Response to Reviewers for "*Characterizing variability and vertical structure of water vapor in the extratropical lower stratosphere*" by Emily Tinney, William Randel, and Jun Zhang

We thank the reviewers for the overall positive comments on the manuscript and the useful suggestions for improvements. We have addressed all the reviewer comments and include individual replies below (line numbers refer to the revised manuscript).

One note is that we have now included a third author (Jun Zhang from NCAR), as she has made important contributions regarding the WACCM model simulations in the paper.

Reviewer comments are shown in **black** and our responses are given in **blue**.

Reviewer 1

The paper by Tinney and Randel introduces a novel tropopause-relative coordinate to analyze the variability and vertical structure of stratospheric water vapor. This framework effectively removes a large portion of the variability associated with the seasonal cycle of tropopause height, allowing the variability of water vapor itself to be examined more clearly. Using this coordinate system, the authors investigate several aspects of lower stratospheric water vapor, which is the most radiatively important region. The manuscript is well written, clearly organized, and provides several interesting new perspectives on the variability of lower stratospheric water vapor.

Thanks for the positive feedback.

My main concern relates to the potential long-term drift in MLS measurements in the tropopause.

Hurst et al. (2016) reported a drift in MLS H_2O since about 2010 using version 4 data. Later, Livesey et al. (2021) identified a drift associated with the 190 GHz sideband fraction and released an updated version of H_2O and N_2O (v5). Although the drift was substantially reduced, MLS H_2O still shows significant drift compared with frost-point hygrometer measurements, and N_2O still exhibits relatively large drift.

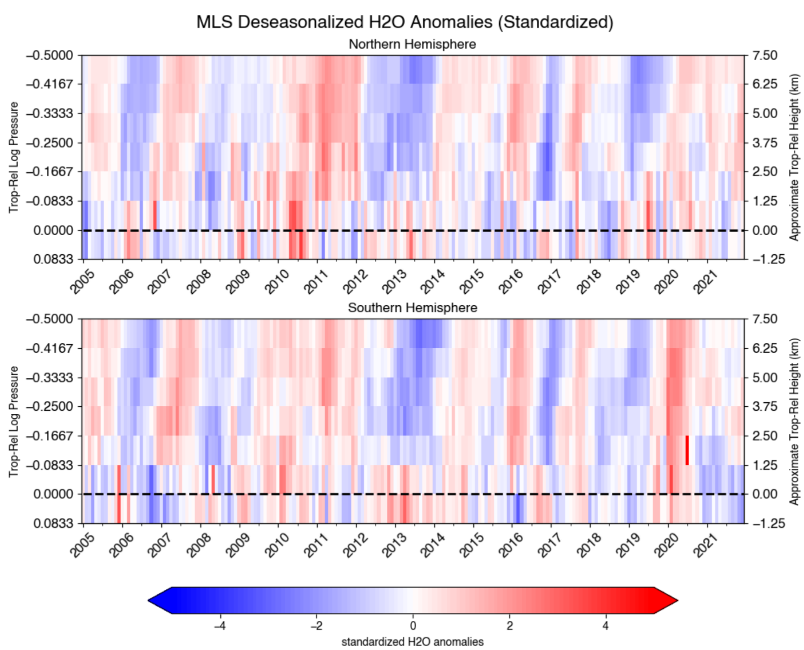
The large negative trend in the UTLS observed by MLS appears inconsistent with current observational evidence (as noted in the manuscript), theoretical expectations based on the Clausius-Clapeyron relationship, and model results. In addition, this negative trend does not seem to be explained by either ENSO or QBO as presented in the manuscript. ENSO shows a positive trend, while the negative QBO influence (Figure

15a) occurs at higher levels than the region of strongest negative H_2O trend (roughly one layer above and below the tropopause). Therefore, the possibility that the negative trend in this region is influenced by residual drift in MLS measurements cannot be excluded. I suggest that the authors cite the studies mentioned above and discuss this issue. It would in fact add scientific value if the manuscript explicitly raised this important issue and discussed whether the observed trend could partly reflect instrumental drift based on the current analysis; however, even a brief mention of this possibility would be helpful to provide broader context.

