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)

1 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.

Reply. We are very grateful indeed to the reviewer for very detailed comments. We have followed reviewer's comments, point by point to revised the manuscript. In this first part of the reply, we have uploaded the revised manuscript and replies as shown below. The complete replies and final revised manuscript will be uploaded in the next two weeks. Thank you very much indeed for your time and comments that have significantly increase the clarity of the manuscript.

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.

Reply. We sincerely thank Reviewer 1 for the thorough and constructive assessment of our manuscript, "Air Pollution in the Upper Troposphere: Insights from In-Situ Airplane Measurements (1991–2018)". The comments have greatly helped us identify areas where additional clarifications, data analysis, and textual revisions were needed. Below, we address each point individually (or in groups, if overlapping) and outline the corresponding changes made to the manuscript. Throughout our revisions, we have:

1. Revised the figures for improved clarity, legibility, and labeling.

2. Strengthened the discussion of uncertainties and significance of trends, including estimates of statistical confidence (where possible).

3. Expanded the literature review to place our work in the broader context of CO trend studies.

4. Enhanced the conclusions and abstract to highlight the main findings clearly.

5. Corrected typographical/technical issues and refined the manuscript for readability.

All revised or new sections are clearly indicated in the manuscript, and we provide line/section references below for easy cross-checking.

2   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.

   Reply. We have include a new Figure 2 to summarize all the datasets (except MLS) used in this work. We have included a new Table 2, which summarizes of previous studies on long-term CO trends, including the data sources, periods, and key findings. Figure 2 is a timeline Infographic, showing the chronological order of datasets. This timeline infographic illustrates the chronological progression of various CO measurement datasets utilized in this study. Each dataset is marked along the timeline with distinct colors and shapes representing the measurement methods employed (e.g., NDIR, QCLS, WS-CRDS, Microwave Limb Sounder). Error values are depicted as horizontal lines extending from each marker, indicating the associated uncertainties. This visual representation aids in understanding the temporal and methodological context, ensuring clarity in interpreting the observed CO trends. By implementing a well-structured timeline infographic, we can significantly enhance the readability and transparency of our manuscript, addressing the reviewers' concerns effectively. This visual aid will help in succinctly conveying the complex interplay of multiple datasets, their measurement periods, methods, and associated uncertainties, thereby strengthening the overall presentation of our work.

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.

   Reply. We have included a new Table 2 to summarize the errors from the each of the dataset used in this work. We have include new Figure 4 and new Figure 5 for verifying the impacts of obseravation efforts on the calculations of CO trends. We have new Figure 3 and new Figure 6 to include statistical test of linear trends and 95% confidence intervals of trends in the figures. In new Figure 9, we have included profile

of statistical tests of linear trends and 95% confidences intervals, and compare the distribution of confidence intervals in various regions. We have included discussions of statistical significance and 95% confidence levels in comparting and understanding the long-term trends of CO in various regions and altitudes.

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.

   Reply. We have improved the quality and redability of figures, following reviewer's comments to make the optimal use of space in a page. The white space has been reduced so that the figure size can be increased to make the figure readable.

4. The Abstract should include the main results in some way; it is too vague as it stands.

   Reply. We have revised the abstract, and the title of this work.

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.

   Reply. We will complete this revision soon.

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.

   Reply. We have include a new Table 1 to include previous studies on CO trends. Table 1 shows data used in the studies, data period, and key findings over difference regions.

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.

Reply. We are very grateful indeed to the precious time and effort the reviewer has devoted to carefully read and comments on our manuscript. We fully understand and greatly indebted to the reviewer's kindness and professional help. We have taken great effort to revise our manuscript following reviewer's comment. Thank you very much indeed.

3   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".

We have included statistically significance of the linear trends, including P-values and 95% confidence intervals in the revised figures and manuscript. Yes, indeed the GTE data from 1991 and 2011 are very valuable for understanding the long-term trends of CO from 1990 to 2020 over the North Pacific. We admire very much the visions of people who were able to take measurements to help us understand the trends. We have revise Figure 2 (new Figure 3) to include 95% confidence levels of the CO trends analyzed.

