

Referee #1

We would like to thank the Reviewer for extensive and constructive comments and recommendations to improve the manuscript.

Our point-by-point answers are found below. For the sake of clarify, the reviewer comments are in bold italics whereas our replies are in normal font.

**Specific comments:**

***1. The abstract already has a very strong focus on results (second paragraph), in contrast with more common practice. It would be recommended to include a notion of the comparison studies that are addressed, cf. subsection titles of section 5, instead.***

We extended our abstract in order to summarize our studies presented in the paper.

***2. The authors should motivate (probably in the introduction) why only limb and occultation measurements are considered for comparison, and not any other nadir viewing instruments, which have a more similar geometry and vertical resolution.***

We have performed such analysis because of the following reasons. First, the limb-viewing instruments have a significantly better vertical resolution than nadir-looking instruments, which allows probing the uncertainties related with the coarse vertical resolution. Second, the selected limb-viewing instruments practically do not use a priori information in the inversion. Finally, the selected limb-viewing instruments provide ozone profiles with a very good accuracy, and they have a small bias with respect to the ground-based measurements (e.g., [Adams et al., 2014]); these measurements have been used in the recent WMO ozone assessment. In short, the objective of our paper is to validate the GOME-2 ozone profiles measurements against the best satellite ozone profile measurements.

In the original manuscript, this motivation was summarized in page 7666, starting from line 6. In the revised version, we extended the description of the motivation for the analysis as explained above.

***3. One or more references are missing for the LIDORTA radiative transfer model. Consider Spurr et al. (JQSR, 2001) or van Oss and Spurr (JQSR, 2002).***

Thank you. We added the reference van Oss and Spurr (JQSR, 2002).

***4. One or more references are missing for the vertical resolution of the nadir ozone profiles retrieved by OPERA. Look for example at van Peet et al. (AMT, 2014) – mentioning a resolution of 7 to 15 km – and references therein.***

Thank you. We have now specified the vertical resolution to be between 7 and 15 km in text part and Table 1. We added also references to Peet et al., 2014 and Meijer et al., 2006.

**5. In Section 3, no motivation whatsoever has been provided for the spatial and temporal collocation criteria that have been applied for reference measurement selection. The authors should indicate (a) whether these criteria have been taken from literature (including proper references) or own considerations, (b) what geometrical, instrumental, statistical, or other considerations these criteria are based on, (c) thereby explaining why different criteria are used for the different reference instruments, and (d) whether the resulting spatio-temporal data sampling allows for an appropriate interpretation of the comparison results. For the latter an indication of the latitude-time collocation distribution would be most instructive, accompanied by either sampling plots or references to Figs. 1 and 5-7 that are informative on this aspect as well.**

In the revised version, we include new subsection Sect 3.4, where we collected the information about collocation criteria and their discussion.

For each reference instrument, the collocation criterion is a compromise of having smallest spatio-temporal separation whilst having sufficient amount of collocations. The separation in time is dictated by the local time of measurements. For spatial separation with GOMOS (separation less than 400 km), we used the effective horizontal resolution of the considered limb/occultation measurements (e.g., [Sofieva et al., 2013]). More dense measurements by OSIRIS and MLS allow tighten the collocation criterion. Formally, the same collocation criterion (as for GOMOS) can be applied for all reference instruments, but this will lead to multiple collocations. In order to get statistically independent pairs, only the closest in time or in space should be used in the analysis. Thus, even if the collocation criteria would be formally the same, the real spatio-temporal difference will be smaller in collocations with denser samplers. Similarity of biases with respect to all reference instruments indicates that the selected collocation criteria are adequate for the validation of GOME-2 profiles.

In addition to this discussion, the new Sect. 3.4 contains the table with the collocation criteria and a figure showing the location of collocated measurements, as recommended by the reviewer.

**6. The MLS data discussion in Section 3.3 seems somewhat disproportionate with respect to the GOMOS and OSIRIS descriptions in Sections 3.1 and 3.2, respectively. The MLS section could therefore be extended, especially on the retrieval technique and settings.**

In the revised version, we have extended the MLS description.

**7. The comparison methodology as currently outlined in Section 4 looks rather poor: (a) The first sentence is vague, whereas critical information on unit conversion and regridding operations should be included for all reference instruments, e.g. see Keppens et al. (AMT, 2015); (b) Despite the reference to Rodgers and Connor (JGR, 2003), the prior correction for retrieval comparisons as discussed therein has not been mentioned. The authors should indicate whether this correction method is applicable and why it has (not) been used; (c) The second equality in Eq. (1) is mathematically trivial and could be omitted. It seems more appropriate to limit Eq. (1) to the expression that has actually been implemented (the other might yield slightly different results due to computational limitations); (d) No indication has**

***been provided on the satellite data screening, whether before or after collocation. This should be added either here or in Sections 2 and 3.***

We have acknowledged these recommendations and modified the manuscript accordingly. In particular (addressing your comments),

(a) We added more details on representation the profiles in the same vertical resolution  
(b) We applied the smoothing using the averaging kernel by Eq.(1). The smoothing low-resolution profiles with the averaging kernel of high-resolution profile, if follow formally (Rodgers and Connor, 2003), is not needed (is not effective) because the limb profiles have significantly finer vertical resolution.

We have added this information in the revised manuscript.

(c) As recommended, we keep only the part of Eq.(1), which was implemented.

(d) We added the information on satellite data screening into Sect. 2 for GOME-2 and Sect.3 for limb instruments.

