Response to Reviewers – "Water vapour and ozone in the upper troposphere – lower stratosphere: Global climatologies from three Canadian limb-viewing instruments" by Paul S. Jeffery et al.

We'd like to thank the Reviewers for their helpful comments. Here we address the comments of each Reviewer, with their comments in black and our responses indented in blue.

Reviewer 1

The manuscript presents ozone and water vapour climatologies for the upper troposphere and lower stratosphere, derived from measurements by three Canadian satellite instruments. The climatologies use special vertical and meridional coordinates - potential temperature relative to the lowest tropopause and equivalent latitude. These coordinates are chosen to reduce smearing out effects from vertical and horizontal transports, and should provide a tighter and more relevant climatology.

The manuscript is generally well written. The underlying assumptions, methods and data appear sound and plausible. Overall the manuscript is very comprehensive and detailed, but also quite long. I make a few suggestion below, to help bring out main results a bit clearer. Still, this is a good scientific paper, using a novel coordinate system, and gives a good overview of water vapor and ozone data from three Canadian satellite instruments. It is acceptable for publication in ACP with just a few minor modifications.

> Thank you for the summary and feedback. Several small edits were made throughout in order to condense portions of the text to address the length. Please see our response to Reviewer 2 concerning condensing the text for the details of these changes.

Figs. 1, 2, 4, 5, 7: It would be good to have lines for a few isentropes (=potential temperature isolines) in these plots. That would help greatly to see where isentropic transport and/or mixing can spread water vapor and ozone, and how that is consistent with the observed climatological values.

> This is an interesting suggestion, but because the data are presented in tropopause-relative potential temperature coordinates it is difficult to represent isentropes given the variable potential temperature of the tropopause itself.

Sections 4.2 and 5.2 compare at great length the climatologies from the different instruments / data versions. Reading this, though, I tend to get lost and/or confused. I never find a concise take-home message. I think it would be very helpful to have one table for water vapor, and one table for ozone that sums up the key differences / messages. That table could have the main latitude bands, main regions below / above the tropopause, major differences / biases / features for each instrument / data set. Also a more concise summarizing paragraph at the end would be good.

> We have summarized the model-measurement comparisons in two Tables, and have changed the text in Sect. 4.2 and 5.2 accordingly, including the summary paragraphs. Changed:
>
> **L538–590**:
> **From** "Having quantified the influence of the instruments' sampling patterns, focus can turn to comparing the measurement climatologies using the subsampled CMAM39-SD climatologies, in Fig. 2, as a reference. Over nearly the entirety of the upper troposphere during the DJF (JJA) period, the subsampled CMAM39-SD climatologies yield elevated levels of water vapour, with average differences of approximately 69 % (73 %) compared to ACE-FTS v4.2, 65 % (69 %) compared to ACE-FTS v3.6, and 65 % (66 %) compared to MAESTRO. Overall, the three measurement climatologies agree to within 4 % in DJF, and 7 % in JJA, with better agreement found between ACE-FTS v3.6 and both other datasets than between ACE-FTS v4.2 and MAESTRO. The remainder of the regions in which CMAM39-SD displays greater water vapour VMRs than the measurement datasets can be subdivided into the spans north and south of approximately 45° from 0–70 K above the tropopause. The southern of these regions yields, in DJF (JJA), differences that are 42 % (23 %), 36 % (28 %), and 51 % (20 %) larger in the model than ACE-FTS v4.2, ACE-FTS v3.6, and MAESTRO. North of 45°, these differences in DJF (JJA) are 22 % (34 %), 24 % (33 %), and 21 % (50 %). Over these two regions and periods, the agreement is variable, with some comparisons showing agreement to within 5 % in DJF and as large as 17 % in JJA (e.g., in the northern lower stratosphere). Altogether, these findings indicate that the satellite datasets are in

generally good agreement across these regions. Note that there is a seasonal influence on the differences, with the measurements showing better agreement with CMAM39-SD and each other during the winter season in each hemisphere.

In the tropical lower stratosphere, the model-derived climatologies, shown in Fig. 2, display smaller water vapour concentrations than those derived from the measurement datasets; this region also corresponds to where the poorest agreement is found between the satellite datasets. The average DJF (JJA) model-measurement deviation for the three measurement datasets is 17 % (19 %) and 25 % (15 %) for ACE-FTS v4.2 and v3.6; however the comparison with MAESTRO yields much greater differences of 83 % (41 %). The cause of this increased water vapour concentration in the MAESTRO dataset is tied to the upper altitude bound of the MAESTRO retrieval, which is lower when the lower stratosphere is drier, and vice-versa, leading to observations of wetter conditions primarily populating higher altitude locations, such as the tropical lower stratosphere. For ACE-FTS, the model-measurement comparison differences in the stratosphere during JJA are greatest over a region approximately 25° wide spanning 40 K in the vertical centered near 30° S and 40–50 K above the tropopause. In DJF, the greatest difference between ACE-FTS v3.6 and the model occurs over an approximately 45° and 30 K region centered around the tropical tropopause, while for ACE-FTS v4.2 only a few grid cells in this region show large differences. Because of the greater consistency in JJA than in DJF, the two versions of ACE-FTS are in better agreement in the former period than in the latter, with the former agreeing to within 5 %, while the latter shows differences of 8 %. In contrast, the large model-measurement differences displayed in the MAESTRO comparisons are not nearly as spatially limited, and align poorly with the two versions of ACE-FTS studied. This apparent overabundance of MAESTRO water vapour in the tropical lower stratosphere, as compared to both CMAM39-SD and ACE-FTS, varies between the two seasons shown, with more extreme differences observed over a larger portion of the lower stratosphere in DJF than in JJA. Thus poor agreement, with differences of 22–66 %, is found between the two versions of ACE-FTS and MAESTRO over this region in both seasons, but within this large range of differences, the agreement in the JJA climatologies is significantly better than that in the DJF climatologies by about 40 %.

Overall there is a clear discrepancy between the model and satellite datasets, with the former yielding elevated water vapour concentrations over the troposphere and polar stratosphere, while showing drier conditions over much of the remaining stratosphere. Despite this, the CMAM39-SD subsampled model climatologies are valuable for assessing the impact of sampling. The measurement datasets show mixed relative performance, with the two ACE-FTS versions showing very good agreement throughout most of the UTLS, with differences of less than 10 %, while the MAESTRO climatologies show varied agreement with the two ACE-FTS versions, with differences ranging from less than 5 % to over 80 %, varying based on location and season. In considering the entire UTLS, from 50 K below to 100 K above the tropopause, the ACE-FTS v4.2 climatologies have an average absolute difference from CMAM39-SD of 28 % (29 %) in DJF (JJA), while ACE-FTS v3.6 comparisons yield a 35 % (28 %) difference, and MAESTRO comparisons a 69 % (44 %) difference. Part of the large difference seen in the MAESTRO comparisons result from the upper bound of the MAESTRO water vapour measurements near 22 km, where measurements tend to possess a high degree of uncertainty due to the worsening quality of MAESTRO water vapour retrievals with altitude. The altitudes included in these tropopause-relative climatologies are higher in the tropics than at the poles because the tropical tropopause is at higher altitude, so the lower-quality MAESTRO measurements from near its upper bound would have a larger influence in the tropics than at the poles in the lower stratosphere. The wet bias in MAESTRO water vapour found here agrees with estimates by Lossow et al. (2019) and Hegglin et al. (2021). While there is overall poor agreement between MAESTRO and ACE-FTS climatologies in DJF, with differences of approximately 40 %, in JJA the overall difference is about 16 %, which is in much better agreement. Some regions of the UTLS show even better agreement, namely the upper troposphere and extratopical lower stratosphere, where the influence of the upper limit to the MAESTRO water vapour product is less impactful. Thus the MAESTRO dataset still provides valuable insight into water vapour in the UTLS, a finding supported by prior validation efforts (e.g., Lossow et al., 2019)."

To: "Having quantified the influence of the instruments' sampling patterns, focus can turn to comparing the measurement climatologies using the subsampled CMAM39-SD climatologies, in Fig. 2, as a reference. The average differences between the measurement and model differences are summarized in Table 1, separated by instrument and the regions discussed below, with the regions corresponding to swaths of the UTLS where the measurement-model comparisons display generally consistent behaviour. Over most of the upper troposphere the subsampled CMAM39-SD climatologies yield elevated levels of water vapour compared to the three instrument climatologies, with differences of 65 % or greater. The three measurement climatologies were found to agree to within 4 % in DJF, and 7 % in JJA, with better agreement found between ACE-FTS v3.6 and both other datasets than between ACE-FTS v4.2 and MAESTRO. The other regions in which CMAM39-SD has higher water vapour than the measurement datasets are the ranges poleward of approximately 45° in each hemisphere and from 0–70 K above the tropopause. Over these two regions and periods, the agreement is variable, with agreement ranging from within 5 % in DJF to as large as 17 % in JJA (e.g., in the northern lower stratosphere). Altogether, these findings indicate that the satellite datasets are in generally good agreement across these regions. Note that there is a seasonal influence on the differences, with the measurements showing better agreement with CMAM39-SD and each other during the winter season in each hemisphere.

In the tropical lower stratosphere, the model-derived climatologies display smaller water vapour concentrations than those derived from the measurement datasets; this region also corresponds to where the poorest agreement is found between the satellite datasets. While the model-measurement differences for the ACE-FTS comparisons are between 15–25 %, the comparison with MAESTRO yields differences 2–4 times greater. The cause of this difference in the MAESTRO dataset is tied to the upper altitude bound of the MAESTRO water vapour retrieval, which is lower when the lower stratosphere is drier, and vice-versa, leading to observations of wetter conditions primarily populating higher altitude locations, such as the tropical lower stratosphere. For ACE-FTS, the model-measurement comparison differences in the stratosphere during JJA are greatest over a region approximately 25° wide spanning 40 K in the vertical centered near 30° S and 40–50 K above the tropopause. In DJF, the greatest difference between ACE-FTS v3.6 and the model occurs over an approximately 45° and 30 K region centered around the tropical tropopause, while for ACE-FTS v4.2 only a few grid cells in this region show large differences. Because of the greater consistency in JJA than in DJF, the two versions of ACE-FTS are in better agreement in the former period than in the latter, with the former agreeing to within 5 %, while the latter shows differences of 8 %. In contrast, the large model-measurement differences displayed in the MAESTRO comparisons are not nearly as spatially limited, and align poorly with the two versions of ACE-FTS studied. This apparent overabundance of MAESTRO water vapour in the tropical lower stratosphere, as compared to both CMAM39-SD and ACE-FTS, varies between the two seasons shown, with more extreme differences observed over a larger portion of the lower stratosphere in DJF than in JJA. Thus poor agreement, with differences of 22–66 %, is found between the two versions of ACE-FTS and MAESTRO over this region, but within this large range of differences the agreement in the JJA climatologies is significantly better than that in the DJF climatologies by about 40 %.

