# Response to Reviewers – “Validation of the version 4.5 MAESTRO ozone and NO2 measurements” 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. Due to some suggestions and requested recalculations from both Reviewers, the plots and tables included in the manuscript, along with some of the text from the Abstract, Sect. 4, and the Conclusion, have been updated. For clarity, the changed figures, tables, and the text relevant to this are provided in their own separate section after both sets of Reviewer comments and our responses.

**Reviewer 1: Robert Damadeo**

The authors present the results of the newest version (v4.5) of the MAESTRO O3 and NO2 data. Simple coincident event comparisons are made with 10 other platforms as well as the ACE-FTS instrument operating on the same spacecraft. The paper is well-written, the methodology of the comparisons is very straightforward, and the results are described in detail. This paper’s subject matter is well suited for this journal. My only concern is that the conclusions about some of the comparisons are likely misrepresenting the actual data quality because of the different sampling patterns of the instruments. I recommend this paper for publication after the following concerns are addressed.

> Thank you for the feedback Dr. Damadeo. Please see below for our responses to your comments.

One simple omission, unless I missed it, is what is the end of the date range of data for instruments that are still operating used for this study.

> We have specified the start and end date for all of the measurement datasets used in this study. We also corrected a typo for the start date of GOMOS operations. These are as follows.
>
> Added on L90:
> “The version 4.5 dataset used covers the period from February 2004 to December 2023.”
>
> Added to L117:
> “The ACE-FTS data used in this study cover the period from February 2004 to December 2023.”
>
> Added to L154:
> “The version 3.0 Odin-SMR data used in this study cover the period from February 2004 to September 2022.”
>
> Added to L177:
> “The OSIRIS data used in this study cover the period from February 2004 to December 2023. ”
>
> Changed L205:
> **From:** “Scientific operation of GOMOS began in April 2004 and ended in April 2012.”
> **To:** “Scientific operation of GOMOS began in March 2002 and ended in April 2012.”
>
> Added to L215:
> “The GOMOS data used in this study cover the period from February 2004 to April 2012.”
>
> Added to L251:
> “The MIPAS data used in this study cover the period from November 2004 to April 2012.”
>
> Added to L278:
> “The SCIAMACHY data used in this study cover the period from February 2004 to April 2012.”
>
> Added to L312 for OMPS-LP:
> “This dataset covers the period from February 2012 to December 2023.”
>
> Added to L347:
> “The Aura-MLS data used in this study cover the period from August 2004 to December 2023.”

> Added to L382:
> “The SAGE II data used in this study cover the period from August 2004 to August 2005.”
>
> Changed L425:
> **From:** “In this study, the version 4 SAGE III/M3M products are used, which extend to the end of December 2005.”
> **To:** “The version 4 SAGE III/M3M data products are used in this study for the period from February 2004 to December 2005.”
>
> Added to L460:
> “The SAGE III/ISS data used in this study cover the period from July 2017 to December 2023.”

Another clarification is regarding the event type. When the authors talk about MAESTRO SRs and SSs, does this refer to the spacecraft event type or the local event type? Obviously, given the nature of the analysis, it makes sense for the separation to be on the local event type and not the spacecraft event type. As far as I am aware, neither the MAESTRO nor the ACE-FTS data products specifically inform the user of the local event type, leaving it as a required calculation by the user. However, the orbit of SCISAT is somewhat unique and the local and spacecraft event types are very often not the same.

> The separation by sunset (SS) and sunrise (SR) is intended to be by local event type, not spacecraft event type. We have reviewed the determination of local event type and have corrected an issue with calculation of this. This change has updated some small number of comparisons within each dataset. Along with other comments from Reviewer 1 and 2, this has resulted in updated versions of the manuscript’s plots and tables, which are included in the last section of this response. Additionally, the recalculated comparisons have led to some small differences in the text of the Abstract, Sect. 4, and the Conclusion of the manuscript, which are also included in the last section of this response.

Lastly, the authors state that they separate coincidences based on the MAESTRO event type (SRs and SSs). However, when making coincident event comparisons with other solar occultation instruments, are the coincidences ensuring the same separation in the coincidences? If not, then there is likely too often a mixing of different kinds of airmasses (e.g., comparing a MAESTRO SR to a SAGE SS) and this analysis would need to be redone accordingly. My remaining comments are predicated upon the assumption that the stated comparisons show MAESTRO SRs/SSs compared to other solar occultation SRs/SSs respectively.

> As suggested by Reviewer 2, we have now applied diurnal scaling for ozone as well as $NO_2$, thereby allowing us to better compare airmasses sampled with different measurement techniques and at different locations/times of day. This treatment has been applied for all datasets, except for ACE-FTS, for consistency in the results. Please see below for the changes to the Figures, Tables, and text.
>
> Changed L517:
> **From:** “The scaling factors, which are functions of altitude, latitude, and solar zenith angle, allow for the scaling of $NO_2$ concentrations to sunrise and/or sunset and have been applied to the non-solar-occultation measurement datasets ahead of comparisons.”
> **To:** “The resulting scaling factors are functions of altitude, latitude, and solar zenith angle and allow for the scaling of $NO_2$ concentrations to local sunrise and/or sunset. These scaling factors have been applied to scale all coincident measurements from the non-solar-occultation instruments used in this study. While it is possible to compare the three SAGE instruments to MAESTRO without the use of diurnal scaling, so long as local sunrise measurements are compared to local sunrise measurements and local sunset to local sunset, this limits the number of potential coincidences that can be examined due to differences in the orbits of these instruments and the short overlap period between MAESTRO and that of SAGE II and SAGE III/M3M. To maximize the number of comparisons with the SAGE instruments, rather than force sunrise-sunrise and sunset-sunset comparisons, the diurnal scaling factors from Strode et al., (2022) have been employed.”
>
> Changed L520:
> **From:** “Finally, a known issue with solar occultation instruments is that there is a difference in observed ozone values between sunrise and sunset measurements (e.g., Sakazaki et al., 2015; Wang et al., 2020). This effect leads sunset measurements of ozone to have larger VMRs than measurements made during

sunrise, and is attributed to the effects of vertical transport of atmospheric tidal winds (Sakazaki et al., 2013, 2015). This difference between the sunrise and sunset measurement values, and the resulting bias between the two, has been noted in previous MAESTRO validation efforts (Kar et al., 2007). To minimize the effects of this difference between the two types of measurements, the MAESTRO sunset and sunrise measurements are treated independently for the calculation of the above metrics in this study. ”

**To:** “Ozone has also been shown to experience a diurnal cycle (e.g., Prather et al., 1986). During the day, molecular oxygen is photolyzed to produce odd oxygen ($O_x = O + O_3$ ) species which then undergo subsequent reactions. Due to the influence of pressure on these reactions, odd oxygen is preferentially converted into ozone in the stratosphere during the day; however, at higher altitudes, more odd oxygen is stored as atomic oxygen during the day. Thus, the concentration of stratospheric ozone peaks in the afternoon, and that of the mesosphere peaks in the night when all atomic oxygen recombines. This diurnal cycle is largest in the upper stratosphere and mesosphere, but still exceeds 2 % in the middle stratosphere (Prather et al., 1986; Sakazaki et al., 2013). Combined with the effects of vertical transport by atmospheric tidal winds, this leads to a distinct difference in observed ozone values between sunrise and sunset measurements for solar occultation instruments (e.g., Sakazaki et al., 2013, 2015; Wang et al., 2020). This difference between the sunrise and sunset measurement values, and the resulting bias between the two, has been noted in previous MAESTRO validation efforts (Kar et al., 2007). To minimize the effects of this difference between the two types of measurements, the MAESTRO sunset and sunrise measurements are treated independently for the calculation of the above metrics in this study. Additionally, diurnal scaling factors for ozone from Strode et al., (2022) have been applied at all altitudes to all comparison datasets, except for ACE-FTS, as done for $NO_2$.”

Changed L535:
**From:** “Comparisons between MAESTRO sunrise and sunset Vis.-ozone data are shown in Figs. 1 through 4.”
**To:** “Comparisons between MAESTRO sunrise and sunset Vis.-ozone data against diurnally-scaled (where required) coincident measurements are shown in Figs. 1 through 4.”

Changed L666:
**From:** “The comparisons between the MAESTRO sunrise and sunset UV-ozone data and the coincident ozone measurements are shown in Figs. 5 through 8.”
**To:** “The comparisons between the MAESTRO sunrise and sunset UV-ozone data and the diurnally-scaled (where required) coincident ozone measurements are shown in Figs. 5 through 8.”

Changed L772:
**From:** “Comparisons between the MAESTRO sunrise and sunset $NO_2$ data against the diurnally-scaled (where required; using the diurnal scaling factors of Strode et al. (2022)) coincident measurements from the other datasets are shown in Figs. 10 through 13.”
**To:** “Comparisons between the MAESTRO sunrise and sunset $NO_2$ data against the diurnally-scaled (where required) coincident measurements from the other datasets are shown in Figs. 10 through 13.”

Because MAESTRO is a solar occultation instrument with sparse sampling associated with that technique and because the analyses performed here are based on coincidences between instruments, I believe that many of the comparisons are adversely affected by sampling biases created when analyzing these coincidences. This would be evident in the comparisons with other instruments that have their own sampling biases, creating comparisons that are not representative of the atmosphere as a whole and/or creating systematic differences in sampling locations. I have looked at this specifically for the SAGE instruments, but less noticeable sampling biases are also possible for GOMOS (stellar occultation) or Odin instruments (i.e., SMR and OSIRIS) that I recall having a hemispheric asymmetry in the overall sampling.

One indication of potential sampling biases is easily seen when looking at the mean SR/SS comparisons. For O3 throughout the lower and middle stratosphere, the impact of diurnal variability is minimal. If the events are generally evenly sampled in time and latitude, then the expectation is that the mean SR and mean SS profiles would overlap, as they do in the ACE-FTS comparison (naturally since every event is coincident) as well as with very dense samplers such as MLS, MIPAS, SCIAMACHY, and OMPS-LP. The fact that the mean SR and mean SS profiles just from the MAESTRO instrument begin to deviate from each other in the other comparisons is the first sign of potential sampling biases. The same is true for NO2 comparisons where, although diurnal variability is expected to be noticeable throughout the stratosphere, the scale of the diurnal variability just between the MAESTRO SR/SS profiles changes between different instrument comparisons.

For SAGE II, sampling biases created in coincident event comparisons are the most egregious. This is because not only is SAGE also a solar occultation instrument, but SAGE II was operating at a 50% duty cycle during the time of operational overlap with MAESTRO. I looked into the temporal and spatial distribution of coincident events (<8 hours, <1000 km) between SAGE II and ACE-FTS (which is data I had on-hand), assuming there would be almost identical sampling between ACE-FTS and MAESTRO, and found the following:

For SSs, all comparisons are basically confined to two small groupings: high southern latitudes (60-70) in late 2004 and high northern latitudes (60-70) in early 2005, noting that all of the southern latitude comparisons have SAGE II observations taking place at a systematically more northern latitude than MAESTRO observations. This systematic offset in latitude could create overall biases in the comparisons. In both cases, this means observations were taking place at a time and place of higher vortex variability, which would likely result in different standard deviations and reduced correlations.

For SRs, all comparisons are basically confined to another two groupings: high southern latitudes (50-70) in late 2004 and another semi-global patch (50N-40S) in early-to-mid 2005, noting that almost ALL of these comparisons have SAGE II observations taking place at a systematically more northern latitude than MAESTRO observations, in some cases exceeding a difference in latitude of 5 degrees. This systematic bias in spatial sampling likely contributes to the systematic bias in O3 seen in comparisons with SAGE II SRs.

For SAGE III/M3M, sampling biases created in coincident event comparisons are also somewhat unique because of the combination of a sunsync orbit with a solar occultation instrument. The effect of this is that all of SAGE III/M3M spacecraft SSs/SRs are observed in the northern/southern hemisphere. However, there is a distinction between spacecraft event type and local event type. For SAGE II, which was in a mid-inclination orbit, the two are almost always the same. For SAGE III/M3M, all of the spacecraft SRs are actually local SSs, and most of the spacecraft SSs are local SSs with the exception of polar winter where they are local SRs. This means that the distribution of coincidences with SAGE III/M3M SSs do not have much of a sampling bias, but a significant one exists for SRs. All coincidences between SAGE III/M3M and MAESTRO SRs occur within a small grouping at high northern latitudes (55-75) in early 2005, with all SAGE III/M3M observations taking place at a systematically more northern latitude than MAESTRO with a minimum offset of 5 degrees in latitude. Additionally, I compute a ratio of coincident SS events to SR events of nearly 10:1 (commensurate with the total number of local SSs versus SRs in the SAGE III/M3M dataset), which is very different from the 3:1 ratio the authors show. This makes me wonder if the authors really are not considering the different event types for comparison solar occultation instruments as I can get a similar number of coincidences if I ignore the SAGE III/M3M event type. If so, then this whole analysis really does need to be redone (at least for all SAGE instruments).

> In combination with suggestions from Reviewer 2, diurnal scaling has been applied to all coincident ozone and $NO_2$ comparisons, with exception for ACE-FTS since it measures at the same local time, to scale the data to the local time of the MAESTRO measurements. This allows for a potentially larger number of measurements to be made than by only comparing sunrise to sunrise measurements and sunset to sunset measurements, and ensures the comparison data is treated similarly. Thus, the ratio of comparisons will differ from what Reviewer 1 suggests.

For SAGE III/ISS, there are again sampling biases from the combined orbital sampling of ISS and SCISAT. Strangely, all of the SR comparisons are in the northern hemisphere, while most of the SS comparisons are in the southern hemisphere, but the latitudinal extent gets broader as the years go by and start to expand into the southern hemisphere. While this could potentially be problematic if looking into drifts between the instruments, I don't see any obvious source of bias in coincident event comparisons.

> The authors appreciate the detailed discussion/insight provided by Dr. Damadeo of the impact of sampling bias on the comparisons performed in this study. Without wishing to remove the instruments with sparser sampling from this study, we have instead opted to incorporate this feedback in a discussion on the potential impact of sampling bias on comparisons in the Methodology section of the manuscript, and have referred back to the potential for sampling biases to impact the comparisons throughout the Results section.

