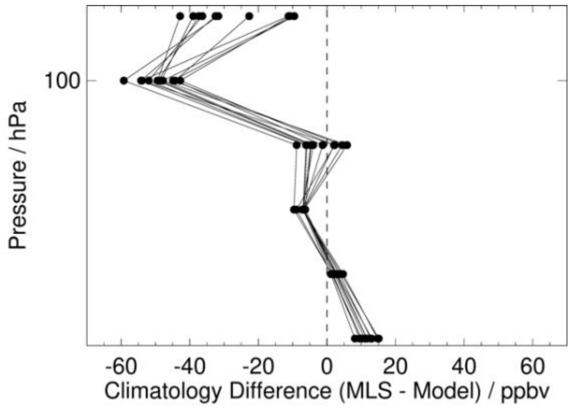
L177: "typically of order a few percent or less" → How few? An estimation would be useful, else it remains too vague.
Answer: Yes, we now more specifically state *(L201)* "typically 1–3%".

L187: Is this reduction of the uncertainty (linked to precision) based on the hypothesis of a zonal homogeneity? It is just for a clarification on this estimation, I do not intend to criticise it.
Answer: The fields need not be zonally homogeneous to try to average down the noise in the retrievals. These zonal or lat./long. estimates of the retrieved precisions are based on Gaussian statistics for random variations/noise, using the retrieval single profile precision estimates (largely based on MLS radiance noise propagated through the retrievals) as a starting point - and these do not vary very much between each profile within a given latitude band. It can be somewhat of a lower limit on the average results' precisions, as it assumes zero correlation between profile random errors/precisions, but this is also just a first-order useful sort of estimate. Also, the trend estimates are really not dependent on this small variability (compared to atmospheric variations), and the measurement precisions practically do not change versus season or year.

L199: The references about the wave-one pattern, as given later in the paper, would be relevant here as well.
Answer: Sure, we have added those relevant references here as well *(L231–232)*.

L293: "differences of this order [5 - 10%] are also observed in the mapped fields"
→ I agree for WACCM, but this statement does not fit with CAM-chem-CEDS.
Answer: OK, we have changed the "5–10%" mention to "5–15%" *(L374)* to make this more correct overall (and without adding even more complexity to the text). Being more precise is a good thing, nevertheless.

L316: What are these "vertical oscillations in zonal mean LMS UTLS O3 profiles"? Where does it come from? Why is it not a normal thing? Nothing is said about it in the paper.
Answer: 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 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."

L319: "these biases should not have a significant impact [...]" → Why?
Answer: Please refer to the answer above.

L323: "the zonal mean model biases are usually less than 10 - 20 %" → I disagree with this statement. I would agree on 20 %, but not 10 %. "Usually around" would be more correct.
Answer: Yes, "usually around 10–20%" is fine, we have made this small change *(L412)*.

L326: "The model CO biases shown in Fig. 2 are about -5 to -15 %" → On average? If not, then I disagree as well.
Answer: Yes, we mean "on average" for these biases, and one can see what the dominant colors are, in the percent difference plots. We deleted the mean bias values here, as they are already mentioned

(with -10% to -20% being a better representation) two sentences prior; we simply adjusted the sentence to state “The model mean CO biases shown in Fig. 2 are well within…”.

L357: “with excellent agreement with MLS from CAM-chem-CEDS” → For the average trends, but also for the pattern, seeing the higher best estimates in the northern hemisphere than in the southern hemisphere. This positive point can be highlighted as well.

Answer: We do agree that there are some interesting tendencies/patterns in the comparisons that one can note, although one should be careful not to over-interpret such things either… (going back to the point from referee 1 about statistical significance, and see below as well). We have added the following sentence: “This agreement is also apparent in the latitudinal pattern, with larger trends in the NH than in the SH, even if the error bars are large enough that there is no statistically significant difference between the hemispheres.” *(L490–492)*

L359: This phrase is hard to understand: it took me time to get its meaning and I still did not get its point: what is the goal of giving this detail? Thus I suggest to simplify it.