In spite of the discrepancy between water vapour in the model and satellite datasets, the CMAM39-SD subsampled model climatologies are valuable for assessing the impact of sampling. In this assessment, the measurement datasets show mixed relative performance, with the two ACE-FTS versions showing very good agreement throughout most of the UTLS, with differences of less than 10 %, while the MAESTRO climatologies show varied agreement with the two ACE-FTS versions, with differences ranging from less than 5 % to over 80 %, varying based on location and season. Part of the large difference seen in the MAESTRO comparisons result from the upper bound of the MAESTRO water vapour measurements near 22 km, where measurements tend to possess a high degree of uncertainty due to the worsening quality of MAESTRO water vapour retrievals with altitude. The altitudes included in these tropopause-relative climatologies are higher in the tropics than at the poles because the tropical tropopause is at higher altitude, so the lower-quality MAESTRO measurements from near its upper bound would have a larger influence in the tropics than at the poles in the lower stratosphere. The wet bias in MAESTRO water vapour found here agrees with estimates by Lossow et al. (2019) and Hegglin et al. (2021). In considering the entire UTLS, from 50 K below to 100 K above the tropopause, there is overall poor agreement between MAESTRO and ACE-FTS climatologies in DJF, with differences of approximately 40 %, but in JJA the overall difference is about 16 %, which is in much better agreement. Some regions of the UTLS show even better agreement, namely the upper troposphere and extratopical lower stratosphere, where the influence of the upper limit to the MAESTRO water vapour product is less impactful. Thus the MAESTRO dataset still provides valuable insight into water vapour in the UTLS, a finding supported by prior validation efforts (e.g., Lossow et al., 2019).”

L695–749:
From “Differences in instrument climatologies from the CMAM39-SD subsampled climatologies (Fig. 5) show that the subsampled model yields smaller ozone concentrations than the measurements in the upper troposphere, and larger concentrations outside of the tropics in the lower stratosphere. Below the tropopause, the model-measurement differences are largest for comparisons with MAESTRO, and smallest for comparisons with the two sets of ACE-FTS climatologies. Quantitatively, the relative difference below the tropopause between the DJF (JJA) subsampled CMAM39-SD ozone distributions and those from the instruments is approximately 15 % (26 %) for ACE-FTS v4.2, 12 % (23 %) for ACE-FTS v3.6, 106 % (91 %) for MAESTRO, and 19 % (37 %) for OSIRIS. The two versions of ACE-FTS are found to agree closely, to within 3 %, over the upper troposphere, while ACE-FTS v4.2 agrees to within 11 % of OSIRIS, and ACE-FTS v3.6 to within 14 % of OSIRIS. In JJA the agreement between ACE-FTS and OSIRIS worsens somewhat, such that the differences are approximately twice as large as those in DJF. In contrast to this general agreement, the MAESTRO ozone climatology agrees poorly with the others below the tropopause, with the smallest difference being 54 %.

In the lower stratosphere, the subsampled model climatologies show generally larger ozone concentrations than the four measurement datasets, with some exceptions in part of the tropics during DJF. Specifically, the subsampled model climatologies show smaller ozone VMRs than the ACE-FTS v4.2 and MAESTRO climatologies in DJF in the tropics between the tropopause and approximately 50 K above the tropopause. In comparison to the ACE-FTS v3.6 DJF climatology, the subsampled model yields smaller VMRs up to about 100 K above the tropopause; however this difference is largest between the tropopause and about 50 K above the tropopause. For these three datasets, the differences occur primarily between about 5°S and 25°N, and successively greater ozone concentrations are seen in the ACE-FTS v4.2, ACE-FTS v3.6, and MAESTRO climatologies. The average model-measurement differences over this region in DJF are 12 %, 21 %, and 32 % for ACE-FTS v4.2, ACE-FTS v3.6, and MAESTRO respectively, translating to ACE-FTS v3.6 agreeing with the other two datasets to within 11 %, while the MAESTRO and ACE-FTS v4.2 climatologies show a 20 % general difference. The model comparison with the OSIRIS DJF climatology over this region of the tropical lower stratosphere leads to a model-measurement difference of 13 %. These model-measurement differences are in poor agreement with those of the other instruments as the model is found to possess smaller ozone VMRs than OSIRIS over this span, with the closest agreement, of 25 %, found between OSIRIS and ACE-FTS v4.2. In JJA the model yields generally greater ozone VMRs than the three ACE datasets, outside of a few grid cells, as it does in both seasons as compared to OSIRIS.

Outside the tropical lower stratosphere in DJF, the subsampled model climatologies generally display more ozone than the four measurement datasets in the lower stratosphere, with exception for the region 0–20 K above the tropopause south of 60°S, where the model exhibits less ozone than the measurements. For the regions where there is greater model ozone, model-measurement differences are larger in the southern hemisphere than in the northern, with the MAESTRO comparisons showing the largest difference from the model while the two ACE-FTS versions show the smallest differences, in agreement with Hegglin et al. (2021). When averaged over the lower stratosphere, with exception of the span between 5°S and 25°N noted above, the model-measurement differences are 17 % (15 %) for ACE-FTS v4.2, 18 % (14 %) for ACE-FTS v3.6, 31 % (33 %) for MAESTRO, and 26 % (19 %) for OSIRIS in DJF (JJA). Model-measurement differences in the lower stratosphere of the southern hemisphere are 22 % (17 %) for ACE-FTS v4.2, 22 % (17 %) for ACE-FTS v3.6, 36 % (33 %) for MAESTRO, and 29 % (23 %) for OSIRIS in DJF (JJA). In the lower stratosphere of the northern hemisphere, model-measurement differences are 11 % (13 %) for ACE-FTS v4.2, 12 % (11 %) for ACE-FTS v3.6, 24 % (34 %) for MAESTRO, and 24 % (16 %) for OSIRIS in DJF (JJA). These model-measurement comparisons imply agreement to within 2 % for the two ACE-FTS versions. The MAESTRO comparisons with the two versions of ACE-FTS yield differences of 13–23 %, with better agreement in DJF than in JJA by 2–10 %. OSIRIS typically agrees better with MAESTRO in DJF and with ACE-FTS in JJA; the DJF differences between MAESTRO and OSIRIS are less than 5 % and those between OSIRIS and ACE-FTS are 8–13 %, while in JJA these differences are 10–18 % and 3–6 % respectively.

In summary, the ozone climatologies from the satellite datasets show similar features, but there are moderate differences in their comparisons to CMAM39-SD, and by extension each other. In the upper troposphere, the subsampled CMAM39-SD climatologies used for comparisons yield smaller ozone VMRs than those for the instrument climatologies, with the two versions of ACE-FTS displaying the least ozone, followed by the OSIRIS and then the MAESTRO climatologies, which yield the greatest VMRs. This is reversed over much of the stratosphere, except for a portion of the tropics where the model shows less ozone than three of the measurement datasets in DJF. With few exceptions there is agreement to within 10 % between the OSIRIS and ACE-FTS climatologies, with average model-measurement differences over the UTLS in DJF (JJA) of 16 % (16 %) for ACE-FTS v4.2, 16 % (15 %) for ACE-FTS v3.6, and 25 % (21 %) for OSIRIS. The MAESTRO ozone climatologies were found to agree fairly poorly with those from ACE-FTS, and only somewhat better with OSIRIS, with average model-measurement differences of 49 % (47 %) in DJF (JJA) over the entire UTLS. The disagreement of the MAESTRO ozone climatologies with those of the other datasets, coupled to the high degree of variability observed for this product, supports the conclusion that the v3.13 MAESTRO dataset should be used with caution, as it shows the least consistency with other instruments. In contrast, the results for the other three datasets demonstrate their value for UTLS ozone studies"

To: "Focus can now turn to the comparisons between the instrument and the CMAM39-SD subsampled climatologies, shown in Fig. 5. The average differences between the measurement and model climatologies for regions of interest are presented in Table 2. It is immediately evident that the subsampled model yields smaller ozone concentrations than the measurements in the upper troposphere and larger concentrations in the lower stratosphere outside of the tropics. Below the tropopause, the model-measurement differences are largest for comparisons with MAESTRO, and smallest for comparisons with the two sets of ACE-FTS climatologies, with the two versions of ACE-FTS agreeing to within 3 %, and OSIRIS agrees to within 11 % of ACE-FTS v4.2 and 14 % of ACE-FTS v3.6. In JJA the agreement between ACE-FTS and OSIRIS worsens, such that the differences are approximately twice as large as those in DJF. In contrast to this general agreement, MAESTRO ozone agrees poorly with the others below the tropopause, with the smallest difference being 54 %.

In the lower stratosphere, the subsampled model climatologies show generally larger ozone concentrations than the four measurement datasets, with exceptions within the tropics during DJF. Specifically, the subsampled model climatologies show smaller ozone VMRs than the ACE-FTS v4.2 and MAESTRO climatologies in DJF in the tropics between the tropopause and approximately 50 K above the tropopause, as well as with the ACE-FTS v3.6 DJF climatology up to about 100 K above the tropopause; however this difference is largest between the tropopause and about 50 K above the tropopause. For these three datasets, the differences occur primarily between about 5°S and 25°N, and successively greater ozone concentrations are seen in the ACE-FTS v4.2, ACE-FTS v3.6, and MAESTRO climatologies. The average model-measurement differences translate to ACE-FTS v3.6 agreeing with ACE-FTS v4.2 and MAESTRO to within 11 %, while the MAESTRO and ACE-FTS v4.2 climatologies show a 20 % general difference. The model comparison with the OSIRIS DJF climatology over this region leads to an oppositely signed difference, and thus the OSIRIS climatologies are in poor agreement with those from the other instruments, with the closest agreement, of 25 %, found between OSIRIS and ACE-FTS v4.2. In JJA the model yields generally greater ozone VMRs than the three ACE datasets, outside of a few grid cells, as it does in both seasons as compared to OSIRIS.

