Interactive comment on "Quality assessment of the Ozone_cci Climate Research Data Package (release 2017): 1. Ground-based validation of total ozone column data products" by Katerina Garane et al.

Anonymous Referee #2

Received and published: 18 December 2017

The manuscript presents the validation study for the merged satellite total ozone record obtained from the GOME-type satellite ozone sensors. In this work the new dataset is validated against the ground-based network of Dobson, Brewer an SAOZ instruments. Presented study fits well to the scope of the problems covered in the AMT. The manuscript is well written and organized.

Major comments:

1. **Comment:**
   Most of the results presented in the study are shown for the hemispheric monthly means. However, the description of the methodology to compute these hemispheric means is not provided. These means could be computed in a number of ways (e.g. with/without weights), thus it would be important to provide a brief description in the text or in the appendix/supplement. Scientific results should be reproducible, and the clear description of the methods is an important component to ensure the robustness of the results.
   REPLY:
   Thank you for this remark. We agree and we have added a few sentences explaining the methodology of our calculations in Section 2.3, where Figure 3 is commented.

2. **Comment:**
   In the Tables in the Supplement, please, indicate the time periods for which the data from individual ground-based stations were used in this study.
   REPLY:
   We have added the time period for each ground based station in the Tables S1 – S3.

3. **Comment:**
   Page 9, lines 15-18: I don't quite understand the reason for not showing a plot with the SZA dependence for Dobson comparisons. Authors stated that Dobson observations depend on the stratospheric temperature, which should produce an artificial dependence on SZA. At the same time, when authors discuss results for comparisons with Brewer and SAOZ, they claim that the observed dependence on SZA for satellite measurements doesn't matter because all satellites show consistent patterns relative to ground-based stations. If so, why don't to show results with Dobson if the goal is to check consistency among satellite records.

REPLY:
Thank you for the suggestion. We agree with the comment and we have added two plots (SH and NH) in Figure 4, showing the dependence of the satellite to Dobson comparison on SZA, and the respective comments in section 2.3.

4. **Comment:**
Page 12, line 15: authors mentioned that the results of the comparison with measurements at Izana station were adjusted to account for the station's elevation. Have these adjustments been applied to any other station? This should be clearly described in the text.
REPLY:
The mentioned correction is an ERA-Interim-based estimate of the column below the instrument altitude in the immediate vicinity of the island and/or mountain top (see Verhoelst et al, 2015, for further details). For the SAOZ/ZSL-DOAS network, Izana and Jungfraujoch are the only stations for which a significant missing column was derived with this methodology (about 2.8% and 3.2% respectively, with some seasonal variation), due to their isolated mountain-top locations. Some more information is added in the manuscript, section 2.3.

5. **Comment:**
Page 12-13, Table 1: I found that the quantities shown in Table 1 are not well described. I think this part needs some major revision, including terms that are used in the table. For example, I would recommend using a term "mean bias" instead of "monthly mean bias" and mentioning in the text that the mean bias was computed from monthly mean differences, because to me "monthly mean bias" would mean the bias in a specific month, while, if I understood this correctly, the biases shown in Table 1 were computed over the entire data record. I don't quite understand what is "monthly mean variability" in Table 1. The 1–sigma standard deviations for biases are shown along with the biases. In the text this quantity is explained as "the variability of the monthly mean standard deviation values". Is that the variability of the standard deviations of differences in individual months? I would also recommend replacing "Seasonality" with "seasonal bias". The two last lines in the table are very confusing (Latitude and SZA). At first, I thought they show the mean differences in latitude and sza between satellite and ground based observations. But according to the text these are the mean differences in ozone between satellite and ground-based observations, calculated by averaging all points shown in Figure 4 and Figure 5. I understand that if you try to bin differences in smaller bins (like you did with SZA) you can uncover some dependence on SZA. Then it would make sense to name these quantities as "Latitudinal biases" and "SZA biases". But why do you expect results to be different from the mean biases if you average over the entire latitude range? Please, explain.
REPLY:
Thank you for the comments and the suggestions. Please find below our replies:
- The "Monthly mean bias" term was corrected in the text as suggested and a phrase explaining how it was calculated is added.
- The "monthly mean variability" term is indeed the variability of the monthly mean standard deviation values and the text is modified so as to be clearer.

- "Seasonality": We consider the "bias" as a term to be the deviation of comparisons from the 0% line, which would mean 0% bias. In this statistic quantity we calculate the peak-to peak range in values, so we think changing the term would imply a different statistical quantity.
- Latitude and SZA statistics: we agree that we do not expect to see any major differences to the mean differences of the monthly means, but these statistical quantities have to be provided as requirements by the Users of the ESA CCI project (they will be uploaded soon at: http://www.esa-ozone-cci.org/?q=documents#). We have added the reference and changed the text accordingly. We also named the quantities "Latitudinal mean bias" and "SZA mean bias".