Answer: This point actually is relevant to the comments above…We now explain this somewhat better, hopefully, but the point is that nothing will change drastically if one uses broader latitude bins, in terms of the significance of the different results. We have now written *(L500–507)*: “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 ranges). 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., from random noise). Thus, we cannot readily obtain much more significance in these trend comparisons by just averaging over broader regions.” We hope that this actually clarifies this issue for the referee(s), and especially for any reader who might want more details about the statistical analyses and significance, as we find this to be of some relevance in this context.

Nevertheless, we agree that certain broad patterns can still have some meaning, even if the pattern differences are encompassed by the error bars. We also feel that trying to go into this in more depth is beyond the scope of this work, and it may well be that many (most) authors are guilty of oversimplifying statistical comparison results (and trend analyses). We do like to be somewhat conservative in our interpretation of results (a viewpoint of the first referee), but we also see the point made here, and we have tried to address at least part of this via broader latitude bin tests/discussion. We will probably have to leave this topic at that, also since this manuscript’s two referees do not exactly agree, based on their (initial) comments. There is no doubt in our mind that actual statisticians can and do write many pages about such things (see the reference mentioned earlier by this referee), and each analysis/result will be somewhat different as well. Interesting topic, but this could also be a never-ending quest for better wording and/or statistical analyses - and for many scientific articles, not just for our manuscript…

L364: “showing the relative insensitivity of the results to the choice of the time period” → I totally disagree on this statement for CAM-chem-CEDS in the southern hemisphere, whatever the pressure level. I agree only for MLS, and except for the northern hemisphere at 147 hPa: here, the ozone trend is obviously sensitive to the 2005-2006 years.

Answer: Yes, we have added the word “MLS” in this sentence; these MLS trend differences, even in the NH at 147 hPa, are well within the error bars, but the wording “relative insensitivity” is there to make it clear that this is not complete insensitivity to the time period chosen. We have also added some nuances regarding the model sensitivity in the sentence that comes afterwards (per a similar comment made elsewhere by this referee). *(L509–514)* “This is also true for the CAM-chem-CEDS trends in the NH tropics, although there is more ozone trend sensitivity to the time period choice in this model’s results over the SH tropics. The WACCM-CEDS tropical UT ozone trend results versus time period (not shown here) lead to a spread in the SH tropical trends that is about halfway between the small MLS trend spread and the larger CAM-chem-CEDS trend sensitivity shown in Fig. 4."

L389: The weaker peaks in CO from the models might still explain their weaker trend "uncertainties", as the latter is linked to the interannual variability. It depends on where the peaks are located in the time series. How do the MLS confidence interval behave during this sensitivity test? Do they become comparable to the ones derived from the models?
Answer: Yes, we did mention that it matters where the largest peaks are located, and fortunately, they are not really at the edge of the current series. This was not a test to try to reduce the overall variability to be as small as the simulated variability, so the reduction in error bars does not reach as low as the case obtained in the simulated series analyses, as only a few of the largest peaks are reduced somewhat in size. Also, we have pointed out that there are places where the model variability is just too low, without even counting the larger excursions. We do not see much more value in pursuing such hypothetical scenarios any further.

L414: "R is negative" → So close to 0, the sign of the R coefficient does not truly have a meaning (one cannot say that it is anti-correlated, it is only not correlated). A less confusing description would be something like "the correlation is weak".
Answer: Alright, we now just specify that "R is very small". [note: one can easily observe some periods of fairly strong anti-correlation – but cancelled out overall by better phase agreement elsewhere]

L432: "the lack of correlation [...] stems from the poorly modeled double peak structure"
→ To me, it is more that the model has difficulties in reproducing the maximum itself.
Answer: There is a double maximum in the MLS data; if there was a 2nd maximum in the model curves, the correlations would improve significantly. We can disagree somewhat here in the details, but regardless, the correlation is poor. We have decided to leave this part of the text unchanged *(L575–576).*