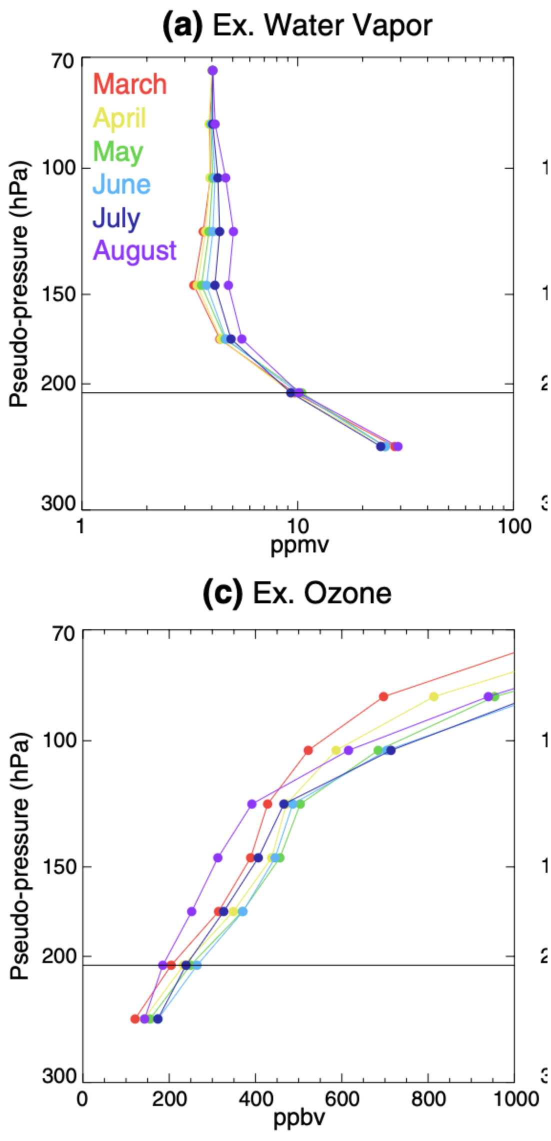
We agree with this suggestion to include more background information on drifts in the MLS H2O data, and we have included additional discussion and suggested references in Section 2.1.

Line 195: The authors state that "There is a distinct separation of patterns in the vertical structure occurring around ~2.5 km above the tropopause in both hemispheres" for H_2O (Figure 7), whereas O_3 shows a "coherent vertical structure in each hemisphere" (Figure 10). Since both H_2O and O_3 exhibit strong gradients near the tropopause, I would not have expected such markedly different behavior. The potential MLS drift mentioned above might contribute to this difference. I would be interested to see whether the structure in Figure 7 remains the same if the long-term trend is removed, and whether the vertical structure then becomes more coherent.

This is a good question. We have calculated a version of Fig. 7 using detrended data at each altitude level:



It looks like some vertical coherence is restored with the trend removed, but overall there is still a clear separation between the overworld and ExTl region that is not present in the ozone observations; we have added this detail in lines 216-217 and included this figure in Supplementary Material. While ozone and water vapor both exhibit strong gradients across the tropopause, we think the difference here is the location of the gradient relative to the tropopause. This is a figure taken from Tinney and Homeyer 2021, which shows vertical profiles of water vapor and ozone in the extratropics (specifically over North America in this study):



These plots show a strong gradient in H2O directly across the tropopause (note the logarithmic x-axis), whereas ozone's strongest gradients exist closer to 100 hPa. The strength of the gradients are also different, with H2O decreasing an order of magnitude across the tropopause, while ozone approximately doubles across the tropopause. Therefore, we believe these differences in water vapor and ozone are realistic. We have expanded on this discussion in lines 240 to 247.