Added on L488:
“As a solar occultation instrument, MAESTRO has relatively sparse spatial and temporal sampling so when employing coincidence criteria for comparisons, as done here, the potential exists for sampling biases to impact the results. This is likely to occur when comparison instruments also provide sparse or seasonally-varying coverage, with the biases resulting from comparisons that do not wholly capture the state of the atmosphere or which result in systematic differences in sampling locations. In this study, a number of instruments with both sparse and dense sampling are employed. The latter, which includes Odin-SMR, SCIAMACHY, MIPAS, Aura-MLS, and OMPS-LP yield comparisons with minimal potential for sampling biases to impact the results of the analysis as the measurement comparisons are generally evenly distributed across space and time. ACE-FTS, while itself also a solar occultation instrument with the sparse sampling that entails, shares a line-of-sight with MAESTRO and so every measurement made by MAESTRO is coincident with one from ACE-FTS, avoiding any systematic differences in measurement locations.

In contrast, OSIRIS is a densely sampling instrument that possesses a seasonal asymmetry in its coverage, generally only covering one hemisphere at a time, while the remaining four instruments employed in this study, GOMOS and the three SAGE instruments, all provide sparse sampling. The sparse sampling of these last four instruments is due to the limitations, addressed above, of the solar occultation technique employed by the SAGE instruments, and the limited number of viable stellar occultation measurements made by the former. Thus, for OSIRIS, GOMOS, and the SAGE instruments there exists the possibility that any comparisons made with them will be affected by sampling biases. This is particularly true for SAGE II since throughout the overlap period of SAGE II with MAESTRO, SAGE II was only operating at a 50 % duty cycle, which, when combined with the orbits of ERBS and SCISAT, causes all coincident measurements to be largely confined to a few narrow groupings, often near the edges of the polar vortex where variability is high.

Despite the potential for sampling biases, this study includes these sparse sampling/seasonally asymmetric datasets for the assessment of the MAESTRO version 4.5 products to allow for an overview of the MAESTRO data in comparison to a diverse suite of measurements made using multiple techniques, with the caveat that some of these comparisons might be affected by sampling biases and should be considered as part of an ensemble of comparisons rather than independently.”

Added on L578:
“Notable across the six sets of comparisons discussed so far is that the comparison datasets are from those least likely to be affected by sampling biases, due to the density of their sampling or their shared line-of-sight with MAESTRO, reinforcing the good agreement found with the MAESTRO Vis.-ozone product.”

Added on L579:
“In the lower and middle stratosphere it is expected that the sunrise and sunset profiles should generally agree with each other due to the small diurnal cycle of ozone at these altitudes. Thus the observed difference between the sunset and sunrise profiles is likely influenced by some form of sampling bias associated with the sparse coverage and few coincidences found between MAESTRO and SAGE III/M3M.”

Added on L587:
“As with the SAGE III/M3M comparisons, this indicates the potential for sampling bias to play a role in the OSIRIS comparisons; however, given the greater degree of agreement between the sunrise and sunset profiles observed here as compared to those for the SAGE III/M3M comparisons, it is likely that it is a more limited effect.”

Changed L592:
**From:** “The remaining datasets all show larger differences from MAESTRO, as well as generally larger differences between their sunset and sunrise profiles.”
**To:** “The remaining datasets all show larger differences from MAESTRO, as well as generally larger differences between their sunset and sunrise profiles that potentially arise from sampling biases. ”

Added to L737:
“Due to the sparse sampling of the first two datasets, there is a strong likelihood that those comparisons

are influenced by sampling bias, contributing to the poor correlation observed."

Added to L792:
"Unlike with the ozone profiles, a more pronounced difference in the sunrise and sunset profiles due to the diurnal cycle of $NO_2$ is expected; however, the scale of these differences between the different sets of comparisons show the potential for sampling biases to impact these comparisons in the less densely sampled, non-ACE-FTS datasets."

Changed L900:
**From:** "Likely influenced by the few coincident profiles, the SAGE II average sunrise correlation is only 0.49 for this range, and the sunset comparisons are found to be slightly uncorrelated, with an average correlation coefficient of -0.03."
**To:** "Likely influenced by the few coincident profiles, as well as a systematic difference in sampling location, the SAGE II average sunrise correlation is only 0.49 for this range, and the sunset comparisons are found to be uncorrelated, with an average correlation coefficient of -0.03."

**Reviewer 2: Anonymous**

This paper presents a validation of version 4.5 O3 and NO2 from the MAESTRO instrument. Both UV and visible O3 products are considered. The MAESTRO observations are compared to observations from many other satellite limb instruments. In general, MAESTRO visible O3 agrees well with the other datasets between 20 and 50 km, while MAESTRO UV O3 has good agreement from 15-40 km. MAESTRO NO2 is biased low.
A thourough validation of the MAESTRO observations is important for anyone who wishes to use the data. The paper is well written and suitable for publication after some minor additions.

> Thank you for the feedback. Please see below for our responses to your comments.

Abstract: add a sentence discussing what has changed since the previous MAESTRO data version.

> We have included a brief summary of the major changes made to the MAESTRO retrieval algorithm.
>
> Changed L5:
> **From:** "The latest ozone and $NO_2$ profile products, version 4.5, have been released, which nominally cover the period from February 2004 to December 2023. Due to the buildup of an unknown contaminant, the UV-ozone and NO2 products are only viable up to June 2009 for NO2 and December 2009 for UV-ozone."
> **To:** "The latest ozone and $NO_2$ profile products, version 4.5, have been released, which initially cover the period from February 2004 to December 2023 and which will continue to be updated. The version 4.5 retrieval algorithm represents an improvement from previous versions, with changes including updated pressure and temperature input information, an improved algorithm for high-sun reference spectrum calculation, improved Rayleigh scattering modeling, and the change to a Twomey-Tikonov inversion algorithm from a Chahine relaxation technique. Due to the buildup of an unknown contaminant, the UV-ozone and NO2 products are only viable up to June 2009 for NO2 and December 2009 for UV-ozone."

Section 2.1.1: How were the values used for data filtering determined? Why filter using both vmr thresholds and a MAD filter?

> The philosophy behind the applied data filtering was to first remove obviously unrealistic data and then apply a more constrained filter on the remaining dataset, with a preference to keep some data that is potentially physically unrealistic rather than remove valid measurements. The most stringent criteria is applied between 5 and 50 km, where the MAESTRO retrieval is well constrained, as that is the region with the highest data quality. The threshold values used here were empirically chosen to remove profiles that did not satisfactorily converge to a realistic profile. Outside of this range, less stringent VMR thresholds are used to remove only highly unphysical values. Rather than remove whole profiles, this level of the filtering is applied independently for each altitude level. Finally, the MAD filter is applied to each altitude level, after the threshold filtering so as to prevent the minute possibility of any impacts of erroneous data on this filter, to provide a degree of sensible data filtration for the remainder of the profile.

Line 179/180: The v7.2 OSIRIS retrieval does not include the MAD filter, or any manual inspection.

> We have corrected this.
>
> Changed L178:
> **From:** "The involves screening the limb radiance measurements for clouds or cosmic rays, screening the trace gas profiles using a five standard deviation and ten MAD filter, and inspecting the profiles to flag and remove outliers (Adams et al., 2013)."
> **To:** "The involves screening the limb radiance measurements for clouds or cosmic rays (Bognar et al., 2022)."

Section 2.6.1: The SAGE II NO2 from sunrise occultations is affected by a thermal shock and so these data are considered a "research product" (Damadeo et al. 2013). This is probably worth mentioning since you are using the data. This could also be a reason that SAGE II sunrise NO2 is the only dataset with less NO2 than MAESTRO in figure 10.

We have made mention of the impact of the thermal shock on the SAGE II $NO_2$ dataset.

Added to L795:
"The differing behavior observed with the SAGE II sunrise $NO_2$ dataset can be at least partially attributed to the thermal shock the instrument experiences during measurement events, which can be readily accounted for in sunset measurements but requires a correction to be applied for the sunrise measurements (Damadeo et al., 2013)."

Changed L798:
**From:** "Both of the SAGE II profiles, as well as the SAGE III/M3M sunset profile, show their largest standard deviation values between 40 and 50 km; however, this is likely associated with retrieval-boundary uncertainty effects"
**To:** "Both of the SAGE II profiles, as well as the SAGE III/M3M sunset profile, show their largest standard deviation values between 40 and 50 km; however, this is likely associated with retrieval-boundary uncertainty effects and the aforementioned thermal shock for the former [SAGE II]."

Changed L854:
**From:** "Thus, the comparisons with SAGE II should still be treated cautiously."
**To:** "This, in addition to the thermal shock effect discussed above, indicates that the comparisons with SAGE II should still be treated cautiously due to the potential impact of sampling biases (Damadeo et al., 2013). "

Line 481: Why look all the way down to 0 km? I don't think these datasets are expected to be reliable in the troposphere.

The vertical axes of the plots have been altered to span from 5 to 80 km now. Please see the last section here for the revised plots. We have updated the text to reflect this.

Changed:
**From:** "As the MAESTRO version 4.5 retrievals are only weakly constrained above 80 km, this study focuses between 0 and 80 km where most of the retrieved profile information is located."
**To:** "As the MAESTRO version 4.5 retrievals are only weakly constrained above 80 km, this study focuses between 5 and 80 km where most of the retrieved profile information is located."

Section 3: Why didn't you consider diurnal variations in O3? This becomes relevant in the upper stratosphere, above 40 km (e.g. Strode et al. 2022). MAESTRO O3 might agree better with the non-solar occultation datasets at these higher altitudes if you include diurnal scaling. It would be good to apply the Strode et al. (2022) scaling factors to one of your O3 comparison datasets to see if this makes a difference.

As noted above in the response to Reviewer 1's comments, we have now included diurnal scaling factors from Strode et al., (2022) for ozone in the comparisons.

Changed L520:
**From:** "Finally, a known issue with solar occultation instruments is that there is a difference in observed ozone values between sunrise and sunset measurements (e.g., Sakazaki et al., 2015; Wang et al., 2020). This effect leads sunset measurements of ozone to have larger VMRs than measurements made during sunrise, and is attributed to the effects of vertical transport of atmospheric tidal winds (Sakazaki et al., 2013, 2015). This difference between the sunrise and sunset measurement values, and the resulting bias between the two, has been noted in previous MAESTRO validation efforts (Kar et al., 2007). To minimize the effects of this difference between the two types of measurements, the MAESTRO sunset and sunrise measurements are treated independently for the calculation of the above metrics in this study. "
**To:** "Ozone has also been shown to experience a diurnal cycle (e.g., Prather et al., 1986). During the day, molecular oxygen is photolyzed to produce odd oxygen ($O_x = O + O_3$ ) species which then undergo subsequent reactions. Due to the influence of pressure on these reactions, odd oxygen is preferentially converted into ozone in the stratosphere during the day; however, at higher altitudes, more odd oxygen is stored as atomic oxygen during the day. Thus, the concentration of stratospheric ozone peaks in the afternoon, and that of the mesosphere peaks in the night when all atomic oxygen recombines. This

> diurnal cycle is largest in the upper stratosphere and mesosphere, but still exceeds 2 % in the middle stratosphere (Prather et al., 1986; Sakazaki et al., 2013). Combined with the effects of vertical transport by atmospheric tidal winds, this leads to a distinct difference in observed ozone values between sunrise and sunset measurements for solar occultation instruments (e.g., Sakazaki et al., 2013, 2015; Wang et al., 2020). This difference between the sunrise and sunset measurement values, and the resulting bias between the two, has been noted in previous MAESTRO validation efforts (Kar et al., 2007). To minimize the effects of this difference between the two types of measurements, the MAESTRO sunset and sunrise measurements are treated independently for the calculation of the above metrics in this study. Additionally, diurnal scaling factors for ozone from Strode et al., (2022) have been applied at all altitudes to all comparison datasets, except for ACE-FTS, as done for $NO_2$."
>
> Changed L535:
> **From:** "Comparisons between MAESTRO sunrise and sunset Vis.-ozone data are shown in Figs. 1 through 4."
> **To:** "Comparisons between MAESTRO sunrise and sunset Vis.-ozone data against diurnally-scaled (where required) coincident measurements are shown in Figs. 1 through 4."
>
> Changed L666:
> **From:** "The comparisons between the MAESTRO sunrise and sunset UV-ozone data and the coincident ozone measurements are shown in Figs. 5 through 8."
> **To:** "The comparisons between the MAESTRO sunrise and sunset UV-ozone data and the diurnally-scaled (where required) coincident ozone measurements are shown in Figs. 5 through 8."
>
> Changed L772:
> **From:** "Comparisons between the MAESTRO sunrise and sunset $NO_2$ data against the diurnally-scaled (where required; using the diurnal scaling factors of Strode et al. (2022)) coincident measurements from the other datasets are shown in Figs. 10 through 13."
> **To:** "Comparisons between the MAESTRO sunrise and sunset $NO_2$ data against the diurnally-scaled (where required) coincident measurements from the other datasets are shown in Figs. 10 through 13."

Is it reasonable to use scaling factor based on 2017-2022 to scale observations from 2004-2009? I know the Strode et al. (2022) paper says that the interannual variability is small enough that it can be neglected. But I feel like this is hard to claim based on only 4 years, and it does not actually look that small to me in their paper (fig. 6). I understand it is not feasible to calculate scaling factors for other years yourself, but you could test the sensitivity to some extent using the existing values. I suggest scaling one of the datasets with the max/min scaling factors based on the 4 years available and seeing how that affects the results. Possibly the effect will average out in the mean profiles.

> To examine the sensitivity of the scaling factors, we recalculated the scaling factors for ozone and $NO_2$ in two ways based on your suggestions. First we calculated new monthly scaling factors using only the maximum value at each latitude/altitude/solar zenith angle for the gas fields provided. Then we calculated new monthly-mean scaling factors by replacing one of the years of data, chosen arbitrarily as 2017, with the maximum value at each latitude/altitude/solar zenith angle for the gas fields provided. We then compared the resulting scaling factors to the original scaling factors and determined the relative deviation between the two. The results of this are shown in Figs. R1 through R4 for the June sunrise scaling factors at an altitude of 31 km. From these plots it is evident that the variability in the scaling factors is on the order of 5% for $NO_2$ and 1% for ozone, and as such we feel the use of these scaling factors is justified even though they were calculated from model data that only cover a portion of the MAESTRO measurement period.