L443 - 446: "The MOPITT values are obtained from V9J…" → This description of the tool should be in Sect. 2, not in the middle of the results.
Answer: This suggested change sounds fine, so we added the following sentence to the observations section (2.1) instead: "We also compared the simulations with Terra/MOPITT, obtained from multispectral retrievals (V9J) Level 3 dry air total column data, or X_{CO} in ppbv (Deeter et al., 2022). Simulated CO are smoothed by using the MOPITT a priori column as well as the 10 layers a priori and averaging kernel profiles, as recommended for a quantitative comparison of modelled and MOPITT X_{CO}."

L450: "although R is slightly smaller at 12°N than at 12°S" → It is smaller, not "slightly smaller".
Answer: Although this depends on what "slightly" means, we agree to change this to "smaller" – not a big issue.

L459-463: This sentence is too heavy, with these brackets.
Answer: Agreed, we have broken this into two sentences, with no brackets.

L460: "related to emissions, convection, and/or transport" → Which transport? Convection is also a transport processus.
Answer: We now just mention "transport" (but it could be slow advective transport rather than faster convective transport).

L480: "not statistically different from zero" → I am not sure that "statistically different" is meaningful. And as said before, saying "not statistically significant" is not sufficient.
Answer: We have referred to some of the discussions mentioned by this referee. "Hatched bins indicate trends for which the 2σ uncertainty range encompasses the zero trend value which is often interpreted as a low level of "statistical significance", although one should be cautious (see the previous Section) regarding the strict application of such a criterion or wording." (L637–640). We refer back to the text we added on this topic earlier (L494–498)): We note that statisticians have been working to guide or adjust "common practices" regarding statements of "significance", and one should be sensitive to some of the broad differences that occur even within the so-called formal criteria (such as 2σ or a p-level of 0.05), which could sometimes be interpreted in too stringent a way (Wasserstein et al., 2019), and as pointed out by Y. Cohen (private communication, 2024).

L481: "The largest MLS trends are observed [...] and Africa" → I disagree about Africa. On the northwest coast of Africa, maybe, but not on the continent.
Answer: Agreed, we have removed "over Africa".

L485: "with an overall better/good agreement between CAM and MLS mapped O3 trends" → to be developed
Answer: We have made this comment more specific (also in answer to a comment by referee 1) as follows *(L654–657)*: "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."

L486: "Broad regions" → Which ones? It is not obvious.
Answer: We have now specified the regions as follows: "Broad regions with positive tendencies are observed in both model trend results; these regions include Southeast Asia, Indonesia, northern Australia, the Atlantic, and northern Africa, with some, but not exact agreement with the regions mentioned above for the larger MLS trends. *(L643–654)*

L487: "The error bars are large enough that the level of trend discrepancy is very rarely statistically significant." → I insist on it, but it does not suggest that there is no difference. This interpretation is an oversimplification. It is visible that there are important discrepancies in this figure, notably on the easternmost longitude band.
Answer: We will go along with this specific comment. The statement has been reworded along the above lines as follows: "The mapped trend discrepancies between the simulations and MLS are rarely outside the 2σ error bar ranges. Nevertheless, some of the discrepancies are worth noting, especially when they cover multiple adjacent bins; in particular, the easternmost longitude band shows MLS trends with (significant) positive values, in contrast to the simulation results, with binned trends that are often small and/or negative." *(L657–661)*

L489-506: This paragraph is only descriptive. What is the deduction that we can make from it?
Answer: We have added a sentence with some interpretation regarding these interesting, yet imperfect ozone trend correlations between the lower and upper troposphere: "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." *(L683–684)*

L493-497: This methodology description takes a third of the paragraph volume. I suggest to move it into the methodology section.
Answer: At least one of these 5 lines is still needed to mention the basic Figure specifics being discussed, and since this MLS vs TCO comparison is still a fairly small sub-section, we prefer to keep the whole description here, rather than breaking up a few lines to be mentioned sort of in a vacuum in the early Section. This is subjective preference, at some point; we have therefore made no change here.