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.

We have revised the Introduction and Method sections so as to include details about the use of datasets and their uncertainties in this work.

Lines 165/166: "The 25th, 50th, ... CO levels are impacted by emissions, with latitudes south of 45N being more affected by Asian emissions ..." 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.

Reply. We thank the reviewer for this important comment. In response, we have revised the text to clarify the basis and time frame for our conclusions. The revised paragraph now reads as follows:

*"In summary, key findings from the in-situ aircraft measurements in the upper troposphere over the middle North Pacific, downwind of Asian emissions (see Figure ??), are as follows. Over the 27-year period from 1991 to 2018, all observed CO concentrations increased—with the exception of the minimum and maximum levels. The minimum CO values, which represent background CO, decreased over time. In contrast, the 25th, 50th, and 75th percentiles of CO levels, which are influenced by industrial emissions, are higher at latitudes south of 45N compared to those north of 45N These findings are consistent with previous studies (Stohl et al., 2002, Hsu et al., 2006) that have documented the long-range transport of Asian pollutants across the North Pacific.

The 95% confidence intervals for the trends in CO concentrations indicate predominantly positive trends for the upper percentiles. Specifically, the confidence intervals are as follows:

maximum, -0.502 to 1.142; 75th percentile, -0.031 to 0.235; 50th percentile, -0.044 to 0.321; and 25th percentile, -0.065 to 0.218. In contrast, the minimum CO levels exhibit a 95% confidence interval of -0.147 to 0.048, reflecting a predominance of negative trends. Overall, these results indicate that CO concentrations over the North Pacific have generally increased over the study period, except for the declining background levels."* This revised description clarifies that our conclusion regarding the greater impact of Asian emissions at latitudes south of 45N is based on the following points: Data Analysis and Time Period: Our analysis is based on in-situ aircraft measurements collected over a 27-year period (1991–2018). We have now explicitly stated this time period to ensure clarity. Statistical Basis: The conclusion is drawn from our statistical analysis of the CO concentration distribution. We examined various percentiles (25th, 50th, and 75th) and observed that, while the minimum CO levels (representing background CO) decreased over time, these percentiles increased. The elevated high-percentile values and the increased scatter in the data south of 45N indicate episodic enhancements attributable to Asian emissions. Consistency with Prior Studies: We also note that these observations are in agreement with earlier studies (Stohl et al., 2002; Hsu et al., 2006), which document the long-range transport of Asian pollutants across the North Pacific. We trust that this revised text and explanation sufficiently address the reviewer's concerns regarding the basis of our conclusions and the specific time period under investigation.

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.

Reply. Thank you for your valuable suggestions regarding the technical improvements to Figure 2. In response, we have made the following changes:

Increased Font Size and Figure Layout: We have increased the text font size and stretched the figure to occupy the full available width of the page to improve legibility. Enhanced Figure Caption: The caption for Figure 2 has been revised to clearly indicate the meaning of the various symbols used. We now include a brief description of the different datasets, specifying which symbols correspond to which data sets. Additionally, the caption now explicitly states the full range of years (1991–2018) covered by the data. We believe these modifications significantly enhance the clarity and accessibility of Figure 2, ensuring that all symbols and curves are sufficiently described for 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.

Reply. We have revised Figure 3 (now new Figure 6), following reviewers's comments.

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.).

Reply. There is an internal bug in the NCAR Graphics library that were used to create this dashed line. We have written a new code so that the two colors has become a single color in a single and straight dashed line. Thank you.

Figure 4

Please explain why there are different numbers of points (and fitted equation details) for the UT CO in the 2nd 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 2nd 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).

Reply. We have revided Figure 4 (now new Figure 7 and Figure 8) to account for the points pointed out by the reviewer. The revised results are now consistent between the vaules shown on the linear regression equations each plot.

