

Referee #2

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.

General

Abstract: please specify version numbers of the used algorithms.

We have added in Abstract the used OPERA version numbers (1.14-1.24). The limited reprocessed GOME-2 data set was generated with a more recent OPERA version (1.32). The algorithm version numbers for GOMOS, OSIRIS and MLS are given in the Sections 3.1, 3.2 and 3.3 respectively.

Relative differences: have you looked at the median relative differences besides the mean relative differences (more sensitive to outliers)?

Yes we have looked both, means and medians. In the manuscript Figs. 5-7 show monthly median relative differences.

The differences between mean and median relative differences occur mainly in the lower altitude layers due to outlier values.

Please describe the rationale of having different collocation criteria for the three reference instruments. I suppose that you tried to be as strict as possible whilst keeping sufficient statistics, but it has to be explained and you should say something about the sensitivity of the results to the two criteria.

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), were 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 indicate that the selected collocation criteria are adequate for the validation of GOME-2 profiles. All these issues are discussed in more details in the revised version. For clarity, we collected the information about collocation criteria and their discussion in new subsection Sect 3.4.

Also, how do you deal with measurements on two sides of the polar vortex?

The collocated GOME-2 pixel with OSIRIS data pixel have been classified to be inside or outside polar vortex by using the 475 K potential vorticity associated with OSIRIS data. We have now added this clarification in the manuscript section 5.4. We have considered the measurements separately having pixels inside and outside polar vortex only when studying the case of Arctic ozone depletion in spring 2011 (Section 5.4).

Data selection: do you apply any further data filtering based on quality (e.g. data flags/uncertainties/unrealistic ozone concentrations)?

- We have accepted GOME-2 ozone profiles with Quality Processing Flag indicating successful retrieval i.e. overall convergence was reached (see Product User Manual for the Near Real Time and Offline Ozone Profile). We have now added this clarification in the revised version (Section 2).
- For GOMOS data we have done the data filtering by selecting profiles retrieved using hot and medium to bright stars as stated in Sect. 3.1.
- We have accepted OSIRIS profiles from morning (sunset) measurements, as stated in page 7669 line 24 (Sect. 3.2). We also omitted unrealistic ozone concentration values. We added this information in the revised version (Section 3.2).
- For MLS data we have restricted the data solar zenith angle to be less than 80° . We used additional SZA restriction since MLS has a dense data coverage. We added this information in Sect. 3.3.

Figures and table

Fig. 1: Please change "Nof" in the titles of the lower row to "N of" or "# of" or similar – with the current font size it looks like it says 'Not'. Also, can you assure the figures retain the same size (for each top and bottom pair, so that the latitude bins are aligned)?

We have now changed "Nof" into "# of" for clarification. We have added the colorbars to the first two figures (uppermost panel) to get the aligned latitude bins.

Figs. 2-4: Is the number of collocations valid for all altitudes (i.e. only full profiles are considered)?

The number of collocations is not necessarily valid for the lowest altitude layers. We have now added this clarification in the Figure caption (Fig.4 in revised version).

Fig. 8: Is this figure the same for all latitudes? Please label the indicated local solar times and include also MLS and OSIRIS. Can you extend the figure down to 15 km to match with your other figures?

Fig. 8 contains the latitudes between 20° and 30°N (and one month data). The diurnal ozone variation in the upper stratosphere is similar for all latitudes. We

have now added in the figure caption: "... between the latitudes 20° and 30°N ...". As suggested, we have added MLS and OSIRIS local solar times and labels in the figure. Below around 20 km the ozone is dominated by the dynamics instead of photochemistry. Thus, it is not very useful to show this region in a figure showing the diurnal cycle. The focus of Fig. 8 (and objective for including such figure) is variations in the upper stratosphere above 40 km.

Fig. 9 and 10: please make clear in the figure captions that the SZA refers to the GOME-2 observations. Add the unit (°) to the x-labels.

Done.

Fig. 9: I do not understand the grey marking of the "outliers" at 30 km as they overlap with the other data.

We apologize that the description of the grey dots was misleading in the figure caption. The grey dots represent the pixels having "outliers" at around 45 km (bottom right panel). But these same pixels do not have exceptionally high or low bias around 30 km (upper right panel). The idea is to show that these profiles do not have exceptional high disagreement at all altitudes.

We have now rewritten the last sentence in the figure caption to make this more clear. We have also modified the related text in page 7675 lines 4-6 in the original manuscript.

Fig. 10: there is no need to mention 'at four altitude layers' as they are the same altitudes as in Fig. 9 (it suggests that the altitudes would be different). There is a group of outliers (>40% difference) for solar zenith angles between about 65 and 80° visible at all four altitudes which has not been coloured grey to indicate them being outliers, but they are described in the text. Please mark them for consistency with the description.