Outside the tropical lower stratosphere in DJF, the subsampled model climatologies generally display more ozone than the four measurement datasets in the lower stratosphere, with exception for the region 0–20 K above the tropopause south of 60°S, where the model exhibits less ozone than the measurements. For the regions where there is greater model ozone, model-measurement differences are larger in the southern hemisphere than in the northern, with the MAESTRO comparisons showing the largest difference from the model while the two ACE-FTS versions show the smallest differences, in agreement with Hegglin et al. (2021). When averaged over the lower stratosphere, with exception of the span between 5°S and 25°N noted above, the model-measurement comparisons imply agreement to within 2 % for the two ACE-FTS versions. The MAESTRO comparisons with the two versions of ACE-FTS yield differences of 13–23 %, with better agreement in DJF than in JJA by 2–10 %. OSIRIS typically agrees better with MAESTRO in DJF and with ACE-FTS in JJA; the DJF differences between MAESTRO and OSIRIS are less than 5 % and those between OSIRIS and ACE-FTS are 8–13 %, while in JJA these differences are 10–18 % and 3–6 % respectively.

In summary, the ozone climatologies from the satellite datasets show similar features, but there are moderate differences in their comparisons to CMAM39-SD, and each other. In the upper troposphere,

> the subsampled CMAM39-SD climatologies used for comparisons yield smaller ozone VMRs than those for the instrument climatologies, with the two versions of ACE-FTS displaying the least ozone, followed by the OSIRIS and then the MAESTRO climatologies, which yield the greatest VMRs. This is reversed over much of the stratosphere, except for a portion of the tropics where the model shows less ozone than three of the measurement datasets in DJF. Over the entire UTLS, from 50 K below to 100 K above the tropopause, with few exceptions good agreement is found to within 10 % between the OSIRIS and ACE-FTS climatologies, while the MAESTRO ozone climatologies were found to agree fairly poorly with those from ACE-FTS, and only somewhat better with those from OSIRIS, showing differences between 31–33 % from ACE-FTS, and between 24–26 % from OSIRIS. While this disagreement is influenced by the large differences in the upper troposphere, MAESTRO is overall in the poorest agreement with the other datasets. This disagreement, coupled with the large variability observed for this product, supports the conclusion that the v3.13 MAESTRO dataset should be used with caution, as it shows the least consistency with other instruments. In contrast, the results for the other three datasets demonstrate their value for UTLS ozone studies."

L17, 18, 20, 21 (and possibly other places): 8% overall difference. To me it is not clear whether this means an 8% difference between the means (which one is higher then?), or if that means RMS or average absolute differences of about 8%. Please clarify / use better wording.

> Clarified by changing:
> **From** "In turn, this permits a consistent evaluation of the measurements of these two gas species. For the water vapour climatologies produced, a less than 8 % overall difference was found between the climatologies generated from the two versions of ACE-FTS, while comparisons with the MAESTRO climatologies were found to yield 15–41 % overall differences, depending on the version of ACE-FTS and the season. When considering the ozone climatologies, those constructed from the two ACE-FTS versions agreed to within 2 % ..."
> **To** "This in turn permits a more consistent evaluation of the distributions of these two gas species, as assessed through the differences between the model and measurement climatologies. For water vapour, the average absolute relative difference between CMAM39-SD and ACE-FTS differed between the two versions of ACE-FTS by less than 8 %, while the MAESTRO climatologies were found to differ by 15–41 % from ACE-FTS, depending on the version of ACE-FTS and the season. When considering the ozone climatologies, those constructed from the two ACE-FTS versions agreed to within 2 % ..."

L23, "consistent": I don't think two data sets are consistent when they differ by as much as 30 to 35%. Rephrase.

> Changed:
> **From** "Overall these findings indicate that this set of Canadian limb sounders provide consistent water vapour and ozone distributions in the UTLS"
> **To** "These findings indicate that this set of Canadian limb sounders yields generally similar water vapour and ozone distributions in the UTLS, with some exceptions for MAESTRO depending on the season and gas species."

L45: maybe add "and the large vertical and horizontal trace gas gradients at the tropopause" after "itself"

> Along with comments from Reviewer 2, changed:
> **From** "The frequency of these localized mixing processes around the tropopause, along with the inherent variability of the tropopause layer itself, leads to the high degree of variability characteristic of the UTLS."
> **To** "The frequency of these localized mixing processes around the tropopause, along with the inherent variability of the tropopause and the large vertical and horizontal trace gas gradients at the tropopause, leads to the high degree of variability characteristic of the UTLS."

L55 and 56: in both brackets, I think you need ozone and water vapor: you need ozone to generate O1D which then with water vapor generates OH. Both ozone and water vapor are radiatively important in the UTLS.

> Changed first mention to water vapour only for the ice surfaces, and removed the species specification from the other brackets:
> **From** "Outside of their role as greenhouse gases, these two species collectively influence the chemical composition of the UTLS by providing ice surfaces for heterogeneous chemical reactions (water vapour;

> Brasseur and Solomon, 2005), producing hydroxyl radicals that rapidly react with other gases (water vapour; Brasseur and Solomon, 2005), and by absorbing ultraviolet and infrared light to moderate the energy balance of the UTLS (ozone; Lacis et al., 1990)."
> **To** "Outside of their role as greenhouse gases, these two species collectively influence the chemical composition of the UTLS by providing ice surfaces for heterogeneous chemical reactions (water vapour; Brasseur and Solomon, 2005), producing hydroxyl radicals that rapidly react with other gases (Brasseur and Solomon, 2005), and by absorbing ultraviolet and infrared light to moderate the energy balance of the UTLS (Lacis et al., 1990; Brasseur and Solomon, 2005)."

Line 78: somewhere around here you might want to say something about the difficulties for satellite measurements in the UTLS. Nadir looking instruments have wide weighting functions which do not resolve UTLS features well. LIMB looking instruments have very long path lengths, which also tend to average out finer structures.

> Added on L80: "However, satellite instruments can still encounter difficulties making measurements in the UTLS because, for example, of the coarse vertical sensitivity of nadir-viewing instruments, which can prevent vertical UTLS features from being well resolved, or because of the long path lengths of limb-viewing instruments, which can smear out finer structures in UTLS gas distributions."

L113, 119: "have been shown capable of reducing" -> "can reduce", "have been shown capable of grouping" -> "can group"; there may be more places like this. Also "have been found to ..." can almost always be replaced by "are" or "can". Much less wordy, much less hedging. If it looks like a spade and handles like a spade, you might as well call it a spade.

> Along with comments from Reviewer 2, we have changed the following:
> **L113**:
> **From** "While these instruments have been employed in previous climatology studies (e.g., Neu et al., 2014; SPARC-DI, 2017; Hegglin et al., 2021), the climatologies produced here employ tropopause-referenced vertical coordinates, which have been shown capable of reducing the effects of geophysical variability by parcelling the data based on the processes which drive this variability, thereby ensuring only similarly driven data are assessed or compared (e.g., Pan et al., 2004; Hoor et al., 2004; Hegglin et al., 2008; Manney et al., 2011)."
> **To** "While data from these instruments has been used in previous climatology studies (e.g., Neu et al., 2014; SPARC-DI, 2017; Hegglin et al., 2021), the climatologies produced here use tropopause-referenced vertical coordinates, which can reduce differences in climatologies arising from different sampling of geophysical variability by parcelling the data based on the processes which drive some of this variability, thereby ensuring only data driven by similar geophysical processes are assessed or compared (e.g., Pan et al., 2004; Hoor et al., 2004; Hegglin et al., 2008; Manney et al., 2011)."
>
> **L119**:
> **From** "The meridional coordinate employed in these climatologies is potential vorticity derived equivalent latitude, which has been shown capable of grouping air masses based on dynamical conditions, thereby reducing the effect of sporadic meridional transport in the UTLS."
> **To** "The meridional coordinate used in these climatologies is potential vorticity derived equivalent latitude, which can group air masses based on dynamical conditions, thereby reducing the effect of sporadic meridional transport in the UTLS on comparisons."
>
> **L768**:
> **From** "It should be noted that ACE-FTS v3.6 showed better agreement with MAESTRO than ACE-FTS v4.2 did; but this is influenced by the MAESTRO product using the v3.6 temperature and pressure information in its retrievals, which has been shown to lead to a drift in the ACE-FTS v3.6 data."
> **To** "It should be noted that ACE-FTS v3.6 showed better agreement with MAESTRO than ACE-FTS v4.2 did; but this is influenced by the MAESTRO product using the v3.6 temperature and pressure information, which has been found to be the source of a drift in the ACE-FTS v3.6 data, thus leading to an expected drift in the MAESTRO data similar to the v3.6 data."

L123: can it really be “well constrained”. I have my doubts, and suggest “constrained better”

> Changed:
> **From** “In spite of these limitations, through the combination of these coordinates the inherent variability of the UTLS can be well constrained in the resulting climatologies.”
> **To** “In spite of these limitations, through the combination of these coordinates the inherent variability of the UTLS can be constrained better in the resulting climatologies.”

L358: suggest to drop “minimizes the internal variability of the model’s circulation”. The internal variabilty of the model is what it is. By nudging you might modify it, but more importantly you push the model towards a real observed situation (with sometimes significant side-effects, particularly if the model does not like it.)

> Changed:
> **From** “To address this deficiency, Newtonian relaxation, known as nudging, can be applied to constrain temperature and circulation fields to a reanalysis dataset. This minimizes the internal variability of the model’s circulation and allows for more direct comparisons.”
> **To** “To address this deficiency, Newtonian relaxation, known as nudging, can be applied to constrain temperature and circulation fields to a reanalysis dataset, pushing the model conditions toward that of the observed atmosphere and allowing for more direct comparisons.”

