Referee report (ACPD) for manuscript by Kuo-Ying Wang et al., Air Pollution in The Upper Troposphere: Insights from In-Situ Airplane Measurements (1991–2018)

General Comments

This manuscript shows analyses focusing on a subset of upper tropospheric CO in situ aircraft data from various sources (commercial plane flights as well as dedicated non-commercial missions) over the past three decades. The analyses are supplemented by satellite CO data (from MOPITT), which are mostly sensitive to the lower troposphere, but also have an upper tropospheric retrieval product, and ground-based (high altitude) in situ data from Mauna Loa, Hawaii. Other related data discussions include in situ upper tropospheric aircraft data for O_3 and H_2O. A 3-D model of the troposphere (the Integrated Modelling System, IMS) is used for comparisons and inferences about relevant processes affecting lower and upper tropospheric CO.

This work's goals are worthwhile: to use various data sets (mentioned above) and model results to draw conclusions about upper (and lower) tropospheric CO behavior in certain regions of the globe, with some inferences about processes controlling the tropospheric CO distribution (some of which is already known). There is a good amount of work in this fairly long manuscript, in terms of comparisons. MOPITT data are also used and this helps to validate some of the other data analyses and results. The model is useful for some explanations (or potential explanations) of observed CO variations and inferences regarding "trends" - or more qualitatively, some inferred tendencies.

However, I consider that this work needs some fairly major revisions before it can be considered ready for publication, as explained below in more detail. I do recommend further work towards publication, as I believe that this type of analyses fits within the scientific goals of ACP.

This work requires more clarity overall, including a better discussion of uncertainties (or why they might be difficult to establish, at the very least, if this is somehow the thinking), as well as a better summary of the main results in the Abstract, main text and conclusions. Improving the Figures should help in a very significant way, although the accompanying text also needs clarifications. If some of the results, especially those having to do with trends, cannot be supported in a robust (mathematical) way, I recommend removing the stated conclusions, or at least indicating that more work (future efforts) would be needed to better establish or confirm what looks like tentative results, for now.

I do hope that these comments, with more specifics below, will be viewed constructively by the authors, towards a revised manuscript, and I trust that at least some of these results can be displayed better and explained more clearly.

Specific Comments

Despite the valuable goals of this work, I summarize here a number of difficulties and issues:

1. The data are somewhat sparse in time or in space, at least for several of the in situ datasets other than IAGOS; this makes it more difficult to derive conclusions or especially trends in a quantitative way.
2. Error bar estimates are unfortunately absent from this work, overall, while a significant set of correlations are poorly described or not very significant, statistically. Scientific results without some sort of error bar estimates (formal regression errors and maybe some discussion of sensitivity to various assumptions, e.g., data sparseness issues in early years) are almost meaningless. As a related comment, only the simplest form of linear regression is used, and no attempt is made to remove an annual cycle, for example, which could actually reduce the error bar estimates (if they were to be calculated); also, some p values are shown, but with essentially no discussion.
3. Importantly, the quality and visibility of many of the Figures needs to be improved, with less white space so that the reader can see the curves or points for different data sets and their comparisons; this can be done by reducing text around some panels and placing it in the plot panels and/or in the captions, and by producing rectangular panels rather than square panels, to fill in the page more efficiently, for better visibility. If needed, as an option, a single panel, or double panel Figure could be made or added in some cases to better illustrate a significant result, although hopefully that is not needed after revisions of several Figures. The poor quality of several Figures makes it difficult for a reader to follow (or believe) some of the stated results, even if many results are, presumably, correct to first-order.
4. The Abstract should include the main results in some way; it is too vague as it stands.
5. The Conclusion section also needs to more clearly specify the main results, with specific references to which Figure(s) the conclusions are based on and more explanations and context, considering that many of the readers will not try to read every line of this (revised) article.
6. More reference to past work on CO trends should be made, in the Introduction and at least in the Conclusions, to try to put this work, and its results, in the context of related past publications (in particular, regarding past trend or variability results for tropospheric CO). This holds even if "exact comparisons" are often difficult to make, given past results over somewhat different regions, altitudes, and/or time periods.
7. Finally, there are a number of necessary minor technical comments at the end of my report; this underscores that not much proof-reading was carried out by the author and co-authors before the submission of this manuscript, which also makes a reviewer's careful work somewhat more tedious.