It is important to note that, regardless of whether such long-term drift exists, it does not diminish the novelty or importance of the manuscript. The focus of this study is primarily on short-term variability (seasonal and interannual), which would not be significantly affected by a slow instrumental drift.

There are also several minor comments:

> Figure 6: Point B (seasonal maximum) occurs in the same season in both hemispheres, whereas point A (seasonal minimum) does not. Could this be related to the fact that the CPT is lowest around January, resulting in lower water vapor concentrations in the following months?
>
> Yes, exactly. We have added additional text (in the paragraph beginning on line 173) to clarify this behavior.
>
> Figure 9: Why is a fixed two-month time lag used here instead of allowing a latitude-dependent lag? One might expect air in the tropics to have a shorter lag and air at higher latitudes to have a longer lag.
>
> Yes, the H2O correlation patterns systematically change with time lag as the signal propagates from the tropics to higher latitudes, as shown explicitly in Randel and Park (2019). Correlations in the extratropical LMS are maximum at around 2 months lag, which is why we show that. We have added text to clarify this detail at line 228.
>
> Line 210: Why is this described as an "artifact"? Figure 9c appears broadly similar to 9b and 9d, and the difference between Figure 9a and 9c seems to reflect a different physical behavior rather than an artifact.
>
> We term this an 'artifact' because the large upper tropospheric correlations are introduced purely from using H2O in tropopause coordinates (i.e. there is an artificial change in tropical upper tropospheric H2O from the coordinate transformation). The results do not reflect actual physical correlations with tropospheric H2O.

There are also a few technical points:

- Line 67: The manuscript states that 2° latitude bins are used for MLS. However, MLS Level-3 products are typically provided in 4° latitude bins, so I assume this analysis uses Level-2 data. It would be helpful to clarify this in the methods section.

  We use the orbital level 2 MLS data, as now clarified in Section 2.1

- Line 139: Formatting issue in “30°N and S”.

  Changed

- Line 300: The sentence appears to be incomplete after “different from”.

  This is corrected.

Reviewer 2

Review of “Characterizing variability and vertical structure of water vapor in the extratropical lower stratosphere” by Emily N. Tinney and William J. Randel

The manuscript examines the variability of water vapor in the extratropical lower stratosphere using geometric and tropopause-relative coordinates. For this purpose, the authors use MLS, ACE-FTS, and WACCM-SD data. The hemispheric coherence is an interesting result. I suggest major revisions, particularly the inclusion of ACE-FTS and WACCM results in all analyses and figures. However, once the authors address these points, the manuscript should be ready for publication.

General concerns:

Why is the study limited to 2005-2021? It seems the model was run up to 2022. Is this due to the possible influence of Hunga in the results? If it is, this needs to mention in the text. If there is another reason, please include it in the text.

Yes, data are analyzed for the pre-Hunga period before 2022. We have added text to clarify in Section 2.1

Why are you using PTGT? This is a new definition, please provide some background and add some details about it. What is the advantage of this tropopause over WMO or cold point tropopause? Perhaps the cold point tropopause will be a better tropopause for water vapor studies?

Background and motivation on using the PTGT tropopause is discussed in Tinney et al (2022). In practice, the PTGT is very close to the WMO lapse rate tropopause (LRT) outside of winter polar regions. As noted in the text, our overall results are very similar using LRT tropopause instead of PTGT. As our analyses focus on extratropics, a lapse rate tropopause definition is more physically meaningful than the cold point.

When using tropopause-relative coordinates to analyze water vapor variability, it is important to note that the natural variability of water vapor cannot be reduced by any coordinate transformation. Coordinate systems can only reduce the artificial variability introduced during data binning. This artificial variability arises from averaging data originating from different dynamical regimes. Although this distinction is subtle, it is important to reflect it clearly in the text; therefore, please revise the manuscript accordingly.