L499: "variations of a factor of 2 to 3 [...] between the western and eastern hemispheres for both sets of trends " → The factor can be much lower, especially for the TCO at 12°S and 0°where it can even equal 1. The "2 to 3" range is thus incorrect, unless I misunderstood the meaning.
Answer: We were just noting that several variations of order two to three are observed between East and West, as these do stand out. One should probably not try to argue over small details in too many sentences - even though we really appreciate the great care that was taken here in reading this manuscript and we tend to agree with most of the comments, as seen in our responses. Here, we have slightly amended the wording to read: "variations of a factor of two to three are observed, mostly in the northern half, between the western and eastern hemispheres for both sets of trends,…" *(L672–674)*

L505: "given the sensitivity of the derived TCO trends" → It has to be precised that it deals with the sensitivity to the upper-tropospheric variations.
Answer: Yes, although this was implied/understood, in our view. The new wording reads: "…TCO measures the entire column whereas MLS measures trends in a vertical region about 5 km wide in the upper troposphere."

L513: "generally near zero but often slightly positive trends" → Once again, seeing the figure, the best estimate is not generally near zero. Half of the bins shown by CAM-chem have a trend higher than 0.3%/yr, which is not near zero.
Answer: We were too imprecise, agreed, and we have decided to expand the discussion of these comparisons. We point the referee to the (now expanded) paragraph *(L689–707)* where we show how the model trend comparisons to the (mostly negative) MLS trends are improved significantly by the use of the CEDS emissions, although MLS still shows the most negative trends; whether this is still a "significant" difference could be argued, since the use of statistics can indeed be complex, but we think this discussion has been improved, without trying to get too caught up in categorical statements.

L516: "although the vast majority of the model CO trends obtained here are not statistically different from zero" → The fact that the great majority of the bins show a negative trend should not be neglected just because their confidence interval include the zero value. The sign of the trends still shows consistent geographical structures, and not only a random noise around zero over the map.
Answer: Yes, and our expanded discussion (mentioned just above) now touches on this point in a better and more complete way than before.

L529: "at least over most of the Pacific" → The "at least" can be removed, as the described feature only concerns the Pacific.
Answer: Alright, we deleted "at least".

L537: "ENSO correlations in the CCM are often weaker than observed for the MLS ENSO R^2"
→ The figure still shows stronger correlations with ENSO for the model, in most of the western hemisphere. I would need a histogram to be convinced by this statement.

Answer: Indeed, we did point out the better agreement between model and data in the western hemisphere. Regarding the eastern hemisphere and the (slightly) weaker model ENSO R^2 versus MLS, we have decided to delete this part, as it is also not really that apparent at 215 hPa (see the added Figure in the Supplement).

L543: "large ENSO-related peaks [...] which the regression model, as designed, can only imperfectly match" → And not for ozone? I do not understand this argument: why is it specific for CO?
Answer: There is larger variability and extrema in the CO time series over the last two decades, with more readily observed disturbances for CO pollution than for O_3. There is only so much variability that the ENSO term can match, and a regression usually leads to an average response, without proper fitting of the very largest extrema. It may be that another proxy could be found to provide a better match, beyond the ENSO proxy (e.g., with a more direct relation to convection and upward transport?); differences in surface loss for these species may contribute as well (so another proxy for this might help). For this already long manuscript, we have not considered additional work on more or better proxies, or studies along these lines - and the potential for improved fits. This is as much as we can currently answer here.

