# Response to reviewers: Stratospheric and mesospheric HO2 observations from the Aura Microwave Limb Sounder

December 16, 2014

We sincerely thank the reviewers for their thoughtful comments on the previous draft, we hope this new version is more suitable for publication.
In doing the corrections of all reviewers we added the following 3 major changes:
(1) A paragraph at the beginning of the results sections explains that the averaging kernels were applied to all comparisons:
In this section we compare the offline HO2 dataset with balloon-borne and other satellite measurements, as well as, with global climate and photochemical model simulations. In making these comparisons, i.e. when showing the absolute or percentage differences between the datasets, the MLS averaging kernels has been applied to properly compare them. Furthermore, when comparing the global climate or the photochemical model simulations, its high vertical resolution has been reduced to the MLS one using a least square fit as described by Livesey et al. (2011, Sect. 1.9). In these comparisons, no altitude extrapolation has been applied to any dataset.

(2) the discussion about the impact of the O2 and H2O cross section was deleted, the discussion about the mesospheric discrepancies now reads:
These discrepancies might be due to a variety of reasons, for example: (1) our understanding of middle atmospheric chemistry may not be complete, (2) there might be due to differences between recent solar spectral irradiance (SSI) satellite measurements (Snow et al., 2005; Harder, 2010) and most parameterizations. These SSI measurements display a larger variability in

solar UV irradiance which can-not be reconstructed with SSI models, including the model of Lean et al. (2005), used in this SD-WACCM run (Marsh et al., 2013). These SSI measurement-model differences have been proven to affect the HOx photochemistry (Haigh et al., 2010; Merkel et al., 2011; Ermolli et al., 2013); more UV irradiance leads to an enhancement of O3 photolysis as well as H2 O photodissociation, which leads to more HOx production through (Reactions R4 to R8). Further, Wang et al. (2013) showed that using a solar forcing derived from these SSI measurements the modeled OH variability agrees much better with observations. Lastly, (3) these discrepancies might be related to the WACCM representation of the mean meridional circulation which has been shown to have some deficiencies (Smith et al., 2011; Smith, 2012), suggesting that the gravity wave parametrization needs to be modified. In addition, Garcia et al. (2014) has shown that adjusting the Prandtl number, used to calculate the diffusivity due to gravity waves, significantly alters the CO2 SD-WACCM simulations improving its agreement with satellite measurements. Such adjustment should also affect the H2O and hence the HOx chemistry.

(3) the photochemical model discussions now reads:
As shown in Fig. 12, in the upper mesosphere (pressures smaller than 0.1 hPa), the Kinetics 1 simulations do not reproduce the magnitude of the measured peak, underestimating it by as much as 60%. On the other hand, Kinetics 2 shows an improvement in the modeling of this peak, reducing the underestimation to less than 40%. These discrepancies coincide with the ones discussed in the previous section strongly suggesting that they are related to the model assumptions rather than to measurement errors. As with the SD-WACCM simulations, several factors could be the reason for this discrepancy: it might be due to limitations in our current understanding of middle atmospheric chemistry and/or due to the deficiencies in the model solar spectral irradiance used, in this case Rottman (1982). Also, considering that Kinetics 2 (the run testing the HOx partitioning) represents the measured HO2 better, these simulations might suggest that, the modeling problems are related to the HOx production and loss balance rather than the HOx partitioning. In the upper stratosphere and lower mesosphere (between 1 and 0.1 hPa) for the most part the photochemical model underpredicts HO2 by around 20% concurring with the SD-WACCM simulations as well as with previous studies (Sandor et al., 1998; Khosravi et al., 2013) but contradicting the result of the study by Canty et al. (2006).

Below are our responses to the reviewers comments in red.

**Reviewer 1**

General comments:

This manuscript presents algorithm and results for a new offline HO2 retrieval from MLS/Aura limb observations. The main difference between this new retrieval and the MLS standard retrieval is that the retrieval is not applied to individual MLS limb measurements, but to zonally averaged data. The resulting noise reduction allows covering a significantly enhanced altitude range and a wider latitude range. I find the paper in general well written and relatively easy to follow. The paper is in my opinion suited for publication in ACP, but I ask the authors to consider the comments given below. My main criticism concerns two aspects:

a) The model-measurement comparison presented in section 4.4 does not really allow any conclusions to be drawn, as far as I can tell. Therefore, one may question the necessity of this section. In my opinion the section should at least be improved to better describe the assumptions made for the 2 model scenarios (see specific comment below) and by adding a more detailed discussion of the implications of the comparisons performed.