6. **Comment:**
Page 15 lines 25-31 and Page 16 lines 1-2: In this part of the manuscript authors describe the correction factors against OMI that were used to correct individual satellite time series. It is stated that "we apply correction factors using the seasonal mean differences…". Then it mentioned that the drift in GOME 2A have been accounted for. Is it a static correction that depends on lat/lon and month of the year only? Or have you implemented time-dependent corrections? Please, explain that in the text.
REPLY:
For GOME and SCIAMACHY it is a static correction that depends on latitude and month of the year. For GOME-2A and GOME-2B it is a fully time- (and latitude-) dependent correction. We have improved the corresponding explanation in the text, Section 3.1.

7. **Comment:**
Page 17, lines 10-20: you need to explain what was done in the merging process when data from two or more instruments are available. Did you simply average all available data? Did you use some weights?
REPLY:
The merging is done on a daily basis. When data from two or more instruments are available, we average all and use as weights the number of measurements per day and grid box for the corresponding sensor. We added the explanation in Section 3.1.

8. **Comment:**
Looking at the results showing in Figure 11, it seems to me that the merged dataset almost fully overlaps with OMI. This is expected since all individual datasets have been corrected against OMI, and OMI has a very dense spatial and temporal coverage. My question here: what is the value of using GOME 2A or GOME 2B in the merged product? Please, provide an explanation in the text.
REPLY:
We agree with the reviewer that it is not unexpected that the results of the level-3 validation are dominated by the very dense OMI measurements. As already seen in the ground-based validation of the previous version of the merged product (Coldewey-Egbers et al., 2015) the quality of monthly mean level-3 data strongly depends on the spatio-temporal coverage of the input (level-2) measurements. Therefore, we think that it is more beneficial to include all

available measurements in order to further improve coverage and, thus, statistics and representativeness of the monthly mean values. We added the explanation in Section 3.1.

9. **Comment:**
Figure 7: I would suggest to keep the range for the time scale (X-axis) the same for all 4 panels;
REPLY:
We changed the range for the time scale as suggested and we added a gray background to better distinguish between missing data and values close to zero.

10. **Comment:**
Figure 9: it's very hard to see blue letters on the green background. I would suggest to move satellite timelines either to the top or bottom of the figure.
REPLY:
We changed the color of the letters to 'black' and moved the timelines to the middle of the plot in order to improve visibility.

11. **Comment:**
Figure 10, left panel: why there is no point for Level 3 product in 80-90N latitude bin even though data for all individual instruments are shown;
REPLY:
Thank you for noticing this issue. The reason is that there is one ground-based station nearly at the limit between the 70°- 80° bin and the 80°- 90° bin, namely Eureka, at 79°.89 North. As a result, when allowing spatial collocations to the satellite central pixel within a certain radius, some of the collocations were allocated to the 80°- 90° bin.
We have updated our validation chain to take this into account and changed Figure 10 accordingly.
When analyzing the ground-based level-2 into gridded level-3 this issue did not come up, so there was from the start no data in the 80° -90° bin.

12. **Comment:**
I am puzzled why the Level 3 value in 70-80S latitude bin is higher than for any given individual instrument.
REPLY:
The time series shown in this Figure are not common collocations, as you can well imagine, since the different Level-2 instruments have a different time span, which again differs from the level-3 data. Hence, the variability shown by all six time series in the -70° to -80° S bin may be due to a number of factors, including ground-based data availability, as well as dynamic issues that may affect different years in a different manner, such as the polar vortex for example.

Minor comments:

1. All abbreviations should be spelled out when used for the first time in the text. For instance, there are many abbreviations in the Abstract and Introduction that are not explained: P. 1, line 15: "GOME-type" – please, spell out "GOME"; P. 1, line 18: "GODFIT", "ERS", "OMI",

"SCIAMACHY" –please, spell them out; P. 2, line 33: please, spell out "SAOZ"; P. 3, line 7: please, spell out "BIRA-IASB" and "DLR"; P. 3, line 10: please, spell out "LIDORT"

REPLY:
We have spelled all abbreviations in the main text (mostly in the introduction), as suggested. The satellite names in the Abstract are left as they were because we think that it would make it too extensive, but they are also spelled in the introduction.

2. P.3 lines 25-28: It is not quite clear from the context which quantity has "been estimated to rise up to +/-2%": systematic uncertainty in the ozone cross sections or ozone itself? Please, consider re-wording this statement.

REPLY:
We agree that it was not quite clear in the text that the quantity that is biased by +/-2 % due to the use of different cross sections, is ozone. The manuscript was rephrased. Thank you for the suggestion.

3. P. 14, line 23: I guess it should be "NOAA 18" (not NOAA 16) to match with the labels on the right panel of Figure 6.

REPLY:
Yes, thank you, it should be NOAA 18. We have made the correction.

4. There are several places in the manuscript where authors use words "excellent", "exceptional" etc. I would recommend to avoid these statements in the scientific publication and rather provide quantitative results like "the stability within +/-1%" or "biases less than 2%".

REPLY:
Thank you, we have scanned through the text and made all the necessary alterations.