We are unsure what the Reviewer is asking for here. The goal of tropopause-based coordinates is to organize variability in a more physically-based coordinate system compared to geometric coordinates, given the knowledge that H2O structure typically follows the extratropical tropopause. We acknowledge additional H2O variability remains in tropopause coordinates, and part of the goal of this study is to identify and quantify such variability.

Please specify how are you computing the trend errors. The anomalies shown in Figure 1 will suggest big errors in the trends, but figure 13 shows statistical significance everywhere (at least for MLS). Note there is autocorrelation in the time series so you will need to take that into account.

We thank the reviewer for this comment. In the original manuscript, trend uncertainties were calculated from the standard errors of the regression coefficients obtained from the multiple linear regression model. Following the reviewer's suggestion, we evaluated the lag-1 autocorrelation of the regression residuals and accounted for serial correlation using an effective sample size correction:

$$N_{eff} = N\frac{(1 - r_1)}{(1 + r_1)}$$

where r_1 is the lag-1 autocorrelation of the residuals. The standard errors and p-values of the regression coefficients were recomputed using N_{eff} and the significance stippling in Figures 13, 14, and 15 have been updated accordingly. This detail is now clarified in lines 286 - 292, equation 2, and the figure captions. While the autocorrelation correction slightly reduced the spatial extent of statistically significant trends in some regions (tropical lower stratosphere for MLS, NH overworld for ACE and WACCM), the overall pattern and conclusions of the MLS trend analysis remain unchanged. The ENSO signal significance changed the most across all three datasets, where significance is primarily found/shown in the tropical upper troposphere. Finally, the spatial extent of the MLS QBO regression significance was reduced slightly. These differences do not impact the overall results, but we sincerely appreciate this catch to improve our paper.

Specific comments are:

L2 Add comma after long-term trends

added

L36 Tao et al 2023 and Yu et al 2022 are not the correct references for this, the importance of methane oxidation has been know for a while please add more appropriate references, such as, 10.1029/JZ055i003p00301, among others.

The Tao and Yu references are recent papers that explicitly estimate CH4 impacts on stratospheric H2O and we think these are appropriate. We have added one additional older reference on this topic (LeTexier et al., 1988). The paper that you cited does not discuss CH4 effects.

L40 I don't think "alternatively" is the right word, the complexity of the transport in the LMS is not an alternative to the simplicity of the overworld, I think "in contrast" will work better

changed

L48 add e.g., before Charlesworth et al 2023

changed

Section 2.1. Are you using quality screening as recommended in Livesey et al 2022 to screen out bad data in the MLS record? If you are please mentioned it here.

Yes, we use the standard MLS quality flags and this is now noted in Sect. 2.1.

Please summarize the water vapor drift here (which may influence your results). See 10.5194/acp-21-15409-2021

We have added additional text in Sect. 2.1 regarding the MLS H2O drifts

Also summarize the MLS water vapor validation efforts

This is discussed in the data document Livesey et al (2022) cited in the text

L69-70: using the version 5.2 retrieval -> using version 5.2

changed

L70: You didn't provided the spectral coverage of the MLS instrument. If you think this is important for ACE-FTS please add it as well for MLS. Also summarize the ACE-FTS water vapor validation efforts

This is part of the standard description for ACE-MLS data and choose to keep it as-is. We have added a reference to the ACE-FTS H2O validation in Sect. 2.2

L81 What are the other problems? The impact of sampling biases could be mentioned here 10.1002/jgrd.50874, 10.5194/acp-18-4187-2018

We don't know what other problems there might be in addition to clouds. We think that we have been clear regarding the data sampling used in this paper. The two suggested references do not specifically pertain to our 3-month sampling of ACE-FTS data (the first regards monthly sampling of ACE-FTS, and the second only discusses MLS and MIPAS satellites).

Section 2.3 What is the temporal resolution of WACCM? Is there any study validation the WACCM water vapor? Is it representative of the atmospheric state despite the wet biases?