L549: "The ENSO-related correlation patterns are broadly similar to the ozone case"
→ In which way is it similar? Is it simply the higher correlation over the Pacific than elsewhere?
Answer: We have specified our broad view/comment as follows: "…in that there is larger variance in the more extreme longitudes of both western and eastern sides." *(L751)*

L575: "For O_3, the models display a slight underestimate [...] at 215 hPa."
→ Not only slight. The underestimation reaches -20% near the equator.
Answer: Alright, we now state: "…the models underestimate the mean MLS values…" *(L789)*

L576: "at 147 hPa [...], the models are biased high by at least 20 %"
→ This "at least" is not clear. Either it refers to an intermodel comparison of the maxima near the equator (then it is true that the model biases reach a maximum of at least 20 %), or it refers to the whole meridian profile, but then the models are rather biased of 12 %, at least.
Answer: We have limited our discussion to 147 hPa and now state: "models are biased high by about 15–25%..." *(L790)*

L577 again: "For CO, the model underestimates the MLS UT values by 10 - 20%"
→ Which model? If it refers to WACCM, then I agree. If it refers to both models, then the 10 - 20% are not correct for CAM-chem-CEDS in the southern tropic, which negative bias is rather at 0 -10%.
Answer: Alright, to cover both models more precisely for all latitudes, we have modified this by stating "by up to 20%". *(L792)*

L597: "we note that ACE-FTS UT CO monthly zonal mean time series track those from MLS."
→ And what does it imply?
Answer: We have added: "; this helps to validate the UT time series and variability from MLS." *(L825)*

L618: "excellent agreement with the above result from CAM-chem-CEDS O_3 zonal mean trends"
→ Which metric is in excellent agreement? On average through all the latitudes, then ok. But if it is latitude bin per latitude bin, then this statement is not true. Some precision must be made in this formulation.

Answer: Yes, we do mean the averaged result over latitudes. We have now specified "the (averaged)"…CAM-chem-CEDS O_3 zonal mean trends…, although our own impression is that this was clear enough (given the numbers being compared and the sentence before this one stating "averaged").

L628: "CO is retrieved using the same radiometer as the MLS standard ozone product"
→ And is this radiometer similarly sensitive to the CO and ozone spectral signatures?
Answer: As mentioned in Section 2.1, the single profile precisions are within a factor of two for these products, with the CO retrievals being less sensitive than ozone. However, as long as we do enough averaging (and monthly zonal means or even the gridded retrievals do so), the signal strength is not an issue, and trends can be detected, with atmospheric variability being the important factor regarding the trend uncertainties. We see no reason to expect significantly poorer instrumental sensitivity to trend detection in the CO trends versus O_3 trends. We also see no strong need to really change this part of the text either.

L630: "The largest MLS-derived mapped ozone trends are observed over [...] the Atlantic and Africa."
→ For the Atlantic, it is true for the northern hemisphere only. For Africa, I disagree.
Answer: Agreed, we were too broad in this statement; we have deleted "Africa" and specified the "northern Atlantic region". *(L857)*

L631: "The Pacific region exhibits small or slightly negative trends" → Not the west half of the Pacific (especially not for the 147 hPa level), where the trends are essentially positive.
Answer: Yes, we agree that there is not a lot of consistency over the Pacific; we have thus decided to delete these comments about the Pacific, here and in most of the main text.

L634: "In terms of the mapped model ozone UT trends, they [the TCO trends] broadly match the MLS-based UT trends, albeit with somewhat smaller variations." → It is ok for the broad "green area" where the relative trends are similar between the two products. But it is worth mentioning the extrema as well. The peaks are similar in the North Atlantic and Southeast Asia. For MLS however, the west Pacific positive trends are centered around the equator, whereas it is shifted north for the TCO trends. In the North Pacific, the central and eastern parts show negative trends with MLS only.
Answer: We have clarified the mapped trend comparisons here, while including some of the comments made by this referee. Also, the revised version separates the TCO comparison more clearly now *(L856–869):*
"The largest MLS-derived mapped O_3 tropical trends (up to $+1.4\%yr^{-1}$) are observed over Indonesia and East of that region, as well as over the northern Atlantic region. In terms of the mapped model O_3 UT trends, they broadly match the MLS-based UT trends, albeit with somewhat smaller variations. The significant model maxima over Southeast Asia and the North Atlantic are similar to the significant MLS patterns in those regions. More qualitatively, the Indonesian region displays smaller model O_3 trends than those derived from MLS data; parts of the western Pacific region exhibit some negative trends in the MLS and model trends, but not with good spatial correlation.