This section reinforces the results shown in the previous section (the WACCM section) as well as suggest further that the model discrepancies are real model errors rather than measurement errors. This has been stated in the new manuscript. See below.

b) Figure 6 shows a comparison between MLS offline and FIRS-2 balloon HO2 observations for the 1 to 10 hPa pressure range. The MLS profile is a daytime/nighttime average, despite the fact that earlier in the paper it was stated that the retrievals between 1 and 10 hPa are affected by systematic biases. For this reason the daytime measurements presented for this pressure range are differences between daytime and nighttime measurements. If there are known biases, a comparison of daytime/nighttime averages to FIRS data does not appear to be a valid comparison.

See below

Specific comments:

Page 22907, line 10: ”in the Lyman-alpha and the Schumann-Runge bands This phrase implies that the Lyman-alpha signature is also a band, which is not the case.
changed to: in the Lyman–Alpha region and the Schumann–Runge bands

Page 22907, line 22: a problem known as the HOx dilemma From the following description of differences between observations and model simulations its not fully clear what the HOx dilemma is. Is it the low bias of OH measurements compared to model simulations reported by Summers et al. (1997) or is it the general disagreement between models and measurements, with the latter being sometimes higher and sometimes lower compared to the models?
The text was rearranged to make it clear that it was the general disagreement between models and measurements. See text

Page 22908, line 2: lower that the values – lower than the values
Done

Page 22908, line 11: Furthermore, models have consistently under-predicted the amounts of O3 at such altitudes, an issue known as the O3 deficit problem Some of the references cited are 3 decades old. The most recent one is already 10 years old. Im wondering, whether more recent studies find better agreement between modelled and observed O3?
I added the Siskind (2013) - Comparison of a photochemical model with observations of mesospheric hydroxyl and ozone, reference that still shows the deficit.

Page 22909, line 17: It covers between 82S and 82N - It covers latitudes between 82S and 82N
Done

Page 22910, line 13: in (Livesey et al., 2006) - in Livesey et al. (2006)
Done

Page 22910, line 19: indecipherable - indistinguishable ?
Done

Page 22910, line 21: with a 10deg latitude typical precision Im not entirely

sure what you mean here. Probably the typical precision for measurements zonally averaged and binned in 10 deg latitude bins? I suggest stating this more explicitly.
Done, changed to: typical precisions varying from 0.15 ppbv ($5210^6$ cm$^{-3}$ ) at 10 hPa to 3 ppbv ($510^6$ cm$^{-3}$ ) at 0.046 hPa for measurements zonally averaged and binned in a $10^o$ latitude bin.

Page 22911, line 20 with day-night differences used as a measure of daytime HO2 for pressures between 10 and 1 hPa where the nighttime values exhibit non-zero values indicative of biases. Im wondering, whether this special treatment between 1 and 10 hPa leads to discontinuities at the 1hPa level? It would be good to provide a quantitative estimate on the jump or discontinuity at 1hPa or an upper threshold. In Figure 2 such a discontinuity is not visible, but this may just be because of the finite width of the vmr bins.
We added in the offline retrieval section: Note that, a visual inspection of the 10o bin monthly average profiles have shown, overall, no sings of a discontinuity at 1 hPa when using this approach.

Page 22913, line 6: this retrieval Suggest to replace this by the retrieval presented in this study, to avoid confusion with the standard retrieval, which is also mentioned in the previous sentence.
Done

Page 22913, line 22: truth model atmosphere - true model atmosphere ?
Done

Page 22914, line 13: The MLS HO2 profiles is a 20 deg latitude bin This statement is certainly not correct, a profile is not a latitude bin. Suggest replacing by, e.g.: The MLS HO2 profile corresponds to a 20 deg latitude bin
Done

Page 22916, line 4: within half an hour of the MLS measurements You mean half an hour in terms of local time, not UT, right? I suggest mentioning this explicitly.
Changed to: Note that only SMILES measurements made within half an hour local time of the MLS measurements were used in this comparison.

Page 22917, line 5: a strong zonal latitudinal gradient I dont quite understand what you mean by zonal latitudinal gradient. Please clarify.
Deleted: strong zonal latitudinal

Page 22917, same sentence: gradient from the summer pole towards the winter pole The gradient generally points from low values to high values, i.e. if you speak of the gradients direction (and not just the fact that there is a gradient), theres a gradient from the winter (NH) pole to the summer (SH) pole and not vice versa. I suggest omitting the statement on the direction of the gradient and just state that there is a gradient.
Changed to: As can be seen, both display similar VMR structures with a gradient from the winter pole towards the summer pole.

