Reply to reviewer #1

We thank anonymous reviewer #1 for his/her constructive review that would improve the contents of our paper. The review comments by anonymous reviewer #1 are numbered and repeated below as *in italic letters*, followed by our answers. In the new draft with corrections (supplement file), red, purple, and blue corrections are the revisions suggested by reviewers #1, #2, and #3, respectively. Yellow-marked sentences were also added in response to Short Comment #1 by Dr. Adrian Tuck.

GENERAL COMMENTS

(1) More context for the FTIR data should be given. A few questions come to mind.
- Is this the first time that the Syowa FTIR data has been retrieved/written up? Or are there other papers that describe this instrument and the data?

Yes, this is the first time that the Syowa FTIR data were retrieved and submitted to a journal paper.

(2) - Why is there only 2007 and 2011 data? Was the instrument not deployed in other years?

The FTIR was installed at Syowa Station in March 2007, and we made one year of operation in 2007. After that time, there were no one who could operate the FTIR at Syowa Station. In 2011, another FTIR operator wintered at Syowa Station, and made one year of FTIR operation. Then, after a few more operation in 2016, the FTIR was brought back to Japan in 2017. This sequence was now apparently described in Section 2.1.

(3) - Are the data publicly available?

Yes, we put the Syowa FTIR data in our data repository of our institute and put DOI information for them. The MIROC3.2 CCM results are stored at the CCMI site of BADC. This information was described in the "Data availability" now.

(4) - How to these data fit in with other ground-based FTIR datasets collected at high latitudes? Is the data quality similar? Have other FTIR instruments been used to study trace gases during spring-time ozone depletion?

As far as we know, there have been only two other FTIR instruments in Antarctica. One is at South Pole Station (90°S), and the other is at New Zealand's Scott Station in Antarctica. We added description related to South Pole FTIR in Section 2.1. The data quality should be similar. However, due to the location of Scott Station (78°S), they cannot measure solar infrared spectra until September when ozone hole is already in progress. I don't know any publications which focused spring-time ozone depletion using the other stations'

FTIR data so far. Therefore, the Syowa Station's FTIR has an advantage to study ozone-hole-related issues.

(5) The validation sections should be more specific in terms of how validation methods are applied (see specific comments for details). It would also be useful to answer:
- What is the expected uncertainty in the FTIR measurements? Is this something you retrieve?

We added typical retrieval error of each species in old Table 2 (new Table 1).

(6) - How do your validation results compare with validation of other ground-based FTIR instruments? Is the instrument at Syowa performing similarly to other FTIR?

By comparing our results with other FTIR validation results by Schneider et al. (2008), they look very similar. We described this issue in Section 3.

(7) - Are there any other factors that could affect comparison results? Do any of these species vary diurnally? Is there a chance that the satellite and station are measuring very difference air masses (e.g., inside/outside the vortex) for some coincidences? If either of these are factors, how does this affect interpretation of the data in Sect. 4?

Among the chemical species we analyzed, only ClO has diurnal variation. Therefore, we used only daytime ClO data for the analysis. Although we set relatively strict collocation criteria (within 300 km radius and +/- 6 hours) for validation, there is a chance that satellite and station are measuring rather different airmasses when Syowa Station was located near the polar vortex edge. In order to check whether such a situation occurs or not, we looked at equivalent latitudes (EL) of measurement locations of MLS and Syowa Station for the collocated coincidence pairs. The result is shown in Figure A (below). As a result, differences of EL are always within 10 degrees in all cases. We now picked up the collocation pairs whose difference in EL are within +/- 5 degrees. The results are shown in Figures B and C (below). As you can see in these figures, the comparison results show almost similar features. Therefore, we concluded that the difference in EL within 10 degrees does not affect the validation results.