We have corrected the Fig. 10 caption in the revised version. We have now coloured these pixels by grey, as recommended.

Fig. 11: Please add also the outside-inside polar vortex plot for the smoothed OSIRIS data for reference. Secondly, it looks like outside the polar vortex below 25 km, the a priori actually agrees better with OSIRIS than the GOME-2 retrieval. Please comment.

We have now added the monthly average of outside-inside polar vortex relative difference for smoothed OSIRIS data as requested (right panel). Yes, in the outside polar vortex case OSIRIS mean profile agree better with GOME-2 a priori than GOME-2 mean profile does but we have no explanation for this. In the GOME-2 retrieval this deviation from a priori indicates that in this altitude range (10-22 km) the information comes mainly from the measurement. We have added this comment in the revised version (Section 5.4).

Table 1: Please add used algorithm versions. It might be useful to add the altitude ranges covered by the instruments . Since you mention “ozone unit”, you might as well add which vertical grid is the native one.

We have now added the suggested information to the table.

Text

Page 7664 Line 25: space born -> spaceborne

Done.

Page 7665 line 8: specify version number for “earlier”

We deleted “earlier version” in the text since the OPERA version number is not relevant here and to be consistent with other text in Section 1.

Page 7665 line 23 target value + ‘of’

Done.

Page 7665 lines 24 – 27 “apart from certain altitude regions”. You mention only one? Perhaps rephrase to: In the stratosphere the 15% target value is met below 37 km, above which the differences are increasing... or clarify otherwise

We have rephrased the sentence as suggested: “In the stratosphere the 15% target value is met below 37 km, above which the differences are increasing...”

Page 7665 line 28: “after November 2008”. Are the >15% differences with reference datasets above 37 km also present for GOME-2 on Metop-B directly since its launch in 2012 (or is it a GOME-2A degradation issue)? Please clarify the text (since in line 17 you mention the validation to have been done for both 2A and 2B).

This text refers to GOME-2/Metop-A ozone profile validation only. We have now removed “Metop-B” from the text since this paper considers only Metop-A and hence discussion about GOME-2/Metop-B validation is not relevant here.

Page 7667 line 2 “1.14-1.24” how many versions does that cover?

(Page 7667 line 22) This covers nine product software versions. We have added this in the text.

Page 7667 lines 28-29 “the vertical resolution... is between 8 and 11 km at its best”. I read this as that for some parts of the profile, you will have a lower resolution, yet in Table 1 you specify the vertical resolution to be 8-11 km. Please clarify. Also, can you add a graph specifying the information content from the observation as a function of altitude for a typical observation?

We have now specified the vertical resolution to be between 7 and 15 km (in text and in Table 1 and added the reference to van Peet et al., AMT 2013 (as suggested by Anonymous Referee #1).

We have also added new figure that shows averaging kernel rows and cumulative DFS for one example retrieval. The cumulative DFS (right panel) specify the increment of the information content from the observation as a function of altitude, as requested. We have added description of this figure in the end of Section 2.

Page 7668 line 26 "nighttime". Please clarify its definition (e.g. product flag 'dark', SZA limit)?

We have rewritten as "nighttime (with solar zenith angle > 107°)".

Page 7669 lines 1-4. How do you consider differences in space between the GOMOS ground coordinates and those at the studied altitude range 15-60 km? I guess that for more oblique observations it does matter whether you collocate at ground level or at, say, 40 km? Also clarify how you deal with this for the other limb sensors where your spatial collocation criteria are tighter (see also comments in the related sections).

Thank you for note. This was typo. We have corrected this sentence. The GOMOS geolocation is taken at 30 km altitude. Also OSIRIS geolocation is taken around 30 km altitude. For MLS, please see our reply below in the related section.

Page 7669 line 8 "OSIRIS also has <> infrared imager". Insert "an".

Done.

Page 7669 line 13 "latest OSIRIS Level 1 data". Specify version number.

Done.

Page 7669 line 15 FMI has been used in the abbreviated form before you define here.

We have now modified the lines 12-15 and defined FMI before the first usage.

Page 7669 line 17 Provide a reference for "OSIRIS SaskMART".

We have added a reference "Degenstein et al. ACP 2009".

Page 7669 lines 18-21 What about performance outside the tropics?

We have added more references and information about FMI-OSIRIS ozone retrieval performance in the revised version (Section 3.2).

Page 7669 line 23 “.. GOME-2 pixel maximum 6 h in time”. Change to “.. GOME-2 pixel and a maximum difference of 6 h in time”.

Done.

Page 7669 line 23. Is the maximum distance of 200 km considered for all altitudes?

The geographic distance is considered between the ground GOME-2 pixel and the location of the OSIRIS profile at around 30 km altitude. We have now clarified this in the revised version.

