Atmos. Meas. Tech. Discuss., 8, C1331–C1337, 2015
www.atmos-meas-tech-discuss.net/8/C1331/2015/

Atmospheric Measurement Techniques
Discussions

Interactive comment on "The GOME-type Total Ozone Essential Climate Variable (GTO-ECV) data record from the ESA Climate Change Initiative" *by* M. Coldewey-Egbers et al.

Anonymous Referee #2

Received and published: 2 June 2015

General comments:

1. Does this manuscript address relevant scientific questions within the scope of AMT?

Yes, this is relevant, as the paper presents the creation of a merged total ozone column data record, as an "essential climate variable" (ECV).

2. Does the paper present novel concepts, ideas, tools, or data?

Yes.

3. Are substantial conclusions reached?

Yes, although some clarifications could be added.

4. Are the scientific methods and assumptions valid and clearly outlined?

Yes, for the most part.

5. Are the results sufficient to support the interpretations and conclusions?

It appears so, but some clarifications are desirable.

6. How traceable are the results?

This is often hard to answer well, but the results should be traceable in theory (given the retrieval algorithms, raw datasets, etc...).

7. Is proper credit given to related work and are the new contributions clear?

Yes, this appears to hold.

8. Is the title clear?

Yes.

9. Is the abstract concise and complete?

It is concise more than it is complete.

10. Is the overall presentation clear?

Yes, for the most part.

11. Is the language fluent/precise?

Yes.

This manuscript seems suitable for publication in AMT, after some clarifications, see below. The manuscript describes a merged long-term total ozone column global data record, based on homogenized retrievals from GOME, SCIAMACHY, and GOME-2 observations from March 1996 to June 2011. The retrievals are briefly described, along with the dataset adjustments that are performed relative to GOME data (as a reference). Validation of the resulting Level 3 data is presented through comparisons relative to Level 2 data and independent ground-based data, as well as other satellite data. The public availability of this merged product will be useful to others with interest in ozone layer issues (e.g., ozone recovery, model comparisons).

The results seem good enough to conclude that this new dataset will be a useful merged global record.My main comments have to do with how much better or worse the new dataset really is, compared to the older merged product (which is mentioned somewhat too quickly), or versus other independent data (also briefly mentioned). Maybe this means providing somewhat more quantitative references to other published work (already mentioned in this manuscript), without repeating all the details. This would help place the work in context for the reader, even if this manuscript is mostly supposed to present the dataset creation and validation results. Also, the Lerot et al. (JGR, 2014) paper describes in a fair amount of detail the procedures used to create the dataset mentioned in this manuscript, and there are some overlapping types of Figures between these two papers. This results in duplication of some of the descriptions and comparisons, and one might expect, therefore, that this manuscript could be even more succinct regarding some of these aspects, while emphasizing the validation/comparisons even more. This is not a huge problem, but nevertheless, some additional thought should be given to this issue, and as this paper is the "third article on the ESA-CCI total ozone ECV", quoting the authors. Having an international team collaboration on such a project is great, but keeping the number of papers describing approach and results distinct enough is also useful, in this world of busy researchers seeking succinct, timely, and clear messages from added publications.

Overall, this manuscript is fairly short and I am not recommending major changes. However, some of the specific comments below are not completely "minor" either.

Specific Comments:

- Introduction: Since WMO (2014) appeared in December, 2014, the WMO Report (2011) reference could be enhanced by pointing to this newer reference (or mentioning both). The same holds for Section 4, where the "last WMO report" is mentioned (but is no longer the last one).

- pg. 4611, line 12, I suggest changing "an homogeneous" to "a homogeneous".

- pg. 4612, line 9: "GOME lost its global coverage". Could you provide more specifics regarding what this means for the subsequent time period? Did the coverage decrease continuously and how many days are needed in the end to obtain global coverage? Or point to a later part of the manuscript if this is better described later on. Not every reader is as familiar with the details as the authors are.