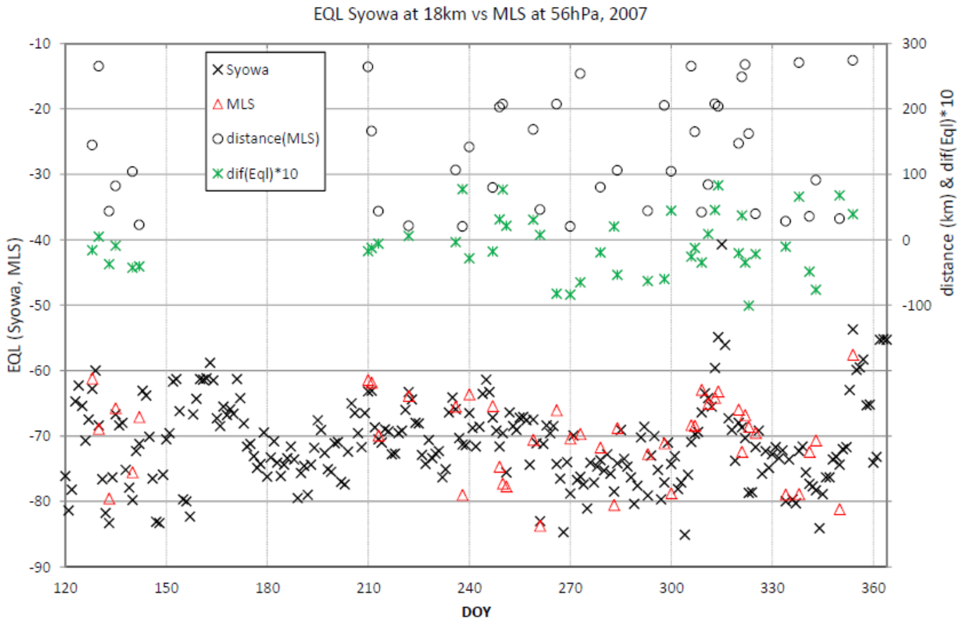

Figure A. Equivalent latitudes (EL) of Syowa Station (black crosses) at 18 km, EL of MLS (red triangle) at 56 hPa, distance between Syowa and MLS measurement locations (black circles), and differences in EL (green crosses) in 2007. Note that the values of differences in EL are multiplied by 10.

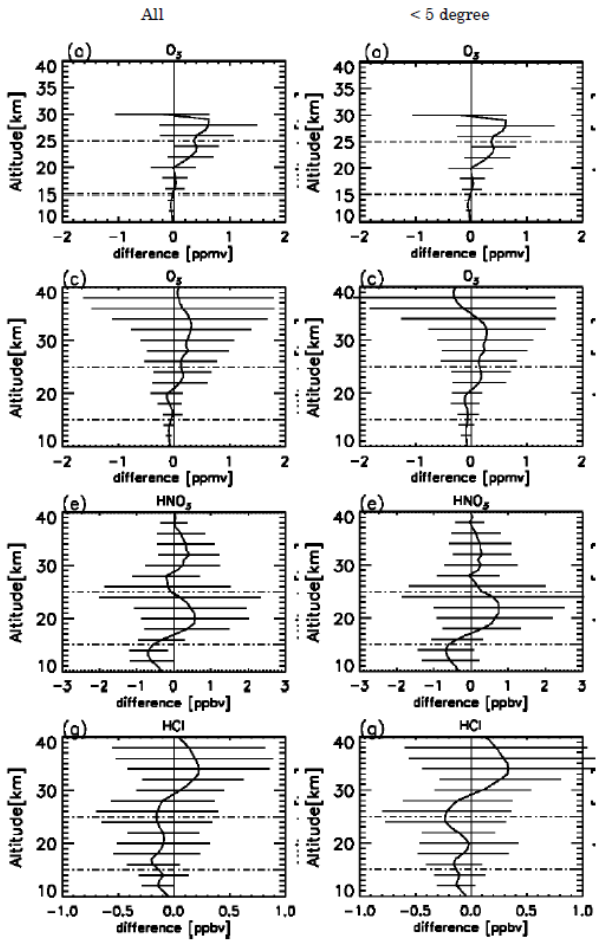

Figure B. Absolute differences with all collocation pairs (left columns) and those for differences in EL < 5 degrees (right columns).