WACCM is sampled once daily, as noted in the text. This specified dynamics run for WACCM is new and no specific validation studies have been performed. In general, WACCM is a state-of-the-art model, with many relevant citations in the Gettelman et al (2019) description.

L109 the Gelaro et al citation seems to be out of place. Was it supposed to be "For ACE and MLS, a reference tropopause is computed from three-hourly MERRA-2 reanalysis data (Gelaro et al., 2017) and interpolated in space and time to each profile location" Gelaro et al does not interpolate to measurements locations.

This has been changed

L131-134: Please quantify the reduction of the H2O variance? You could simply show the standard deviation or the normalized standard deviation. (either in the text or in the manuscript). Note that this "reduction on variability" has been noted before 10.1029/2008JD009984 Please cite accordingly

We now cite a reduction in standard deviation by ~30% near the extratropical tropopause. We have also included a citation to the Hegglin 2009 paper regarding reduction of H2O variance using tropopause coordinates.

L139 degree symbol not 30o

changed

L138-145: It seems to me that in tropopause coordinates, the variability associated with the jets is binned below the tropopause (the two maxima around 30S and 30N). Also, the variability in polar regions seems to be clearly related to the tropopopause height, which will explain the large variability in tropopause relative coordinates in the polar regions.

The enhanced upper tropospheric H2O variability near 30S and 30N in tropopause coordinates is related to ENSO variations in the tropopause itself. And yes, the tropopause coordinate does impact H2O variability in polar regions. We have added some additional text to clarify in the paragraph beginning on line 148.

I will not cause those results artifacts, it is simply a consequence of averaging the data differently and it gives you a sense of where the variability is coming from.

It is an artifact in the sense that the variability is introduced by the use of tropopause coordinates.

That said, the variability around the equator is surprising, where is it coming from. Could this be a consequence of the PTGT definition, did you explore the thermal or cold point tropopause, is that something you could consider?

We find similar results using LRT, and the cold point is not appropriate for extratropics. The primary source of the equatorial tropospheric H2O variability in tropopause coordinates is QBO-related variability in the tropopause (similar to ENSO effects near 30S and 30N).

L142 What is “low frequency changes”? Do you mean day to day changes or month to month, please clarify.

As noted in the text, these are related to low frequency ENSO and QBO variations of the tropopause.

Figure 2 Please consider adding the ACE-FTS and the WACCM-SD results to see how robust the MLS results are?

The ACE-FTS 3-month sampling is very different, but the results are clearly seen in Fig. 1. WACCM results are very similar to Fig. 2 (as noted) and we don’t see the value in explicitly showing this.

Figure 3: presumably panels a,b,and c are in pressure and not in pseudo pressure.

Labels have been changed

What is the meaning of the curly brackets in the color bar? They are not mentioned in the text. Consider removing them.

The curly brackets in the colorbar are intended to highlight the non-linear color scale to the reader, as this is an important detail that might otherwise be missed. Text has been added to the figure caption to describe the purpose of them. Our goal with these were to better communicate the results to the reader, so if it is still adding confusion with the new figure caption, we would be happy to remove them.

L152 the equation should be (WACCM-MLS)/MLS*100

Agreed and changed

Figure 4 panel a not pseudo. The results are no fractional, they are in percent. The equation should be (WACCM-MLS)/MLS*100

These are changed

Figure 5 remove curly brackets from the colorbar. Please show the WACCM and ACE-FTS results as well to show how robust the MLS results are

Curly brackets - same as Figure 3 above. ACE-FTS and WACCM (annual average) results are shown in Fig. 3. The seasonal ACE-FTS and WACCM results are similar overall to Fig. 5 and there is not much reason to explicitly show them.

Figure 6 ordinate -> coordinate

Ordinate is the correct word here since we are referring to the y-axis in a cartesian coordinate system.