More Specifics

I provide detailed comments here, following the order of the Figures and related text/results, especially regarding trends and their robustness, given the limited discussion of these issues. I offer some suggestions for improvements, as well as regarding Figure legibility issues.

Figure 2:

This is where we start to see simple linear regression fits but there is no discussion about error bars or the significance of the results, or the impact on the (potential) conclusions. One issue is that the single point before 1992 does not exemplify the likely scatter that would exist if there was more data available in that time period (or from 1991 to 2008). Giving formal (2 sigma) error bars from the fits is probably a sufficient response to this question. It also depends what scientific conclusions one is trying to make, and how one tries to convey the importance of any results based on such a series of plots. Because of the data sparseness, my top-level impression from the p50 panels is that it will be difficult to state categorically that there has been an increase between 1991 and 2015–2018. The same probably holds for the implied decreases from the pmin panels. As another illustration of my concerns, if one did not have the post 2012 data yet for panels p50 and p25 45N-60N, one might well try to conclude that there were decreases between 1991 and 2011, whereas the slopes drawn after the fairly sudden increases post-2012 could be interpreted as a slow, gradual increase from 1991 to 2018. Also, if one did not have the data point(s) from 1991, one would conclude that the increases (or decreases) are much steeper than shown currently for the 2008-2018 period. My point here is a somewhat qualitative one, but the lack of more continuous data can lead to uncertainties in, or over-interpretation of inferred “trends”.

Thus, if many trend results are not convincing (mathematically), I would tone down the statements that are made on lines 156 to 166, in particular. This is even more true for Fig. 6 showing trends. The general comments about larger values in a certain region versus another can probably be supported based on standard errors in the means, but even these statements are worth checking for significance, if this is viewed as important enough. I am also ignoring the accuracy estimates for the data points shown here, assuming that this plays a smaller role than other issues. However, the introductory remarks about the data sets, especially for the in situ CO aircraft data, should state something about the error bars in the data points that are displayed in various plots; readers should not be expected to know this information, even if you have (broadly) pointed to some relevant past publications.

Lines 165/166: “The 25^{th}, 50^{th},… CO levels are impacted by emissions, with latitudes south of 45N being <u>more affected by Asian emissions</u> …” What leads to this conclusion, specifically? Is this based on scatter, large values, or what exactly? Which time period does this refer to? Please be more specific.

Figure 2 technical improvements: The text font size is rather small, so stretching this Figure to occupy as much space as possible across the page would be good. The caption should indicate what the various symbols mean, i.e. which data set do they correspond to. While this can be inferred from the years shown, a brief sentence or two would be worthwhile in my opinion (most readers are not as close to the various data sets as you are). Captions are there to ensure that symbols and curves shown in each Figure have been described sufficiently well. Also, the full range of years should be mentioned in the caption, to further assist the reader.

Figure 3:

Again, although this one has fewer panels than Figure 2, using most of the page width in a final manuscript would be beneficial (and this may well be the Journal's responsibility, I understand). Also, it seems that the y-intercept value is given by the first number in the linear fit equations listed above each panel; however, for the upper tropospheric blue result in pmean, a y-intercept value of 85.4 ppbv does not seem to match the plotted dashed line at year 1991. Similar comment regarding the p50 panel, where 62.3 ppbv does not seem to match the plotted y-value for the blue dashed line in 1991. Were some of the fits performed with a different approach (about a year within 2000-2008 maybe) rather than for an expected y-intercept in 1991? Please clarify. Also, I do not understand why some of the dashed lines show as partly red and partly blue, when one expects two lines, one red and one blue, for the different altitude regimes. In terms of which measurements are used for this Figure, this should be specified (even briefly) in the caption as well (e.g. lower tropospheric data are from …, etc…).