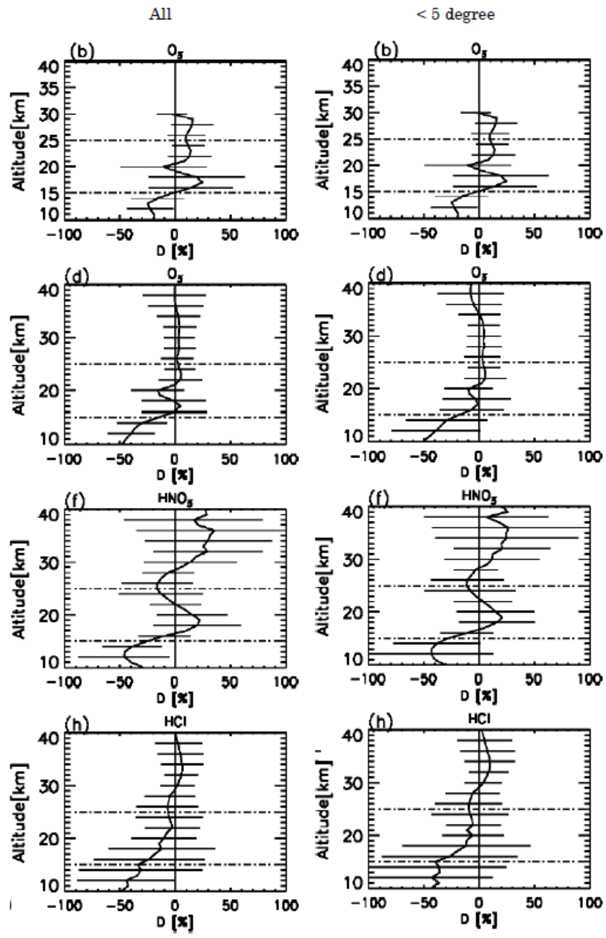

Figure C. Percentage differences with all collocation pairs (left columns) and those for differences in EL < 5 degrees (right columns).

(8) The discussion of results is repetitive and at times is difficult to follow. There are many timeseries figures (Fig. 4-9; Fig. 11-12). I found it hard to track through these and determine what exactly the authors were trying to highlight. I also had a hard time connecting all of these pieces to the concluding arguments around transport, chlorine deactivation, etc. A few possible solutions:
- Determine what you are trying to show, perhaps starting from the statements made in the conclusion. Only include figures/discussion that are relevant to what you're trying to show. Move additional timeseries figures into an appendix or supplement. Some tables (Table 1, Table 4) could also be moved out of the main body of the paper.

We modified several descriptions in Sections 4. Table 1 is now moved to Appendix, and Table 4 was deleted from the draft.

(9) - For Fig. 8-9 is there a more concise way of showing the relationships between the chlorine species than using the timeseries (e.g., through scatter plots or something else)?

We modified Figures 8 and 9. Now, partitioning of chlorine species are expressed by bar graphs instead of scatter plots in order to gain visibility. Also, descriptions of these figures were modified accordingly in Section 4.2.

(10) - Can the comparison with the modelling data be merged in with the validation of the FTIR (Sect. 3)? Can the modelling timeseries figures be folded into the timeseries figures presented in previous sections?

We showed modelling data in order to show other chlorine species which cannot be measured by FTIR and/or satellites, not for validating our measurements. Model data validation is out of the focus of this paper. Therefore, the modelling data section is kept as it is.

(11) I found it hard to follow the discussion around the mechanisms behind the decrease of HCl. It would be helpful if the discussion was expanded, to help answer the following:
- What hypotheses are there for the decrease in HCl in the literature? Is there just Solomon et al. (2015) and Grooss et al. (2018)? Are these hypotheses in conflict with each other or is it possible that they both contribute to the decrease in HCl together?

Also, there was several comments by reviewer-#3 on the decrease of HCl. The discussion on the decrease of HCl is much expanded. In the current modified manuscript, we showed two hypotheses in parallel; one is the $ClONO_2$ transport from the sunlit lower latitude towards the core of the vortex based on Solomon et al. (2015), and the other is the heterogeneous reaction between HCl and HOCl based on Müller et al. (2018). Please see the revised Section 4.6.

(12) - For the existing hypotheses, what data/evidence were used to develop the hypotheses? How does the data that you collected add to the existing supporting data/evidence?

The main data we used are based on our MIROC3.2 CCM results. However, our measurements of ClO, HCl, and $ClONO_2$ gains insights for these hypotheses.

(13) - Why does the sporadic increases in $ClONO_2$ in the model data support the transport mechanism? Do you have any other data to support that $ClONO_2$ is being transported (e.g., maps of $ClONO_2$, tracers showing transport patterns, evidence in the satellite data)? Are there alternative mechanisms that could explain your observations, such as chemistry?

Although there was no direct evidence of $ClONO_2$ transport, if $ClONO_2$-rich airmass was transported from the vortex boundary towards the vortex core region, it could decrease HCl by heterogeneous reaction (R1). Another alternative mechanism to decrease HCl was the heterogeneous reaction with HOCl (R4), which is now shown by another increase pattern on July 7 (day 188) and July 20 (day 201) in Figure 15(e).