Page 7670 line 11 “gave the overall 5-10% agreement” à “resulted in an overall agreement of 5-10%”

Done.

Page 7670 lines 12-14. Change to “The collocation criteria used here allow the maximum distance between the GOME-2 pixel center and MLS to be 100 km and the maximum time difference between the sensing times to be 6 h”.

Done.

Page 7670 lines 12-14. The maximum distance of 100 km is considered for all altitudes or do you only check the collocation criteria at the ground level?

We would like to note that the distance in the atmosphere can differ from ground distance by 20-30 km (Livesey et al., 2011). We added the note in to the revised version.

Section 3.3 is quite brief in length in comparison to 3.1. Perhaps add some more details on the ozone retrieval method.

We have extended the MLS description in the revised version.

Page 7670 lines 17-18. Please add details on the procedure and (when applicable) which external information (e.g. ECMWF pressure?) has been used to do the conversion. Do you also do this for the reported uncertainties?

We have now given more detailed description of the conversion procedure in the revised version.

The conversion has been done only for the reference ozone profiles.

Page 7671 lines 23-24 “at lower altitudes below ~20 km” change to “at altitudes lower than ~20 km” or skip ‘lower’.

Page 7672 line 10 “... there was also detected overestimation ...” change to “... overestimation was also detected ...”

Done

Page 7672 line 16 “MLS data in summer”. It is hard to verify the statement about the 55-65 °S results from Fig. 1 in Figs. 3-4 where 60° is at the edge of

two classes and the covered time period is not the same. Also, please indicate that the summer is the southern hemispheric (local) one or specify the months to avoid confusion.

We agree. This statement (i.e. connected to local summer months) cannot be verified from Fig.1.

We have rephrased the text accordingly.

***Page 7672 line 16 “the positive bias” change to “a positive bias”
Page 7672 lines 17-18 ‘this bias’. Clarify that this refers to the OSIRIS comparison as clearly for MLS it persists.***

Done.

Page 7672 lines 24-25. Clarify why you have decided to reduce the time frame to two years.

We have added the following clarification in the revised manuscript: “A period of two years enables a good representation of latitudinal and seasonal dependent relative differences without excessive averaging.”

Page 7672 from 26 – page 7673 line 1. What does the low seasonal variation in relative differences have to do with the larger relative differences in DJF? What is the cause of the increased bias?

We have rephrased the two sentences as “At tropical latitudes (30° S–30° N) the seasonal variation of the mean relative difference is low, which corresponds to the low variation in ozone in this latitude zone. Yet, at the tropical latitudes the negative peak of relative difference just above 40 km is on average about 10% deeper in DJF than in the other seasons. The reason for this seasonal increase in the relative difference is not yet understood.”

Page 7673 lines 13-14 “These differences in the results are due to temporal sampling...”. I guess this is also true in other months, for instance in MAM for the northern high latitudes? To what part can the observed differences be attributed to two sensors sampling different air masses as there are larger spatial variations in polar springs? Are the GOMOS data taken in full dark (no straylight/twilight contamination)? If not, do you think this could play a role (besides the temporal sampling) in the different relative difference shape observed?

- Yes, this is true in other seasons also. For instance in MAM for the southern high latitudes (we assume that reviewer meant southern high latitudes) collocated data for MLS includes March and April, whereas for GOMOS and OSIRIS only March is included. We have added this notation in the revised version.

- We have now added the following note in the revised manuscript: “The different air masses, being inside or outside polar vortex, can also contribute to the different mean relative difference values in polar spring.”

- At high latitudes, some of GOMOS data are with straylight contamination. However, we use only the data from sufficiently bright stars, therefore this contamination have a small influence on the data quality. We believe, the spatio-temporal sampling is the main reason for the observed differences.

Page 7673 lines 17 and 18. Please also discuss the deviations at the bottom of the profiles (rapid increase in the 1σ deviation), as well as differences for GOMOS around 30 km in the 30-90 °S region in the SON period and differences for MLS between 20 and 25 km in the 60-90 °N region in the period MAM-SON.

- We have added the following notation in the revised manuscript: "At the lower altitudes, the accuracy of the ozone profiles from limb-viewing instruments worsens due to lower signal-to-noise ratio thus causing a wide standard deviation around the mean relative differences."
- We have added the notation in the revised manuscript about differences of relative differences for GOMOS and MLS in the 30-90° S and 60-90° N respectively.

Page 7673 line 25 "In <> tropics ..." insert "the"
Done.

Page 7673 line 27 – page 7674 line 2. Why haven't you reprocessed the full timeseries with the same version if you notice differences between versions? If not possible, you should at least mention the version numbers and indicate in the figures (5-7) where you switch with a vertical line (with labels).