L363: here is an example where you can easily drop “have been found to”

> Quoted and similar text has been removed throughout document. Please see our response to Reviewer 2 concerning condensing the text for the details of these changes. Changed:
> **From** “Prior nudged specified dynamics runs of CMAM, such as CMAM30-SD, which focused on the 1980 to 2010 period, have been found to agree well with Aura-MLS observations of meteorological fields.”
> **To** “Prior nudged specified dynamics runs of CMAM, such as CMAM30-SD, which focused on 1980 to 2010, agree well with Aura-MLS observations of meteorological fields.”

L428, Eq. 2: I am surprised to see X(i,j) in the denominator, not Y(i,j). Or, I am surprised to have X-Y in the enumerator, not Y-X. Is there a specific reason to use this unusual notation? Please clean up or explain.

> This particular notation is intentional. We found that the discussion was clearest when put in terms of whether the model climatologies were higher or lower than the measurement climatologies (model minus measurement), and with the model as the reference for this comparison (denominator). This resulted in the formulation expressed in Eq. 2.

L480: might want to add “troposphere” after “hemisphere”

> The seasonal asymmetry is noted up to about 20 K above the tropopause, but we have emphasized the troposphere. Along with further changes from Reviewer 2, line is changed:
> **From** “Seasonal enhancement of water vapour is evident between the two sets of seasonal climatologies, with the summer hemisphere (northern for JJA, southern for DJF), showing elevated water vapour concentrations compared to the winter hemisphere as a consequence of the seasonal dehydration in winter. Lower wintertime temperatures are associated with a decrease in saturation vapour pressure, resulting in a reduced carrying capacity for water vapour in the winter hemisphere. This effect is evident in the troposphere and the lowermost 20–30 K of the stratosphere, a region where the summer hemisphere displays a consistently greater water vapour concentration than the winter hemisphere.”
> **To** “Seasonal enhancement of water vapour, predominately in the troposphere but also over the lowermost 20–30 K of the stratosphere, is evident between the two sets of seasonal climatologies, with the summer hemisphere showing elevated water vapour concentrations compared to the winter hemisphere. This is influenced by the seasonal cycle in tropical cold point temperatures, and the decrease in saturation vapour pressure associated with lower temperatures in the winter that result in a reduced carrying capacity for water vapour in the winter hemisphere.”

Reviewer 2

This paper co-locates water vapor and ozone mixing ratio output from the nudged Canadian Middle Atmospheric Model (CMAM39-SD) with measurements by three limb-viewing satellite instruments, ACE-FTS, MAESTRO and OSIRIS (ozone only), to evaluate the consistency between the three sets of satellite data. The model output and satellite measurements are 2-dimensionally binned using equivalent latitude (5° bins) in the horizontal and potential temperature relative to the tropopause (10 K bins) in the vertical. "Zonal" mean values over 3-month periods are computed for each bin using 14 years of satellite data (2004-2018) to create "zonal mean" climatologies for water vapor and ozone in different seasons, although only summer and winter data (DJF/JJA) are presented. Model data are sub-sampled to correspond to the times and locations of measurements by each of the satellite instruments, with model vs model comparisons indicating that differences between some of the measurement climatologies are due to the different instrumental sampling patterns. Differences between model and measurement climatologies are also presented for DJF and JJA, but these can (should) only be used to evaluate the consistency of the 2-dimensional differences across all three (two for water vapor) satellite instruments, not the differences themselves because the model output itself may be (is likely) biased.

Thank you for the summary and following detailed comments.

General Comments:

My major sticking point with this method of comparing the measurements of ACE-FTS and MAESTRO is that there should be no need to involve climatologies or a model since the two instruments "share the same sun tracker and optical bore". Why can't the spatially and temporally coincident measurements of these two instruments be compared directly? I was looking for an explanation of this in the paper but did not find one. If there is a good reason why direct comparisons cannot be made between the two SCISAT-1 instruments, please add it to the paper. The complexities of determining climatologies and including a model are more warranted in comparisons with OSIRIS, which is on a completely different satellite and samples the globe in a very different way from the other two instruments.

This paper is not aimed at comparing measurements as a satellite validation paper would, but at examining and comparing global climatologies, as such a direct comparison between the ACE-FTS and MAESTRO measurements would not fit within the scope of this paper. The direct comparisons would still show bias as the SCISAT-1 instruments, despite sharing a sun tracker and optical bore, do not make measurements with the same vertical resolution. Furthermore, not all measurements yield successfully retrieved profiles for the trace gases of interest. Thus any attempt at direct comparisons would still have some of the bias, due to differences in their sampling, that we are trying to avoid. To this end, we have developed the model-comparison method to facilitate these comparisons, and have used this technique for all datasets to ensure consistency in our methods.

I fully support plotting relative differences in the Figures because of the large dynamic ranges of water vapor and ozone mixing ratios in the UTLS. However, I find that the color scales used in these Figures don't provide enough resolution to show the greatest amount of detail. Can other colors (like green) be incorporated in the color scales to increase the visual resolution of the plotted data? Also, the color scale for the CMAM – satellite relative differences include white, and large areas of these plots are white, which could lead some readers to mistakenly assume that there are no biases in those areas instead of there being inadequate data populations in those areas.

We selected the colour schemes for these figures based on guidance from the journal about palettes that are well-suited for readers with colour vision deficiencies. To add clarity and address your second point, concerning the CMAM – satellite relative difference plots, we have added hatching to all of the plots and the following sentence to their captions:
"Note the addition of hatching to the plots where there are no data available."

Figure captions should identify the content of the graphs so they can be better understood - like which panels show "ACE-FTS v4.2 (top row), ACE-FTS v3.6 (middle row), and MAESTRO (bottom row)", not the body text (Line 504). The body text should provide explanations of what the graphs indicate through their interpretation. For Figure 1, there's no need to include "ACE-FTS v4.2 (top row), ACE-FTS v3.6 (middle row), and MAESTRO (bottom row)" in the caption because each panel already identifies the instrument and season.

Both mentions in the body text of figure row have been removed. To ensure thorough figure descriptions

> we prefer to retain the row descriptions in the figure captions. Changed
> **L504:**
> **From** "While many of the prominent features are present in the climatologies constructed from each dataset, the specifics can vary, as shown in Fig. 2, which compares the climatologies constructed from CMAM39-SD, subsampled to the times and locations of the ACE-FTS v4.2 (top row), ACE-FTS v3.6 (middle row), and MAESTRO (bottom row) measurements. While the CMAM39-SD subsampled climatologies (Fig. 2 first and second columns) generally show patterns of water vapour consistent with the measurement climatologies, there are differences found in all of the subsampled climatologies (Fig. 2 third and fourth columns)."
> **To** "While many of the prominent features are present in the climatologies constructed from each dataset, they can vary, as shown in Fig. 2, which compares the climatologies constructed from CMAM39-SD, subsampled to the times and locations of the ACE-FTS v4.2, ACE-FTS v3.6, and MAESTRO measurements. The CMAM39-SD subsampled climatologies (Fig. 2, first and second columns) generally show patterns of water vapour consistent with the measurement climatologies, but there are large relative differences between the measurement and model-subsampled climatologies (Fig. 2 third and fourth columns)."
>
> **L662:**
> **From** "Figure 5 shows DJF and JJA ozone climatologies generated from CMAM39-SD sampled to the times and locations of the ACE-FTS v4.2 (top row), v3.6 (second row), MAESTRO (third row), and OSIRIS (bottom row) measurements, as well as the comparison between these and the associated satellite climatologies."
> **To** " Figure 5 shows DJF and JJA ozone climatologies generated from CMAM39-SD sampled to the times and locations of the ACE-FTS v4.2, ACE-FTS v3.6, MAESTRO, and OSIRIS measurements, as well as the comparison between these and the associated satellite climatologies."

In certain sections of the paper (e.g., Lines 541-548; Lines 727-736) the flow of the presentation gets very bogged down by an overload of numbers (statistics) that appear in the text. A remedy for this problem would be to include tables where these statistics can be presented, allowing the implications of these numbers to be presented in the text instead of the numbers themselves.

> As per the suggestion of both Reviewers, we have summarized the model-measurement comparisons in two Tables, and have updated the text in Sect. 4.2 and 5.2 accordingly, including the summary paragraphs. Please refer to the above comment by Reviewer 1 for the exact text change.

There are ample opportunities to condense multiple sentences into one, reducing the "wordiness" of the paper. Here are two examples, but there are many more places in the paper where repetitive text can be reduced. This would make the paper more concise and easier to read.

Lines 479-482: For each satellite instrument, the JJA and DJF water vapor climatologies differ, mainly due to the seasonal cycle in tropical cold point temperatures that determine the amount of water vapor entering the tropical lower stratosphere.

> Thank you for the suggestions, we have attempted to condense the text. The lines were condensed, but not as much as suggested given other comments from both Reviewers. Changed:
> **From** "Seasonal enhancement of water vapour is evident between the two sets of seasonal climatologies, with the summer hemisphere (northern for JJA, southern for DJF), showing elevated water vapour concentrations compared to the winter hemisphere as a consequence of the seasonal dehydration in winter. Lower wintertime temperatures are associated with a decrease in saturation vapour pressure, resulting in a reduced carrying capacity for water vapour in the winter hemisphere. This effect is evident in the troposphere and the lowermost 20–30 K of the stratosphere, a region where the summer hemisphere displays a consistently greater water vapour concentration than the winter hemisphere."
> **To** "Seasonal enhancement of water vapour, predominately in the troposphere but also over the lowermost 20–30 K of the stratosphere, is evident between the two sets of seasonal climatologies, with the summer hemisphere showing elevated water vapour concentrations compared to the winter hemisphere. This is influenced by the seasonal cycle in tropical cold point temperatures, and the decrease in saturation vapour pressure associated with lower temperatures in the winter that result in a reduced carrying capacity for water vapour in the winter hemisphere."