Page 22917, line 11: zonal latitudinal gradient Please change (see comment above) Deleted: strong zonal latitudinal

Page 22917, line 20: eg. - e.g.
Done

Page 22918, line 23: as well as a strong zonal latitudinal gradient from the summer to the winter pole' See comments above
Changed to: as well as a gradient from the winter to the summer pole.

Page 22918, line 26: the H available is the one generated at sunlit latitudes, transported at high altitudes poleward, where it descends and reacts with O2 at night. I suggestion mentioning explicitly that this applies to the winter, i.e., the northern hemisphere in this case.
The sentence now reads: however in this case, due to the lack of photodissociation of H2O, the H available is the one generated at sunlit latitudes (in this case, in the northern hemisphere), transported at high altitudes towards the winter pole where it descends and reacts with O2 at night (in this case, in the southern hemisphere) (Pickett et al., 2006)

Page 22918, line 27: Pickett et al. (2006) - (Pickett et al., 2006)
Done

Page 22919, line 17: which adds a constraint to MLS OH to mostly .. dont understand what this means. What kind of constraint is that. Does it

simply mean that you use MLS OH profiles? Please clarify.
expanded to: in addition, constrains the model using MLS OH measurements

Section 4.4: I think this section is the weakest part of the paper, because the implications of the model-measurement comparisons are not clear. If there are no conclusions to be drawn from this comparison, one may question, why section 4.4 is necessary at all. If this section remains in the paper, the implications of the differences between the model runs need to be explained better and in more detail, in my opinion. What exactly do we learn from the fact that the agreement to measured HO2 is improved if OH is also taken from MLS measurements?
We learn that the error is more related to the HOx production than the HOx partitioning as said in the text. Also, we added when talking about the mesospheric differences: These discrepancies coincide with the ones discussed in the previous section (WACCM section) strongly suggesting that they are related to the model assumptions rather than to measurement errors.

Can robust conclusions be drawn if the uncertainties of the MLS data products are considered?
The errors are plotted in the figure and the differences are bigger.

Page 22920, line 12: These offline HO2 has - This offline HO2 dataset has
Done

Page 22920, line 19: from 10 to 0.0032 - from 10 to 0.0032 hPa Done

Next line: from 1 to 0.0032 - from 0 to 0.0032 hPa
Done

Page 22921, line 3: in the low side - on the low side ?
Done

Page 22921, line 11: as much as 60% but probably - as much as 60%, which is probably ?
Done

Figure 4, caption, line 1/2: daily, weekly .. yearly 10 deg latitude bin I find this phrase odd, because the latitude bin is not a daily, weekly etc.

latitude bin
Changed to: Expected precision for a daily (D), weekly (W), monthly (M) and yearly (Y) MLS HO2 offline data averaged over a 10 latitude bin.

Figure 5, caption, line 3: Magenta lines There are no magenta lines on my screen (nor on the printout). To me it looks more like violet.
Changed to: Purple lines

Figure 6, caption, line 2: The MLS data correspond to the daytime-nighttime average of the 15 to the 25 September .. Earlier in the paper you wrote about possible biases affecting both daytime and night time measurements at altitudes below the 1 hPa level. Because of the bias you reported differences between daytime and nighttime HO2 between 1 and 10 hPa. For the FIRS-1 comparison you use the daytime-nighttime average, which leads to the conclusion that the comparison shown in Fig. 6 is not a valid comparison. Can you use FIRS daytime measurements only to compare to your bias-corrected daytime measurements? This issue needs to be addressed.
We did use FIRS daytime data only as explained in the text. We also added in the caption: The FIRS profile corresponds to the one with the closest SZA to the MLS (daytime only) data.
Even though this is not the best comparison, since we are taking the day-night difference and an averaging over a latitude bin to compare against a single profile; this is the best we can do with noisy products. See for example, Kovalenko (2007) Validation of Aura Microwave Limb Sounder BrO observations in the stratosphere or Stachnik (2013) Stratospheric BrO abundance measured by a balloon-borne submillimeterwave radiometer or Pickett (2008) Validation of Aura Microwave Limb Sounder OH and HO2 measurements

Figure 6, caption, line 5: Suggest changing The differences shown are to The differences shown in the bottom panels are
Done

Figure 8, caption, line 2: shown on April - shown for April
Done

Figure 10: this figure also shows daytime data, right? This should be mentioned in the caption.
Done

Figure 12: the dashed lines in row 2 and 3 are barely visible.
Changed to a darker gray

Figure 12, caption, line 5: has been use to - has been used to
Done

Baron et al. (2009) reference: All last names end with a k. Theres something wrong. Also the second Urban should be Murtagh, right?
Corrected

Kikuchi et al. reference, line 3: Is Susukik correct? This should probably read Suzuki?
Corrected