Figure 4:

Please explain why there are different numbers of points (and fitted equation details) for the UT CO in the 2^{nd} from top left panel versus the top left panel, whereas the descriptions of the regions used appear to be identical; N is 171 versus 254, but the CO data points appear to match in these panels. I also do not completely understand the O_3 fitted line in the 2^{nd} panel from the top (left panel), as its line intercept is -22.6; while this may be fine, it seems that the y-intercept approach (equation) is different in some cases. One would also expect a few comments about R values (significance, or why shown if not useful?) and p values. While it is stated that the CO trends show some increases, the fitted lines seem to give a very small slope, with a zero trend very likely included in the uncertainties (2-sigma error values are really needed to help one understand this more quantitatively, but this is my guess). One cannot robustly determine the validity of the statements on lines 183 to 188 without more quantitative information about fitted trends and* uncertainties (error bars). I expect that many statements are not supported by the error bars, so the authors should probably modify the strength of such conclusions, since (unfortunately) I would expect only marginal significance in many/most of the conclusions regarding trends. One way to reduce the uncertainties in the trends would be to fit an annual cycle (as a sum of sine and cosine functions with one year period, multiplied by constant terms to be fitted) and a linear trend term to datasets with enough points to do so; the explanation of such variability will help to reduce the underlying trend's uncertainty (error bar). This should produce more robust results, even if it may still be difficult to state that a zero trend is highly unlikely in some cases (especially those with fewer data points or shorter series). As stated in the manuscript, some (or most) CO tendencies are "close to zero" indeed. Trend values and error bars could be useful for O_3 and H_2O, if there is enough significance (at least beyond one sigma, say), with ppbv or ppmv per year, or %/yr units - of potential interest to the tropospheric community (and for future comparisons to other studies, for example).

In terms of Figure 4 improvements, trying to stretch the panels to be more rectangular (longer on the x-axis than on the y-axis) would be useful as well.

The possibility that "vertical pumping processes" can explain the out-of-phase relationship between upper tropospheric and lower tropospheric CO is interesting; this might be worth checking against the results of the IMS model, if possible. Does this suggest a several month timescale for such a process? A reference to the paper by Schoeberl et al. (Geophys. Res. Lett., 2006, doi:10.1029/2006GL026178) on the CO tape recorder at tropical latitudes would probably be appropriate as well, although tropical upward transport and convection processes happen on a faster timescale than for the regions this manuscript focuses on (extra-tropics for the most part). More comments about the model comparisons, in any case, would be useful as well, if possible; is the model capable of matching some of the observational inferences, including correlations/anti-correlations between CO, O_3 H_2O, especially those with enough significance?

Figure 5:
Regarding the comments on lines 226, 227, and the "pumping processes", please provide past references and evidence, to strengthen these sorts of statements, which are an indirect inference based on some observed profiles; if there is an informative model comparison, you could/should mention this as well. In terms of legibility, there are a lot of curves in these panels, which makes it somewhat difficult to track the different years or the evolution, if any significant changes ("trends") can be detected somehow, but this is more a question regarding Figure 6. The statement on lines 230/231 should be explained better, namely how do the measurements "align with" the reduction in surface emission sources? Please be more specific and mention which panel you are commenting about.

Figure 6 and lines 239 to 258 (these results do not currently show enough statistical support):
This could be a very interesting and significant Figure, but only if there is enough actual significance in at least some of the trend results. However, without error bars, this is essentially impossible to determine; unfortunately, this can be viewed as a major problem regarding the description and robustness of almost every "trend result" in this manuscript, which is why I am asking for major revisions in the analyses and/or in the description of the robustness of the results. Without error bars, you should only state that there are "indications" of a decrease or an increase and* state clearly enough the reasons for not trying to obtain, or discuss, such error bars. Possibly, this is why you introduce the plot results as "tendencies" rather than "trends", but wording like "distinctive positive trends" cannot carry scientific weight without a discussion of trend values and* their uncertainties; the plots show trends (or tendencies), but what the about the 2nd part of what is needed here (error bars)? On line 250, you mention "the most significant positive trends" – but you need to point specifically to what, mathematically, supports this statement, without expecting readers to try to figure on their own.