The mapped MLS-based UT O_3 trends and TCO trends for the same period (see [old] Fig. 13, based on the analyses of Ziemke et al., 2019) provide good correlations in parts of the tropics, with similar values and longitudinal patterns. However, the MLS UT O_3 trend maxima over the western Pacific are symmetric about the equator, whereas the TCO maxima in that region are found in the northern part only. Since the TCO measurement weighting does not favor the UT region, we would not necessarily expect a really high correlation versus the MLS UT trends."

*L673: "Zhang et al. (2016) and Wang et al. (2022) have ascribed the positive sign of **post-2000 tropical ozone trends** to an equatorward redistribution of surface emissions over the years."*
→ I have two comments here.
1/ "post-2000" can bring confusion, as the period starts in 1980 for Zhang et al. and 1995 for Wang et al.
2/ They show that the equatorward redistribution of surface emissions is the main factor, but the other factors are still not negligible.
Answer: 1) We changed the wording to state "over the last four decades". 2) We do not think there is a need to change any other detail beyond the main point about redistribution towards the tropics.

L677: "Most of the UT model ozone trends shown in our work are significantly (> 30 - 50%) larger in the NH tropics than in the SH tropics." → Concerning the "> 30 - 50%" estimation, I agree for WACCM but not for CAM-chem-CEDS.
Answer: The only pressure level where this does not hold is at 147 hPa for CAM-chem-CEDS (Fig. 3); we are just modifying this sentence slightly as: "The UT zonal mean model O_3 trends shown in our work are typically larger (by ~30–50%) in the NH tropics than in the SH tropics."

L686: "These authors [...] (mostly in the extra-tropics)."
→ Is it still related to this paper then, if it is mainly an extratropical feature?
Answer: Not that much, agreed, except for some advective transport-related influence (possibly). It still may be worth pointing out, for more general information; in addition, the sentence after this one mentions that VOC changes are less likely to be relevant for the tropical regions anyway. No change in the statements has been made.

L699: "the zonal mean CO trends do not differ in a significant way from the 2005-2020 results"
→ I totally disagree with this affirmation, even more than for ozone. Figure 6a and 6c show substantial differences at every latitude, and Fig. 6b as well in the North hemisphere. We cannot conclude to a low CO trend sensitivity to the first and last years, especially for the last years with MLS.
Answer: Agreed, we have changed the text here and in the discussion of Fig. 6 to state *(L959–961)* that "larger MLS CO abundances in 2020 explain why the MLS CO UT trends are more negative if one stops the analyses in 2018 or 2019." That is indeed the largest sensitivity over these particular years; the error bars are fairly large – but the sensitivity/tendency exists.

L727: "(this species also having OH as a major sink [...])" → … and being theoretically not emitted anymore.
Answer: While there are certainly some uncertainties about OH trends and assumptions about methyl chloroform-based estimation of OH trends, we do not see any problem with our sentence here.

Technical comments

The purpose here is not to make a correction of the typo or language mistakes, firstly because I do not pretend to have a sufficient English level for it compared to many native English speakers, and secondly because it should be the role of the journal itself rather than that of the reviewers or scientific editor, unless their institutes are paid for this task.

L55 and 56: "changes in ozone precursor emissions" is repeated in two consecutive sentences, it is quite redundant for the reader.
Answer: Yes, we have deleted this from the first sentence and kept it in the second, more general, sentence.

L57: "(CO)" → Please avoid the imbricated brackets. And here, there is no need to precise the CO chemical formula, as it is done in the next mention of CO.
Answer: We have chosen to give the chemical formula once, here, where "carbon monoxide" first appears in the manuscript; we have also avoided the double parentheses.