Snow et al. (2005) reference: Mcclintock - McClintock
Corrected

**Reviewer 2**

The manuscript Stratospheric and Mesospheric HO2 Observations from the Aura Microwave Limb Sounder by Millán et al. presents new HO2 data derived from the AURA/MLS. Compared to the standard MLS version, these new data provide significant improvements with in particular information at high altitudes (up to 0.003 hPa) covering the mesospheric peak of HO2. These data are important to address opened issues such as the underestimation by the models of middle atmospheric O3 and day-time HO2. The error analysis and the comparison with two other instruments (FIRS-2 and JEM/SMILES) demonstrate the good quality of the dataset. Also the comparisons with WACCM (3-D climate model) and the Caltech/JPL 1-D chemical model confirms the problems to reproduce HO2 mesospheric abundance with current chemistry models. They also discuss the origin of this problem and conclude that the issue is related to the source and sink of HOx and not its partitioning. I think the manuscript can be published in ACP but before I would like the authors to check an issue related to Figure 12:

For me, the kinetic-2 and MLS profiles (upper panels) are in very good agreement at about 0.02 hPa (HO2 peak). The agreement is much better

than the value of 20% seen in the lower subplots. Also even considering that the upper panel profiles are not smoothed (is-it correct ?), I dont understand why the absolute and relatives differences show that the maximum difference is around 0.02 hPa. I may have misunderstood something, please check.
That's correct, the difference is due to the smoothing. We added the following sentence at the beginning of the results comparisons to make it clearer: In this section we compare the offline HO2 dataset with balloon-borne and other satellite measurements, as well as, with global climate and photochemical model simulations. In making these comparisons, i.e. when showing the absolute or percentage differences between the datasets, the MLS averaging kernels has been applied to properly compare them. Furthermore, when comparing the global climate or the photochemical model simulations, its high vertical resolution has been reduced to the MLS one using a least square fit as described by Livesey et al. (2011, Sect. 1.9). In these comparisons, no altitude extrapolation has been applied to any dataset.

I have also some minor comments that I have listed below:

P22909, Line 5: Are these data currently publicly available? Will they become part of the standard MLS dataset?
In the introduction we added this sentence: To date, this dataset provides ten years of data and, in the near future, it will be publicly available for download in a daily based hierarchical data format (HDF).

P 22911, Line 6: Are the selected radiances include bands 28 and 30? Are you including other bands? Have you compared only band 28 vs bands 28+30 ? Using band 30 should increase the contamination from the O3 line and, hence, increases the sensitivity of the retrieval to uncertainties on the O3 VMR, temperature and spectroscopy.
We added: The best retrievals were found when doing a jointly band 28 and 30 retrieval as oppose to doing retrievals using only band 28 or only band 30, even despite the O3 line influencing band 30.

Are these errors taken into account in the measurement covariance matrix? A comment about this issue could be added in the paper.
In the Error assessment section, we added in the systematic uncertainties list: the contaminant species errors, such as the O3 line influencing band 30 ...

P 22912, Line 15: I understand that only HO2 VMR is retrieved and other relevant atmospheric parameters (e.g., temperature, O3) are fixed (based on the standard MLS products). Is-it correct ? Are any other parameters retrieved to correct instrumental baseline ?
We added: Furthermore, at each pressure level, a constant baseline is retrieved for each band to correct any instrument baselines as well as to take care of the water vapor continuum contribution.

P 22915, Line 5: Would it be possible to provide the order of magnitude of the differences between a single profile at the FIRS-2 position and the zonal mean profile? For instance, the authors could use a model like WACMM or the MLS water vapor profiles a proxy.
Done: Overall, the two instruments agree on the HO2 vertical structure, with increasing HO2 with height (in the VMR representation) however there seems to be an bias between them, with the MLS data in the lower bound. This might be due to the differences between the FIRS-2 single profile and the MLS zonal mean profile. In SD-WACCM (see Sect. 4.3 for the model description), these differences (i.e. comparing a single profile versus a zonal mean profile at this location) are around 30%.

The smoothing of the FIRS-2 profile and the day-night difference MLS profile indicated in the caption of Fig. 5 could also be indicated in the text.
We added a paragraph at the beginning of the result section stating that all the comparisons used the MLS averaging kernels.

The night-time MLS profile could also be plotted in Fig. 5.
After careful consideration, we decided not to add the night MLS profile to avoid to confusion since this is a daytime comparison.

What is the highest altitude of the FIRS-2 profile (before smoothing)? Is the FIRS-2 profile extrapolated for the interpolation at the higher altitudes?
In the paragraph at the beginning of the result section we also stated that there are no altitude interpolation.