Lines 596-599: The ozone climatologies for different satellite instruments (Fig. 4, columns 1 and 2) share common features, especially the large mixing ratio increases from a relatively uniform 0.05-0.1 ppmv below the tropopause to values nearly ten times that above 70 K.

Updated this, and the following sentence. Changed:
From "Many features of the equivalent latitude zonal-mean multiyear-mean ozone climatologies (first and second columns in Fig. 4) are consistent among the datasets, with the most prominent feature being the clear gradient in ozone extending from a nearly uniform minimum below 30–50 K above the tropopause to VMR values nearly an order of magnitude greater near the top of the UTLS. All the ozone datasets show relatively uniformly small values, approximately 0.05–0.1 ppmv, below the tropopause. This feature arises because the tropopause acts as a transport barrier, largely confining higher ozone values to the stratosphere, while the reactivity of ozone as an oxidizing agent prevents it from becoming well-mixed over time."
To "Many features of the ozone measurement climatologies (first and second columns in Fig. 4) are broadly consistent, most notably the large gradient in ozone VMR from a relatively uniform 0.05–0.1 ppmv below the tropopause to values nearly an order of magnitude greater above 70 K. This gradient arises because the tropopause acts as a transport barrier, largely confining higher ozone values to the stratosphere, while the reactivity of ozone as an oxidizing agent prevents it from becoming well-mixed over time."

The following lines were also condensed as per the above suggestion. Changed:
L207:
From "Bognar et al. (2019) found ACE-FTS v3.6 ozone over the high Arctic to be about 5 % larger in the upper troposphere and 5 % smaller in the lower stratosphere, than OSIRIS, and between 5 and 10 % larger than MAESTRO over the UTLS. In addition to their work in evaluating the drift in the ACE-FTS products, Sheese et al. (2022) also determined that ACE-FTS v3.6 ..."
To "Bognar et al. (2019) found ACE-FTS v3.6 ozone over the high Arctic to be about 5 % larger in the upper troposphere and 5 % smaller in the lower stratosphere, than OSIRIS, and 5–10 % larger than MAESTRO throughout the UTLS. In addition to their work in evaluating the drift in the ACE-FTS products, Sheese et al. (2022) determined that ACE-FTS v3.6 ..."

L240:
From "This reliance on the ACE-FTS v3.6 pressure and temperature does however introduce the possibility of a drift in the MAESTRO products because of the systematic CO_2 modeling error discussed above and in Sheese et al. (2022)."
To "This reliance on the ACE-FTS v3.6 pressure and temperature introduces the possibility of a drift in the MAESTRO products because of the systematic CO_2 modeling error discussed above and in Sheese et al. (2022)."

L343:
From "All MAESTRO profiles need an associated successful ACE-FTS retrieval to provide temperature and pressure information and current versions of the MAESTRO products use the ACE-FTS v3.6 information for these fields. Therefore the JETPAC product produced for ACE-FTS v3.6 has also been used for the MAESTRO datasets."
To "MAESTRO profiles need an associated successful ACE-FTS retrieval to provide temperature and pressure information, and as the current MAESTRO products use the ACE-FTS v3.6 information for these fields, the JETPAC product for ACE-FTS v3.6 have also been used for the MAESTRO datasets."

L366:
From "The vertical resolution, which varies with altitude, increases with altitude from approximately 900 m at 300 hPa to 1500 m at 30 hPa (Scinocca et al., 2008)."
To "The vertical resolution, which coarsens with altitude, ranges from approximately 900 m at 300 hPa to 1500 m at 30 hPa (Scinocca et al., 2008)."

L386:
From "To analyze the distribution of water vapour and ozone in the UTLS, equivalent latitude zonal-mean climatologies have been generated. In Sect. 3.1 the method used to generate these climatologies is

outlined, which is followed by the comparison methodology in Sect. 3.2."
To "To analyze the distribution of water vapour and ozone in the UTLS, equivalent latitude zonal-mean climatologies are generated. Section 3.1 outlines the method used to generate these climatologies, followed by the comparison methodology in Sect. 3.2."

L410:
From "Following this, the data are divided into 5° equivalent latitude bins at each vertical level. Both the tropopause and equivalent latitude information come from the JETPAC, or JETPAC-like in the case of CMAM39-SD, products generated for each dataset."
To "The data are then divided into 5° equivalent latitude bins at each vertical level. Both the tropopause and equivalent latitude information come from the JETPAC, or JETPAC-like for CMAM39-SD, products generated for each dataset."

L432:
From "As the CMAM39-SD model simulation has the highest spatial density and most global coverage, climatologies derived from it have been chosen as the reference, with all other climatologies being compared to them."
To "As the CMAM39-SD model simulation has the highest spatial density and most global coverage, climatologies derived from it are chosen as the reference, with all other climatologies being compared to them."

L439:
From "To account for the impact of this sampling bias, comparisons have been made between the satellite climatologies and climatologies generated from model data sampled along measurement profile pathways through the atmosphere."
To "To account for the impact of this sampling bias, comparisons are made between the satellite climatologies and climatologies generated from model data sampled along measurement profile pathways through the atmosphere."

L622:
From "As shown in the subsampled CMAM39-SD climatologies in Fig. 5 (first and second columns), there is little variation in ACE-FTS and MAESTRO sampling of this feature, indicating that this difference is most likely related to the ozone products themselves. This is supported by the comparison of the CMAM39-SD subsampled climatologies in Fig. 6, which emphasizes that there is little difference between the ACE-FTS and MAESTRO sampling patterns for ozone."
To "As shown in the subsampled CMAM39-SD climatologies in Fig. 5 (first and second columns) and the comparison of the CMAM39-SD subsampled climatologies in Fig. 6, there is little variation in ACE-FTS and MAESTRO sampling of this feature, indicating that this difference is most likely related to the ozone products themselves."

L779:
From "In addition to assessing the agreement of the measurement climatologies, the subsampled model climatologies allowed for evaluation of the influence of sampling patterns on the representation of water vapour and ozone in the UTLS."
To "In addition to assessing the agreement of the measurement climatologies, the subsampled model climatologies permit investigation into the influence of sampling patterns on the representation of water vapour and ozone in the UTLS."

My final general comment is to question if ACP is the correct journal for this satellite measurement comparison paper. I think AMT would be a better fit, but I will leave it to the authors and the ACP editor to decide this.

We would like to note that this paper is not focused on satellite profile comparisons, but the preparation, utilization, and comparison of water vapour and ozone climatologies, hence fitting within the scope of ACP. We have updated the text in the abstract, on L15, to emphasize this. Changed:
L15:
From "Specifically, this method of using a subsampled model addresses the impact of each instrument's

measuring pattern, and allows for the quantification of the influence of different measurement patterns on multiyear climatologies. In turn, this permits a more consistent evaluation of the measurements of these two gas species."
To "Specifically, this method of using a subsampled model addresses the impact of each instrument's measuring pattern, and allows for the quantification of the influence of different measurement patterns on multiyear climatologies. This in turn permits a more consistent evaluation of the distributions of these two gas species, as assessed through the differences between the model and measurement climatologies."

Other comments:

L2: What does this study include data only through May 2018 since it is now July 2022?

The CMAM39-SD model run was for the period 1980 to 2018. Hence we kept to the period for which we have model output.

L21: consolidate to "by 30-35% and 25%, respectively"

Changed:
From "The MAESTRO ozone climatologies were found to differ from ACE-FTS and OSIRIS by about 30–35 % for the former, and 25 % for the latter, albeit with regions of better agreement within the UTLS."
To " The MAESTRO ozone climatologies differ from those from ACE-FTS and OSIRIS by 30–35 % and 25 % respectively, albeit with regions of better agreement within the UTLS."

L23: One might ask why are all three limb sounders are necessary?

A major focus of this paper was to examine Canadian satellite datasets as a contribution to the OCTAV-UTLS project, hence the inclusion of all three instruments (ACE-FTS, MAESTRO, and OSIRIS).

L29: It is difficult for model output and observational datasets to be "similarly composed" since the data origins are completely different. Please explain further what you mean here.

For clarification, this term was included to refer to the fact that climatologies should generally be compared only if they share the same coordinate system. Changed:
From "Additional uses of climatologies include evaluating systematic differences between similarly composed model or observational datasets (e.g., Eckstein et al., 2017; Kolonjari et al., 2018), and acting as a priori information for the retrieval of trace gas profiles from observational measurements (e.g., Vigouroux et al., 2008; Sofieva et al., 2014)."
To "Additional uses of climatologies include evaluating systematic differences between model or observational datasets (e.g., Eckstein et al., 2017; Kolonjari et al., 2018), evaluating trends (e.g., Deeter et al., 2018), and acting as a priori information for the retrieval of trace gas profiles from measurements (e.g., Vigouroux et al., 2008; Sofieva et al., 2014)."

L30: How about the use of climatologies for trend evaluations?

This was added as shown in the line above (L29).

L42: "mixing processes that do arise' is vague. Please be more precise.

For clarity and to specify the processes listed in the prior sentence, changed:
From "Many of the mixing processes that do arise tend to be variable in nature, leading to rapidly-changing small-scale gas features that further impact the trace gas distributions around the UTLS."
To "Many of these mixing processes are variable in nature, leading to rapidly-changing small-scale gas features that further impact the trace gas distributions around the UTLS."

L45: please briefly define "tropopause layer"

The tropopause has finite depth, and can be referred to as a layer which has characteristics of both the UT and LS in terms of VMR and seasonal variation. For clarity we have changed "tropopause layer" to "tropopause" on L45 and L68.

L52: why are the strengths of their "greenhouse gas efficiency" so high in the UTLS?

Added a brief explanation. Further detail can be found in the provided references. Changed:
From "Atmospheric temperatures are at a local minimum in the UTLS, which maximizes the influence that radiatively active gases have on atmospheric forcing (Forster and Shine, 2002; Gettelman et al., 2011; Riese et al., 2012). This holds for water vapour and ozone, two of the most important greenhouse gases, as both exhibit their greatest impact on surface temperature from their UTLS distributions where their greenhouse gas efficiency is largest (Lacis et al., 1990; Forster and Shine, 1997; Solomon et al., 2010)."
To "Atmospheric temperatures are at a local minimum in the UTLS, maximizing the thermal contrast of the upwelling infrared radiation absorbed as compared to that emitted, which maximizes the influence of gases that absorb in the infrared on the atmospheric forcing of surface temperature (Forster and Shine, 1997, 2002; Gettelman et al., 2011; Riese et al., 2012). This holds for water vapour and ozone, which this effect makes two of the most important greenhouse gases (Lacis et al., 1990; Forster and Shine, 1997; Solomon et al., 2010)."