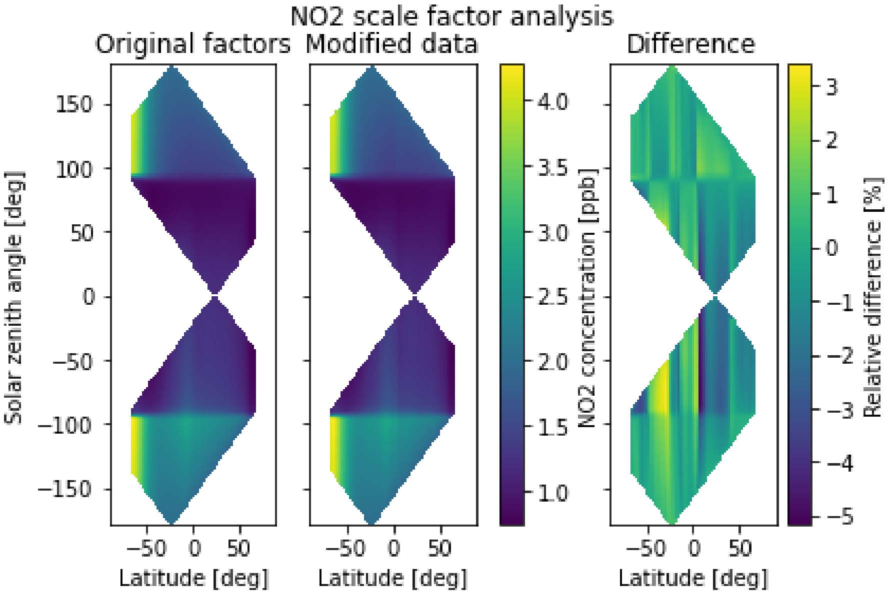

Figure R1: Original (left) and modified (center) monthly $NO_2$ sunrise scaling factors, along with the relative deviation between the two (right). The modified scaling factors were calculated using the maximum $NO_2$ values at each latitude/altitude/solar zenith angle rather than the mean of the data. The relative deviation was calculated as the original factors minus the new ones divided by the original. Data are shown for the June scaling factors at an altitude level of 31 km.

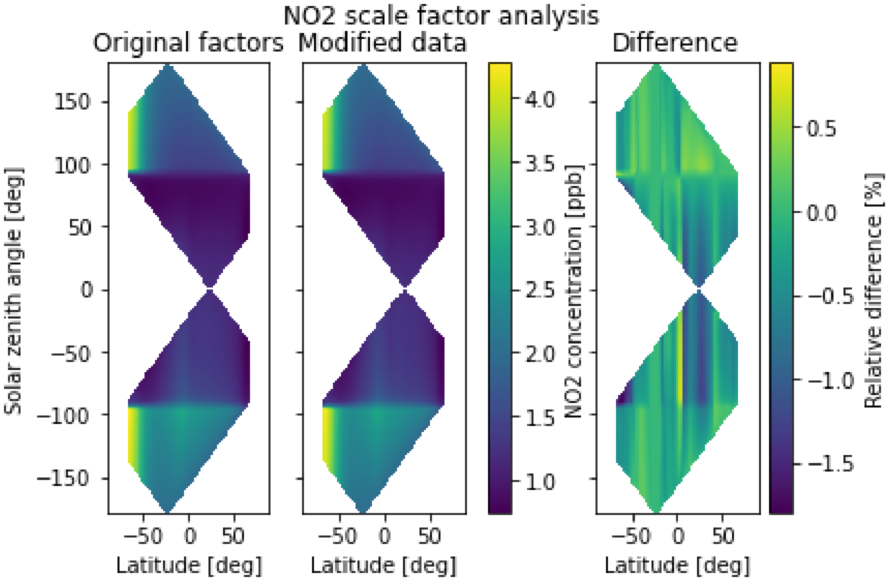

Figure R2: Same as Fig. R1 but the modified monthly scaling factors were calculated by replacing one year of data (2017) with the maximum values at each latitude/altitude/solar zenith angle before taking the mean of the data.

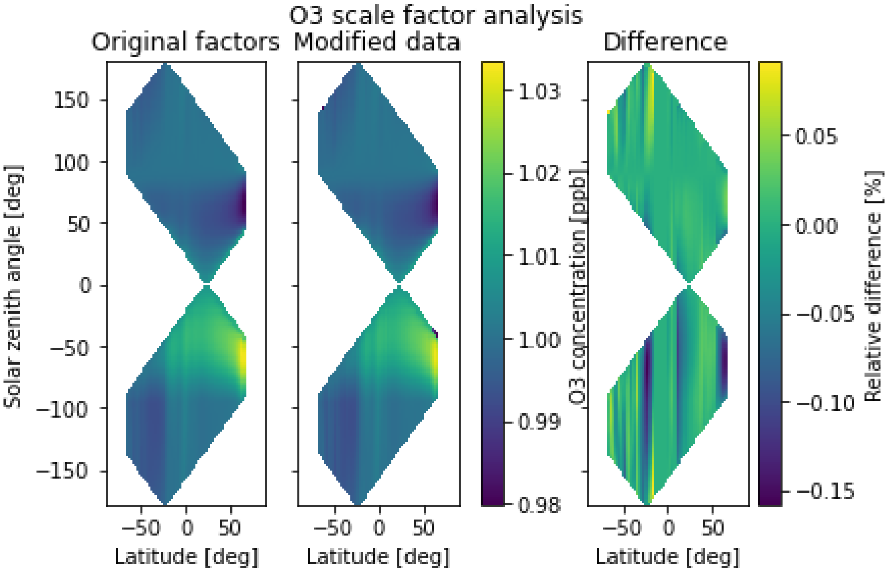

Figure R3: Same as Fig. R1 but for ozone.

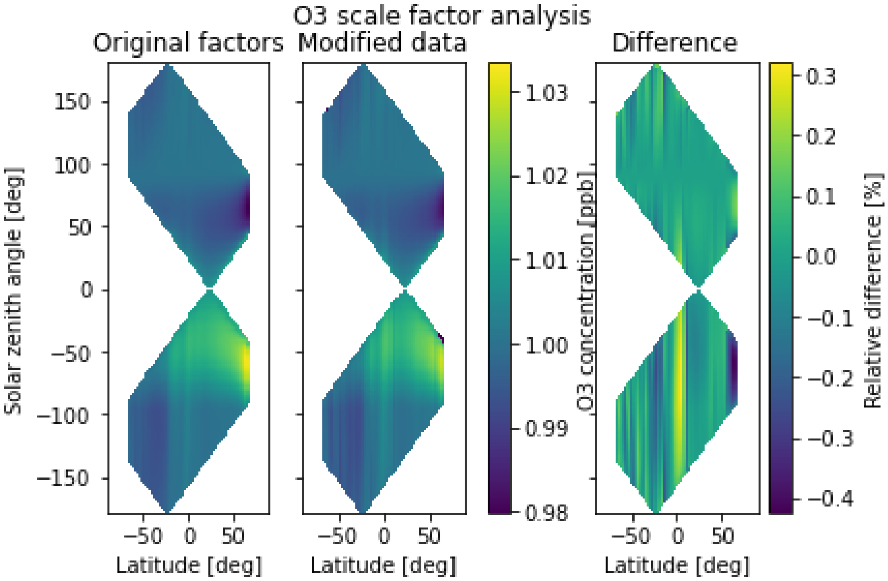

Figure R4: Same as Fig. R2 but for ozone.

Section 4.2.3: It would be useful to include some discussion about possible reasons for the differences between the two MAESTRO ozone products.

> We have incorporated a brief discussion of how the UV channel signal loss may contribute to the differences between the two MAESTRO ozone products.
>
> Changed L768:
> **From:** "Between the sharper decrease in ozone above 50 km and the somewhat more limited range over which good agreement and high correlation is found between the MAESTRO UV-ozone product and that of the other datasets, altogether it is found that the Vis.-ozone product is better suited for use in scientific applications."
> **To:** "The loss of signal from the UV channel of MAESTRO likely contributes to the observed differences between the two MAESTRO ozone products. As stated above, the gradual buildup of an unknown contaminant reduced the throughput of the MAESTRO UV channel, such that past 2009 there was insufficient signal for the retrieval of viable products from this channel. However, this was a gradual change over time, rather than a sudden event, and the 2009 end date is empirically determined based on the quality of data retrieved from MAESTRO measurements. As such, there was a gradual decrease in the quality of the products over time, and while the version 4.5 UV products have been thoroughly vetted, this degradation may contribute to the observed differences between the UV-ozone and the Vis.-ozone. The larger standard deviation profiles of the former support a gradual change in the product over time. Between this gradual loss, the sharper decrease in ozone above 50 km and the somewhat more limited range over which good agreement and high correlation is found between the MAESTRO UV-ozone product and that of the other datasets, overall it is found that the Vis.-ozone product is better suited for use in scientific applications."

Figure 12: Some of the large differences at lower altitudes could be because diurnal variations along the instruments' line of sight are neglected in many of the retrievals (e.g. Dube et al. 2021). This is especially a problem with measurements near the terminator and could be mentioned as an additional source of bias.

> We have made mention of this as a potential source of the large differences at low altitudes.
>
> Added to L861:
> "Neglected diurnal variations along the line-of-sight in the retrievals of the instruments examined may contribute to these low altitude differences (e.g., Dube et al., 2021)."

Line 900: "slightly uncorrelated" is not very precise. It looks like they are not correlated.

> We have updated the text to clarify the meaning. We have also added mention of sampling bias here, following suggestions from Reviewer 1.
>
> Changed L900:
> **From:** "Likely influenced by the few coincident profiles, the SAGE II average sunrise correlation is only 0.49 for this range, and the sunset comparisons are found to be slightly uncorrelated, with an average correlation coefficient of -0.03."
> **To:** "Likely influenced by the few coincident profiles, as well as a systematic difference in sampling location, the SAGE II average sunrise correlation is only 0.49 for this range, and the sunset comparisons are found to be uncorrelated, with an average correlation coefficient of -0.03."

Conclusion: An additional paragraph the clearly states when/where the MAESTRO O3 and NO2 can confidently be used for scientific purposes would be useful.

> We have added a paragraph to the conclusion outlining the regions/periods where the MAESTRO products should be used.
>
> Added to L937:
> "Overall, the findings presented in this work support the use of the MAESTRO version 4.5 dataset for stratospheric studies. The Vis.-ozone product is viable from the start of the ACE mission (February 2004) through to the present and their usage should principally be confined to between 20 and 50 km. This Vis.-ozone product is the preferred MAESTRO ozone product for general applications, due to the

UV-ozone products only being viable until December 2009 and over a slightly narrower stratospheric range from 20 to 45 km; however, for studies focusing on UV-derived measurements of ozone, it is a valid dataset for consideration. Finally, the MAESTRO $NO_2$ product is found to be only viable from the start of the mission until June 2009 and general agreement with other datasets is only achieved between 20 and 40 km. So long as applications for this product are able to work within this limited range of viability, then the MAESTRO $NO_2$ product should be suitable for scientific applications."

## Updated Figures, Tables, and Text

Due to suggestions from Reviewer 1 and 2, several small changes have been made to the comparisons, resulting in updated versions of Figures 1 through 13 from the manuscript, Tables 1 through 3 from the manuscript, and the text from the Abstract, Section 4, and the Conclusion pertinent to these comparisons. The updated Figures and Tables are shown below, with their captions omitted for brevity, instead being presented with a reference to their Figure/Table number in the manuscript.

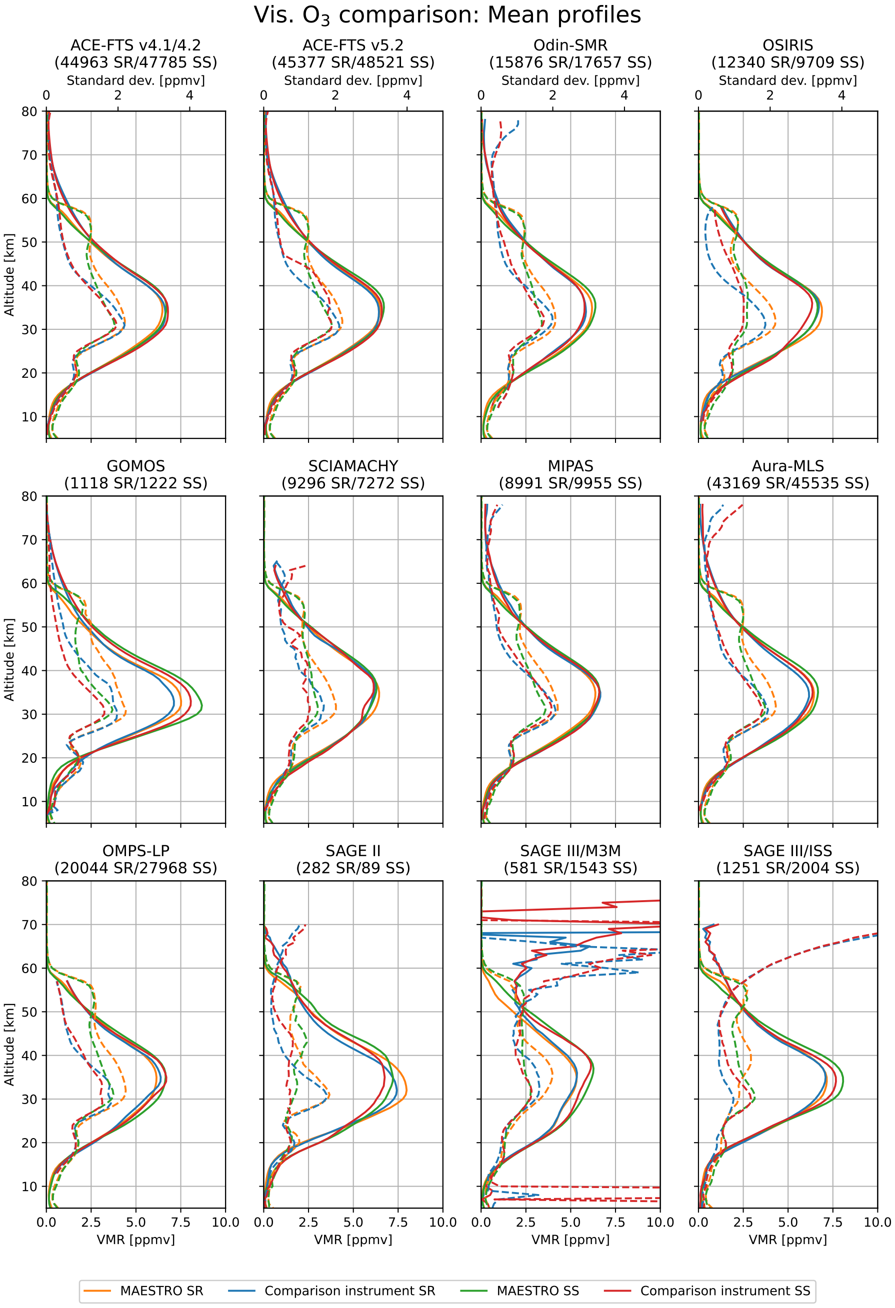

Figure R5: Revised paper Fig. 1.

