Author reply to Referee #1

Ground-based assessment of the bias and long-term stability of fourteen limb and occultation ozone profile data records

Hubert *et al.*, Atmos. Meas. Tech. Disc., 2015, 8, 6661-6757.

Thank you for your commitment and for your very interesting report. Our response to your questions, comments or suggestions can be found below, with different text formatting for referee comments and author replies.

This paper is an extensive review of the accuracy of 14 ozone profiling instruments that use either the occultation or limb scattering measurement techniques by comparison with ground based sonde and lidar data records. The paper is unusually well written with few typos, but at 56 pages of actual text with 33 co-authors, this assessment is excessively long. It would benefit greatly if it were shorter and more focused.

Indeed, the length of our paper may require an important effort from the reader, which is why we tried our best to provide a well-structured document with different access points for readers with different (levels of) interests. Sections 4-6 provide the information for those interested in the data quality of a particular data set. Sections 7-8 are of interest to those using multiple limb/occultation data sets (see also Section 8). As such, some information is inevitably redundant, but we hope it is the best we could do to serve different audiences.

However, one section where trimming is possible is Section 2, as also noted by Referee #5. We will shorten the instrument description section parts, since most of the information is already available in the literature or can be presented in a more condensed form.

On the plus side, the limitations of the various instruments becomes clear in the detailed comparison plots. I particularly like the attention to the importance of the auxiliary data in doing such comparisons. The fact that this adds uncertainty is a point frequently neglected.

Thanks for the compliment. This is indeed unfortunately often a missing piece of information in many validation and intercomparison studies, although especially relevant for the merging of records with different native representations.

The discussion section could be clearer as to which instruments have really useful data and which should probably not be used. Particularly the section on merging schemes should address which data sets are usable and which would need large corrections to be included.

Our main goals for this work were (a) an objective documentation of the data quality of individual limb/occultation ozone profile records, and (b) subsequently an objective study of the consistency between satellite data sets. An evaluation of which record is more suitable than another depends on the application and on the requirements of the user, which can obviously vary. Moreover, even for one given application, it is not always clear what the research community requires. E.g. in Section 8 we considered climate applications with requirements on decadal stability varying between 1 and 3% per decade. A judgement of usefulness is – by construction – tied to subjective criteria, which we believe is better to leave to the user.

The one serious problem I see with this paper is the use of lidar data for the comparisons in the middle and upper stratosphere. While the rationale for excluding a few sonde stations with limited sampling is reasonable, excluding 9 out of 13 possible lidar stations is problematic. In accepting only 4 stations are you using only the data that gives results that you expect? I am tempted to say either use all the lidar stations or none of them. This is a serious weakness.

This is a valid remark. We revisited the comparison time series and decided to loosen the selection of lidar sites. We now accept a few polar and the Hohenpeissenberg lidar (the latter only for satellite records ending prior to 2007). This change only has a minor impact on the outcome. The drift estimates of some satellite records changed somewhat, but not significantly (i.e. within the drift uncertainty). For other records the drift uncertainties changed somewhat, with a few drifts now slightly more significant, others less significant. All of our conclusions remain valid however.

We updated Sections 4.2 to reflect this change, together with Figs. 3, 5, 9 and 12, and Table 4. Any quantitative discussion of the lidar-based drift estimates in Sections 4.3, 6, 7 and 8. In addition, we added an appendix which illustrates how our final results alter if the entire lidar network would be used.

Minor comments:

p 6667 line 21-22 “We report the general tendencies...” Need to reword this sentence.

Done.

p 6670 line 5 This is the first mention of the ERA-Interim reanalysis. Need a short description here.

This will be added.

p 6670 line 6 No instrument is truly “self calibrating”

The connection to the following phrases (which challenge the “self calibrating” nature) will be made clearer.

p 6671 line 22 A bias of 8-10% is large in a paper claiming general agreement of +/-5% for most instruments in the 20-40 km region.

“biased low” may indeed be interpreted as small bias. We actually meant it as a “negative bias”, this is now made clearer.

p 6672 need to note that SAGE III has data at high latitudes only

The text (p 6672, l 8) and Table 3 mention 50-80°N (sunset) and 30-50°S (sunrise). So profiles in the Southern hemisphere are located only at mid-latitudes. This is clear from e.g. the meridian plot in Fig. 6.

p 6677 “... optical” should be “... optimal”?

Done, thanks for the catch.

p 6678 line 17 to avoid confusion say: (altitudes below 1 hPa)

Done. We have now done this throughout the manuscript.

p 6682 what is the wavelength range used by the VIS spectrometer?

This was added in the text.

Fig. 1 this figure shows MLS down to the surface, yet the MLS section notes that it is useful for scientific use only down to 262 hPa. Are any of the data useful down to the surface??

This may indeed be confusing to the reader. We removed the grid levels in Fig. 1 that were not used the analysis.

Fig. 2 What are the three different ground based instruments? Also, for both figures could you make the lidar points easier to distinguish from all the sonde points?

Each of the three panels in Fig. 2 shows a different combination of satellite and ground instrument. This was labelled in the figure itself. This will be clarified in the caption in the revised version. Marker style will be kept, since there is no "mix" of different correlative instrument types in each individual panel.

p 6707 line 16 "... measurements mostly agree..."

Done.

Section 8.2 You should mention the NASA SBUV v8.6 merged ozone data set, which is possibly the longest merged data set available. The Tummon paper provides some perspective relative to the occultation and limb data sets that are the subject of this paper.

We will add a reference to the merging of the SBUV nadir profile records.

p 6717 line 27 The performance of GOMOS is not as good as one might expect. Is it possible that GOMOS profile data derived using bright star data only might perform better?

The GOMOS Quality Working Group continues to work on improving the star selection. So far, no clear pointers were found as to why there would be a negative drift of GOMOS ozone data in the lower stratosphere. The (year-dependent) star selection used in this work is the current recommendation by the GOMOS QWG, and the result of extensive studies of the impact of star properties on data quality. This remains an ongoing activity, and better selections may appear in the future.

all Have you considered using integrated profile total column ozone compared to ground based as an element of quality control?

No, such an approach (also suggested by Referee #4) falls outside the scope of the paper. A detailed study of spatio-temporal mismatch uncertainty for all instruments is an analysis in its own right.