P 22915, Line 7: Is-it the monthly mean of the differences (SMILES-MLS) or the difference of the monthly mean profiles?
We added: Figures 7 and 8 show daytime and nighttime comparisons (monthly

means and the differences of the monthly means), respectively.

P 22916, Line 8: How the SMILES profiles have been smoothed in the upper altitude range of the retrieval (need of altitude extrapolation) ?
In the paragraph at the beginning of the result section we also stated that there are no altitude interpolation.

It should be indicated that unlike MLS, SMILES data are not regularly distributed over a month. This could explain some of the differences seen between SMILES and MLS in the mesosphere since the HO2 mesospheric peak shows large month to month variability.
We stated: The retrieval top level differences will need to be explored further, to investigate if they are due to retrieval artifacts (both retrievals are more sensitive to the apriori at these levels), calibration uncertainties or sampling differences (unlike MLS, SMILES data are not regularly distributed); this will require a joint effort from the MLS and SMILES teams.

P 22918, Line 14. It should be also mentioned that the underestimation of daytime HO2 in the model is seen above 1 hPa at all latitudes. This is consistent with previous studies given in the introduction of the paper (Sandor et al., 1998 and Khosravi, 2013).
We added: Figure 9 also shows that SD-WACCM underpredicts HO2 by about 20–30% between 1 and 0.1 hPa. This result agrees with previous studies (Sandor et al., 1998; Khosravi et al., 2013) but contradicts the result of the study by Canty et al. (2006).

I havent seen any comment on the large overestimation of the model near 0.2 hPa.
The apparent model overestimation near 0.2 hPa is probably related to the MLS change in vertical scan pattern near this pressure level.

P 22918, Line 15. I would rather consider the range 10 to 1 hPa since a quasi-systematic underestimation of the WACCM HO2 occurs above 1 hPa.
See P 22918, Line 14 response

P22919, Line 12. Does Fig. 12 shows the daytime MLS profiles? This information should be indicated in the caption or in the text.
Done

The size of the lowest subplots could be increased.

In the final paper, there will be bigger. No ACPD banner and one column (instead of two column) figures.

What is the meaning of the dashed lines?

In the caption we added: The dashed gray lines show the MLS precision as well as the 20 and 40% percentage regions

P22919. It would like to have a short comparison with the conclusions from previous studies. Are these results in contradiction with the study by Canty et al. who used the standard MLS data in the lower-mesosphere ?

See P 22918, Line 14 response

**Reviewer 3**

Review of manuscript "Stratospheric and mesospheric HO2 observations from the Aura Microwave Limb Sounder" by Millán et al.

General comments:

This work presents a new HO2 dataset derived from Aura MLS measurements using an offline retrieval algorithm. The product, retrieved from averaged radiance profiles, presents several advantages over the standard v3.3 product, as the extended altitude range, the coverage of the polar regions, and also provides nighttime values for a wide altitude region. The manuscript describes the algorithm and the characterization of the retrieved quantity and assess the different error sources.

Comparisons with balloon-borne measurements and satellite measurements are also presented as well as a comparison with a 3D chemistry climate model and a 1-D photochemical model. It is claimed that this dataset can be useful for a better understand- ing of the mesospheric O3 and HOx chemistry. In particular they found that the HOx partitioning in the retrieved HO2 and OH from MLS are compatible with our current understanding of the mesospheric chemistry. However, the absolute values of mesospheric HO2 are significantly underestimated by the models. Possible reasons for this underestimation are mentioned/listed although not really discussed or addressed.

I think that this new dataset of HO2 measurements add significant extra information to the standard product (e.g Fig. 2) and hence worth to be published. The result on the HOx partitioning of MLS products is also a significant contribution from the scientific (not only methodological) point of view. The other scientific result is just to point out to a models/MLS measurements disagreement which is not addressed. Then, it is not clear for me if the paper should be published in AMT or in ACP. I suggest that the authors give some more details and discussions on the possible causes of the disagreement (see below). This would make easier its publication in ACP.

Major comments:

Page 22913. Lines 21-24. I do not understand the meaning of the "retrieval numerics" error. My first guess would be that they are the "forward model" error, but this is considered in a separate contribution. Would that be what is normally called "smoothing" error? I.e., the effects of the regularization used in the retrieval? In the sentence "It is calculated as the retrieved value from the unperturbed radiances and the truth model atmosphere, i.e. that used for computing the synthetic radiance.", was that retrieval done with or without adding the noise to the synthetic radiance? I think it is important to clarify this error, since it is the major uncertainty in the region of HO2 maximum, above 0.1 hPa (Fig. 5).
We modified the sentence to state: The comparison between the unperturbed noise-free radiances run, and the true model atmosphere estimates the errors due to the retrieval numerics, which, in other words, is a measure of error due to the retrieval formulation itself, in this case, mostly an smoothing error.