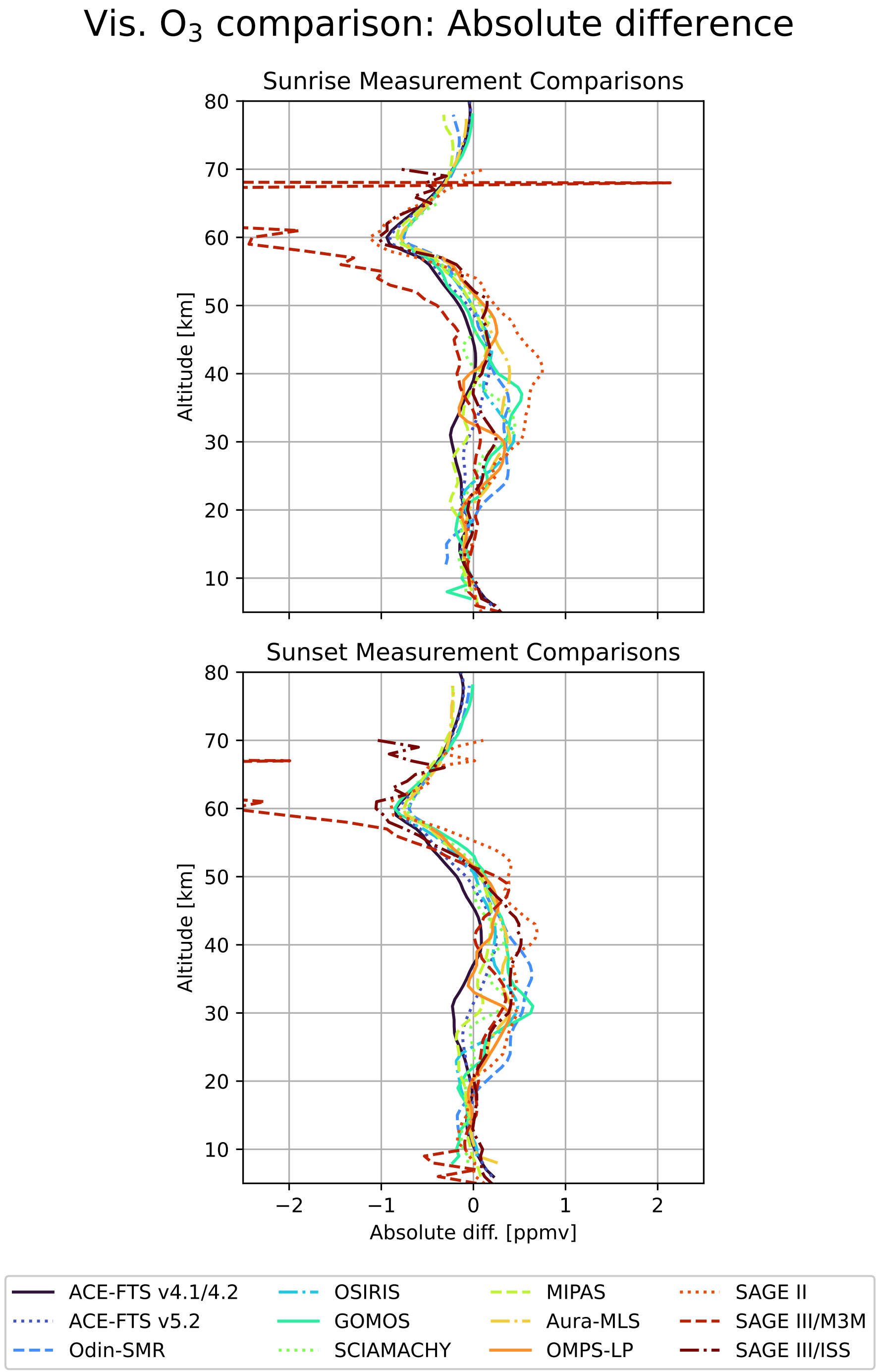

Figure R6: Revised paper Fig. 2.

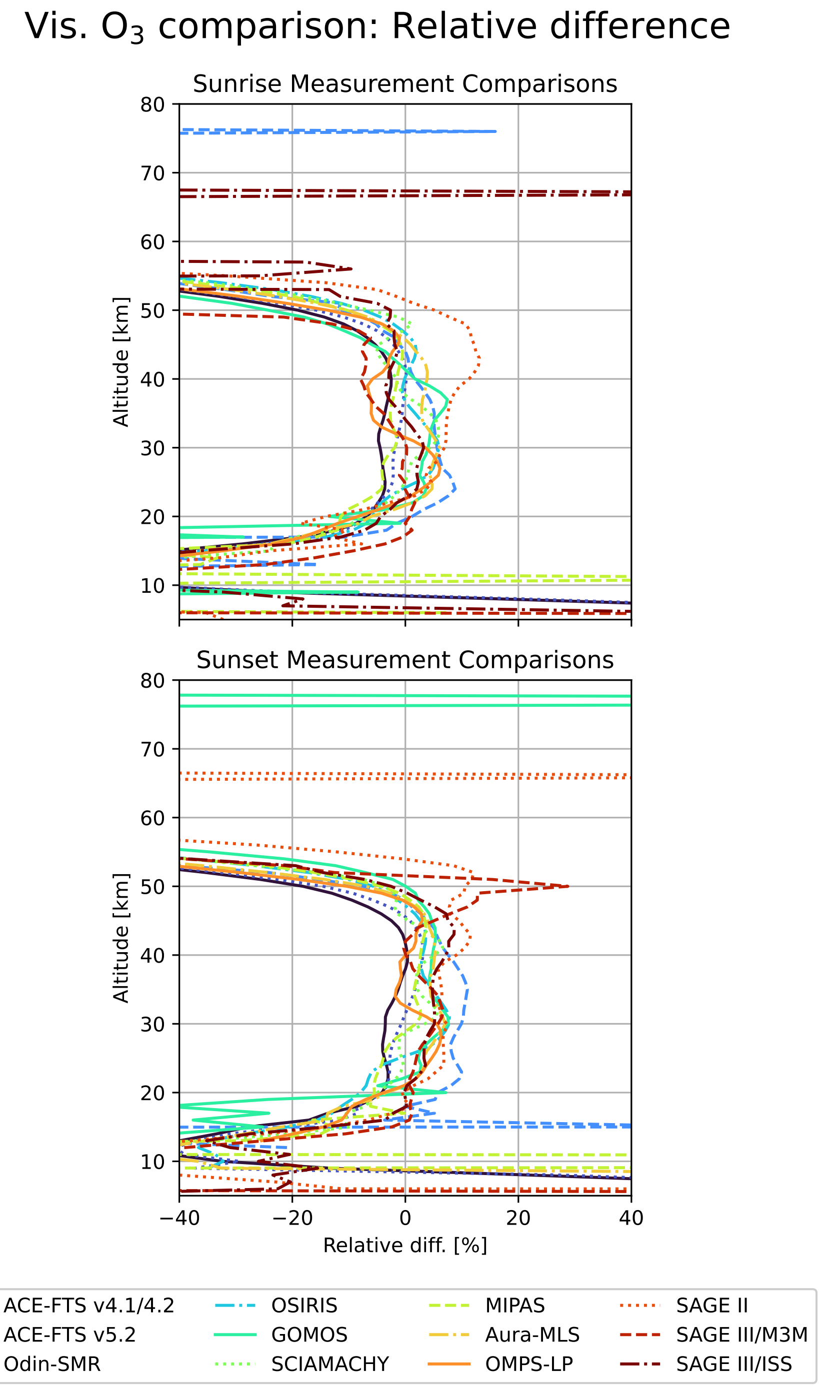

Figure R7: Revised paper Fig. 3.

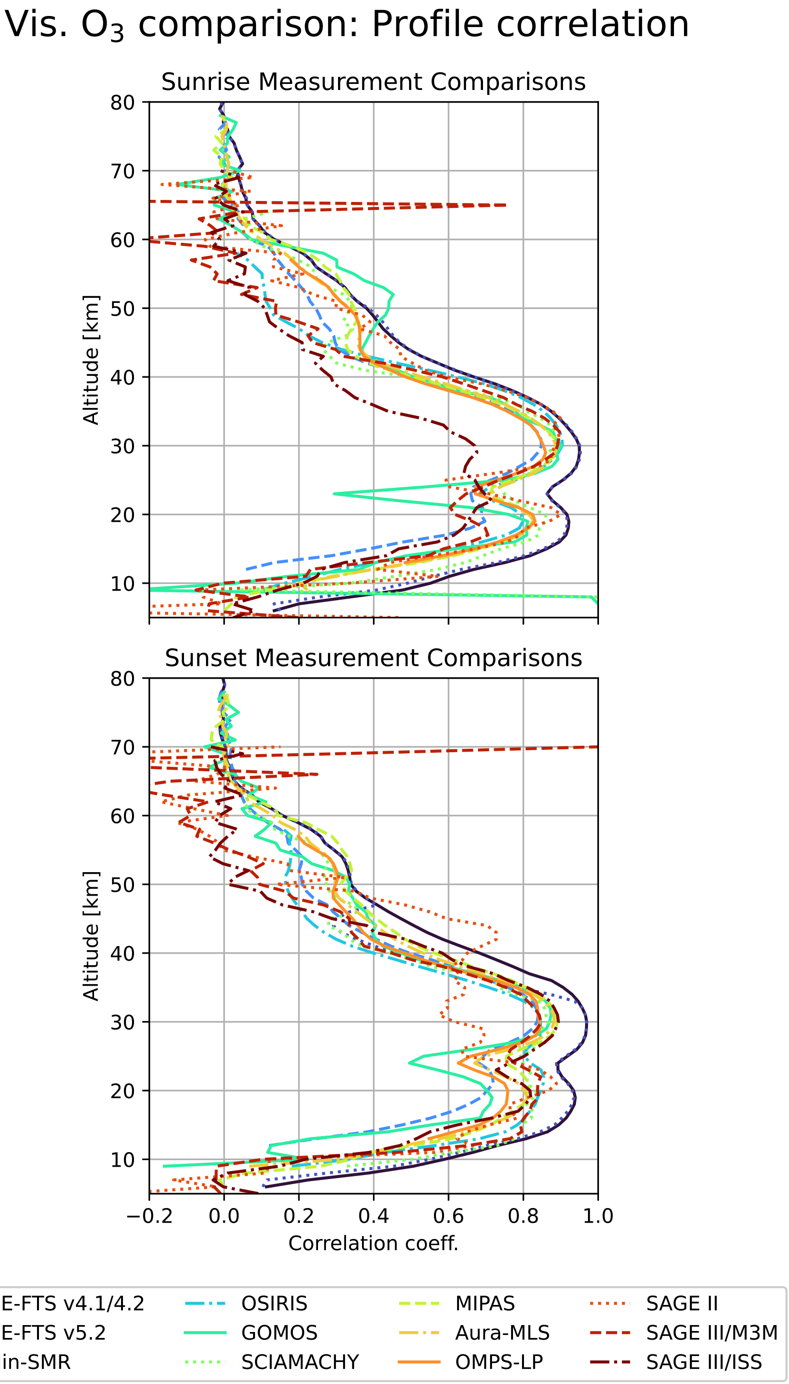

Figure R8: Revised paper Fig. 4.

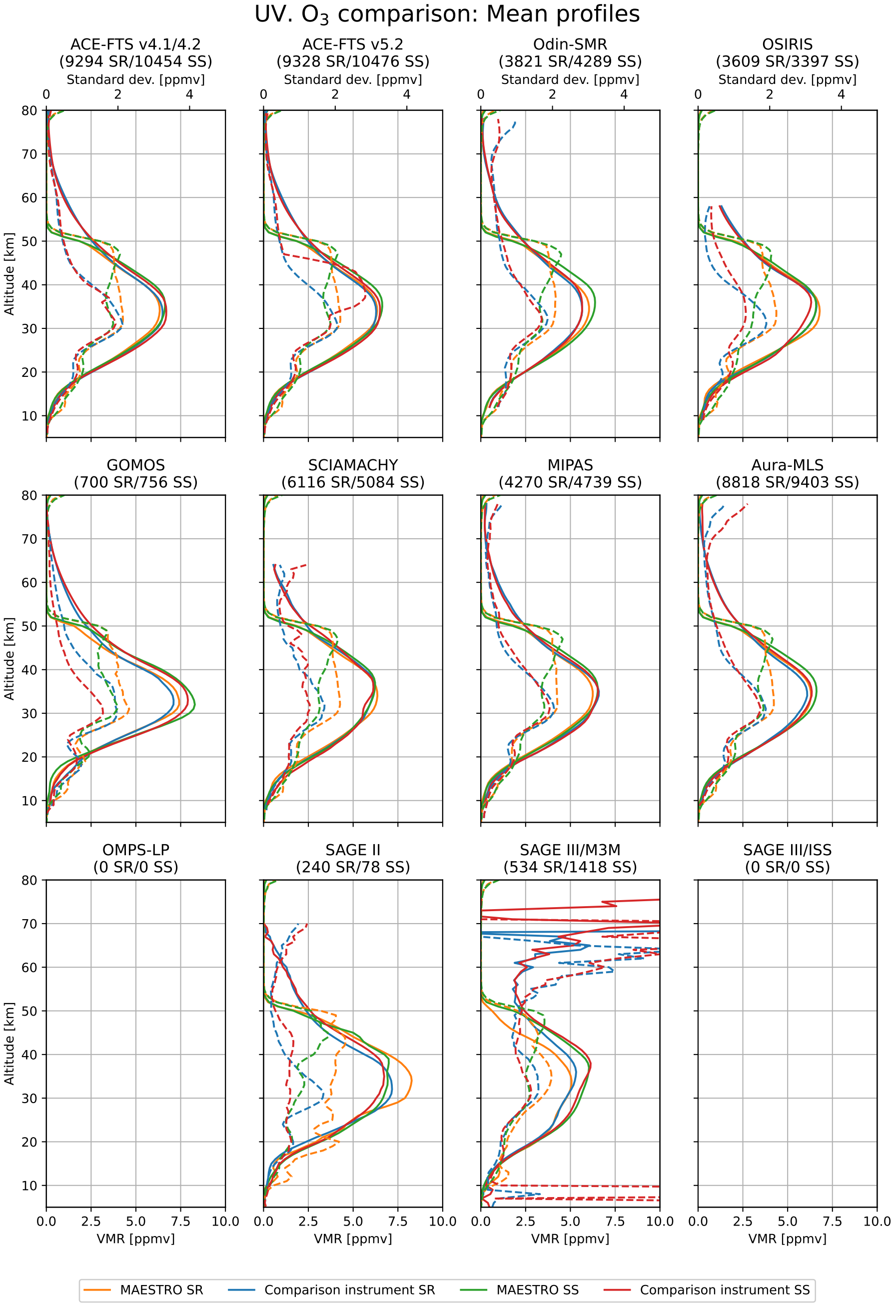

Figure R9: Revised paper Fig. 5.