Remove the curly brackets

See response regarding Figure 3 above.

Add ACE-FTS

Because of the 3-month sampling ACE-FTS results are less interesting and we choose not to include them in Fig. 6.

L217 ExTL , (delete extra space)

changed

Figure 7 Please show the ACE-FTS and the WACCM results as well.

We now include these results in Supplementary material

Figure 9 Please show the ACE-FTS and the WACCM results as well.

The parallel calculations cannot be done on the 3-month sampled ACE-FTS data. For WACCM we now include these results in Supplementary material.

Figure 10 Please show the ACE-FTS and the WACCM results as well.

Since ozone is only being shown to highlight the notable structure of H2O, we feel that showing these results would be superfluous and potentially detract from the narrative we aim to build.

Figure 11 Why is panel (a) the fourth layer above the tropopause rather than the third as in panel b and in figure 8 panel a?

Because ACE-FTS data are sampled at 1-km layers while MLS/WACCM are on 1.25 km grids.

Figure 12, Could you add the same analysis for ozone? Since that is discussed in figure 10

There is no NH-SH coherence for ozone

L269-270 & Figure 13: I think based on Figure 13 this sentence is not justified. ACE-FTS display non statistical significant trends in the region, even suggesting negative trends in the tropics (in the tropical regions where it can actually measure there are some hints of blue) and WACCM-SD magnitudes are so positively bias that it is hard to trust them.

We have slightly changed the wording (now referring to 'systematic, large-scale negative trends'), but maintain our conclusion that the MLS trends are probably spurious, as discussed in Section 4.

That said, I think the authors could paraphrase what they have in the conclusions here, that is: The MLS decreases in the tropical upper troposphere seen in Fig. 13 also disagree with other infrared and microwave satellite observations, which show consistent specific humidity increases in this region over the recent decades (e.g., John et al., 2025; Allan et al., 2022). Likewise, results for ERA5 reanalysis show moistening of the tropical upper troposphere since 2000 (e.g., Li et al., 2024; Allan et al., 2022).

We think the extended discussion of MLS trends belongs in Section 4 and should not be repeated in the main text.

The authors could add 10.1029/2021GL097609 to the ERA5 references.

We have added this citation

L316-318: perhaps something like: As a result, we view the negative trends with caution, as they may represent a potential retrieval artifact in the MLS data that should be further investigated.

Wording changed as suggested

Figure 15 Why are you only showing MLS, please show ACE and WACCM as in Figure 13 and 14

The QBO regression results are nearly identical to MLS as stated in the text, and we don't think it necessary to show them explicitly.

L292 extra space between (and transition

changed

L300 incomplete sentence

fixed

Reviewer 3

Review of "Characterizing variability and vertical structure of water vapor in the extratropical lower stratosphere" by Emily N. Tinney and William J. Randel

The authors effectively highlight the importance, current complications and numerous open questions in characterizing water vapour in the extratropical lower stratosphere, providing a strong motivation for their investigation of vertical water vapour profiles, intercomparing model (WACCM) and satellite data (MLS & ACE-FTS). Notably, their results demonstrate that interannual ExTL water vapour variability is distinct from overworld variability and shows patterns that disagree between the datasets (specifically the MLS trend).

The paper is well-written and thoughtfully structured. The presented results take a meaningful step towards understanding LMS water vapour structure and variability and also provide multiple incentives for further investigations. Therefore, I think the paper is clearly of interest to ACP readers.

Thanks.

I have only a few comments/suggestions:

- Line 142: "We find that much of the enhanced H_2O variance in this region is due to low frequency changes in tropopause height tied to ENSO and QBO variations (Randel et al., 2000)." Please add that these findings are shown/discussed further in section 3.4.

It is not the ENSO and QBO effects on H2O itself that causes the enhanced variance, but rather ENSO and QBO variations in the low latitude tropopause that introduces H2O variance in the upper troposphere when using tropopause coordinates. We have tried to clarify this point in the text in the paragraph beginning on line 148.