L56-58: replace "ground" with "surface" since water covers 75% of Earth. Add "following the lapse rate" after "upwards to the tropopause". Replace "away" with "poleward".

Replaced "ground" with "surface" on L56 and L232. The other two suggested changes have been made. Changed:
L56
From "Water vapour is primarily concentrated near the ground, with a strong negative gradient extending upwards to the tropopause, while ozone is mainly concentrated in the stratosphere, with a gradient extending away from its maximum concentration in the tropical middle stratosphere."
To "Water vapour is primarily concentrated near the surface, with a strong negative gradient extending upwards to the tropopause following the lapse rate, while ozone is mainly concentrated in the stratosphere, with a gradient extending poleward from its maximum concentration in the tropical middle stratosphere where the majority of ozone production occurs."

L232
From "One measurement sequence consists of 60 spectra taken between the cloud tops and 100 km above the ground, as well as 20 spectra taken between 100 km and 150 km for use as reference spectra."
To "One measurement sequence consists of 60 spectra taken between the cloud tops and 100 km above the surface, as well as 20 spectra taken between 100 km and 150 km for use as reference spectra."

L60: add "in the tropical middle and upper stratosphere" after "the oxidation of methane"

Text updated to reflect this oxidation being in the middle and upper stratosphere. Changed:
From "The limited stratospheric water vapour arises mainly from upwards transport at the equator, though these air parcels are largely dehydrated when passing through the low-temperature tropical tropopause region, with the oxidation of methane serving as a secondary source."
To "The limited stratospheric water vapour arises mainly from upward transport at the equator, though these air parcels are largely dehydrated when passing through the cold tropical tropopause region, with the oxidation of methane, largely in the middle and upper stratosphere, serving as a smaller secondary source."

L70: Typical "radiosondes" struggle to measure RH in the UTLS because it is very cold and relatively dry. Perhaps you instead mean "frost point hygrometers". Omit "of trace gases" since water vapor and ozone are the focus of this sentence.

As shown by Miloshevich et al. (2009) there can be good agreement with bias-corrected radiosonde (such as those from GRUAN (GCOS (Global Climate Observing System) Reference Upper-Air Network; Dirksen et al., 2014) and frost-point hygrometer (FPH) profiles in the UTLS. However, we concede that frost-point hygrometers are more widely used for studies of UTLS humidity, and have made the change:

From "Many UTLS-focused studies of water vapour and ozone have employed aircraft and balloon-borne instruments, including ozonesondes and radiosondes, to make highly vertically resolved measurements of trace gases."
To "Many UTLS-focused studies use aircraft and balloon-borne instruments, including ozonesondes and frost-point hygrometers, to make highly vertically resolved measurements of water vapour and ozone."

Miloshevich, L., Vömel, H., Whiteman, D., and Leblanc, T.: Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements. J. Geophys. Res., 114, D11305, https://doi.org/10.1029/2008JD011565, 2009.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463–4490, https://doi.org/10.5194/amt-7-4463-2014, 2014.

L77: I think "characterized" is too soft here. Model simulations must be "validated" against measurements.

We have chosen to change to move to a more quantitative term in "evaluated," as opposed to "characterized," as "validated" can be a problematic view of models as per Oreskes et al. (1994). Changed:
From "Chemistry-climate models can aid investigations into climate forcers, such as greenhouse gases, by providing detailed simulations of chemical and transport processes; however these must be characterized against measurements to ensure their veracity."
To "Chemistry-climate models can aid investigations into climate forcers, such as greenhouse gases, by providing detailed simulations of chemical and transport processes; however, these must be evaluated against measurements to ensure their veracity."

Oreskes, N., Shrader-Frechette, K., and Belitz, K.: Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences, Science, 263, 641–646, https://doi.org/10.1126/science.263.5147.641, 1994

L93: A paper by Read et al. was recently published that compares frost point hygrometer and satellite measurements in the UT.

Read, W. G., Stiller, G., Lossow, S., Kiefer, M., Khosrawi, F., Hurst, D., Vömel, H., Rosenlof, K., Dinelli, B. M., Raspollini, P., Nedoluha, G. E., Gille, J. C., Kasai, Y., Eriksson, P., Sioris, C. E., Walker, K. A., Weigel, K., Burrows, J. P., and Rozanov, A.: The SPARC Water Vapor Assessment II: assessment of satellite measurements of upper tropospheric humidity, Atmos. Meas. Tech., 15, 3377–3400, https://doi.org/10.5194/amt-15-3377-2022, 2022.

Reference is added to L100.

L103: "can possess water vapor differences >10%". Do climatologies "possess" differences"? Differences from what?

Changed:
From "Toohey et al. (2013), as part of the SPARC Data Initiative, found that sampling biases can lead some measurement climatologies to possess water vapour differences in excess of 10 % and differences in ozone exceeding 20 % around the UTLS, emphasizing the need to constrain this error in future climatology comparisons."
To "Toohey et al. (2013), as part of the SPARC Data Initiative, found that sampling biases can lead to differences around the UTLS in excess of 10 % between water vapour measurement climatologies, and in excess of 20 % between ozone climatologies, emphasizing the need to better constrain this error in future climatology comparisons."

L109: add "down" before "into the UTLS"

Changed:
From "All three instruments measure into the UTLS with high vertical resolution, have long data records, and provide global coverage over the course of the year, making them good fits to study this highly variable region."
To "All three instruments measure down into the UTLS with high vertical resolution, have long data

records, and provide global coverage over the course of the year, making them good fits to study this highly variable region."

L114: add "some of" before "this variability". What are "similarly driven data"?

The term refers to measurements of air parcels impacted (driven) by similar drivers of geophysical variability. Along with comments from Reviewer 1, the text has been updated. Changed:
From "While data from these instruments has been used in previous climatology studies (e.g., Neu et al., 2014; SPARC-DI, 2017; Hegglin et al., 2021), the climatologies produced here employ tropopause-referenced vertical coordinates, which have been shown capable of reducing the effects of geophysical variability by parcelling the data based on the processes which drive this variability, thereby ensuring only similarly driven data are assessed or compared."
To "While data from these instruments has been used in previous climatology studies (e.g., Neu et al., 2014; SPARC-DI, 2017; Hegglin et al., 2021), the climatologies produced here use tropopause-referenced vertical coordinates, which can reduce differences in climatologies arising from different sampling of geophysical variability by parcelling the data based on the processes which drive some of this variability, thereby ensuring only data driven by similar geophysical processes are assessed or compared."

L123: I think "well constrained" is too strong. "reduced" is more appropriate.

Along with comments from Reviewer 1, the text has been updated. Changed:
From "In spite of these limitations, through the combination of these coordinates the inherent variability of the UTLS can be well constrained in the resulting climatologies."
To "In spite of these limitations, through the combination of these coordinates the inherent variability of the UTLS can be constrained better in the resulting climatologies."

L170: I have no idea what "and has an improved rate of change" means in this context.

This statement refers to the trend in the CO_2 input used for ACE-FTS retrievals. Changed wording to more closely match Boone et al. 2020:
From "As described in Boone et al. 2020, the key difference between these versions is that v4.2 retrievals use a CO_2 input for the temperature and pressure retrievals that explicitly accounts for seasonal and latitudinal variations and has an improved rate of change compared to the v3.6 CO_2 input, as well as consideration for the age of air in the stratospheric distribution of CO_2."
To "As described in Boone et al. 2020, a key difference between these versions is that v4.2 retrievals use a CO_2 input for the temperature and pressure retrievals that explicitly accounts for seasonal and latitudinal variations and has an improved temporal variation compared to the v3.6 CO_2 input, as well as consideration for the age of air in the stratospheric distribution of CO_2."

L198: Aren't the biases between all instrument pairs dependent on the comparison instruments?

Here we are emphasizing that not all biases have the same sign, but have updated the text to remove the ambiguity. Changed:
From "However, comparisons with the Stratospheric Aerosol and Gas Experiment III instrument on the International Space Station (SAGE III/ISS) by Davis et al. (2021) showed a wet bias for ACE-FTS v3.6 exceeding 20 % in the upper troposphere, and at about 5 % in the lower stratosphere, indicating a dependence of the bias on the comparison instrument."
To "However, comparisons with the Stratospheric Aerosol and Gas Experiment III instrument on the International Space Station (SAGE III/ISS) by Davis et al. (2021) showed a wet bias for ACE-FTS v3.6 exceeding 20 % in the upper troposphere, and at about 5 % in the lower stratosphere."

L200: FTIR vertical resolution is, at best, extremely coarse. How can a UTLS bias be determined using FTIR retrievals when the water vapor gradient is so strong in this region?

Weaver et al. (2019) compared measurements of water vapour from the ground-based PEARL FTIR to those from satellites. These measurements were compared by smoothing the satellite profiles using the PEARL FTIR averaging kernels, and making profile comparisons at the altitudes of the PEARL FTIR retrieval grid between 6.4 and 12.0 km. The PEARL FTIR water vapour retrieval maintains sensitivity above 0.5 up to 12.0 km, and has 2.9 degrees-of-freedom-for-signal on average, resulting in the capability to resolve independent levels of the UTLS at around 3–4 km resolution.

L202: See comment for line 70.

Weaver et al. (2019) compared ACE-FTS and MAESTRO against GRUAN-processed radiosondes (Dirksen et al., 2014), which Miloshevich et al. (2009) showed compare well against frost-point hygrometers.

Miloshevich, L., Vömel, H., Whiteman, D., and Leblanc, T.: Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements. J. Geophys. Res., 114, D11305, https://doi.org/10.1029/2008JD011565, 2009.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463–4490, https://doi.org/10.5194/amt-7-4463-2014, 2014.

