Ref #2

A general reply to Ref #2:

We do not intend the main focus of the paper to be a complete and comprehensive validation study of the new data set of mesospheric O3. Rather in this paper we focus on presenting the retrieval technique, and the utility of the OSIRIS limb emission profiles as a sample data set for this type of high resolution mesospheric O3 data product. This primary goal is mentioned in the end of Sect.1 and beginning of Sect.3. Indeed, if the entire ~20 year Odin-OSIRIS data is processed, it would comprise a very valuable data set as the reviewer acknowledges. The processing of the entire 20 year OSIRIS data set with the new instrument corrections and using the inversion technique to the Level 2 product is a substantial computational undertaking and must be addressed in future work beyond this first paper.

We would like to thank to referee #2 by raising several important discussion questions that have helped us to improve the manuscript and the data processing method, namely the equilibrium assumption issue accompanying 1.27 µm emission being the proxy of O3, and the absorption corrections on the VER retrieval.

The manuscript does fall into the scope of AMT. It deals with a potential new data set of daytime mesospheric O3 and it is therefore very important. Eventually it can be sufficiently sound to be posted in its discussion forum but, in my opinion, not in its current form. I think it needs a substantial revision before it can be posted.

I would recommend to the authors the following actions:

The first part of the paper is well written, clear and well focused. This section, however, still requires some actions as:

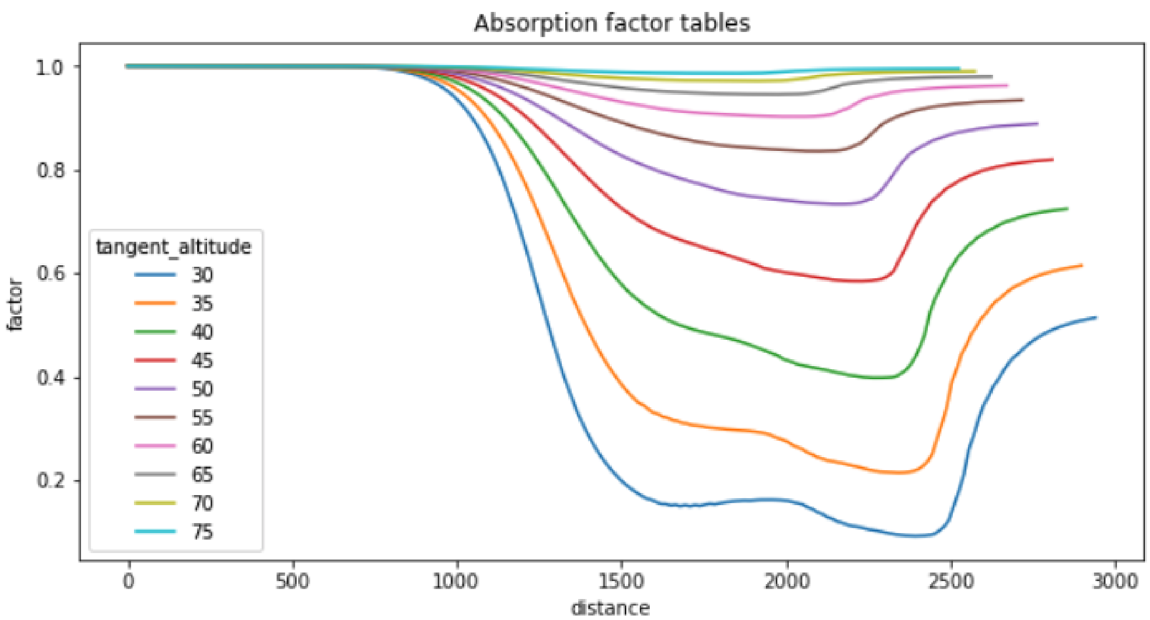
1) Include a Table listing all error sources and their estimated values. It is true that in order to verify if the estimated systematic errors are plausible (for example that of the stray-light, particularly near 80 km when the atmospheric signal is very small but radiance for the bright region below might be important) one needs to perform a thorough validation. Hence, I propose to present that in the second part of the paper (see below).

We agree that a table summarising all error sources and their estimated values would improve the paper. However, quantification some of the error sources can only be done after more data being processed, or require a comprehensive modelling study such as in Zhu et al. (2007), thus it will be left to the future validation study. We have however provided a table of possible error sources, quantifying those that can be quantified and estimation others at the end of Sect. 2.

2) To include a proper radiative transfer (i.e., do not assume the optically thin approach) for the 60-70 km region. Although the authors cite Degenstein (1999) (a Thesis work) as a support for the validity of this approach in that region, Mlynczak et al. (2007) state that "Below 70 km absorption of the O2(1D) emission by O2 itself begins to become important and the weak-line retrieval approach becomes invalid."

I suggest that a) include full RT below 70 km, b) limit the data to 70 km, or c) estimate (quantitatively) the errors of that approach in the 60-70 km region.