For the upper troposphere in particular, which seems to be the focus of this work (also based on the manuscript title), how are readers supposed to interpret the curves? For example, using the p50 panel, what are the typical error bars as a function of altitude – and how fine of an altitude grid can one use for such an analysis (or would averaging over 1 or 2 km help make some trends more robust/significant?)? If the uncertainties are larger than 2 or 3 ppbv/year, it will be difficult to distinguish between 0, +3, or -3 ppbv/year trends. Please respond to this sort of question as well as possible in the revised manuscript. The same holds for the analyses in the low- to mid-troposphere, where some of the trends seem to be large, and thus, potentially significant. One way to illustrate which regions have statistically robust trends could be to make the lines three times as thick in these regions. Showing error bars at all altitudes would lead to plots that are difficult to read, so one would need to do this for a few altitude regions and maybe just for cases where the significance is large enough (e.g., 2-sigma error bar does not include zero), if this happens indeed.

An additional suggestion would be to make a larger separate Figure for p50, where more details can be visible, regarding error bars. Alternatively, use a given altitude range and show the separate percentage cases along the y-axis just for this altitude region; add one or two more regions, with a total of 3 panels, possibly (one panel per altitude regime). I hope that I have made this point sufficiently clear: I expect to see more information about error bars.

Figure 7:

The intent of this Figure and the three panels seems fine. The caption should, however, indicate specifically what the lines mean in each panel, especially for the two lines in panel c, which I think I follow, but help the reader out in the caption. For panel (a), I have the same question as usual, do you detect significant enough trends, and can you state specifically what these are, with* error bars? The Mauna Loa curve should be more representative of the lower troposphere, as the model implies as well. Taking out (fitting) an annual cycle would enable you to come up with a better error analysis for the trend results and comparisons. The linear fits lead to very small trends, it seems, which may also mean that it is difficult to provide a number with enough significance (outside of zero, basically). Please comment if you think there is more that can be said/detected here. Panel c is interesting but it also shows a lot of scatter. Also, It might be more clear to show panel (a) without points, just show lines of different color (or solid lines rather than dashed lines, although data gaps can be problematic); more clarity could also be achieved by having a longer x-axis for panel (a) at least or even a separate Figure. These are mainly suggestions for further thought.

Figure 8:

Line 286 states that the IMS simulations correlate well with the IAGOS data; I think this appears to be correct for the middle panel mainly, but certainly not for panels on the right (R = 0.14). Also, the model CO underestimation of future years (2004–2020) is mostly true for the middle panel(s). As a minor comment, the caption should correct "scattered plot" to "scatter plot" (for the three instances). One question though: why do the model values (dark blue points) not change between the different panels?

Lines 294–296: It is not clear how you can distinguish between emission reductions and an increase in stratosphere-to-troposphere exchange of air (also, what would drive the latter process?). If this sounds interesting but too vague and speculative, you probably need to reword this as a suggestion of possible causes, with no quantification; the reduction in emissions may be believable based on Figure 1, but the air exchange process and importance appear to be quite speculative; please clarify if you can, or if I am missing an explanation somewhere in the manuscript.

Also, for lines 297–299, and the other region mentioned here, the emissions hypothesis seems well founded based on historical estimates of emissions; however, one cannot rule out (as well) some decrease in air exchange between the stratosphere and upper troposphere, can one? More work is needed to help better elucidate such questions regarding processes and budgets.

If you can help to clarify this, however, I recommend that you add some text to make additional arguments - with enough support. I am not asking for more speculation, but admitting to the need for more work is not something one needs to shy away from either, since there can always be more work done to better understand the “exact causes” for certain comparison results.

Finally, I am not too convinced by the statement of “reductions in chemical sources and emissions of CO and hydrocarbons” on line 301, if you could therefore help to clarify the rationale behind this.

Figure 9:

The conclusions here should be taken cautiously, as there is no discussion of possible factors that could lead to biases between the model and the MOPITT data, especially for the less characterized upper tropospheric values. The correlation values would not change, even if an offset to debias the two sets of values was applied. Nevertheless, there are improvements to be made here.

For example, I think that you should mention/confirm (in the caption) what years are used for these comparisons (just the overlap years between these two sets of values presumably?); in other words, this is not trying to use any extrapolated model values, correct?