Related to this point, if they are actually the "smoothing" errors, they would be already taken into account when applying the AKs to the data to be compared and hence, the "bias" would not be as large as the 1 ppbv shown in Fig. 5 but significantly smaller. If this interpretation is correct, I would not consider this error as a "bias" and would not mix with the other systematic (bias) errors.
In this case, we do not apply the kernels to the truth profile, to see the effects of the retrieval (the smoothing).

Related to this point, what is the "scatter" of the errors? What do they indicate?

The standard deviation, the title of the plot was changed from Scatter to Standard Deviation

In Secs. 4.3 and 4.4 the authors mention possible reasons for the discrepancies between the WACCM and 1-D models and MLS HO2 measurements. In particular they refer to our current uncertainty on the knowledge of the solar spectral irradiance measurements and/or its model representations, and the spectral resolution of the absorption cross sections of H2O and O2. Could the authors give some more details on what use the two models for these quantities? Do they have some hints on why they think they are possible causes or is it just speculation?
After careful consideration the discussion about the spectral resolution of the absorption cross sections was deleted because it was based plainly in the photochemical model representation of these values but we didnt change the resolution to corroborate the hypothesis. With respect to the solar spectral irradiance the discussion was expanded to: These discrepancies might be due to a variety of reasons, for example: (1) our understanding of middle atmospheric chemistry may not be complete, (2) there might be due to differences between recent solar spectral irradiance (SSI) satellite measurements (Snow et al., 2005; Harder, 2010) and most parameterizations. These SSI measurements display a larger variability in solar UV irradiance which cannot be reconstructed with SSI models, including the model of Lean et al. (2005), used in this SD-WACCM run (Marsh et al., 2013). These SSI measurement-model differences have been proven to affect the HOx photochemistry (Haigh et al., 2010; Merkel et al., 2011; Ermolli et al., 2013); more UV irradiance leads to an enhancement of O3 photolysis as well as H2 O photodissociation, which leads to more HOx production through (Reactions R4 to R8). Further, Wang et al. (2013) showed that using a solar forcing derived from these SSI measurements the modeled OH variability agrees much better with observations.

Other comments.
- Figures are very small and they have so many panels that are hardly readable in the printed version (I could read them only when zoomed out on the screen).
Figures 5,7,8 and 12 should be pagewidth in the final publication aiding the readability.

In this sense, most of the figures are duplicated presenting the results in vmr and in number density. I cannot see any advantage of presenting additionally the number density figures. I think they could be removed and would help to make the other panels more readable.
After careful consideration we decided to leave the duplication. Eventhough most people are familiar with the VMR unit, in the OH and HO2 community most papers (Pickett 2006,2008, Canty 2006, and Wang 2013) use number density units.

- Fig. 1 and Page 22910 (lines 8 and 9). The text refers to 1K, 2K and 4K limb radiance precision. Is any of these that shown as the noise in Fig. 1? Why do you compare between these three precisions to say that the noise is large and averaging is needed? The signal at the top panel (band 28) for 4.6 hPa is much smaller (particularly at night) than the noise. However it looks as not affected by noise (very smoothed). Is it because the number of measurements averaged is very large? Would be useful to mention that number in the figure caption.
The text was changed to The 1 K HO2 signal is relatively small compared to the individual limb radiance precision which varies from 2 K at the bands edges to 4 K at the band center (gray dotted line), hence ...
The caption does state that this is a monthly radiance average.

Page 22910. Lines 25 and ff. Just for curiosity, are the non-zero nighttime abundances positive, negative, both?
Both, thats why we didnt specify the sing.

You suggest to take the nighttime values as the "zero" for calculating the daytime values. However, the daytime and nighttime measurements are taken on different parts of the orbit (either ascending or descending). For other instruments the offset changes significantly along the orbit. Is that a good approach for MLS or is the uncertainty in the correction of a similar magnitude that the correction itself?
To imply that this is a valid approach for MLS we added: In addition to the MLS HO2 product, this daynight difference approach to ameliorate biases has been used succesfully for the BrO and OH MLS products. (Livesey et al., 2006b; Pickett et al., 2008; Millan et al., 2012)

Sec. 3. First full par. To be safer, I would consider as the daytime scans

those with SZA< 85. Would that make a significant change in the polar regions?
Before setting in the 90-100 SZA we did check other options but they did not make much difference.