L73: This multiple citation is a long interruption of the phrase. Given its length, it could be better to move it at the end of the sentence instead.
Answer: Yes, we have restructured this sentence in a more appropriate way.

L152: "minimize [...] results that might depend more on lower stratosphere"
→ Minimize is not a relevant adjective to qualify "results".
Answer: Yes, we have added "to avoid" [results…].

L171: Heavy sentence here. Why not starting with "In this work, we use time series based on zonal means, and latitude bands subdivided into 12 longitude-latitude bins.", and then describing these two approaches one by one, in separate phrases?
Answer: Alright, we have reworded the first part of the sentence as follows *(L197):* "In this work, we use monthly mean time series based on zonal averages as well as latitude bands divided into 12 longitude bins." The next sentence provides the additional details (with no changes).

L174: "is or order" → "is of order"
Answer: Yes, fixed.

L247: "SSP5-85" → SSP5-8.5
Answer: Yes, fixed.

L294: "(and the percent different fields)" → No need to clarify it, because the absolute values were not mentioned anyway.
Answer: Yes, fixed.

L294 again: "The differences reach about 20%" → It is worth adding the sign of the difference (-20 % then): this way, there is no risk of confusion with the +20 % found for the 147 hPa level.
Answer: Yes, fixed.

L317: "Livesey et al., 2022" → I did not find the document in the link given in the reference section.
Answer: Our mistake, the proper link for the version 5 MLS data quality document (by Livesey et al.) is actually https://mls.jpl.nasa.gov/eos-aura-mls/documentation.php .

L353: The ozone trend given here is for 147 hPa: it seems that the author forgot to precise it.
Answer: No, this is an average trend over three pressure levels in the UT for O_3. The sentence has been reworded for added clarity: "The zonal mean MLS ozone trend, using an average from the three pressure levels at 147, 178, and 215 hPa, for 2005–2020 in the 20°S–20°N UT region is 0.39 ± 0.28 %yr^{-1}."

L547: "over the South Atlantic region [...] biomass burning periods in this region"
→ One could understand that there are biomass burning events in the ocean.
Answer: Yes, we have now specified biomass burning "over Africa".

L557: "the Eastern side of the maps" → This formulation can be replaced by "the Eastern hemisphere".
Answer: Yes, done.

L577: "an average negative bias in the MLS UT values" → Instead of UT, the authors probably mean 147 hPa, unless they do not consider 215 hPa as upper tropospheric.
Answer: Yes, we reworded this to state: "average negative bias in the corresponding MLS values."

L580 (and 582): "a low model CO bias", "low biases" → Does "low" mean "negative"? It can be understood as "close to zero".
Answer: This has been rephrased for enhanced clarity as follows: "…found that model CO values from a (WACCM4) simulation at 147 hPa were smaller than the ACE-FTS (and MLS) CO abundances."

L586: "we find significantly poorer matches at 215 hPa in the NH tropics" → Seeing the text before this citation, it is not clear to me whether the "poorer" adjective compares to the southern hemisphere, or to previous studies.
Answer: We have modified this sentence regarding CO comparisons for the southern versus northern (tropical region): "poorer matches at 215 hPa in the northern tropics than in the southern tropics."

L593: It is worth precising that the variability mentioned here is a temporal variability.
Answer: Certainly – done.