Are there any trends of significance (statistically), especially in the MOPITT data?

Furthermore, the authors have failed to refer to the Worden et al. (ACP, 2013, doi:10.5194/acp-13-837-2013) article regarding MOPITT CO (and other satellite-based) CO trends, as this should be an important reference to add (and more about the MOPITT data characteristics as well).

The meaning of the lines in the two panels on the right side should be mentioned in the caption, namely what points are being fit with what line (just certain seasons it seems)? Having some description in the text does not replace the need for Figure captions to accurately represent what the plots show (and the main text can often be shortened as a result).

As a detail, the MOPITT legend in the left panels shows a closed circle (dot), but all the points are open circles. You should maybe just show IMS (red, blue) and MOPITT (light green, sky blue), or maybe use “turquoise” for the MOPITT colored UT points. For better visibility, you could also consider not using points, or using smaller ones, for the left panels of this Figure; this might show more clearly that the seasonal cycles between the model and MOPITT agree, as implied in part by panels (b) and (f).

More generally, how do your several comparisons using MOPITT data relate to (agree with) past MOPITT analyses/trends?

Figure 10:

Which part of the MOPITT CO data are used in panel (b), just the lower tropospheric part? Since MLO data are more representative of the lower portion, this probably would make sense (and one should show the MOPITT UT data only as general information). Please specify your approach in more detail, and justify it.

Also, given the vastly different weighting function between MOPITT data and MLO in situ data, a one-to-one comparison in terms of absolute values is difficult or impossible. What are the systematic uncertainties in in situ and in MOPITT CO data, even if one were to ignore the vertical footprint issues? Showing the LT MOPITT data tend to decrease at a similar rate as what the in situ MLO data indicate is useful, but again, one would need to know what the error bars on the trends are, if one tries to address this. Otherwise, I would show this as a qualitative comparison, but a conclusion about the seasonal cycle is useful, given the good correlations in panel (b), as well as (a), although making the points smaller and the lines thicker would possibly help the visibility for panel (a); extending the left panel in the x-direction as a more rectangular plot would also help the visibility for the overlapping data sets (black and red especially).

Lines 338, 339, I would argue that without error bars, this is hard to judge. I would maybe agree that there are indications of positive trends in the data sets you discuss, but since this Figure is focusing on MLO, it might be best to focus on the lower tropospheric measurements only (and discuss UT MOPITT data versus IAGOS data elsewhere, such as in Figure 11).

Figure 11:

In this comparison, again, it is difficult to compare in situ data to broadly averaged data from MOPITT, but there is some averaging occurring for IAGOS as well. Nevertheless, IAGOS data represent the upper troposphere, yet the blue MOPITT data for the UT do not correlate as well with IAGOS as the red (LT) MOPITT data. It is not clear what causes the underestimation of IAGOS values by MOPITT CO, especially for the LT MOPITT data, which generally reach larger values than in the UT. There may be issues that likely fall within the systematic errors of both data sets (MOPITT and IAGOS), which is why this should be discussed in more detail as well.

As a detail, the Figure caption should repeat here that red is for the lower troposphere and blue is for the upper troposphere, for MOPITT (correct?). This is done parenthetically in the regression line equations, but I would add this in the caption as well, as a reminder.

Figure 12:

This one is very hard to see clearly. Try to use lines instead of dots (and use much smaller dots if dots are really desired) to make the main points you are trying to make; also, you would have more space to use rectangular plots if the labels on the outside (right side) were eliminated. Also, the linear fits are not visible enough.

I cannot even begin to process any results until this is shown more clearly; blowing up the plots on the screen helps, but we should not have to do this to such an extent.