Near the end of this par., lines 10-13. "interpolated radiances". Apparently the sampling in altitude of MLS is ¡1 km, and the vertical grid used here is 3km. Hence, it is also done some kind of "averaging" rather than "interpolation" in the radiances. Isnt it?
Correct, we changed it to averaged

Lines 19-23. You mention here that "... for pressures between 10 and 1 hPa where the nighttime values exhibit non-zero values indicative of biases." However, Fig. 5 shows that the biases are not particularly large at those pressure levels; actually they are larger at lower pressures (higher altitudes). Shouldnt daytime values be calculated in a similar way above around 0.1 hPa, where the bias is also large?
At 0.1 hPa the night values are expected to be non-zero hence not usable for bias correction. Below 1 hPa they are expected to be zero and hence any non-zero value is a sing of an artifact

Fig. 3. Are the results shown here for a daytime case? Please, state that, if so.
Done

Fig. 4. I would remove the number density plot and would use a log scale for the errors. Log scale Done. About the log scale for the errors, we prefer the absolute to easily emphasize how the increase with height.

The caption refers to a "This profile". Is it the solid black line?
The caption was change to: The black lines show typical HO2 profiles, daytime in solid and nighttime dashed. These profiles are a yearly average over all latitudes of the SD-WACCM model

Last par. in Sec. 3.2, lines 6-8. "For pressures smaller than 0.1 hPa, the main source of bias and scatter are retrieval numerics, which, although unsatisfactory, is understandable given the 14 km vertical resolution in this region." This suggests to me that you are talking about a "smooth" error

(see above). Correct?
Correct

Fig. 5 caption. families of systematic errors - sources(?) of systematic errors
Done

As before, I suggest to remove the panels with the errors in the density. Idem for Fig. 6.
See above

Sec. 4.1 Comparisons with FIRS-2. How many FIRS-2 profiles are available for that day? Just that used? If there are more but taken at other SZAs, and if SZA is very important, they could be corrected with a photochemical model. I think the statistics should be increased.
We carefully thought about this. There are more than one but adding the photochemical correction will also add an extra uncertainty and so, we decided against it.

BTW, in the figure caption is not mention that it is just one FIRS-2 profile.
We added in the caption: The FIRS profile corresponds to the one with the closest SZA to the MLS (daytime only) data.

Page 22916, lines 7-8, "The retrieval top level differences will need to be explored further,..." Given that there are so few HO2 measurements, and the importance of these measurements for the mesospheric chemistry (next sections), should not this be explored further in this work? It is important to clearly state that the models/MLS measurements comparison in the next section is not caused by a bias in MLS HO2 data.
To find the cause we will need a joint effort between the MLS and SMILES teams outside the scope of this study. We added: The retrieval top level differences will need to be explored further, to investigate if they are due to retrieval artifacts (both retrievals are more sensitive to the apriori at these levels), calibration uncertainties or sampling differences (unlike MLS, SMILES data are not regularly distributed); this will require a joint effort from the MLS and SMILES teams.

Sec. 4.3. It is known that WACCM does not reproduce very well the measured temperature and O3 fields and even the meridional circulation (e.g. Smith, 2012; Smith et al., 2011; 2013). Could these be possible reasons to explain the HO2 WACCM-MLS differences? Furthermore, Garcia et al. (2014) has found that the parameterization of the gravity waves (GW), done through the change of the Prandtl number, significantly changes the CO distribution in the upper mesosphere. This might also impact H2O and hence HO2. Has this been explored?
We added the following discussion: Lastly, (3) these discrepancies might be related to the WACCM representation of the mean meridional circulation which has been shown to have some deficiencies (Smith et al., 2011; Smith, 2012), suggesting that the gravity wave parametrization needs to be modified. In addition, Garcia et al. (2014) has shown that adjusting the Prandtl number, used to calculate the diffusivity due to gravity waves, significantly alters the CO2 SD-WACCM simulations improving its agreement with satellite measurements. Such adjustment should also affect the H2 O and hence the HOx chemistry.

In connection with this and the possible reason mentioned in the manuscript about possible inaccuracies in the representation of the absorption cross sections of H2O and O2 around the LymanAlpha region and the SchumannRunge bands, Garcia et al. (2014) has found that an overestimation of the O2 cross-section in the 105121 nm wavelength range was causing a too low CO concentration in the upper mesosphere. The large O2 cross-section assumed in the standard WACCM absorbed the UV radiation at high altitudes, preventing its penetration into lower altitudes and hence the CO production from CO2 photolysis. Although this spectral range is just at the edge of the Lyman-alpha, which affects H2O, this might be a reason for the WACCM/MLS discrepancy. With the reduced O2 cross-section, radiation will penetrates deeper, H2O will be more strongly photodissociated and hence producing more OH and more HO2. It might worth to explore this point.
We decided to leave this discussion out of the study, because it was based plainly in the photochemical model representation of these values but we didnt change the resolution to corroborate the hypothesis.

