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.

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?

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.

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.

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.

<u>Introduction</u>: 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.

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.

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.

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?

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).

Section 3.2: Zonal mean trends

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

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. By the way, Figs. 7 and 8 are not necessary and I suggest to move them into the appendix, or into the supplementary material.

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.

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.

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?

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 troposheric column? In this case, could this tropospheric overestimation of ozone be the cause of the CO underestimation through an underestimation of the CO lifetime?

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.

<u>Issue about the statistical significance criterion:</u>

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 onee 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. 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.

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?

L49: “Global anthropogenic emissions dominate the 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

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

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. See for example Clark et al. (2021):

Clark, H., Bennouna, Y., Tsivlidou, M., Wolff, P., Sauvage, B., Barret, B., Le Flochmoën, E., Blot, R., Boulanger, D., Cousin, J.-M., Nédélec, P., Petzold, A., and Thouret, V.: The effects of the COVID-19 lockdowns on the composition of the troposphere as seen by In-service Aircraft for a Global Observing System (IAGOS) at Frankfurt, Atmos. Chem. Phys., 21, 16237–16256, https://doi.org/10.5194/acp-21-16237-2021 , 2021.

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.

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.

L113: “Decreasing CO emissions from anthropogenic and biomass burning sources” → Is it really decreasing from biomass burning sources?

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.

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

L177: “typically of order a few percent or less” → How few? An estimation would be useful, else it remains too vague.

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.

L199: The references about the wave-one pattern, as given later in the paper, would be relevant here as well.

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.

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.

L319: “these biases should not have a significant impact [...]” → Why?

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.

L326: “The model CO biases shown in Fig. 2 are about -5 to -15 %” → On average? If not, then I disagree as well.

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.

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.

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.

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?

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”.

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.

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.

L450: “although R is slightly smaller at 12°N than at 12°S” → It is smaller, not “slightly smaller”.

L459-463: This sentence is too heavy, with these brackets.

L460: “related to emissions, convection, and/or transport” → Which transport? Convection is also a transport processus.

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.

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.

L485: “with an overall better/good agreement between CAM and MLS mapped O3 trends” → To be developed.

L486: “Broad regions” → Which ones? It is not obvious.

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.

L489-506: This paragraph is only descriptive. What is the deduction that we can make from it?

L493-497: This methodology description takes a third of the paragraph volume. I suggest to move it into the methodology section.

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.

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.

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.

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.

L529: “at least over most of the Pacific” → The “at least” can be removed, as the described feature only concerns the Pacific.

L537: “ENSO correlations in the CCM are often weaker than observed for the MLS ENSO R2” → 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.

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?

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?

L575: “For O3, the models display a slight underestimate [...] at 215 hPa.” → Not only slight. The underestimation reaches -20% near the equator.

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.

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%.

L597: “we note that ACE-FTS UT CO monthly zonal mean time series track those from MLS.” → And what does it imply?

L618: “excellent agreement with the above result from CAM-chem-CEDS O3 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.

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?

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.

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.

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.

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.

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.

L686: “These authors [...] (mostly in the extra-tropics).” → Is it still related to this paper then, if it is mainly an extratropical feature?

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.

L727: “(this species also having OH as a major sink [...])” → … and being theoretically not emitted anymore.

<u>Technical comments</u>

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.

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.

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.

L152: “minimize [...] results that might depend more on lower stratosphere” → Minimize is not a relevant adjective to qualify “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?

L174: “is or order” → “is <u>of</u> order”

L247: “SSP5-85” → SSP5-8<u>.</u>5

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

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.

L317: “Livesey et al., 2022” → I did not find the document in the link given in the reference section.

L353: The ozone trend given here is for 147 hPa: it seems that the author forgot to precise it.

L547: “over the South Atlantic region [...] biomass burning periods <u>in this region</u>” → One could understand that there are biomass burning events in the ocean.

L557: “the Eastern side of the maps” → This formulation can be replaced by “the Eastern hemisphere”.

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.

L580 (and 582): “a low model CO bias”, “low biases” → Does “low” mean “negative”? It can be understood as “close to zero”.

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.

L593: It is worth precising that the variability mentioned here is a temporal variability.

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?

L652: “We refer to fairly general past IAGOS-based results” → Does “fairly general” mean “positive trends for ozone and negative trends for CO”?

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”?

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?

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.

Figures 4 and 6: These figures could be widen.

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

Figure 13: These figures are almost unreadable.
1/ There is enough space in the page width, still the maps (and the text) are tiny.
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.
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 colorbar for the pressure levels instead?

Figure 14: The black crosses are not visible in the dark blue cases.

Figure 15, lines 1407 and 1408: “(top 8 panels)” → top 6 panels instead.

Figure 16: Any justification for the white band in the WACCM QBO-related R^2 ?

Figure 17: A warning for the reader about the colorbar 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.