Again, comparing satellite data versus in situ data is complicated; it is probably expected that smaller variations are to be observed in a broadly vertically-sensitive MOPITT retrieval. You will need to give more information about this MOPITT dataset's characteristics (error bars, weighting, past validation), as I mentioned before, if one is to understand or believe top-level conclusions based on such comparisons. Including more specific inputs from the MOPITT team itself would have been helpful, if this sort of discussion has not been pursued (enough). I would recommend first trying to discuss the dataset comparisons themselves, if any clear enough conclusions can be drawn, despite the biases that undoubtedly exist between MOPITT and IAGOS data; maybe some indications of similar trends can be made, but again, this would require error bars as well, or you would need to state qualitative conclusions and "indications". It is not completely obvious how the model helps, as there will be biases between the model and the datasets, and these will affect conclusions that are based on extrapolated model predictions. It is hard to justify the statements on lines 373–375 unless the issues mentioned above are better addressed, so be very specific in your explanation of those statements, with a pointer to certain panels that make the point the best, once the legibility of the Figure has been improved. This may end up being mostly justified, but this needs more work, especially for legibility (hopefully this is the main issue, but I am not yet convinced).

Figure 13:

My comments about how to show the datasets more clearly are the same for this Figure, even if it is larger – but can be made more rectangular, with smaller (closed) dots and thicker lines connecting the dots (or no dots even). You need to clearly state if IMS is just for the lower tropospheric heights; also is this MOPITT data for the lower troposphere, please state in the caption (and the text). Why not give trend values with error bars, once again? It may not be that any of these trends are distinguishable from each other (and again, one could try to reduce the error bars by fitting an annual cycle as well), but this is OK, if they agree broadly, as one might hope/expect they would.

Line 426: Could it not be that emissions decreased enough that NLO CO decreases are responding to emission decreases alone? Please be more specific regarding this comment.

Figure 14:

I think that this is an interesting and useful look into the model results. In fact, it might be more useful to also add a climatological average view of these two time periods (i.e. average all the years from each panel and create a new set of plots, or make a 4-panel Figure. Also, it would seem better to plot the (panel with) earlier years on the left, just because time usually is shown as increasing from left to right. You should try to explain better what the main changes are between the two panels; e.g., are the black curves identical, why are the green curves changing, and also summarize what references or methods are used to determine the various curves (e.g., 1100 MT/yr).

Line 422; again, one should add some reference to the work by Worden et al. (2013). The usefulness of MOPITT is not something new, nor are some trend estimates using that dataset (and other datasets and analyses in more recent papers).

Line 425… why would negative trends in CO not reflect decreases in CO emissions, even with constant chemical sinks (such as OH abundances)? You need to be clearer about the statements that refer only to chemical sinks dominating over sources, and why the emissions might not play a role (in some regions – are they increasing there?).

Line 437; there are other past references that could/should be cited, see below. I also mentioned Schoeberl et al.

References: Other useful references (or information) could include these (among others), for mention in the text, Introduction, and/or Conclusions

The first one (Cohen et al.) refers to IAGOS data analyses, which, surprisingly, was not cited by this manuscript focusing on IAGOS (and other) in situ data for CO; how are your results different than in this reference?

Cohen, Y., Petetin, H., Thouret, V., Marécal, V., Josse, B., Clark, H., Sauvage, B., Fontaine, A., Athier, G., Blot, R., Boulanger, D., Cousin, J.-M., and Nédélec, P.: Climatology and long-term evolution of ozone and carbon monoxide in the upper troposphere–lower stratosphere (UTLS) at northern midlatitudes, as seen by IAGOS from 1995 to 2013, Atmos. Chem. Phys., 18, 5415–5453, https://doi.org/10.5194/acp-18-5415-2018, 2018.

Park, M., H. Worden, D. Kinnison, B. Gaubert, S. Tilmes, L. Emmons, M. Santee, L. Froidevaux, and C. Boone, Fate of pollution emitted during the 2015 Indonesian Fire Season, Journal of Geophysical Research: Atmospheres, doi:10.1029/2020jd033474, 2021.

Li, Q. B., J.H. Jiang, D.L. Wu, W.G. Read, N.J. Livesey, J.W. Waters, Y. Zhang, B. Wang, M.J. Filipiak, C.P. Davis, S. Turquety, S. Wu, R.J. Park, R.M. Yantosca, and D.J. Jacob, Convective outflow of South Asian pollution: A global CTM simulation compared with EOS MLS observations, *Geophys. Res. Lett. 32*, L14826, doi:10.1029/2005GL022762 , 2005.