Page 22917, par. at lines 17-21. Since the feature discussed in not shown in the presented figures I cannot see the reason for its discussion. I suggest

to remove it.
It is there, in the number density subplots, the text was changed to: In Fig. 9 in the number density subplots, between 10 and 0.1 hPa, both the offline MLS dataset and the SD-WACCM simulations behave in a similar manner both in structure and in magnitude; however, due to the small HO2 signal in the MLS radiances, the offline MLS retrieval is noisier.

Page 22918, lines 10-15. It would be useful to mention which solar flux data is used in WACCM and how other data would change (at least qualitatively) the results. The same applies to the 1-D model described in Sec. 4.4 and it is extensive to the parameterization of the cross-sections (see major comment above).
Done, see paragraph added above

Page 22918, lines 18-20. "For pressure levels smaller than 0.1 hPa, the lack of a clear second peak in the SD-WACCM dataset reflects the smaller mesospheric concentrations in 20 this dataset." To which "second peak" do you refer? I cannot see it (just see the peak at 0.02 hPa).
The text was changed to: The lack of a clear second peak at 0.02 hPa in the SD-WACCM dataset reflects the smaller mesospheric concentrations in this dataset.

Page 22918, line 27. "poleward" - "towards the winter pole"?
Done

Page 22919, lines 1-2. "and overestimating by as much as 50% over the polar winter regions." This might be right but I would not conclude that when comparing second (WACCM with AKs) and third (MLS) panels in the left column of Fig. 11. They both appear with the same light green color. BTW, in the percentage differences panels, which WACCM is being used, with or without the applied AKs? This comment is extensive to all figures where the AKs are applied (e.g. Figs. 7, 8, 9, and 10 as well). We added a paragraph at the beggining of the result section stating that all the comparisons used the MLS averaging kernels: n this section we compare the offline HO2 dataset with balloon-borne and other satellite measurements, as well as, with global climate and photochemical model simulations. In making these comparisons, i.e. when showing the absolute or percentage differences between the datasets, the MLS averaging kernels has been applied to properly compare them. Furthermore, when comparing the global climate or the photochemical model simulations, its high vertical resolution has been reduced to the MLS one using a least square fit as described by Livesey et al. (2011, Sect. 1.9). In these comparisons, no altitude extrapolation has been applied to any dataset.

page 22920, line 13, Typo, extended?
Corrected

Page 22291, line 2, I would call this a "bias" rather than an "offset".
Done

Page 22291, line 10-14. "In the upper mesosphere, we found an underestimation by the model by as much as 60% but probably in part due to the low spectral resolution of the absorption cross sections in the LymanAlpha region and SchumannRunge bands." This reason has just been mentioned in the text as a possible explanation but has not been studied in this work. The reader might be mis-led. It should be re-written.
The discussion of the absorption cross sections was deleted. This section now reads: Using the Caltech/JPL-Kinetics 1-D photochemical model we found similar results. In the upper mesosphere, we found an underestimation by the model by as much as 60,%, and in the upper stratoosphere / lower mesosphere an understimation by about 20%. These results strongly suggest that these discrepancies are related to the model assumptions rather than to measurement errors.

References: Garcia, R. R., Lpez-Puertas, M., Funke, B., Marsh, D. R., Kinnison, D. E., Smith, A. K. and Gonzlez-Galindo, F.: On the distribution of CO2 and CO in the mesosphere and lower thermosphere, J. Geophys. Res., 2013JD021208n/a, doi:10.1002/2013JD021208, 2014.
Smith, A. K., Garcia, R. R., Marsh, D. R. and Richter, J. H.: WACCM simulations of the mean circulation and trace species transport in the winter mesosphere, J. Geophys. Res., 116(D20), D20115, doi:10.1029/2011JD016083, 2011. Smith, A.: Global Dynamics of the MLT, Surv. Geophys, 33(6), 11771230, doi:10.1007/s10712-012-9196-9, 2012.
Smith, A. K., Harvey, V. L., Mlynczak, M. G., Funke, B., Garca-Comas, M., Hervig, M., Kaufmann, M., Kyrl, E., Lpez-Puertas, M., McDade, I., Randall, C. E., Russell, J. M., III, Sheese, P. E., Shiotani, M., Skinner, W. R.,

Suzuki, M. and Walker, K. A.: Satellite observations of ozone in the upper mesosphere, J. Geophys. Res., 118(11), 58035821, doi:10.1002/jgrd.50445, 2013