(14) - Can your data be used to refute any other hypotheses for decreasing HCl?

Our data are consistent with both $ClONO_2$ transport hypothesis and reaction with HOCl hypothesis.

SPECIFIC COMMENTS:

(15) - Page 1, Line 17: You state "This was the first continuous measurements of chlorine species throughout the ozone hole period from the ground in Antarctica" here and elsewhere in the paper. This statement is a bit vague. Have other studies looked at any chlorine species from the ground in Antarctica? Are they looking at fewer species? Or for shorter time periods? It would be helpful if you included a literature review of ground-based measurements (FTIR and maybe other measurements of chlorine, such as OClO from UV-vis?) in Antarctica in your introduction.

We now described history of several ground-based observations except for FTIR (UV-visible and microwave) related to ozone chemistry in the last paragraph of Section 1, and the history of Antarctic ground-based FTIR measurements in the first paragraph of Section 2.1. Other FTIR stations in Antarctica can measure ozone-related species only after September, when ozone hole was at the recovery phase. This was now described in Section 2.1.

(16) - Page 4, Line 25: Are there other ground-based FTIRs in Antarctica?

Yes, there were two other ground-based FTIRs in Antarctica in history; one at South Pole Station, and the other at McMurdo/Scott Station, which is now described in Section 2.1. We think the South Pole FTIR is not in operation now.

(17) - Page 5, Line 1: What is the temporal resolution of the FTIR at Syowa? Does it only take measurements during sunny days? Does it require manual operation?

One measurement takes about 10 minutes and we can make observations only on sunny days. It requires manual operation. This was explained in Section 2.1 now.

(18) - Page 5, Line 18: Is there a way of using the averaging kernel to determine whether the 18 km and 22 km concentrations are independent from each other?

The concentrations of 18 km and 22 km are expected to be almost independent when you look at the averaging kernel in Figure 1 and the degree of freedoms (DOFS) in new Table 1.

(19) - Page 5, Line 25: State where various satellite datasets were obtained from? (E.g., URL?)

We now showed the URLs of satellite dataset pages.

(20) - Page 5, Line 24: The application of selection criteria should be clarified, perhaps with a full paragraph at the end of this section. Was this same criteria applied to all three satellite instruments (MLS, MIPAS, and CALIPSO)? Was the same criteria used for both the validation (Sect. 3) and the discussions of results over Syowa (Sect. 4)? E.g., for the timeseries figures was the 6 h criterion applied?

The selection criteria for MLS and MIPAS are already stated in the draft (within 300 km radius from Syowa Station and within +/-6 hours of the FTIR measurement). The same criteria were used both for the validation and the results. For the CALIPSO data, we selected the closest orbital data of the day, whose maximum distance to Syowa Station is 320 km. This is now stated in the draft.

(21) - Page 5, Line 30: Have you considered applying a selection criteria based on inside/outside the polar vortex to the satellite data along their line of sight instead of using values over the station for all measurements?

No. As is stated in (old draft's) page 6, line-16, we collocated the locations of tangent height of 20 km for the direction of the sun from Syowa Station at the time of the FTIR measurement, and the tangent point at 20 km for the satellites' data.

(22) - Page 6, Line 18: What is the shape of the 5 km-wide slit function? Is this based on the FTIR resolution? Why was this used instead of the averaging kernel? Also, why are the MLS data smoothed? Is the vertical

resolution for MLS much higher than the ground-based FTIR?

It is a 5 km-wide running mean. The description was reworded. 5 km is the typical vertical resolution of FTIR measurements. Since the vertical resolution of satellite measurements are a bit better than FTIR resolution (3-4 km for MLS, and 3 km for MIPAS), smoothing of ozonesondes and satellites' data were performed in order to merge the vertical resolution of each measurement.

(23) - Page 6, Line 25: State the expected uncertainty and any expected biases in ozonesonde measurements

Typical precision and accuracy of the ECC-type ozone sondes are considered to be ±(3-5)% and ±(4-5)%, respectively (Komhyr, 1986). This is now described in the second paragraph of Section 3.

(24) - Page 6, Line 24: I'm a bit confused about the language used throughout this section. The figures show mean relative difference and mean absolute difference versus altitude. Replace "The absolute difference. . ." with "The mean absolute difference. . ." and "The mean relative difference" in this line.