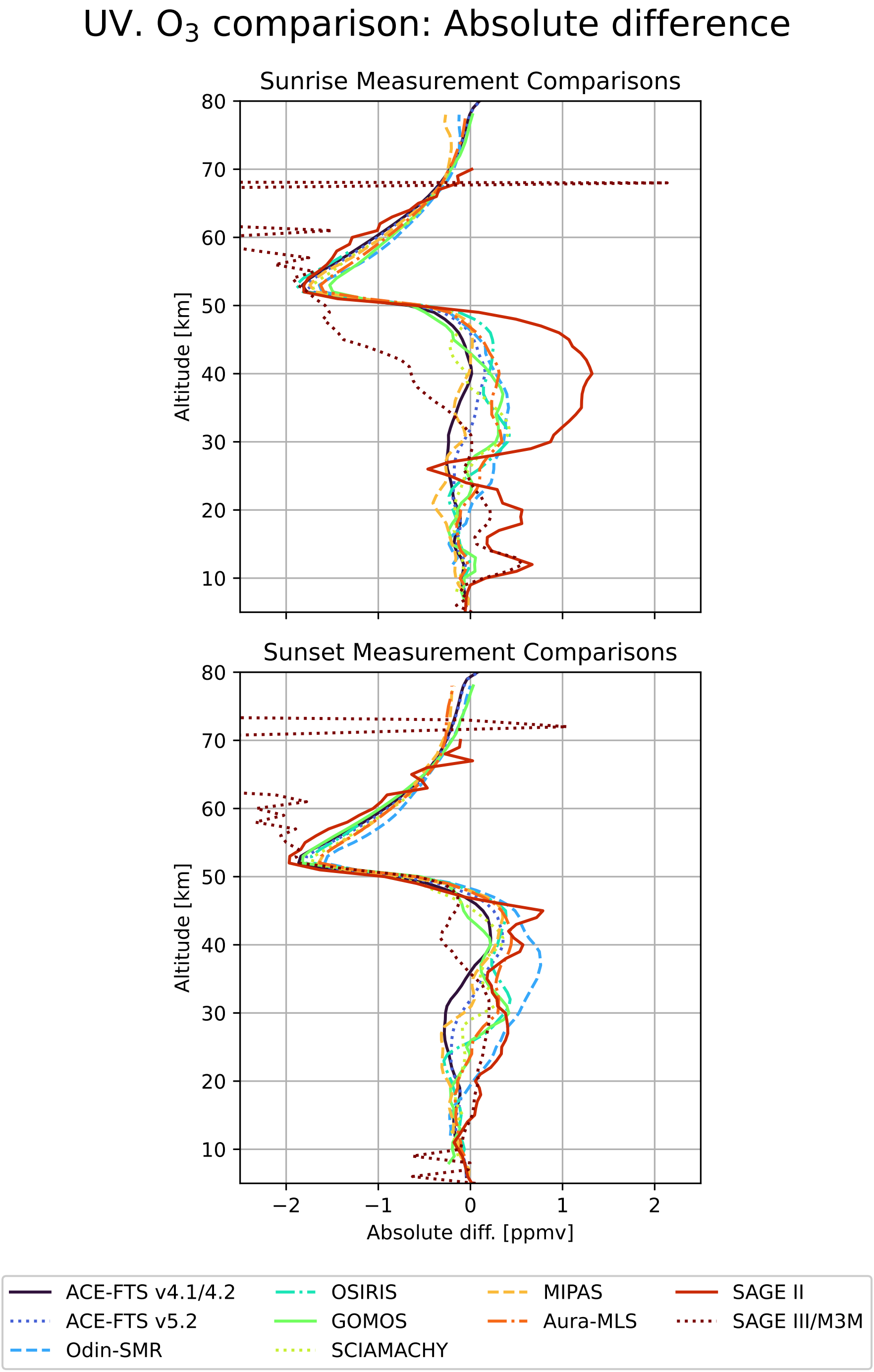

Figure R10: Revised paper Fig. 6.

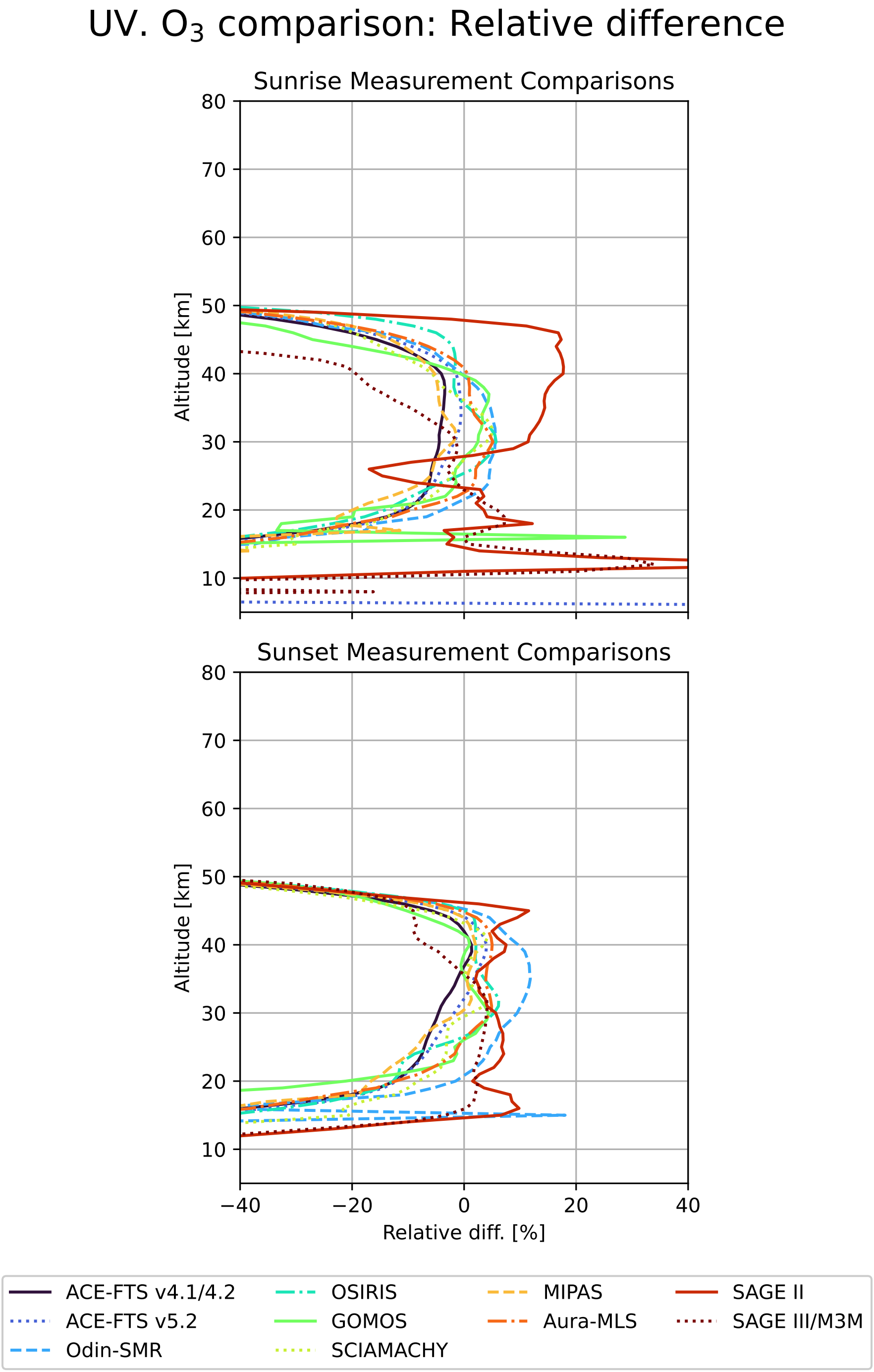

Figure R11: Revised paper Fig. 7.

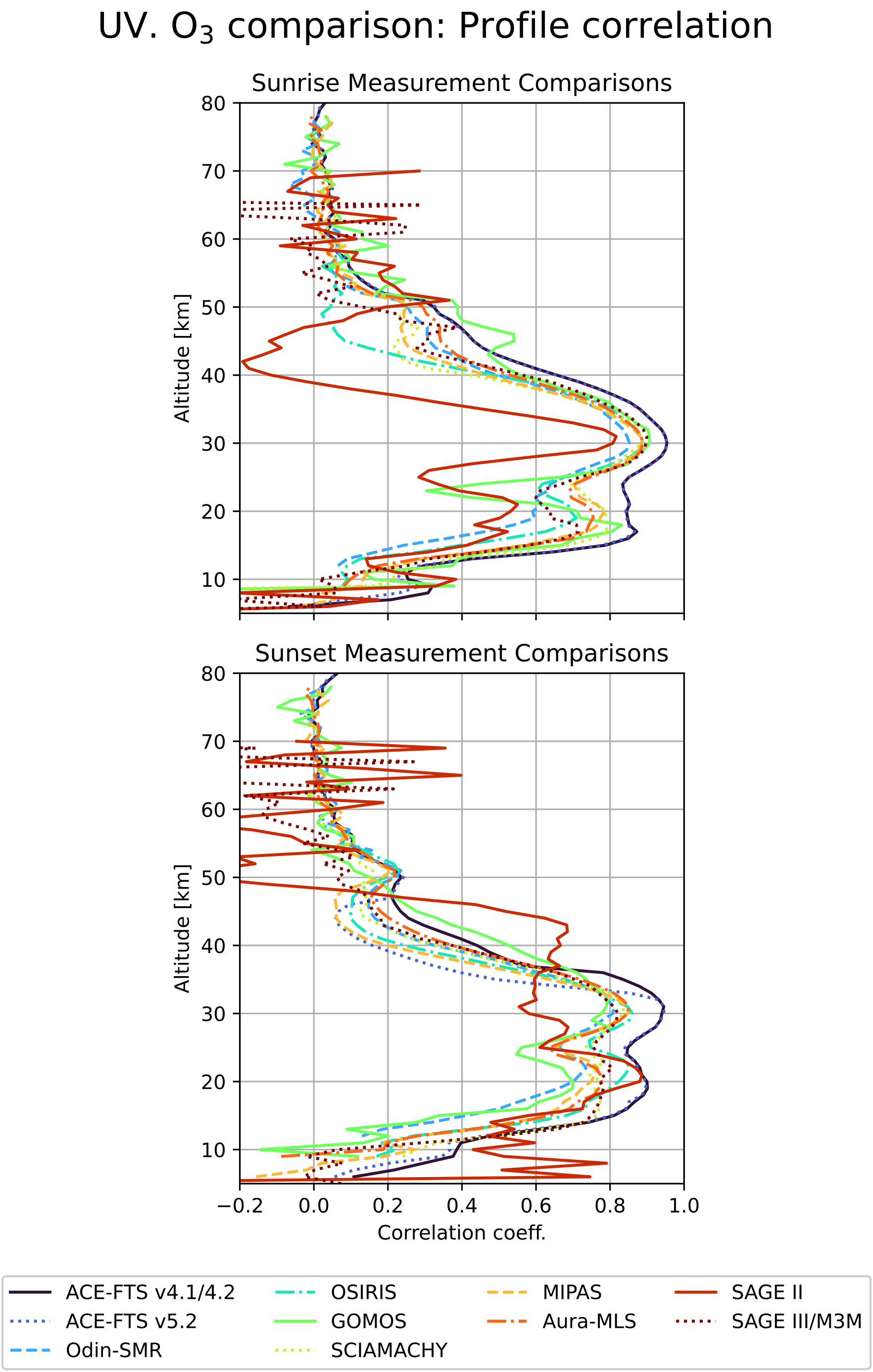

Figure R12: Revised paper Fig. 8.

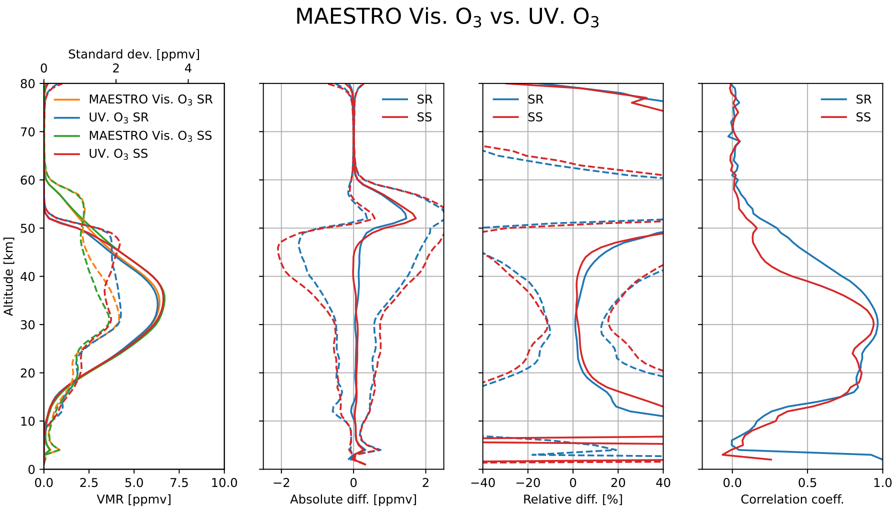

Figure R13: Revised paper Fig. 9.