Conclusions Section:

Regarding CO trends in the UT (e.g.,line 439 on a rise in upper tropospheric CO in some regions (North Pacific UT), or other CO decreases elsewhere) need to be supported by trends with error bars; if not, the reader does not have enough information to scientifically assess the validity or robustness of the assertions made here.

Lines 441–447: these are somewhat vague conclusions; it is nothing new that chemistry plays a significant role for CO. Why do some results show up as CO increases here, whereas there are also some decreases in other plots in this manuscript, also in the UT?

There are not enough specifics in these concluding remarks (or in the Abstract) based on the whole contents of this manuscript, which has a lot of Figures (and work), but falls short in terms of clear statistical analyses with error bars, followed by a reasonable summary regarding the most robust conclusions, especially for IAGOS data.

(Mostly) minor and technical comments

This is a mixture of minor comments that require some slight modifications and improvements, typographical details, and wording suggestions/corrections.

L37–39: Is there a reference (Wang et al., 2024?) for this? Please specify (or does this relate to the current manuscript?). Same for the next paragraph (L40–45).

L45: “68 pptv” should read “68 ppbv”.

L56–58: “By combining …strategies and policies”. This sounds good, but is this “informing” done as a result of this work in any way? If not, and it does not seem to be at the moment, is it worth stating this general goal? Maybe - I am not trying to push hard one way or another here.
Also, how do emissions “impact the dynamics”?

L91: change “and significant growth” to “with significant growth”.

L100: I would rephrase to “regions are characterized by key dynamical processes…”

L105: Please provide the altitude of the in situ data (site) at Mauna Loa.

L109: Need to define “IMS” here (not further below).

L112: Measurements (plural).

L113: Buchholz, not Buchho.
The MOPITT section needs more information about the weighting functions or region of sensitivity since you use lower tropospheric and upper tropospheric data from MOPITT; it also should have more references.

Section 2.4: Please provide the vertical and horizontal resolution of this model, and the height range used (how far into the stratosphere?). Also, how are the biomass burning (not the industrial) emissions characterized? Are they constant also?

Figure 2 caption: Measurements (typo) on first line. Downwind of the Asia emission areas and over the North Pacific… The lower panels show data from the low latitudinal…(typos).

L166: how, specifically, do you know that latitudes south of 45N are more affected by Asian emissions? Please clarify. Is it just geographical (based on latitudes) and according to the prevailing wind direction – or something else?

Figure 3 caption: Time series of long-term CO…

Line 193: Please specify more clearly why “vertical pumping” (through which process here?) creates anti-correlations between LT and UT; this may not be immediately obvious (and it depends on the timescale for the “pumping” process).

L195: process → processes

L197: “are negative”. Not negatives. Monthly CO in the upper troposphere is higher…

L198: Monthly O3 in the UT is higher…

L199: negative O3 indicates less… Also, what does “(ref)” mean? Please add a reference.

L200: O3 is negatively correlated with CO in the upper troposphere.

L204: H2O is positively correlated with CO in the upper…

Figure 5 caption: (a) downstream (typo), areas (typo).

L211–213: Why is horizontal transport not included as a possible mechanism?

Figure 6 caption: this should include the acronym names used in the panels (after the descriptions made in the caption). Also, change “and regions (red)” to “and all regions (red)”.
(e) says “tropical and Southern Hemisphere”, what latitudes exactly (South of 25N? like (d) as well)?

L249–250: Why would negative trends not possibly occur as a result of reductions in emissions (chemical sources)? Why is it only chemical sinks that might control the budget? Please clarify the thinking here. Would one need (for example) an OH increase to lead to a CO increase? What other factors could be in play, theoretically at least?

Figure 8: Why is the equation above each top panel the same? Also, change “scattered plot” to “scatter plot” in the caption.

L286: “correlate well”? This does not seem to be true for every panel shown.

L295: what would cause an increase in UTLS air exchange processes? Provide a reference if possible.

L320: providing tropospheric CO estimates (?) Model is not a measurement per se…

L330: lower troposphere is correlated with…

L345-347: Can use present tense rather than past tense here.

Figure 12: linear regression (typos).

L371: It’s worth → It is worth

L413: zeros → zero