We now added the word "mean" at several places in Section 3. Also, "mean absolute difference" was used instead of "agreement" at several places for clarity.

(25) - Page 6, Line 25: What do you mean by "mean relative difference" here? Is this the mean relative differences averaged again over an altitude range? It might be simpler just to state the mean relative difference values at 18 km and 22 km instead of taking another average over altitude.

First, altitude range is different (15-25 km) and (18-22 km). Second, the "averaged" value in 18-22 km range was calculated. We rewrote "the average of mean relative differences ..." here.

(26) - Page 6, Line 26: What to you mean by "within error bars"? Are you referring to the standard deviation in the mean differences? This isn't really an error – it's the variability in the comparisons. It would make more sense to determine whether the agreement is as expected based on estimated uncertainty in the FTIR measurements (if this exists) and known uncertainties/biases in the satellite data. Also, why is the standard deviation used for the comparisons instead of the standard error?

The typical errors in FTIR measurements are shown in (new) Table 1 (old Table 2). We added some more official error values of satellite measurements in Section 3. We now used standard root mean square errors of FTIR and validation data instead of standard deviation, and rewrote to: "FTIR data agree with validation data within root mean squares of typical errors in FTIR and validation data at the altitude of interest." Also, the figure caption in Figure 2 had a mistake. We now rewrote to: "Horizontal bars indicate the root mean squares of differences at each altitude." Root mean squares of official errors in FTIR and satellite measurements are now added in (new) Table 2 (old Table 3).

(27) - Page 6, Line 31 – Page 7, Line 2: See comments for previous paragraph.

Same as above.

(28) - Page 7, Line 10: "with a precision of 0.2-0.6%" – is this the precision in the MLS measurements? Or is this the precision of the systematic bias?

It was a precision in the MLS measurements, but not 0.2-0.6%, but 0.2-0.6 ppbv, which corresponds to 10-30 % (See Table 3.9.1 in Livesey et al., 2013). We corrected the description here.

(29) - Page 8, Line 6: You describe a set of steps for detecting the inner and outer edges of the vortex. Is this a new method? Is this expected to work better than other established methods for some cases?

When we looked at the actual variation of the isentropic potential vorticity gradient with respect to equivalent latitude, we often saw apparent double-peak structure of polar vortex boundaries at 500-600 K in the case of Antarctic winter (see Figure 5 in Tomikawa et al. (2015)). Therefore, we decided to determine the inner and outer edges of the vortex when there were double peak structures. Actually, this method seems to work better with interpreting the observed data related to polar chlorine chemistry, rather than the conventional single vortex boundary categorization (e.g., Nash et al. 1996). This description was now added in the text.

(30) - Page 8, Lines 11-14: Please break this down into smaller steps – I have a hard time understand what was done. My best guess is that you calculated the isentropic potential vorticity gradient as a function of equivalent latitude. You found the local maxima. You defined the inner/outer edges of the vortex at the local maxima that both (?) exceeded the wind-speed threshold and were at least 5 in equivalent latitude apart? Are there always only two local maxima that meet these criteria?

Your understanding is correct. When there was only one local maxima, the regions are categorized into only two categories (inside and outside the polar vortex, i.e., no boundary region). In order to describe the definition of polar vortex boundaries in more detail, the polar vortex boundary definition part was now moved to Appendix B, and new Figures A1 and A2 were added.

(31) - Page 8, Line 18: Did you filter the MLS data according to an error threshold? Or did you just remove suspect data when you looked at the timeseries? All filters applied to the satellite data (for, e.g., uncertainty, etc) should be described explicitly in Sect. 2.2.

The lack of ClO and HCl data from day 195 to 219, 2007 is due to extremely large error values in the MLS data products (~1.2 ppbv error compared with ~0.2 ppbv error in other periods). The lack of $ClONO_2$ data from day 170 to 216, 2007 are negative values (-999.9) in MIPAS data products. This was now described in the text.

(32) - Page 11, Lines 10-24: Can the model comparisons be merged into Sect. 3? It would be nice if consistent comparison methods were applied. Similar to Sect. 3 – are the comparisons consistent with what is expected based on known model performance and known biases in the satellite instruments?

We showed modelling data in order to show other chlorine species which cannot be measured by FTIR and/or satellites, not for validating our measurements. Model data validation is out of the focus of this paper. Therefore, the modelling data section is kept as it is.