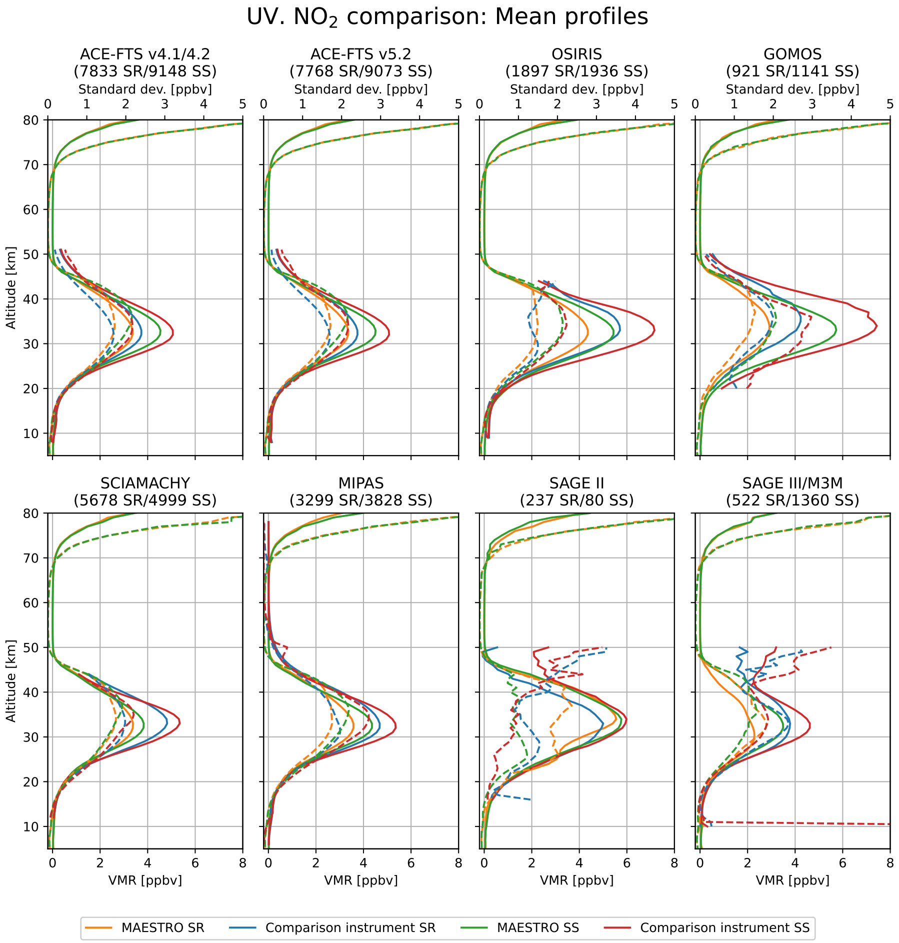

Figure R14: Revised paper Fig. 10.

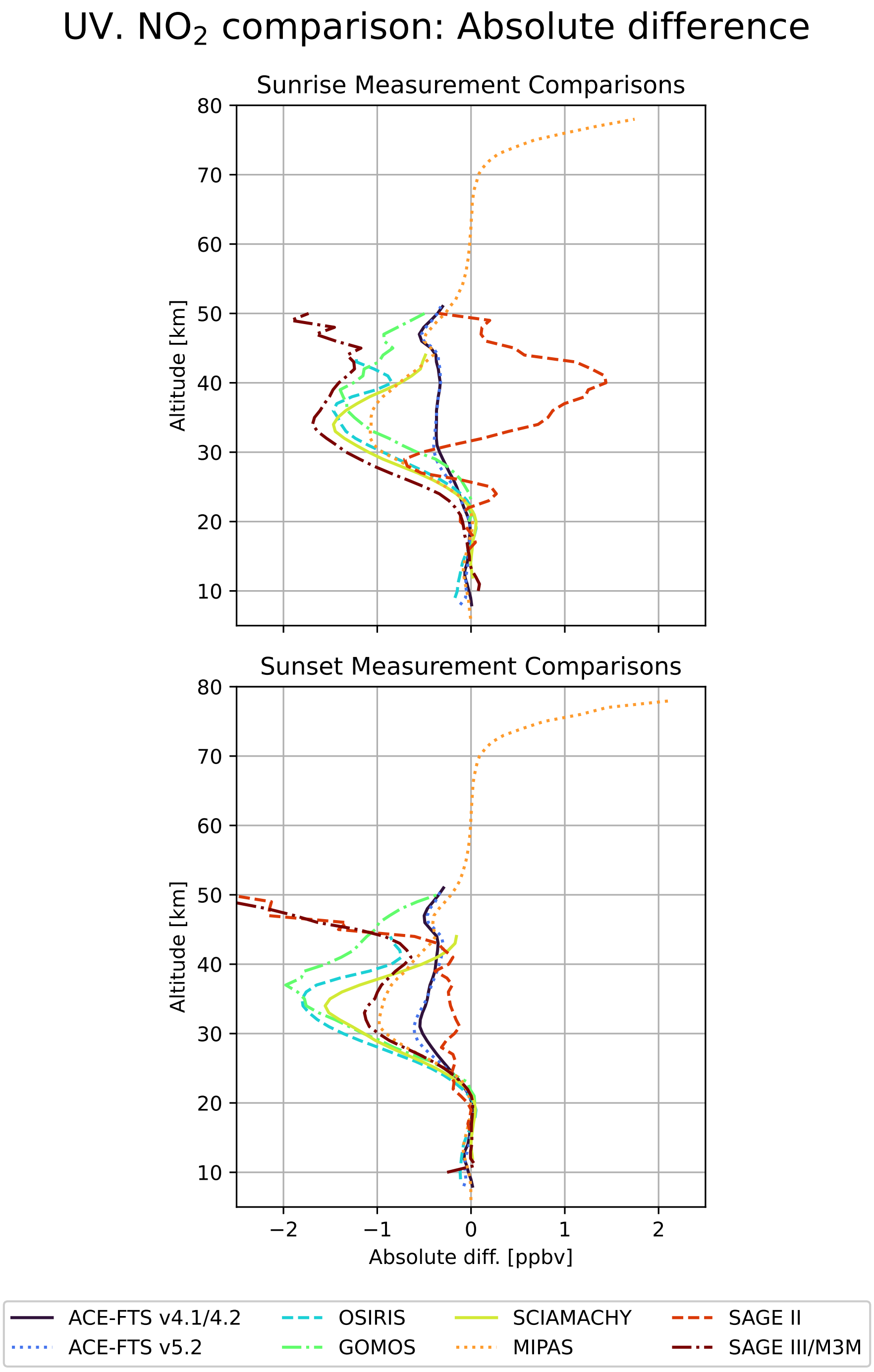

Figure R15: Revised paper Fig. 11.

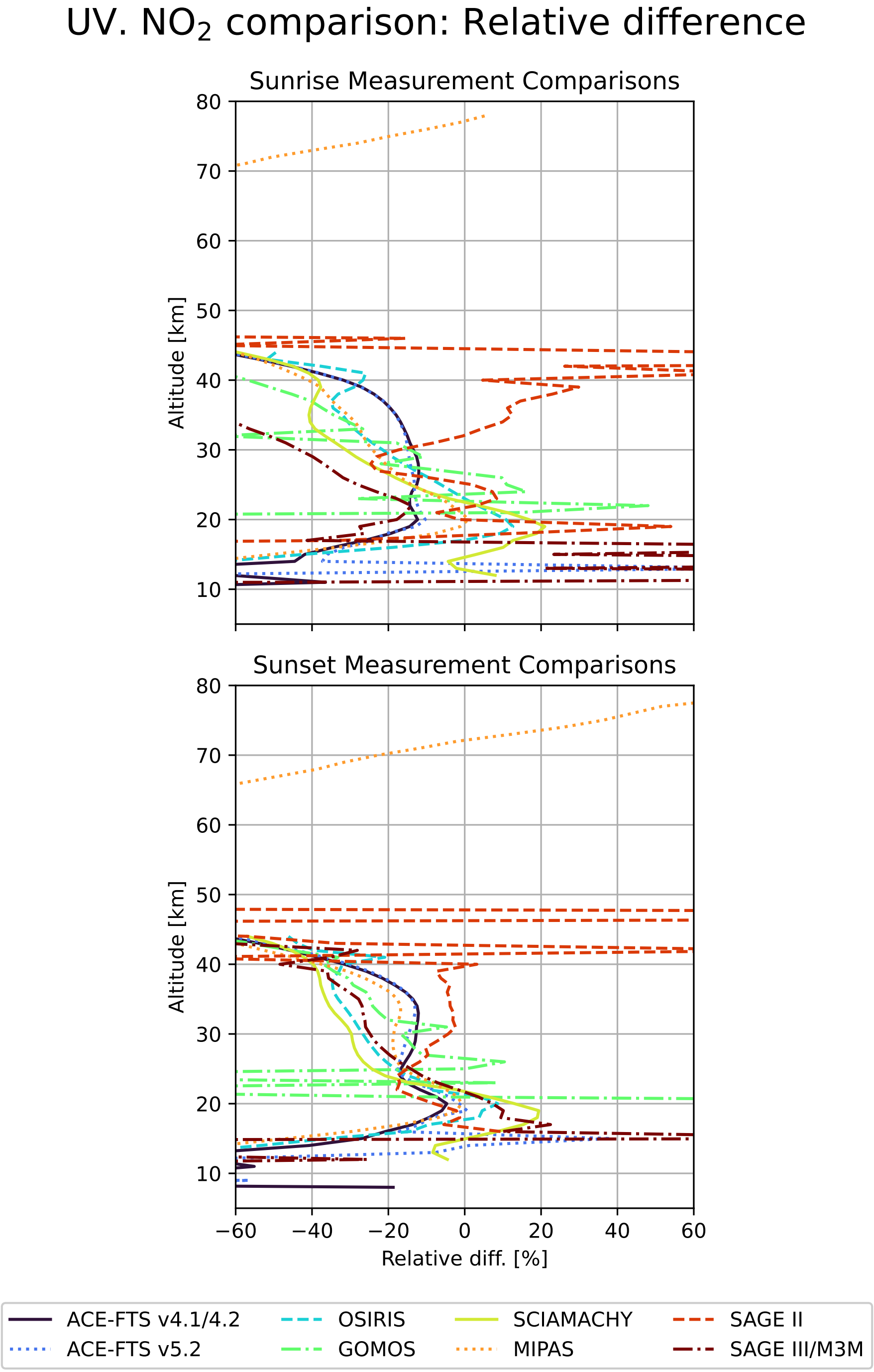

Figure R16: Revised paper Fig. 12.

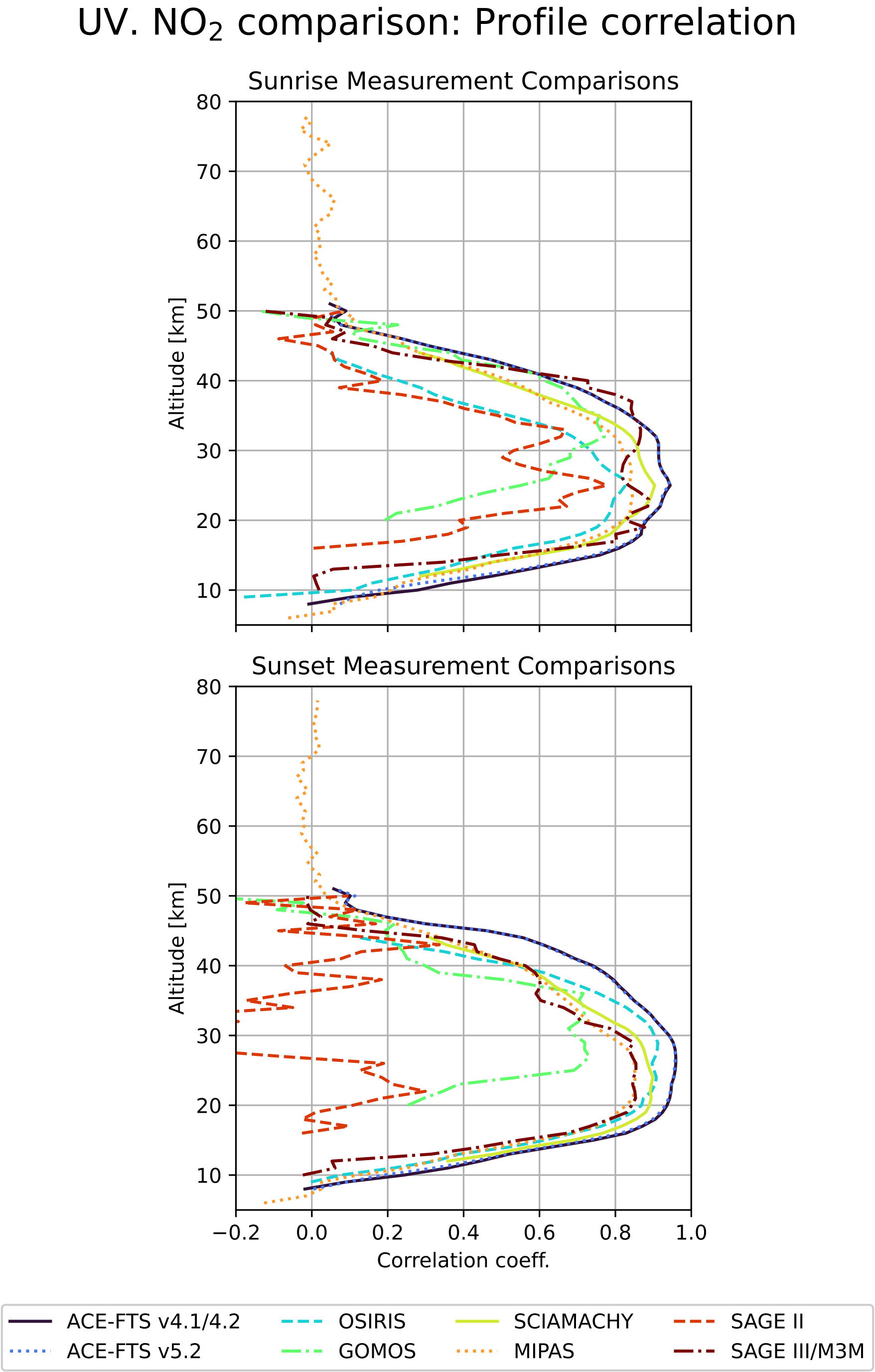

Figure R17: Revised paper Fig. 13.