***8. In Section 5.1 it could be helpful and instructive to make reference to the (meridian dependence of) the tropopause and ozone maximum in the discussion of the dependence of the differences on latitude.***

We have added the mean climatologic tropopause height in the upper panels of Fig.1 (Fig.3 in the revised manuscript).

***9. Figs. 2-4 of Section 5.2 contain many plots that seem to contain more information than what is currently discussed: The difference between AK smoothed results and direct comparisons is only mentioned in one sentence (page 7673, lines 18-19), whereas the standard deviation of the differences is not handled. It might therefore be useful to combine Figs. 2-4 into a single figure, showing the mean AK smoothed differences for all three reference instruments. Such approach also allows the reader to more easily compare the three mean difference statistics. Brief indications of the comparison spread or the effect of comparison without smoothing could then be included in the text without specific reference to a plot.***

As suggested, we combined Figs 2-4 into a single figure and show only means of AK smoothed differences together with the standard deviation around the mean. We extended also the corresponding discussion in (Section 5.2). The mean relative differences when compared with non-smoothed GOMOS, OSIRIS and MLS profiles are described only in the text (Section 5.2), as suggested.

***10. It would be very clarifying and therefore recommended to draw vertical lines and add version numbers in Figs. 5-7 in agreement with the OPERA retrieval algorithm version updates as mentioned at the top of page 7674.***

As suggested, we have now added the vertical lines indicating the software version updates 1.20 and 1.23 that are now mentioned also in the text (Section 5.2).

***11. Similarly, in Fig. 8 it is not obvious from the plot and caption only which grey dashed line relates to which instrument. The overpass times of the other two instruments could be added as well for completeness.***

We have revised the Fig. 8 so that overpass times to all instruments are marked and labelled.

**12. Could the authors provide any reason for the exceptionally high relative differences and their sharp cut-off values discussed in the last paragraph of Section 5.3? If not, that should also be mentioned, e.g. with reference to future work.**

Unfortunately, we cannot provide the explanation for these features. We will investigate this in the future studies. We added this note in the revised manuscript.

**13. The last sentence of Section 5.4 raises the question of significance for the comparison results: (a) Nowhere in the text have the relative differences been compared with the (remaining) error budget of the satellite and comparison uncertainties. It is highly recommended to do so at least in Section 5.2, 5.4, or 5.5, and to add the respective random satellite errors in Table 1 for quick reference; (b) The manuscript does not make mention of the satellites' vertical sensitivities (AK matrix row sums), which could be of importance for the comparative analysis, especially for the nadir viewing instrument under study (see Rodgers (2000) book). Again, the vertical ranges of relevance, i.e. with sensitivity close to unity, could be indicated in Section 5 and Table 1. Fig. 11 for example reveals a sensitivity close to zero for the GOME-2 levels below about 10 km altitude.**

(a) We have now added in Fig. 11 the average profile of the GOME-2 retrieval error for both inside and outside polar vortex situation and related comment in Section 5.4. In Table 1, we have added the information about typical retrieval uncertainties for all satellite data. We have added also discussions on statistical significance of biases.

(b) We have added a new figure (as suggested also by Anonymous Referee #2) showing AK rows, AK row sums and DFS for a one typical observation as an example. We have added description (Section 2) of this new figure showing vertical range of relevance.

**14. A reference should be provided for the sentence on lines 19-20 of page 7676.**

Reference to Munro et al. 2015 added.

**Technical corrections:**

**1. Neither the title nor the abstract mention that this work is on space-based "satellite" observations. Although this is clear for people from within the field who are familiar with the instrument names, it should also be so for other interested readers.**

In the revised version, it is clarified that satellite measurements are considered.

**2. The MetOp-A acronym should be spelled out in the abstract and at first use within the text.**

Done.

**3. In the abstract on line 11 it is not clear whether the “non-degradation corrected” also applies to the previous sentence or not. This is clear for the remainder of the text, but should also (or especially) be so for the abstract.**

When we modified the abstract text according to the Specific comments 1, this vagueness vanished.

**4. Page 7664, line 26: “since the 1970s”**

**5. On page 7665, line 19; the O3M SAF abbreviation should be fully spelled out at first use; the O3M part is missing.**

**6. Page 7665, line 29: “validation with ozonesondes shows better agreement”**

**7. Page 7667, line 4-5: Rewrite “to the fitting the ozone profile to the”**

**8. Page 7667, sentence starting at line 11: “two types of products available” and “consisting of data blocks”**

Done.

**9. Mention version numbers for the OPERA and MOP retrieval algorithms in Table 1. If multiple versions have been used (as for the OPERA updates), they should be indicated as well, both in the table and the relevant text parts (e.g. first two lines of page 7674).**

Done.

**10. Page 7668, line 23: “have been validated”**

Done.

**11. Page 7670, line 6: The indication of the spectral region for retrieval requires two frequencies as edges.**

We have now rewritten the sentence: “Ozone is retrieved using the 240-GHz retrieval.”

**12. Page 7673, lines 25-26: The statement that “In the tropics, there are NO seasonal variations in the GOME-2 biases, as expected.” seems too strong and should be relaxed. Some seasonal variation can even be discerned from the plots.**

Referee is right. We have now corrected this sentence in the revised version.

**13. Page 7676, line 28: Rewrite “in coarse resolution horizontal resolution”**

We have rewritten it as “in coarse resolution”