- The merging procedure describes offsets that vary with latitude and time, but then mentions that there is no temporal drift in the corrections. Does this mean that using a time-dependence in the corrections is not really needed? Or is what is most relevant in the time dependence short-term in nature, so the long-term drift can be small or zero even with substantial short-term "drifts"? This also seems to assume that GOME does not drift in and of itself (versus "truth"), as it is used as a reference (and is this true?). It is is somewhat difficult to understand from the description provided in this manuscript alone (without maybe reading other related publications) why this procedure was chosen rather than an average of SCIA and GOME during their overlap period. Some more clarifications and motivation would have been useful, besides what the merging algorithm itself does, as there could be more than one approach (and this is indeed a change from the previous method/results).

- pg. 4613, line 13: what is meant exactly by "low time resolution" (please specify the resolution)?

- pg. 4616, line 12: please describe more specifically how this standard deviation is obtained? Are the combined (adjusted) datasets used to obtain their combined variability (for what grid size and time step, 1 by 1 degrees and one month?)?

- It is interesting that the standard error is not obtained from something like the standard deviation divided by the square root of the number of included input points, and a simulation is used for this result. I understand that there is a sampling component that one can try to quantify from such a simulation, but if the measurement errors (noise) is excluded (and maybe it is not), is there not an underestimate of this standard error (or is this a negligible effect)? This also assumes that the model variability mimics the actual variability well enough...

- pg. 4619, line 7: not sure what "a limit" means here. Is this a screening criterion? "Optimization reasons" is also quite nebulous. Probably a detail, but if you wish to start to explain this, maybe you should pursue it more fully (or ignore this if it is really not worth the effort, or makes little difference in the end).

- pg. 4620: Statistics for the Northern Hemisphere, you give a bias for the Dobson comparison but no such number for the Brewer comparison. Is that (mean) bias zero (within error bars)?

- pg. 4620, line 7: you could delete the "and" after "(from top to bottom)".

- pg. 4620, line 28: is there a significant trend in the differences between the L3 and L2 comparisons (end of parag., "very satisfactory" is vague language)? Overall, should one try to delete outliers in the L3 results to obtain the same trend as the L2 results? It would be nice to see somewhat more quantitative results or descriptions of the agreements (same for the next page on SH results which are in "near-perfect" agreement). Trends will matter to people looking to such datasets for robust longer-term analyses (and you have other references on this topic), so I wish you could expand or refer more quantitatively to another reference regarding this topic. The qualitative language and the (admittedly nice looking) plots can only go so far in terms of convincing the reader as to how useful this new dataset really can be. I am not necessarily suggesting a complete trend analysis, but a slightly more quantitative result would be more useful. Table 4 has some results, but not necessarily comparing L3 to L2 (?).

- pg. 4622, line 15: change "meets easily" to "easily meets".

- pg. 4627, line 10/11: "CCMs are three-dimensional..." I would delete this well known description.

- Fig. 3: The caption mentions March as the change-over date, whereas Fig. 2 shows April (please correct one or the other for consistency). If Fig. 2 is correct, would one not state that April is the changeover date (to SCIA data for example)?

- Figs. 4/5: Can you specify what GOME 2A means?

- Fig. 8: there is a lot of white space here, the y axis ranges could be shrunk to help see the results better.

- Fig. 9: Not clear that the outliers have been adequately explained in the text - but I gather this is mainly a satellite sampling/gridding issue; would you generally trust the L2 results more (this was maybe not spelled out well enough)?

- Fig. 10/11: again, I find this hard to read because of all the white space but it appears the authors want to keep one scale throughout; at least the scale could start at -10 rather than -15, it would seem...

- I like Fig. 12, but/so something similar for the Brewer results might be worth considering as well (?).

- Fig. 13: The "red open circles" are quite difficult to see. Please try using solid black or blue dots or another clearer choice.

- Fig. 15: please specify the years used in the caption (as done in Fig. 14).

- Since the patterns between Figs. 14 and 15 are pretty similar (but hard to tell unless one has the numbers or overplots), with dips at high latitudes, should one not conclude that the GDP product was/is closer to the SBUV-MOD V8.6 result and the new CCI results are in poorer agreement with SBUV? Please quantify or address this.

- It would have also been nice to see how the comparisons versus ground-based data have changed in the new dataset versus the old one (e.g., showing Table 4-type results for the older (GDP) product?). Do such results not exist for both datasets?

- Table 4: What does the "Latitude" row represent? Are the error bars 1-sigma?

Interactive comment on Atmos. Meas. Tech. Discuss., 8, 4607, 2015.