Table R1: Revised paper Table 1.

| Alt. range | 15–20 km | | 20–50 km | | 50–80 km | |
|---|---|---|---|---|---|---|
| Metric | Mean $\Delta_{abs}$ (ppmv) | Mean $\Delta_{rel}$ (%) | Mean $\Delta_{abs}$ (ppmv) | Mean $\Delta_{rel}$ (%) | Mean $\Delta_{abs}$ (ppmv) | Mean $\Delta_{rel}$ (%) |
| ACE-FTS v4.1/4.2 | 0.09 (0.03) | 22.0 (14.6) | 0.11 (0.11) | 4.8 (3.1) | 0.38 (0.40) | 144.9 (148.4) |
| ACE-FTS v5.2 | 0.08 (0.02) | 19.7 (11.2) | 0.10 (0.12) | 2.5 (2.5) | 0.35 (0.37) | 153.1 (147.5) |
| Odin-SMR | 0.14 (0.09) | 140.1 (14.3) | 0.25 (0.40) | 4.4 (7.6) | 0.32 (0.31) | 141.3 (143.9) |
| OSIRIS | 0.05 (0.09) | 17.7 (14.1) | 0.19 (0.24) | 2.7 (4.4) | 0.26 (0.28) | 40.2 (45.5) |
| GOMOS | 0.17 (0.11) | 74.3 (30.7) | 0.23 (0.31) | 5.3 (4.5) | 0.33 (0.31) | 144.7 (140.3) |
| MIPAS | 0.12 (0.07) | 21.9 (8.3) | 0.12 (0.13) | 3.3 (2.8) | 0.37 (0.36) | 138.5 (141.2) |
| SCIAMACHY | 0.11 (0.07) | 17.7 (9.0) | 0.16 (0.14) | 3.0 (2.7) | 0.41 (0.39) | 101.6 (95.1) |
| OMPS-LP | 0.09 (0.04) | 20.1 (10.5) | 0.17 (0.17) | 4.3 (2.6) | 0.14 (0.20) | 46.0 (50.7) |
| Aura-MLS | 0.08 (0.05) | 20.4 (13.3) | 0.26 (0.26) | 3.7 (4.2) | 0.30 (0.35) | 141.9 (149.6) |
| SAGE II | 0.06 (0.04) | 14.7 (4.8) | 0.45 (0.41) | 8.2 (7.1) | 0.47 (0.43) | 116.7 (111.5) |
| SAGE III/M3M | 0.02 (0.03) | 2.9 (1.2) | 0.11 (0.18) | 4.5 (4.0) | 4.67 (6.95) | 147.8 (136.2) |
| SAGE III/ISS | 0.03 (0.03) | 16.1 (4.0) | 0.11 (0.31) | 2.3 (4.7) | 0.49 (0.62) | 177.8 (124.7) |

Table R2: Revised paper Table 2.

| Alt. range | 15–20 km | | 20–45 km | | 45–80 km | |
|---|---|---|---|---|---|---|
| Metric | Mean $\Delta_{abs}$ (ppmv) | Mean $\Delta_{rel}$ (%) | Mean $\Delta_{abs}$ (ppmv) | Mean $\Delta_{rel}$ (%) | Mean $\Delta_{abs}$ (ppmv) | Mean $\Delta_{rel}$ (%) |
| ACE-FTS v4.1/4.2 | 0.14 (0.14) | 29.4 (31.2) | 0.15 (0.19) | 5.7 (4.2) | 0.64 (0.64) | 162.6 (159.1) |
| ACE-FTS v5.2 | 0.13 (0.12) | 26.8 (28.5) | 0.12 (0.20) | 3.6 (4.0) | 0.60 (0.61) | 158.6 (158.0) |
| Odin-SMR | 0.13 (0.13) | 22.0 (21.9) | 0.27 (0.49) | 4.0 (7.6) | 0.55 (0.55) | 158.9 (159.8) |
| OSIRIS | 0.13 (0.14) | 32.6 (26.8) | 0.23 (0.28) | 3.9 (4.8) | 1.00 (0.95) | 108.0 (110.6) |
| GOMOS | 0.20 (0.20) | 33.4 (63.7) | 0.17 (0.16) | 4.8 (3.3) | 0.58 (0.62) | 160.0 (154.1) |
| MIPAS | 0.25 (0.21) | 31.2 (31.8) | 0.17 (0.21) | 6.6 (4.6) | 0.63 (0.63) | 162.4 (161.0) |
| SCIAMACHY | 0.15 (0.14) | 23.6 (16.6) | 0.17 (0.16) | 5.0 (2.8) | 0.88 (0.90) | 140.5 (141.3) |
| Aura-MLS | 0.13 (0.16) | 26.7 (31.9) | 0.21 (0.25) | 2.8 (4.2) | 0.58 (0.62) | 162.2 (164.7) |
| SAGE II | 0.36 (0.08) | 5.0 (7.3) | 0.83 (0.34) | 11.9 (5.2) | 0.95 (0.91) | 153.8 (153.3) |
| SAGE III/M3M | 0.13 (0.04) | 3.7 (1.9) | 0.30 (0.17) | 10.8 (4.0) | 4.35 (5.42) | 174.9 (162.3) |

Table R3: Revised paper Table 3.

| Alt. range | 15–20 km | | 20–40 km | | 40–60 km | |
|---|---|---|---|---|---|---|
| Metric | Mean $\Delta_{abs}$ (ppbv) | Mean $\Delta_{rel}$ (%) | Mean $\Delta_{abs}$ (ppbv) | Mean $\Delta_{rel}$ (%) | Mean $\Delta_{abs}$ (ppbv) | Mean $\Delta_{rel}$ (%) |
| ACE-FTS v4.1/4.2 | 0.02 (0.01) | 27.3 (15.4) | 0.25 (0.35) | 15.7 (14.3) | 0.40 (0.40) | 102.3 (101.4) |
| ACE-FTS v5.2 | 0.02 (0.01) | 25.0 (15.4) | 0.28 (0.37) | 15.9 (14.4) | 0.39 (0.36) | 102.1 (100.9) |
| OSIRIS | 0.04 (0.04) | 16.7 (13.6) | 0.76 (1.02) | 19.2 (23.6) | 1.04 (0.81) | 38.3 (36.5) |
| GOMOS | - | - | 0.60 (1.00) | 42.3 (42.9) | 0.92 (1.00) | 120.0 (99.0) |
| MIPAS | 0.03 (0.03) | 22.3 (20.4) | 0.65 (0.60) | 20.9 (16.2) | 0.31 (0.26) | 128.0 (127.8) |
| SCIAMACHY | 0.02 (0.03) | 13.1 (12.6) | 0.79 (0.83) | 26.6 (27.2) | 0.59 (0.29) | 47.1 (46.9) |
| SAGE II | 0.03 (0.02) | 90.8 (4.6) | 0.50 (0.20) | 12.0 (8.5) | 0.66 (1.22) | 137.4 (151.3) |
| SAGE III/M3M | 0.05 (0.01) | 52.5 (34.7) | 1.03 (0.66) | 43.4 (21.0) | 1.45 (1.44) | 143.9 (113.0) |

The following lines have been altered following the changes made in the comparisons suggested by Reviewers 1 and 2. Most of these changes are changes in the values reported, with some reflecting small changes in the discussion of these comparisons.

Changed L11:
**From:** "A similar bias, albeit with slightly poorer agreement, is found with the UV-ozone product in the stratosphere, with the average stratospheric agreement between MAESTRO and the other datasets ranging from 2.9 % to 11.9 %. For $NO_2$, general agreement with the comparison datasets is only found in the range from 20 to 40 km. Within this range, MAESTRO is found to have a low bias for $NO_2$, and most of the datasets agree to within 27.5 %, although the average agreement ranges from 8.5 % to 43.4 %."
**To:** "A similar bias, albeit with slightly poorer agreement, is found with the UV-ozone product in the stratosphere, with the average stratospheric agreement between MAESTRO and the other datasets ranging from 2.8 % to 11.9 %. For $NO_2$, general agreement with the comparison datasets is only found in the range from 20 to 40 km. Within this range, MAESTRO is found to have a low bias for $NO_2$, and most of the datasets agree to within 27.2 %, although the average agreement ranges from 8.5 % to 43.4 %."

Changed L558:
**From:** "The MAESTRO standard deviation is found to be larger than that of ACE-FTS between 30 and 60 km, with the largest differences around 55 km. Above this altitude range, the near 0 ppmv MAESTRO standard deviation profiles are smaller than those profiles from ACE-FTS."
**To:** "With exception for the ACE-FTS v5.2 sunset profiles, the MAESTRO standard deviation is found to be larger than that of ACE-FTS between 30 and 60 km. The largest differences in these standard deviation profiles occur around 55 km, where the ACE-FTS v5.2 sunset profile also are found to fall to lower values than the corresponding MAESTRO profile. Above 60 km, the near 0 ppmv MAESTRO standard deviation profiles are smaller than those profiles from ACE-FTS."

Changed L562:
**From:** "The comparisons with MIPAS are largely similar to those with ACE-FTS, with the two mean MIPAS ozone profiles overlapping significantly with each other and with the two mean MAESTRO ozone profiles below 55 km. However, above 60 km, the mean MIPAS profiles are found to show larger concentrations of ozone than observed for the ACE-FTS or MAESTRO, and accompanying these larger mean profiles are much larger standard deviations than for these previous datasets."
**To:** "The comparisons with MIPAS are largely similar to those with ACE-FTS, with the two mean MIPAS ozone profiles overlapping significantly with each other and with the two mean MAESTRO ozone profiles below 50 km, and with similar standard deviation profiles as observed with ACE-FTS. However, above 65 km, the MIPAS standard deviation profiles are found to be significantly larger than those observed for ACE-FTS or MAESTRO."

Changed L566:
**From:** "Generally good agreement is found with the SCIAMACHY, Aura-MLS, and OMPS-LP comparisons; however, only Aura-MLS reaches to the top of the MAESTRO profile, so the other two cannot be used to assess the representation of mesospheric ozone from MAESTRO. For SCIAMACHY, the mean profiles near the stratospheric ozone maximum are observed to flatten somewhat between 30 and 35 km, leading to the high bias of about 0.5 ppmv observed over this range for MAESTRO. Other than that, the two datasets are found to agree between approximately 15 and 55 km. The profiles from Aura-MLS differ from those from MAESTRO by about 0.5 ppmv in the middle stratosphere; however, a more pronounced difference is visible between the the mean sunrise and sunset coincident profiles, which are found to differ from each other to a greater extent than for the previously discussed datasets. In the mesosphere, the Aura-MLS comparisons are found to be similar to those made with MIPAS, with higher ozone VMRs over this range than observed with ACE-FTS and MAESTRO. Lastly, the coincident OMPS-LP profiles are found to yield smaller mean VMRs than MAESTRO between 25 and 35 km and larger mean VMRs between 35 and 40 km, but overall similar agreement is found through most of the profile as that observed for the previous two datasets. "
**To:** "Generally good agreement is found with the SCIAMACHY, Aura-MLS, and OMPS-LP comparisons; however, only Aura-MLS reaches to the top of the MAESTRO profile, so the other two cannot be used to assess the representation of mesospheric ozone from MAESTRO. The profiles from Aura-MLS differ from those from MAESTRO by about 0.5 ppmv in the middle stratosphere; however, a more pronounced

difference is visible between the sunrise and sunset coincident profile, which are found to differ from each other to a greater extent than for the previously discussed datasets. In the mesosphere, the Aura-MLS comparisons are found to be similar to those made with MIPAS, with larger ozone standard deviation and slightly larger mean ozone values over this range than observed with ACE-FTS and MAESTRO. For the comparisons with SCIAMACHY, the largest differences in the mean ozone profiles are found just below the stratospheric ozone maximum, where MAESTRO is found to yield larger ozone VMRs by about 0.5 ppmv. Other than that, the two datasets are found to broadly agree between approximately 15 and 55 km. Lastly, the coincident OMPS-LP profiles are found to yield smaller mean VMRs than MAESTRO between 25 and 33 km and similar to slightly larger mean VMRs between 33 and 40 km, but overall good agreement is found through most of the profile, similar to that observed for the previous two datasets. ”

Changed L595:
**From:** “Between 60 and 70 km the GOMOS profiles have larger ozone concentrations than MAESTRO, by about 1.2 ppmv, in closer agreement to what was observed for many of the prior comparisons.”
**To:** “Between 60 and 70 km the GOMOS profiles have larger ozone concentrations than MAESTRO, by up to about 1.2 ppmv, in closer agreement to what was observed for many of the prior comparisons.”

Changed L606:
**From**: “These two datasets had the shortest overlap period, and only 371 comparisons could be made for the Vis.-ozone product, nearly an order of magnitude fewer comparisons that for the dataset with the next fewest coincident measurements.”
**To:** “These two datasets had the shortest overlap period, and only 371 comparisons could be made for the Vis.-ozone product, nearly an order of magnitude fewer comparisons than for the dataset with the next fewest coincident measurements.”

Changed L614:
**From:** “This last increase is found to be similar to what is observed for the sunrise SCIAMACHY profiles.”
**To:** “This last increase is found to be similar to what is observed for the sunset SCIAMACHY profiles.”

Changed L630:
**From:** “Focusing between 20 and 50 km, where the overall closest agreement is observed, MAESTRO measurements agree best with ACE-FTS version 4.1/4.2, ACE-FTS version 5.2, and MIPAS, which have averaged absolute differences over this vertical range for sunrise (sunset) measurements of 0.12 (0.10) ppmv, 0.10 (0.11) ppmv, and 0.11 (0.16) ppmv respectively. This profile-averaged metric was calculated using the unsigned magnitude of the differences to avoid oppositely signed values from cancelling. These differences translate into profile-averaged relative differences of 5.0 (2.9) % for ACE-FTS version 4.1/4.2, 2.7 (2.4) % for ACE-FTS version 5.2, and 3.9 (3.5) % for MIPAS over this range. Very good agreement is also found with OSIRIS, SCIAMACHY, OMPS-LP, SAGE III/M3M, and sunrise measurements from SAGE III/ISS. With exception for SCIAMACHY, better agreement is consistently found with sunrise measurements than for sunset measurements. Comparisons with Odin-SMR and SAGE II show the poorest agreement over this range, with average absolute differences of 0.24 (0.39) ppmv and 0.45 (0.41) ppmv respectively for sunrise (sunset) comparisons, however these translate into average relative differences of 4.7 (7.6) % for Odin-SMR, and 8.2 (7.1) % for SAGE II, indicative that the MAESTRO Vis.-ozone product is still generally in good agreement in the range of 20 to 50 km.”
**To:** “Focusing between 20 and 50 km, where the overall closest agreement is observed, MAESTRO measurements agree best with ACE-FTS version 4.1/4.2, ACE-FTS version 5.2, and MIPAS, which have averaged absolute differences over this vertical range for sunrise (sunset) measurements of 0.11 (0.11) ppmv, 0.10 (0.12) ppmv, and 0.12 (0.13) ppmv respectively. This profile-averaged metric was calculated using the unsigned magnitude of the differences to avoid oppositely signed values from cancelling. These differences translate into profile-averaged relative differences of 4.8 (3.1) % for ACE-FTS version 4.1/4.2, 2.5 (2.5) % for ACE-FTS version 5.2, and 3.3 (2.8) % for MIPAS over this range. Very good agreement is also found with OSIRIS, SCIAMACHY, OMPS-LP, SAGE III/M3M, and sunrise measurements from SAGE III/ISS. Comparisons with Odin-SMR and SAGE II show the poorest agreement over this range, with average absolute differences of 0.25 (0.40) ppmv and 0.45 (0.41) ppmv respectively for sunrise (sunset) comparisons, however these translate into average relative differences of 4.4 (7.6) % for Odin-SMR, and 8.2 (7.1) % for SAGE II, indicative that the MAESTRO Vis.-ozone product is still generally in good

agreement in the range of 20 to 50 km."