We acknowledge this limitation. We have reprocessed the VER retrieval by introducing an absorption correction factor to address this issue. The correction factor is calculated based on the tangent pressure (see the revised Sect. 2.2). The changes on the resulting VERs are less than 10% in the 60-70 km region, as shown in the figure below.



The figure shows the correction factor along the tangent path compared to the optically thin case along the tangent path. Colours correspond to different tangent heights of the paths in the unit of km. As can be seen from the red line representing line-of-sight tangent of 60km, the absorption would result in approximately 10% underestimation for the band intensity for emission near the tangent point. Furthermore, this factor is formatted in a pressure grid and applied to the retrieval scheme according to the pressure level at the tangent point.

3) If I have understood correctly, the 1.27 μm channel has not been calibrated in flight. It is just assumed that it behaves as other inflight-calibrated channels. I have not seen any error associated with this (lack of) calibration in the manuscript.

We acknowledge this comment. Odin carries no optical calibration sources. As mentioned in Sect. 2.1.4, the absolute calibration relies on the pre-launch value, and for the 1.53 μm channel this has been confirmed in flight as indicated in the paper. As at this point, assessment of the uncertainty due to the in-flight changes in the absolute calibration is difficult. The long term stability of this channel will however need to be assessed when the full 20 years of data is processed. Thus the characterisation of this type of error source will be addressed in the future study. However, we have included this in the error source list as suggested in point 1) in the revised manuscript.

2nd part. Section 3.

This section is a mixture of a kind of a soft (descriptive, not rigorous) validation exercise together with some partial description of the behaviour of O3, incomplete from my point of view. In my view, none of the two aspects are shown with sufficiently sound scientific treatment.

We would again like to emphasise that we do not intend this paper to be a comprehensive validation study of a new data product. We acknowledge the reviewer's point that the comparative statements regarding the measurements by the various instruments are somewhat soft, yet we believe these are useful comparisons (not validation) to show the fidelity of the retrieval technique. We agree with the reviewer that a comprehensive validation that uses the entire Odin-OSIRIS data set would indeed be valuable and we hope to be able to address this in a future paper.

For example, the current validation is only descriptive, a kind of hand-waving comparison, side by side figures, e.g., not showing differences, no co-location criteria has been used (or at least not mentioned). Sentences like "The differences between IRI, ACE-FTS and MIPAS in Fig.12 may be explained by their sampling at different SZA and the underlying assumptions in their retrieval techniques." are very vague and little informative. If sampling is a cause, the differences should be looked at by restricting it. Etc.

We agree that this comparison is descriptive. We mainly intend to demonstrate an Odin internal comparison of their co-incident profiles, as they are based on different physical measurement techniques and overlapping. Yet, these side-by-side figures show that the ozone profiles from three independent Odin-borne instruments complement each other well. Showing the differences between the coincident profiles would not be particularly informative in this case, due to their limited overlapping altitude range.

Regarding the comparison of zonal mean monthly average distribution with other instruments, an updated difference plot is provided in Fig. 14 including several latitude bands in order to make it more complete. We would like to point out again that, as stated in Sect. 3, only a test sample of 5% of all the measurements collected during a one-year period has been processed for this study. This information has been also added to the introduction section in the revised version.

About the second aspect, the study is mainly descriptive and based on partial datasets and considering the O3 number density instead of the O3 vmr. Sentences like: "Thus such a monthly mean profile should be treated with caution since it may not necessarily well represent the spatial and temporal distribution of daytime ozone." again adds little information. If it does not represent the distribution, why then show it? The study on the "Monthly mean ozone" is based on 1 month of one year and of O3 density. I would suggest the authors to refer to other similar studies (e.g. Lopez-Puertas et al. 2018).

We use number density mainly as it is the natural unit for the IRI measurements. Moreover, for SMR O3 at higher altitudes, the lines are mainly Doppler broadened rather than pressure broadened, thus the natural unit is closer to number density rather than VMR. Of course, IRI can be converted to VMR, by introducing external data such as MSIS, as shown in a small portion of Fig. 11. As Smith et al. 2003 mentioned, the different background density to derive VMR may introduce additional uncertainty between O3 profiles from different instruments, we would like to avoid introducing external data. We have addressed this issue in the revised version.

Fortunately, thanks to the updated scheme of ozone retrieval, the zonal mean monthly average ozone demonstrated in Fig. 12 and Fig. 13 agree much better with MIPAS than those in the previous version of the manuscript. We agree that the study is descriptive in regard to the behaviour of O3, indeed based on what we can tell from the limited amount of data we have at the moment. We intended to confirm that IRI data is able to reproduce a general pattern of the O3 distribution in the MLT.

My recommendation for this section 3, would be to focus on a thorough, rigorous validation (following the standard guidelines, see some more comments below) and leave aside from this paper the kind of characterisation of O3 features. Maybe the authors would like to consider future papers as, e.g. the overall OSIRIS O3 data sets, or tackle comprehensive works, including or not other datasets, as seasonal/latitudinal variations, or local time and annual variations, etc.