L645: "The typical trends from Gaudel et al. (2020) are generally in accord with (but somewhat larger than) the average tropical UT ozone trends we obtain from the MLS data"
→ This formulation is too vague: what does "generally in accord" mean, concretely?
Answer: We have now added the average trend value (0.5%yr^{-1}) from their study of (5) tropical regions, translated to units we can use to compare to the MLS trends; we then conclude the following, somewhat more concretely (with caveats) *(L884–889)*: "The above average trend agrees quite well with the average tropical UT O_3 trends (0.39 ± 0.28 %yr^{-1}) we obtain from MLS, which provides more uniform (and daily) tropical coverage. This seems to be unexpectedly good agreement, since there are different sampling characteristics, regions, and time periods for IAGOS versus MLS; given the time period differences, in particular, we should only consider this to be a loose comparison." [The point about ongoing assessments is that efforts can be made to compare at least the time periods more precisely]

L652: "We refer to fairly general past IAGOS-based results" → Does "fairly general" mean "positive trends for ozone and negative trends for CO"?
Answer: Yes, but since we have said it is beyond the scope to give more detailed comparisons in the tropical UT, we have just deleted this last part of the sentence, to simplify things, and with no real need to restate the previous sentences slightly differently – and more vaguely.

L725: "Jiang et al. (2017) implied that uncertainties in global OH would not readily explain global CO decreases" → This formulation is not clear to me. I guess the authors meant something like: "the high uncertainties in global OH do not allow to explain the global CO decreases"?
Answer: We have changed the text slightly: "However, Jiang et al. (2017) implied that changes in global OH abundances could not readily explain global CO decreases, given constraints from methyl chloroform surface data (Montzka et al., 2011)."

L734: "As discussed by others, some non-linearity in CO trends…" → I do not get the meaning of this sentence. What are these non-linearities? What is the "variability in past tropospheric CO trend results"? For the latter, do the authors mean "variability through the past studies", or "geographical variability in past CO trends", or else?
Answer: We have specified that this refers to temporal variability (and non-linearity) in past CO trend studies, see the text *(L1006)*: "… some temporal non-linearity in CO trends may be responsible for some of the differences between past tropospheric CO trend results over different periods."

Figures

Figures 1, 2, 12, 13, 14: The maps are quite hard to read, as the land white lines are hard to see through these changing colors. Providing at least the coordinates (and not only 180 °E and 180 °W), as it is done in Figs. 17 and 18, would be helpful for the reader.
Answer: Alright - the continent lines were made thicker; some longitude coordinates were also added, inasmuch as possible (while trying not to reduce the panel sizes much).

Figures 4 and 6: These figures could be wider.
Answer: This is a very minor point, in our view, and it will depend on the paper layout as well; no changes were made.

Figure 9, L1345: The authors can remove the indication in bracket, as the CAM-chem-CEDS simulation is eventually called with its long name.
Answer: Yes, done.

Figure 13: These figures are almost unreadable.
1/ There is enough space in the page width, still the maps (and the text) are tiny.
Answer: Yes, we have used wider maps and reformatted this Figure.

2/ I do not see why the legend in Figure 13b does not represent the shape of the points instead of some describing text. The figure already has a lot of information, and this format adds a supplementary effort for the reader to connect the different lines to the legend.
Answer: Yes, we have changed the legend as suggested.

3/ Figure 13c is the hardest. The points are tiny, the text also, the lines are not even colored: everything is hardly recognisable. Why not using a cold-to-warm color bar for the pressure levels instead?
Answer: Yes, we have now used colors instead of symbols for this panel, with larger font size as well.

Figure 14: The black crosses are not visible in the dark blue cases.
Answer: We have made this more visible (thicker crosses) now - although we do not want to overdo this either.

Figure 15, lines 1407 and 1408: "(top 8 panels)" → top 6 panels instead.
Answer: Yes, we changed this as suggested.

Figure 16: Any justification for the white band in the WACCM QBO-related R^2 ?
Answer: We are not sure we see this much but we will remove anything in the code that might make this too visible - for the final copy.

Figure 17: A warning for the reader about the color bar asymmetry would be welcome. By the way, I guess that the authors intention is to make the "positive sensitivity cases" directly comparable to those in Fig. 18. If so, I think it should be mentioned clearly.
Answer: We have addressed the color bar asymmetry (added a note for the O_3 Figure); the small point regarding the same range of positive values in both Figures is now made in the caption as well.