- Fig. 5: I think it is currently very difficult to see the dry signals (tape recorder, latitudinal spread etc.) in this colour scheme (everything above 3.5 ppmv looks almost the same to me)...perhaps you could improve this by using different colours or different colour level spacing... For the contour lines: please check that the labels do not overlap (e.g. for the 400K isentrope in JJA).

The colorbar for figure 5 (and figures 4 and 6 has been reworked to cutoff at 3 ppmv, since there were no values showing up in the lower 0.5 ppmv boxes. This allows the < 3 ppmv color to be lighter and more distinct, which we hope fixes some of the issue with seeing the dry layer. We have tried numerous other versions of a colorbar, and think this one shows it best to both regular vision and color-blindness. The contour labels have been fixed where appropriate to the best ability of the plotting software.

- Line 223: "The different behavior for [missing word] is of course due to the large reservoir of below the tropopause, together with the source of overworld variability in the tropics, distinct from ozone." Could you please expand on that? (Do you want to imply that the amount of water vapour in the troposphere influences the importance of tropospheric variability contributions to ExTL variability?)

Missing word = H2O. Yes, the large reservoir of H2O below the tropopause influences the behavior of water vapor in the ExTL. The reworking of this paragraph in response to comments above hopefully helps clarify this point.

- Line 232: "While the overworld variations are consistent among MLS, ACE and WACCM, the ExTL variations are quite different..." Besides the trends not matching the one found in the MLS data, how are the hemispheric mean ACE and WACCM variations different from each other? (Which features differ?)

The trends in MLS overwhelm the variability, so that we can't objectively identify 'other' signals that might be coherent among MLS, ACE-FTS and WACCM (aside from ENSO). We have added this sentiment to the text at line 259-260.

Generally, it would be good if a few more ACE-FTS and WACCM plots could be shown (especially for Fig. 7 and Fig. 9) to allow the reader to directly compare them to the MLS results. Given the number of figures, I think it would suffice to add these plots as supplementary material...

Good suggestion and we have now included ACE-FTS and WACCM versions of Fig. 7 in Supplementary material. We can’t calculate the equivalent of Fig. 9 from the 3-month sampled ACE-FTS data, but we have included a version of Fig. 9 with WACCM results to the supplementary material as well.

I am also currently missing information on the significance of the trends and correlations (what are the confidence intervals? And how were they computed?) - please add this information in some way.

In the original manuscript, trend uncertainties were calculated from the standard errors of the regression coefficients obtained from the multiple linear regression model. Following the suggestion of Reviewer #2, we evaluated the lag-1 autocorrelation of the regression residuals and accounted for serial correlation using an effective sample size correction:

$$N_{eff} = N\frac{(1-r_1)}{(1+r_1)}$$

where r_1 is the lag-1 autocorrelation of the residuals. The standard errors and p-values of the regression coefficients were recomputed using N_{eff} and the significance stippling in Figures 13, 14, and 15 have been updated accordingly. This detail is now clarified in lines 286 - 292.

Technical corrections:

- Caption of Figure 1: "at 45 N" -> " at 45°N"

changed

- Line 125: " 45° N" ->" 45°N"

changed

- Line 139: "near 30o N and S and near the equator" -> "near 30°N and S and near the equator"

changed

- Line 155: "as highlighted in (Charlesworth et al., 2023)." -> "as highlighted in Charlesworth et al., 2023."

changed

- Line 217: "and ExTL , as identified" -> "and ExTL, as identified"

changed

- Line 271: missing word in the sentence that starts with "In general,..."

fixed

- Line 285: "...and hence have less impact on [missing word] (not shown)"

fixed

- Line 297: "In the ExTL and below, variability is tied to tropospheric [missing word]..."

fixed

- Line 300: "As a note, LMS ozone interannual variability is very different from [missing word] ,..."

fixed