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.

Reply. We have updated Figure 4 to new Figure 7 for the long-term time series of CO over the North Pacific, and Figure 8 for the assessments over the North Atlantic. The comparison between CO in the upper and the lower troposphere aims to demonstrate the coupling between the upper and the the lower troposphere through the vertical pumping process. CO is a very unique tracer for demonstrting the vertical pumping process in the troposphere. We have included O_3 and H_2O to help validate the pumping processes. The bahaviour and characteristics of O_3 and H_2O in the upper troposphere have been studied and well known. We have included 95% confidence intervals in the time-series trends and the correlations to make meaningful comparisons. We have stretched Figure 4 (new Figure 7 and Figure 8) as suggested by the reviewer.

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?

Reply. Yes, the IMS model is capabel of matching some of the observational vertical pumping processes over the North Pacific. We have revise this section and Figure 4 (now new Figure 7 and Figure 8) to account for the vertical coupling process in the troposphere through the vertical pumping processes. The IMS model also shows the vertical pumping processes over the North Pacific during the months. We have also cited and discuss Schoeberl in the revised manuscript: "We note that the vertical pumping processes discussed above corrobate the CO tape recorder process (Schoeberl et al., 2006). Schoeberl et al. (2006) discuss the relationship between the seasonal signal of CO and its lower-tropospheric sources. In their analysis of the "CO tape recorder," they point out that because CO has a relatively short chemical lifetime, its vertical profile in the tropical upper troposphere and lower stratosphere retains a clear imprint of the seasonal variations at its source. In particular, they note that the seasonal cycle is strongly influenced by changes in tropospheric CO—much of which is attributed to biomass burning—being convectively injected into the upper troposphere. Thus, the seasonal variability observed in the CO profile is a direct reflection of the variability in its lower-tropospheric sources. Even though CO has a relatively short chemical lifetime—on the order of a few months—the, key to the tape recorder effect lies in the balance between its chemical loss and the rate of vertical transport in the tropical atmosphere. In the tropics, strong convective processes rapidly lift CO from the lower troposphere into the upper troposphere and lower stratosphere. This upward transport occurs on a timescale that is comparable to or even shorter than CO's chemical lifetime. As a result, the seasonal variations in CO concentrations at its source (driven largely by factors like biomass burning and other lower-tropospheric processes) are "recorded" in the vertical profile as the gas ascends. Even though CO is being chemically removed, the quick vertical transport preserves its seasonal signature long enough to be observed as the tape recorder effect in the upper atmosphere."

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.

Reply. We have revised the description. We have cited and discussed previous work by Schoeberl et al. (2006), who has discussed the relationship between the seasonal signal of CO and its lower-tropospheric sources. In their analysis of the "CO tape recorder," they point out that because CO has a relatively short chemical lifetime, its vertical profile in the tropical upper troposphere and lower stratosphere retains a clear imprint of the seasonal variations at its source. In particular, they note that the seasonal cycle is strongly influenced by changes in tropospheric CO—much of which is attributed to biomass burning—being convectively injected into the upper troposphere. In the tropics, strong convective processes rapidly lift CO from the lower troposphere into the upper troposphere and lower stratosphere. This upward transport occurs on a timescale that is comparable to or even shorter than CO's chemical lifetime. As a result, the seasonal variations in CO concentrations at its source (driven largely by factors like biomass burning and other lower-tropospheric processes) are "recorded" in the vertical profile as the gas ascends. Even though CO is being chemically removed, the quick vertical transport preserves its seasonal signature long enough to be observed as the tape recorder effect in the upper atmosphere.

Figure 6 and Lines 239 to 258

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 clearly state 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 about the 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 figure it out 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.

Reply. We hve revised Figure 6 (now new Figure 10). We now compute and display 95% confidence intervals for the linear trend estimates at each altitude and for each region. These error bars are plotted alongside the trend lines so that readers can directly assess the statistical significance of the trends. In the revised text, we now explicitly state that a trend is accompanied by its 95This provides a clear mathematical basis for our statement about "the most significant positive trends."