L256: "flagged with values between 4 and 7" has no tangible meaning for most readers

Changed on L177 and L256 both instances of specifying the filter flag. Changed:
From "These have been applied to the ACE-FTS data used in this work to filter extreme outliers from the data, with all profiles flagged with values between 4 and 7 being removed as recommended in Sheese et al. (2015)."
To "These are applied to the ACE-FTS data used in this work to filter extreme outliers from the data, with all flagged outlier measurements removed as recommended in Sheese et al. (2015)."

From "A set of quality control flags has been calculated for the MAESTRO v3.13 ozone and v31 water vapour products following the method of Sheese et al. (2015). These were applied to the MAESTRO data used in this work, and all profiles flagged with values between 4 and 7 were removed from the water vapour dataset as recommended in Sheese et al. (2015), while all profiles with nonzero flags were removed from the ozone dataset (personal communication with P. Sheese, 16 December 2019)"
To "Quality control flags have been calculated for the MAESTRO v3.13 ozone and v31 water vapour products following the method of Sheese et al. (2015), and in this work all measurements flagged as outliers were removed."

L260: "conflict"? "philosophy"? This dilemma is true of all data filtering, so there's no conflict or philosophy. The more stringent the quality control, the lower the data population available for computing meaningful statistics.

The terminology employed is consistent with Sheese et al. (2015), which outlines the quality flags used for ACE-FTS and MAESTRO. The line has been changed:
From "This is influenced by the conflict between the underlying philosophy of the quality flags, which prioritizes retaining all potentially valid measurements, and the extremely high variability of the MAESTRO ozone product."
To "This is influenced by the underlying philosophy of the quality flags (which prioritizes retaining all potentially valid measurements) coupled with the extremely high variability of the MAESTRO ozone product."

L426 and throughout: "deviation" can easily sway readers to think about variability, while "difference" cannot be misconstrued. "Mean difference" might also be called "bias" that is also very straightforward.

This choice of wording was based on the terminology of the relative metric, as a deviation from a reference not as a bias. However, changed from deviation to difference as appropriate throughout the manuscript.

L433: "most robust" is too strong. Direct comparison would be the "most robust" method because no model output is needed. Your method allows for a greater number of the satellite observations to be part of the comparison, but is not necessarily the “most robust”.

> Within this context this method includes the maximum possible amount of data for the construction and assessment of climatologies, hence it is an extremely robust approach. However, we have updated the text to reflect the existence of other comparably robust approaches. Changed:
> **From** “This choice permits the most robust comparison as it avoids the need to subsample the satellite datasets by limiting the data to only coincident pairs of measurements in order to account for differences in the instruments’ sampling patterns, and so the maximum number of comparisons can be made.”
> **To** “This choice permits an extremely robust comparison as it avoids the need to subsample the satellite datasets by limiting the data to only coincident pairs of measurements in order to account for differences in the instruments’ sampling patterns, and so the maximum number of comparisons can be made.”

L461: add “with altitude” after “VMR”. The first “effect” (dehydration) quite possibly accounts for >95% of setting the stratospheric entry mixing ratios for water vapor.

> We agree that the first effect accounts for the majority of stratospheric entry mixing ratios. Changed:
> **From** “The zonal-mean multiyear-mean climatologies from the measurement datasets (first and second columns in Fig. 1) show a roughly two order of magnitude decrease in water vapour VMR, indicating a strong vertical gradient between the upper troposphere and lower stratosphere.”
> **To** “The zonal-mean multiyear-mean climatologies from the measurement datasets (first and second columns in Fig. 1) show a roughly two order of magnitude decrease in water vapour VMR with altitude, indicating a strong vertical gradient between the upper troposphere and lower stratosphere.”

L469: Why compare in situ water vapor production in the stratosphere to mixing ratios in the troposphere? The water vapor added to the LS at higher latitudes by the transport of water from CH4 oxidation by the downward branches of the Brewer Dobson circulation can be as much 100% (3.5 -> 7 ppmv), so there is a significant influence on the UTLS distribution of water vapor at higher latitudes.

> As the Reviewer asserts, in situ water vapour production is relevant to the LS distribution of water vapour. In context, we are asserting that while this is the case, the amount produced by methane oxidation is much smaller than that observed in the UT, hence it appears less pronounced than the effects of the vertical gradient in water vapour as observed across the entire UTLS, which covers approximately two orders of magnitude.

L473: change “deposition of water vapor from” to “transport of water vapor by”

> Changed:
> **From** “The regions of elevated water vapour in the midlatitudes and polar regions result from deposition of water vapour from the Brewer-Dobson circulation, while the comparatively reduced concentrations centered around the equatorial tropopause result from dehydration of air through the cold-point tropopause and subsequent transport of constituents away from the tropical UTLS region.”
> **To** “The regions of elevated water vapour in the midlatitudes and polar regions result from transport of water by the Brewer-Dobson circulation, while the comparatively reduced concentrations centered around the equatorial tropopause result from dehydration of air through the cold-point tropopause and subsequent transport of constituents away from the tropical UTLS region.”

L481: what is “seasonal dehydration”? Dehydration occurs year-around. Might you be referring to a change in tropical cold point temperatures?

> Updated along with feedback from the other Reviewer. Changed:
> **From** “Seasonal enhancement of water vapour is evident between the two sets of seasonal climatologies, with the summer hemisphere (northern for JJA, southern for DJF), showing elevated water vapour concentrations compared to the winter hemisphere as a consequence of the seasonal dehydration in winter. Lower wintertime temperatures are associated with a decrease in saturation vapour pressure, resulting in a reduced carrying capacity for water vapour in the winter hemisphere. This effect is evident in the troposphere and the lowermost 20–30 K of the stratosphere, a region where the summer hemisphere displays a consistently greater water vapour concentration than the winter hemisphere.”
> **To** “Seasonal enhancement of water vapour, predominately in the troposphere but also over the lowermost

20–30 K of the stratosphere, is evident between the two sets of seasonal climatologies, with the summer hemisphere showing elevated water vapour concentrations compared to the winter hemisphere. This is influenced by the seasonal cycle in tropical cold point temperatures, and the decrease in saturation vapour pressure associated with lower temperatures in the winter that result in a reduced carrying capacity for water vapour in the winter hemisphere.”

L487: change “in regions such as the span above 50 K” to “in the 50-100 K region”

Changed:
From “The MAESTRO datasets show a larger relative standard deviation over much of the stratosphere than those from the two versions of ACE-FTS data, by over 50 % in regions such as the span above 50 K in the summer hemisphere.”
To “The MAESTRO datasets show a larger relative standard deviation over much of the stratosphere than those from the two versions of ACE-FTS data, by over 50 % in the 50–100 K region in the summer hemisphere.”

L492: Why does ACE-FTS v4.2 have “the most tightly confined region...”

This is a statement about what is observed in the relative standard deviation of the different climatologies. Speculating on the origin of this observation is beyond the scope of this work.

L495: Is the cause “variations in vertical transport”, or “variations in tropical cold point temperatures”?

Variations in tropical cold point temperatures play an important role, but there are other variations, such as those affecting the Brewer Dobson circulation, that affect the transport of water vapour into the LS. As a result, we chose the broader description “variations in vertical transport.”

L505: change to “large relative differences”

Changed:
From “While the CMAM39-SD subsampled climatologies (Fig. 2 first and second columns) generally show patterns of water vapour consistent with the measurement climatologies, there are differences found in all of the subsampled climatologies (Fig. 2 third and fourth columns).”
To “The CMAM39-SD subsampled climatologies (Fig. 2, first and second columns) generally show patterns of water vapour consistent with the measurement climatologies, but there are large relative differences between the measurement and model-subsampled climatologies (Fig. 2 third and fourth columns).”

L511: It would be worth adding one sentence here explaining why you’ve chosen to not elaborate on the model-satellite biases in the paper.

As stated on L515, “the main use of CMAM39-SD is to compare the measurement datasets in a consistent fashion by accounting for sampling differences.” Hence we feel this point has been addressed. Additionally, per the scope of the paper, as described in the abstract and elsewhere, the focus of this paper is not on model validation, but toward the development and analysis of UTLS-focused climatologies and the utilization of a technique to use a model as a consistent comparison reference to evaluate climatologies.

L545: change to “ranges poleward of approximately 45° in each hemisphere”

Changed:
From “The remainder of the regions in which CMAM39-SD displays greater water vapour VMRs than the measurement datasets can be subdivided into the spans north and south of approximately 45° from 0–70 K above the tropopause.”
To “The other regions in which CMAM39-SD has higher water vapour than the measurement datasets are the ranges poleward of approximately 45° in each hemisphere and from 0–70 K above the tropopause.”

L600: If not stated before, include the fact that ozone is mainly produced in the tropical middle and upper stratosphere, the main reason there’s a positive gradient in the stratosphere.

This is introduced more explicitly in the introduction on L56, in addition to updating this line. Changed:
L56:
From “Water vapour is primarily concentrated near the surface, with a strong negative gradient extending upwards to the tropopause following the lapse rate, while ozone is mainly concentrated in the stratosphere, with a gradient extending poleward from its maximum concentration in the tropical middle stratosphere.”
To “Water vapour is primarily concentrated near the surface, with a strong negative gradient extending upwards to the tropopause following the lapse rate, while ozone is mainly concentrated in the stratosphere, with a gradient extending poleward from its maximum concentration in the tropical middle stratosphere where the majority of ozone production occurs.”

L604:
From “The tropical ozone peak is associated with the relative proximity of the tropical tropopause to the ozone concentration maximum in the middle stratosphere.”
To “The tropical ozone peak is associated with the relative proximity of the tropical tropopause to the ozone concentration maximum in the tropical middle and upper stratosphere.”

L670: change “does” to “do”

Changed:
From “As with the water vapour comparisons, these differences between the model simulation and the measurements does not limit the use of the model to evaluate the consistency of the measurement datasets.”
To “As with the water vapour comparisons, these differences between the model simulation and measurements do not limit the use of the model to evaluate the consistency of the measurement datasets.”