(33) - Page 11, Line 28: Why was a different method used to determine the polar vortex than at Syowa station?

We now unified the method of polar vortex definition at Syowa Station. The definition method of polar vortex boundaries is now moved to Appendix B. The description of polar vortex boundary in Section 4.5 was modified.

(34) - Page 11, Line 31: What do you mean by "This boundary was located in between the inner and the outer edge of the polar vortex as were defined in Sect. 4.1"? Did you compare the two definitions of the vortex? If so, was this just done at Syowa station for all data or a subset of data? Or is it the case that the max gradient of PV at 475 K is always between the inner/outer edge values by definition?

Now the definitions of polar vortex boundary were unified.

(35) - Table 4: Should describe how you came up with each of the various ranges presented in the table. Here are a few examples of questions that should be answered. How do you define the threshold for a ClONO2 enhancement or ozone starting-ending day of decrease? What are given in "HCl Value after increase" ranges? Is this the min/max of individual FTIR measurements over a given time-period? For the ClO enhancement period, is the 80% of maximum value different in 2007 and 2011 or is a single value used? What does "Variation when HCl 0 ppbv" mean?

We thought this table can be used by readers to see typical values of each species during the development and recovery of ozone holes in 2007 and 2011. However, it is rather arbitrary to pick up each value. Moreover, the contents of this table are not discussed in the paper. Therefore, we decided to omit this table from our paper.

(36) - Figure 3: Can you add a panel describing hours of sunlight or SZA here or to one of your other timeseries? It would be helpful to visualize this through discussions around available sunlight.

We added a panel for showing daytime hours in Figure 3(a).

(37) - Figure 3: Are the gaps in the PSC timeseries because CALIPSO observed no PSCs or because CALIPSO didn't collect any coincident data during this time? Either describe in text and/or add marker for CALIPSO measurements which did not observe PSCs.

The gaps in the PSC timeseries are the days when CALIPSO observed no PSCs over Syowa Station. We looked at the closest CALIPSO orbit to Syowa Station in each day. This daily sampling is now described in the last paragraph of Section 2.2.

(38) - Figures 14/15: Why have you used model data instead of satellite data for the parameters that are available from satellite? Did you check to see if the satellite saw the same patterns as the model?

The comparison between satellite observations and model results were shown in Figures 11 and 12, although comparison or validation of model results with actual measurements are not the main issue of this paper. In Figures 14 and 15, we focused on MIROC3.2 model results, in order to keep the consistency of each chlorine species.

MINOR/TECHNICAL COMMENTS:

(39) - Throughout the text, there are some awkwardly worded sentences and minor grammatical mistakes. I assume that these would be corrected through copy-editing, so have not listed these here, unless the meaning is unclear.

We think grammatical mistakes will be corrected through copy-editing.

(40) - Table 2: Define "PT"

We reworded to "Pressure and temperature profiles".

(41) - Table 3: Should define the fields included in the table. Is D(%) 18-22 km the average of the mean relative differences? Is (Min/Max %) 18-22 km the max/min mean relative difference across the various altitudes? (Might be better just to show mean relative differences +/- standard error at 18 km and at 22 km.)

Yes, your understanding is correct. We modified (new) Table 2 (old Table 3) to explain the validation results more comprehensively. The numbers in this table are now explained in the text in Section 3 as well.

(42) - Table 3: See comment re: precision in Page 7, Line 10

It was modified in the (new) Table 2 (old Table 3).

(43) - Table 3: Replace "Agreement" with "Range of mean absolute differences for 15-25 km"

It was replaced as suggested.

(44) - Figure 2, caption: Replace "Absolute (a) and percentage" with "Mean absolute (a) and mean percentage"

They were replaced as suggested.

(45) - Figure 2, caption: Replace "FTIR measurements and those from ozonesonde" with "FTIR measurements minus those from ozonesonde"

They were replaced as suggested.

(46) - Figure 2, caption: Move "Horizontal bars indicate the standard deviation of differences at each altitude." to the end of the caption, since it applies to all panels.

It was moved as suggested.

(47) - Figure 2, caption: Describe what the horizontal dashed lines indicate.

They indicate the altitude range of our focus (15-25 km). We also described it at the last paragraph of Section 2.1.

(48) - Figure 4, caption: Replace "from N_2O value" with "from Aura/MLS N_2O".

It was replaced as suggested.