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., a 2-sigma error bar that does not include zero), if that is indeed the case. An additional suggestion would be to make a larger separate Figure for the p50 panel, 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.

Reply. We have revised Figure 6 (now new Figure 7) to include error bars in the linear trends. We have noticed that the vertical distribution of the error bars vary closely follow the patterns of the linear trends of CO at each altitude for each region. If the linear trends are negatives, then the ranges of the error bars are dominated by the neative values. On the other hand, if the linear trends are positive, then the ranges of the error bars are dominated by the positive values. Inclusion of the error bars further reinforce the characteristic distribution of the linear trends.

Large variations of the ranges of error bars mostly occur in the low altitudes, reflecting the direct impacts of CO measurements close to the industrial and CO emission areas. The maximum CO (pmax) exhibits large variations in the ranges of error bars, consistent with elevated CO been transported from the heavy emission areas (such as the CO tape recording elevated CO from biomass buring shown in Shoeber et al. (2006)). In contrast, the minimum (pmin) and the 25th percentile of CO shows ranges of error bars more confined to the central values of the CO linear trends. The mean and the 50th percentile of CO, except for the top of the data layer close to 13 km, and in low altitudes below 3 km , CO trends at more vertical

altituides are confined to the central of the linear trends

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 please help the reader by clarifying in the caption. For panel (a), I have the same question as usual: do you detect trends that are significant enough, 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 (i.e., fitting and removing) 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 (i.e., distinguishable from zero). Please comment if you think there is more that can be said or detected here. Panel (c) is interesting but it also shows a lot of scatter. Also, it might be clearer to show panel (a) without points, but rather with lines of different color (or solid lines rather than dashed lines, although data gaps can be problematic); additional clarity could also be achieved by having a longer x-axis for panel (a) or even a separate Figure.

Reply. We have revised Figure 7 (now new Figure 11). We have included 95% confidence intervals of the CO trends on (a). We have also revised discussion for Figure 7.

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 the 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 change "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?

Reply. We have revised Figure 8 (now new Figure 12). We have also include 95% confidence intervals of the linear CO trends from the IMS model and the IAGOS observations on the time-series plots.

Lines 294–296: It is not clear how you can distinguish between emission reductions and an increase in stratosphere-to-troposphere exchange of air (and 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, without quantification; the reduction in emissions may be believable based on Figure 1, but the air exchange process and its

13

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 better elucidate such questions regarding processes and budgets. If you can help to clarify this, I recommend that you add some text to make additional arguments — with enough support. I am not asking for more speculation, but admitting the need for further work is acceptable, since there can always be more research 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. Please clarify the rationale behind this.

Reply. We have revised description for Figure 7 (now new Figure 12).

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 were applied. Nevertheless, improvements are needed.

For example, you should mention/confirm (in the caption) which years are used for these comparisons (presumably just the overlap years between these two sets of values); in other words, you are not trying to use any extrapolated model values, correct? Are there any trends of statistical significance, 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 (it appears to be just certain seasons). Having a clear description in the caption is necessary even if the main text contains some details.

As a detail, the MOPITT legend in the left panels shows a closed circle (dot), but all the points are open circles. You might consider showing IMS (red, blue) and MOPITT (light green, sky blue), or perhaps use "turquoise" for the MOPITT colored UT points. For better visibility, you could also consider not using points, or using smaller points, for the left panels; this might more clearly show that the seasonal cycles between the model

and MOPITT agree, as implied in panels (b) and (f). More generally, how do your several comparisons using MOPITT data relate to (or agree with) past MOPITT analyses/trends?

Reply. We have revised Figure 9 (now new Figure 10).

Reply. Notes: Replies to the rest of comments will be uploaded in next two weeks, together with replies to the comments from reivewer 2 and reivewer 3.