L674: change to “ACE-FTS and MAESTRO measurements”

Changed:
From “In evaluating the differences between the CMAM39-SD subsampled ozone climatologies, the average magnitude of the difference between the model climatologies generated using the times and locations of the ACE-FTS v4.2, ACE-FTS v3.6, and MAESTRO measurements is 1–2 %.”
To “In evaluating the differences between the CMAM39-SD subsampled ozone climatologies, the average magnitude of the difference between the model climatologies generated using the times and locations of the ACE-FTS v4.2, ACE-FTS v3.6, and MAESTRO measurements is 1–2 %.”

L680: consolidate sentences: “Greater differences, 9% on average and as large as 76% in the UT, are found in the subsampled model results and OSIRIS measurements.”

Changed:
From “Greater differences are noted when comparing the subsampled model climatologies made using the OSIRIS measurement pattern to the others. The set of subsampled climatologies generated from the OSIRIS measurement pattern is found to differ by 9 % on average from the other three instrument-subsampled climatologies, but differences as large as 76 % are observed in a subset of the grid cells, mostly confined to the upper troposphere.”
To “The model subsampled climatologies constructed using the OSIRIS sampling pattern were found to differ more greatly from those constructed using the sampling patterns of the other instruments, with differences on average of 9 % but as large as 76 % in the upper troposphere.”

L727-736: This is another overload of numbers that would greatly benefit from a table.

As per the suggestion of both Reviewers, we have summarized the model-measurement comparisons in two Tables, and have changed the text in Sect. 4.2 and 5.2 accordingly, including the summary paragraphs. Please refer to the comment from Reviewer 1 for the exact text change.

L756: change to “some sources of variability”. Change “distribution” to “distributions”.

Changed:

> **From** “These climatologies employ tropopause-relative potential temperature and equivalent latitude coordinates in an effort to best represent the distribution of these two gases in the UTLS by accounting for sources of variability (Pan et al., 2004; Hoor et al., 2004; Hegglin et al., 2008). The distribution of these two gases, and their variability, have been examined for two seasons, DJF and JJA.”
> **To** “These climatologies use tropopause-relative potential temperature and equivalent latitude coordinates in an effort to best represent the distribution of these two gases in the UTLS by accounting for some sources of variability (Pan et al., 2004; Hoor et al., 2004; Hegglin et al., 2008). The distributions of these two gases, and their variability, have been examined for two seasons, DJF and JJA.”

L758: “reference” is too strong, implying that the model output is some sort of “gold standard”, which it is not. See the definition of a “climate reference network”. I recommend omitting “reference” here and throughout the paper for this reason.

> The use of the term “reference” is in regard to the technique employed for comparing climatologies, not as a general reference. To reinforce this, the following lines have been changed:
> **L127**:
> **From** “To characterize the differences between the datasets, the CMAM39-SD model output dataset is employed to serve as a consistent comparison reference.”
> **To** “To characterize the differences between the datasets, the CMAM39-SD model output dataset is used as a consistent comparison reference for this analysis, though it should be noted that this does not imply that the model output serves as a standard outside of this.”
>
> **L760**:
> **From** “These subsampled model climatologies provide a consistent reference to assess the agreement of the instrumental datasets, as well as to explore the influence of different measurement patterns on the datasets.”
> **To** “These subsampled model climatologies provide a consistent comparison reference to assess the agreement of the instrumental datasets, as well as to explore the influence of different measurement patterns on the datasets, but should not be considered to be a standard beyond this use.”

Other Changes

The following are changes made solely based on grammar, word choice, and correcting unit representation. Changed:

L33:
From “In recent years, there has been a push to expand the database of climatologies focused on one of the most challenging regions of the atmosphere to study, the upper troposphere - lower stratosphere (UTLS), in order to better constrain the role the UTLS plays in climate processes through detailed characterization of the constituents of the region.”
To “In recent years, there has been a push to expand the database of climatologies focused on one of the most challenging regions of the atmosphere to study, the upper troposphere - lower stratosphere (UTLS), in order to better constrain the role the UTLS plays in climate processes through the detailed characterization of the constituents of the region.”

L79:
From “The need for well characterized trace gas climatologies that can serve toward the parallel goals of analyzing gas distributions and trends, and of validating model datasets, is emphasized in the reports produced as a part of the Stratosphere-troposphere Processes And their Role in Climate (SPARC) Chemistry-Climate Model Validation (SPARC-CCMVal, 2010) and the SPARC Data Initiative (SPARC-DI, 2017) projects.”
To “The need for well characterized trace gas climatologies that can serve toward the parallel goals of analyzing gas distributions and trends, and of evaluating model datasets, is emphasized in the reports produced as a part of the Stratosphere-troposphere Processes And their Role in Climate (SPARC) Chemistry-Climate Model Validation (SPARC-CCMVal, 2010) and and the Data Initiative (SPARC-DI, 2017) projects.”

L87:
From “To fully characterize the differences between instruments in the UTLS, this report stressed the need for climatologies to be constructed in ways to best minimize the effects of geophysical variability on the diagnostics compared.”
To “To fully characterize the differences between instruments in the UTLS, this report stressed the need for climatologies to be constructed in ways that best minimize the effects of geophysical variability on the diagnostics compared.”

L116:
From “Specifically, tropopause-referenced potential temperature is employed because of its ability to account for the role of the tropopause as a transport barrier, which enables it to represent the large trace gas gradients that arise around the tropopause leading to improved partitioning of data.”
To “Specifically, tropopause-referenced potential temperature is employed because of its ability to account for the role of the tropopause as a transport barrier, which enables it to represent the large trace gas gradients that arise around the tropopause, leading to improved partitioning of data.”

L124:
From “The climatologies presented here cover the 14-year period between 1 June 2004 and 31 May 2018, and are intended to both provide a well characterized representation of water vapour and ozone as measured by the three instruments involved in this study, and to provide a source of data for future work including intercomparison studies and model validation.”
To “The climatologies presented here cover the 14-year period between 1 June 2004 and 31 May 2018, and are intended to both provide a well characterized representation of water vapour and ozone, as measured by the three instruments involved in this study, and to provide a source of data for future work including intercomparison studies and model evaluation.”

L144:
From “The Canadian satellite SCISAT-1 was launched into a circular low-Earth orbit (650 km altitude, 74° inclination) on 12 August 2003.”
To “The Canadian satellite, SCISAT-1, was launched into a circular low-Earth orbit (650 km altitude, 74° inclination) on 12 August 2003.”

L228:
From "The second instrument aboard SCISAT-1 is MAESTRO, a dual spectrograph designed to measure between 285 and 1030 Nm, with the two grating spectrophotometers recording spectra with a wavelength dependent resolution of 1–2 Nm."
To "The second instrument aboard SCISAT-1 is MAESTRO, a dual spectrograph designed to measure between 285 and 1030 nm, with the two grating spectrophotometers recording spectra with a wavelength dependent resolution of 1–2 nm."

L233:
From "Over time, MAESTRO has been affected by the gradual buildup of an unknown contaminant, and since 2015 no light with wavelength shorter than 500 nm is transmitted through the instrument; however the water vapour and ozone retrievals remain operational."
To "Over time, MAESTRO has been affected by the gradual buildup of an unknown contaminant, and since 2015 no light with wavelength shorter than 500 nm is transmitted through the instrument; however, the water vapour and ozone retrievals remain operational."

L288:
From "It consists of a grating optical spectrograph (OS) that records Rayleigh- and Mie-scattered sunlight spectra from 280–810 Nm with 1–2 Nm resolution, and an infrared imager (IRI) measuring airglow."
To "It consists of a grating optical spectrograph (OS) that records Rayleigh- and Mie-scattered sunlight spectra from 280–810 nm with 1–2 nm resolution, and an infrared imager (IRI) measuring airglow."

L348:
From "The free running extension of the Canadian Centre for Climate Modeling and Analysis (CCCma) third-generation atmospheric general circulation model (AGCM3) is the Canadian Middle Atmosphere Model (CMAM), which uses AGCM3 as the underlying basis for its middle atmosphere dynamical and chemistry-climate modelling components."
To "The free-running extension of the Canadian Centre for Climate Modeling and Analysis (CCCma) third-generation atmospheric general circulation model (AGCM3) is the Canadian Middle Atmosphere Model (CMAM), which uses AGCM3 as the underlying basis for its middle atmosphere dynamical and chemistry-climate modelling components."

L354:
From "Free running models are incapable of reproducing day-to-day variations in meteorology, which hinders direct comparisons between observed and simulated trace gas concentrations, because of the inherently chaotic nature of atmospheric circulation."
To "Free-running models are incapable of reproducing day-to-day variations in meteorology, which hinders direct comparisons between observed and simulated trace gas concentrations, because of the inherently chaotic nature of atmospheric circulation."

L367:
From "Fields from CMAM39-SD are sampled every 6 hours, starting at midnight, and are provided on their native hybrid-sigma pressure coordinate grid."
To "Fields from CMAM39-SD are sampled every 6 hours, starting at midnight UTC, and are provided on their native hybrid-sigma pressure coordinate grid."

L464:
From "The second influence is the somewhat permeable barrier represented by the tropopause, which decouples moisture-rich tropospheric air from the overlying regions outside of sporadic synoptic wave-driven mixing processes and limited vertical transport near the equator associated with the Brewer-Dobson circulation."
To "The second influence is the somewhat permeable barrier represented by the tropopause, which decouples moisture-rich tropospheric air from the overlying regions outside of sporadic synoptic wave-driven mixing processes and vertical transport near the equator associated with the Brewer-Dobson circulation"

L492:
From “This dataset also has the most tightly confined region of elevated variability near the the tropical tropopause in both seasons.”
To “This dataset also has the most tightly confined region of elevated variability near the tropical tropopause in both seasons.”

Fig. 1 caption:
From “Gaps in the data occur when there are either no observations made or when fewer than five observations are available for a given bin.”
To “Gaps in the data occur when there are either no observations made or where there are fewer than five observations available for a given bin.”

L594:
From “As with water vapour above, results shown are based on seasonal (3-month) climatologies for DJF and JJA.”
To “The results shown are based on seasonal (3-month) climatologies for DJF and JJA.”