Changed L643:
**From:** "Between 15 and 20 km, near the lower bounds of many of the instrument measurements, most of the datasets continue to show reasonable agreement with MAESTRO, with most of the sunrise (sunset) MAESTRO measurements agreeing with the comparison datasets to within 25 (15) %. The main exception to this are the comparisons with the GOMOS instrument, which show values differing by 98.9 (33.9) % on average for sunrise (sunset) comparisons. Below this altitude range, the comparisons show significant disagreement, often displaying differences at particular altitudes in excess of 50 %. Similarly, above 50 km, the comparisons generally show considerable disagreement, with differences reaching over 100 %."
**To:** "Between 15 and 20 km, near the lower bounds of many of the instrument measurements, most of the datasets continue to show reasonable agreement with MAESTRO, with most of the sunrise (sunset) MAESTRO measurements agreeing with the comparison datasets to within 22.0 (14.6) %. The main exceptions to this are the comparisons with the GOMOS instrument, which show values differing by 74.3 (30.7) % on average for sunrise (sunset) comparisons, and the sunrise Odin-SMR comparisons which show a 140.1 % difference on average. Below this altitude range, the comparisons show significant disagreement, often displaying differences at particular altitudes in excess of 50 %. Similarly, above 50 km, the comparisons generally show considerable disagreement, with differences reaching over 100 %."

Changed L703:
**From:** "Within this range, the best agreement is found with ACE-FTS, with version 4.1/4.2 differing by 0.15 (0.20) ppmv and version 5.2 differing by 0.11 (0.21) ppmv, and with GOMOS, with differences of 0.15 (0.22) ppmv on average from sunrise (sunset) coincident profiles. The relative differences also reflect this good agreement, with average sunrise (sunset) differences of 5.8 (4.3) % for ACE-FTS version 4.1/4.2, 3.7 (4.3) % for ACE-FTS version 5.2, and 4.3 (4.0) % for GOMOS. Due to greater differences near the stratospheric ozone maximum, Aura-MLS is found to have a smaller average relative difference than most of these comparisons, of of 2.9 (4.7) % during sunrise (sunset), while also having larger average absolute differences, of 0.17 (0.26) ppmv. This mixed behaviour emphasizes the need to include both difference metrics in this analysis."
**To:** "Within this range, the best agreement is found with ACE-FTS, with version 4.1/4.2 differing by 0.15 (0.19) ppmv and version 5.2 differing by 0.12 (0.20) ppmv, with GOMOS, with differences of 0.17 (0.16) ppmv, and with SCIAMACHY, with differences of 0.17 (0.16) ppmv, on average from sunrise (sunset) coincident profiles. The relative differences also reflect this good agreement, with average sunrise (sunset) differences of 5.7 (4.2) % for ACE-FTS version 4.1/4.2, 3.6 (4.0) % for ACE-FTS version 5.2, 4.8 (3.3) % for GOMOS, and 5.0 (2.8) % for SCIAMACHY. Due to greater differences near the stratospheric ozone maximum, Aura-MLS is found to have a smaller average relative difference than most of these comparisons, of 2.8 (4.2) % during sunrise (sunset), while also having larger average absolute differences, of 0.21 (0.25) ppmv. This mixed behavior emphasizes the need to include both difference metrics in this analysis."

Changed L711:
**From:** "Considering all of the datasets within the 20 to 45 km range, the majority of the comparisons have average absolute differences between 0.11 and 0.30 ppmv for the sunrise comparisons and between 0.17 and 0.34 ppmv for the sunset comparisons. The exceptions to this are the sunrise SAGE II comparisons which have an average absolute difference of 0.83 ppmv, and the sunset Odin-SMR comparisons with an average difference of 0.50 ppmv. These two datasets also show the highest relative differences, of 11.9 % and 7.9 % on average; however, these average differences indicate still reasonable agreement. The other datasets show span-averaged relative differences between 2.9 and 10.8 % for sunrise comparisons, and between 2.9 and 5.4 % for sunset comparisons, showing that many of the datasets are in excellent agreement with MAESTRO UV-ozone in the stratosphere."
**To:** "Considering all of the datasets within the 20 to 45 km range, the majority of the comparisons have average absolute differences between 0.11 and 0.30 ppmv for the sunrise comparisons and between 0.17 and 0.34 ppmv for the sunset comparisons. The exceptions to this are the sunrise SAGE II comparisons which have an average absolute difference of 0.83 ppmv, and the sunset Odin-SMR comparisons with an average difference of 0.50 ppmv. These two datasets also show the highest relative differences, of 11.9 % and 7.9 % on average; however, these average differences indicate still reasonable agreement. The

other datasets show span-averaged relative differences between 2.9 and 10.8 % for sunrise comparisons, and between 2.9 and 5.4 % for sunset comparisons, showing that many of the datasets are in excellent agreement with MAESTRO UV-ozone in the stratosphere."

Changed L721:
**From:** "In the lower stratosphere, between 15 and 20 km, the MAESTRO UV-ozone comparisons show larger relative differences, with the average difference of most of the datasets falling between 5.0 and 33.0 %. The best mean agreement is noted for comparisons with SAGE II sunrise measurements, which have a relative difference of 5.0 %, and with SAGE III/M3M, which have a 3.7 (1.9) % relative difference during sunrise (sunset). The GOMOS sunset coincident measurements show the largest average relative difference of 55.0 %."
**To:** "In the lower stratosphere, between 15 and 20 km, the MAESTRO UV-ozone comparisons show larger relative differences, with the average difference of most of the datasets falling between 5.0 and 33.0 %. The best mean agreement is noted for comparisons with SAGE II and SAGE III/M3M measurements, which have a relative difference of 5.0 (7.3) % and 3.7 (1.9) % during sunrise (sunset) respectively. The GOMOS sunset coincident measurements show the largest average relative difference of 63.7 %."

Changed L735:
**From:** "The lowest average correlation coefficients in this range come from comparisons with SAGE II sunrise measurements, which have an average value of 0.54, with GOMOS sunset measurements, which have an average value of 0.68, and with Odin-SMR, which has an average coefficient of 0.66 (0.69) during sunrise (sunset). The remaining datasets show average correlation coefficients over this range between 0.73 and 0.79 for sunrise coincident measurements, and between 0.70 and 0.78 for sunset coincident measurements."
**To:** "The lowest average correlation coefficients in this range come from comparisons with SAGE II sunrise measurements, which have an average value of 0.54, with GOMOS sunset measurements, which have an average value of 0.67, and with Odin-SMR, which has an average coefficient of 0.66 (0.69) during sunrise (sunset). The remaining datasets show average correlation coefficients over this range between 0.72 and 0.79 for sunrise coincident measurements, and between 0.70 and 0.79 for sunset coincident measurements."

Changed L794:
**From:** "The largest differences that show this are found with OSIRIS, GOMOS, and SCIAMACHY"
**To:** "This low bias is most clearly illustrated in comparisons made with OSIRIS, GOMOS, and SCIAMACHY."

Changed L820:
**From:** "The first difference is in the magnitude of the mean sunset $NO_2$ profile; both the GOMOS sunset profile and the MAESTRO sunset profile show higher $NO_2$ concentrations around the stratospheric $NO_2$ maximum, by about 2 ppbv and 1ppbv respectively, as compared to the same type of profile from the ACE-FTS datasets. The second difference is in the location of the stratospheric $NO_2$ peak, with all four profiles in the GOMOS set of comparisons peaking approximately 2 km higher than those seen in the ACE-FTS comparisons."
**To:** "The first difference is in the magnitude of the mean sunset $NO_2$ profile; both the GOMOS sunset profile and the MAESTRO sunset profile show higher $NO_2$ concentrations around the stratospheric $NO_2$ maximum, by about 2 ppbv and 1.5 ppbv respectively, as compared to the same type of profile from the ACE-FTS datasets. The second difference is in the location of the stratospheric $NO_2$ peak, with the two GOMOS profiles peaking approximately 2 km higher than those seen in the ACE-FTS comparisons."

Changed L858:
**From:** "Between 20 km and 40 km, a more distinct low bias is noted across most of the datasets in comparison to MAESTRO."
**To:** "Between 20 km and 40 km, a more distinct low bias is noted for MAESTRO as compared to most of the comparison datasets."

Changed L868:
**From:** "The next closest agreement is found with ACE-FTS version 4.1/4.2 and version 5.2, which have sunrise (sunset) absolute differences of 0.26 (0.35) and 0.28 (0.38) ppbv respectively. These translates

into mean relative differences from the MAESTRO sunrise (sunset) measurements of 15.6 (14.4) % for ACE-FTS version 4.1/4.2 and 15.8 (14.4) % for ACE-FTS version 5.2."
**To:** "The next closest agreement is found with ACE-FTS version 4.1/4.2 and version 5.2, which have sunrise (sunset) absolute differences of 0.25 (0.35) and 0.28 (0.37) ppbv respectively. These translates into mean relative differences from the MAESTRO sunrise (sunset) measurements of 15.7 (14.3) % for ACE-FTS version 4.1/4.2 and 15.9 (14.4) % for ACE-FTS version 5.2."

Changed L875:
**From:** "Despite larger average absolute differences, ranging from 0.50 to 0.80 ppbv, the MIPAS sunset, OSIRIS sunrise, and SAGE II sunrise comparisons all show decent agreement, to within 20 %, with MAESTRO as well. For sunset-coincident MIPAS measurements, the average relative difference is 15.9 %, while for OSIRIS sunrise coincident measurements the average relative difference is 19.4 %, and for SAGE II sunrise coincident measurements it is 12.0 %. The OSIRIS sunset coincident measurements show a large absolute difference of 0.99 ppbv, but due to this set of comparisons possessing the highest average VMRs, this is only a 23.3 % difference from what is observed with the coincident MAESTRO sunset measurements. The remaining coincident comparisons all have larger absolute and relative differences, with mean relative differences in excess of 21 % over this 20 km span."
**To:** "Despite larger average absolute differences, ranging from 0.50 to 0.76 ppbv, the MIPAS sunset, OSIRIS sunrise, and SAGE II sunrise comparisons all show decent agreement, to within 19.2 %, with MAESTRO as well. For sunset-coincident MIPAS measurements, the average relative difference is 16.2 %, while for OSIRIS sunrise coincident measurements the average relative difference is 19.2 %, and for SAGE II sunrise coincident measurements it is 12.0 %. The OSIRIS sunset coincident measurements show the second largest absolute difference, of 1.02 ppbv, but due to this set of comparisons possessing the highest average VMRs, this is only a 23.6 % difference from what is observed with the coincident MAESTRO sunset measurements. The remaining coincident comparisons all have larger absolute and relative differences, with mean relative differences in excess of 20.9 % over this 20 km span."

Changed L884:
**From:** "In this range, the ACE-FTS sunset, OSIRIS, MIPAS sunset, SCIAMACHY, and SAGE II sunset comparisons are all found to yield agreement with MAESTRO to within 20 %; with the SAGE II sunset comparisons showing the best overall average agreement to within 4.6 %. However, many of the comparisons fare more poorly, with relative differences ranging from 24.2 % to 90.8 %."
**To:** "In this range, the ACE-FTS sunset, OSIRIS, MIPAS sunset, SCIAMACHY, and SAGE II sunset comparisons are all found to yield agreement with MAESTRO to within 20.4 %; with the SAGE II sunset comparisons showing the best overall average agreement to within 4.6 %. However, many of the comparisons fare more poorly, with relative differences ranging from 22.3 % to 90.8 %."

Changed L894:
**From:** "Within this span, the highest correlation is found with the two versions of ACE-FTS, which have a sunrise (sunset) measurement correlation of at least 0.88 (0.89). Over this range, the OSIRIS sunset, SCIAMACHY, MIPAS, and SAGE III/M3M comparisons all have an average correlation above 0.74. The OSIRIS sunrise comparison is somewhat worse overall, with an average correlation of 0.7, and GOMOS also has lower correlation, at 0.70 (0.60) for the sunrise (sunset) comparison, due to the poor correlation found between it and MAESTRO between 20 and 25 km."
**To:** "Within this span, the highest correlation is found with the two versions of ACE-FTS, which have a sunrise (sunset) measurement correlation of at least 0.87 (0.89). Over this range, the OSIRIS sunset, SCIAMACHY, MIPAS, and SAGE III/M3M comparisons all have an average correlation above 0.75. The OSIRIS sunrise comparison is somewhat worse overall, with an average correlation of 0.65, and GOMOS also has lower correlation, at 0.60 (0.59) for the sunrise (sunset) comparison, due to the poor correlation found between it and MAESTRO between 20 and 25 km."

Changed L920:
**From:** "The UV-ozone product from MAESTRO was found to agree slightly less well with the coincident datasets, with average relative differences between 20 and 45 km of 2.9 to 11.9 %."
**To:** "The UV-ozone product from MAESTRO was found to agree slightly less well with the coincident datasets, with average relative differences between 20 and 45 km of 2.8 to 11.9 %."

Changed L930:

**From:** “Finally, UV $NO_2$ from MAESTRO was found to agree with the majority of the comparison datasets to within 27.5 % between 20 and 40 km, with the relative differences ranging from 8.5 to 43.4 % on average over this span.”

**To:** “Finally, UV $NO_2$ from MAESTRO was found to agree with the majority of the comparison datasets to within 27.2 % between 20 and 40 km, with the relative differences ranging from 8.5 to 43.4 % on average over this span.”