At the time of manuscript writing, a reprocessing of the full data set was not possible due to computational resource constraints. There is a reprocessing planned in the future, which would make revisiting this satellite inter-comparison relevant.

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

Page 7674 lines 13-20. It is not clear to me if the "implemented additive offset" has been implemented here or only in other studies. On page 7671 discussing Fig. 1 you state "which are not corrected for instrument degradation". Please clarify (in the conclusions it is clearly stated, but that is quite late!). Also, "data has become noisier due to the instrument degradation", is that valid for all altitudes or is that altitude dependent? Otherwise, how does that relate to the difference between Van Peet et al. and your results?

- Additive offset refers to all data produced operationally. But in this paragraph we referred van Peet et al. paper. We have rephrased the sentence.
- We have added in the beginning of this section 5.2 notation "The GOME-2 ozone profiles in this figure are based on measurements that are not

corrected for instrument degradation.” In addition, we added in the middle of section 5.2 ” The GOME-2 ozone profiles in these figures are based on measurements that are not corrected for instrument degradation.”

- We changed the validation report reference to ”Kins and Delcloo, 2012” where is stated that the degradation effect was first seen in the higher altitude layers (>40 km). Later (since 2010) at lower altitudes also on northern midlatitude (30-60N). We have rewritten the last sentence of Sect. 5.2 in the revised manuscript as”, ...due to the instrument degradation mainly in the higher altitude layers above 40 km.”
- We do not see inconsistency between our results and results shown in paper Van Peet et al. where the ozone profiles are globally compared to ozone sondes.

Page 7674 line 25. Detail why you chose the southern high latitudes to illustrate the SZA dependence and why the dataset is now limited to 2010.

We chose the southern high latitudes since there can be seen clearly high relative difference values dependence on SZA. The similar SZA dependence is visible over all the years considered. Therefore we restricted to one year results to keep the figures clear. We selected year 2010 since then there is a good representativeness of the collocated GOMOS data points in the southern high latitudes.

Page 7674 line 28 (and onward). “The similar dependence” à “A similar dependence”. This sentence and the rest of the paragraph are not supported by figures (mention that).

Page 7676 line 13 “... the very useful information” remove “the”

Page 7676 line 13 “.. to note, that ..” remove the comma.

Done.

Section 5.5. Does this correction go on top of the previously used algorithms or is it another algorithm version? Please state in the text!

The degradation correction is a factor that is applied to the measurement. Since this is still an experimental correction, it is not included in the general operational baseline but only in the data produced for section 5.5.

In the revised version, we added: ”This degradation correction is applied to the measurement, leaving all other Opera ozone profile algorithmic features intact. The data produced with the degradation correction is only used for comparisons in this section, not in the operational data shown earlier in this paper.”

Page 7676 line 22 “.. some comparison ..” either change to plural (“some comparisons”) or “comparison of some GOME-2 ..”

Page 7676 line 23 “using <> newly developed ...” add “the” or “a”.

Done.

Given that the comparison for March 2008 was done for the uncorrected data (dotted lines) with the oldest algorithm version, would it show such substantial improvements if applied to the latest OPERA version (for instance, for comparisons in March 2011)? I mean, can you quantify to what extent the improvement comes from the degradation correction and to which extent it comes from the algorithm changes?

The data from 2008-2011 that is used in this paper was produced operationally, based on different versions of Level-1b data, and processed with different versions of the Level-2 processor. Over the course of time, several improvements and corrections have been applied at both production levels. One of the corrections mentioned in the paper is the 'additional offset', which is introduced in the operational stream in November 2009. Only data produced operationally after this date has this correction. When considering a comparison of operational data from 2008 to reprocessed data using both the offset and the degradation correction, it is indeed not clear whether the improvement of the comes from the degradation correction or from the additional offset. Unfortunately, without a reprocessing of the 2008 data set without the degradation correction, it is not possible to give accurate quantitative differences. We can however look at Figure 4, where an improvement of a bias of -40% to a range of -20 to -15% can be seen after Nov 2009 in the upper part of the stratosphere (depending on latitude band). A further improvement of the profiles in figure 11 to an average bias of +/- 0 to +10% (depending on the reference instrument) is expected to be attributed to the degradation correction.

Page 7676 line 28. "original offline ozone product" Please state which version this is.

Done.

Page 7677 line 1 "new", again state which version number.

For clarity sake we have removed the word "new". The methodology for the degradation correction is still in the experimental phase and has no version number.

Page 7677 line 4 "under the continuous" remove "the"

Page 7677 line 5 "developement" à development

Done.

Page 7677 last paragraph: Please provide a better introduction to the study case than "the special case ..."

We have rephrased the beginning of the sentence as: "The detailed investigation considering..."