We want to emphasise again that a comprehensive validation study is not the focus of this paper. We acknowledge that a rigorous validation study is indeed valuable for a future study, after the ~20 years data will be processed. Section 3 is intended to provide a first-hand comparison in order to demonstrate the fidelity of the retrieval technique described in Sect. 2, which is our primary focus in this paper. We have described our purpose more clearly in the revised version.

Recommendations on validation:

1) It should be based on collocated data and, whenever possible, the same physical conditions, e.g., based on a coincidence criteria of time, spatial (latitude, longitude) and local time.

This issue can be addressed providing that more IRI data are processed and a sufficient amount of coincidence profiles between MIPAS and IRI are found. However, as already explained above, this can not be done at this stage.

2) Compare the appropriate instruments. That is, there is no altitude overlap of IRI wrt SO. Hence I see no reason for including SO data.

After the reprocessing of the IRI data with a correction on absorption, data can reach as low as 40km (50km after the last 10 grids were removed to avoid edge effect coming from the retrieval), thus IRI and OS has at least 10km of overlapping. Moreover, it is the first time to demonstrate how these three ozone data sets from Odin complement each other so well, despite their intrinsically different underlying physics in terms of measurement techniques. This also shows for the broader scientific community how Odin can cover a large part of the atmosphere using its different instruments. Thus, we do not want to completely remove the OS data set from this study, particularly Fig. 11.

- About MIPAS data, I suggest the author include a more appropriate reference, e.g., Lopez-Puertas et al. (2018). MIPAS middle atmosphere data ranges from ~20 km (not 5 km, Table 1) up to ~100 km. BTW, I believe the authors have used only DAYTIME MIPAS data (not stated explicitly in the manuscript). Mention which pressure/temperature is used (MSIS?, MIPAS?) to calculate O3 density.

We acknowledge the reference. MIPAS night time ozone is screened out as described in line 442. To calculate the density, we use pressure and temperature measured by MIPAS. We have provided this information in the revised version.

- I would be inclined to not include ACE data. O3 shows a large diurnal variation in the mesosphere and ACE is always measuring at the terminator. Hence it is difficult to distinguish systematic differences inherent to the instruments from those due to the solar illumination. BTW, the authors should state early in the paper that the 1.27 mm emission has a radiative lifetime of approximately 75 min and does not provide a representative measure of ozone until 2–3 h after sunrise (Mlynczak et al., 2013). Has this fact been taken into account in the current comparison? This fact automatically should avoid to compare to ACE sunrise occultations.

We agree that, for ACE data, it is difficult to distinguish the differences due the solar illumination from those inherent to the instrument. Thus we have removed ACE data for this study.

Regarding the long radiative lifetime of the 1.27 μm emission, we have added this information early in the revised version of the manuscript. In addition, we address this issue in an updated inversion process of ozone by increasing the uncertainty matrix of the VER, based on an 'equilibrium index',

Index = $1-\exp(-t/\tau)$,

where τ is the total lifetime ($\tau = 1/(A+Q)$) of the 1.27 μm emission, being the combination of the emission lifetime (1/A) and the quenching lifetime (1/Q) at a given altitude. This index indicates how far to the photochemical equilibrium state at a given altitude and time after the local sunrise. The detailed descriptions are provided in the newly added Sect. 2.3.3. In short, the effect of the newly introduced 'equilibrium index' in the O3 retrieval will suppress the underestimation of O3 where the steady state assumption is invalid, and thus these regions will be filtered out by low measurement responses before making comparisons with other ozone datasets.

- I am really missing a validation against SABER. In particular the O3 derived from the O2 1.27 μm channel. This is an instrument that uses the same technique and would therefore be very valuable.

Indeed it will be very valuable to compare SABER 1.27 O3 to IRI O3 in a future study, when the full data set will be processed and we will be able to carry out a validation study rigorously based on the comparison of coincident profiles. However, measurements from SABER drift in local time, unlike all other instruments included in this paper. For this reason, we have decided not to include SABER in this first study.

3) Quantify the differences (of the co-located data) for the different seasons/latitudes/altitudes. That is, as Fig. 11, but including more altitudes/seasons and enough years of retrievals to make the statistic significant.

A new figure including differences of more latitudes is added, see Fig. 14. The comparisons between IRI and OS, SMR are always coincident. Based on the available IRI one-year-data set, it is difficult to have a statistically significant number of coincident profiles with MIPAS. As mentioned earlier, the focus of this study is on the retrieval technique. Processing a statistically significant amount of data for several years would require substantial computation undertaking and thus this should be left for future studies.

About the 2nd part, a description of the O3 characteristics should be presented in a different paper and I would recommend the authors to please cite other previous recent works about this.

Other minor points:

In general several many figures are very small, particularly those with several panels, e.g. Fig. 3, 5, 9 and 11

It is revised in the new version of manuscript.

Fig. 3 Could you show the solar local time?

Solar local time for Odin is relatively constant (6h,18h). Instead, we have added the time after the local sunrise axis in the figure.

a typo: earth -> Earth

It is revised in the new version of manuscript.