Earth System
Science
Data
Discussions
Open Access

Interactive comment on "The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: A long-term database for climate studies" *by* Sean M. Davis et al.

Anonymous Referee #3

Received and published: 20 July 2016

This manuscript introduces a new merged ozone and water vapor data set from satellite limb sounders – SWOOSH – which constitutes a very valuable resource for the study of stratospheric climate variability and change that can be anticipated to be widely used in the community. While the paper is generally well written I miss some key information that I would expect in the documentation of a new database of this importance (these are, to summarize the more detailed comments below, the validation of the dataset and its consistency with other measurement systems, clearer guidance on uncertainty measures), as well as providing better information of how SWOOSH distinguishes itself from e.g. the GOZCARDS (Froidevaux et al., 2015) or Bodeker scientific BDBP (Bodeker et al., 2013) datasets that are already available. I acknowledge that the paper

states its main goal is to introduce the methodology of merging, but without validation (or at least a rough sanity check) of the product, knowledge on the methodology is not very useful since the reader cannot judge the validity of the methodology applied. I hence suggest some major improvements as detailed below before I can recommend publication, which hopefully will help make the manuscript more valuable to the data user.

Major comments

The main problems I see with this manuscript are:

- The authors neglect to put their new database into context with already existing stratospheric databases. I don't agree with the statement in the introduction that 'Despite their chemical and radiative importance in the stratosphere, there have been relatively few attempts at constructing long-term data records of O3 and WV based on vertically resolved satellite limb-based observations of these species'. Given the limited number of satellite instruments covering the time period 1979-2003, how many times can these be merged to the newer satellite instruments (starting 2002 or later) with adding value to previously merged datasets? Providing information on improvements or differences in the merging approaches used is hence a necessity in order not to confuse the data user on which data product should he/she use, but mostly lacking in this document. An obvious omission I see is not citing Bodeker et al. (2013), but then there are other merged datasets that are not mentioned (ESA CCI, Sofieva et al., 2013), wrongly cited (Froidevaux et al., 2015; Hegglin et al., 2014), or just mentioned in passing (Randel and Fu, 2007).

- The manuscript should provide a much better guideline for the end user for what the dataset can be used for and for what not. The information is sprinkled throughout the manuscript, but never brought to a conclusion or into a summary. The inexperienced data user will then just go ahead and use the merged anomaly-filled dataset that is easiest to use (since there are no holes in the record), but which arguably has the largest

uncertainties in reproducing the real atmosphere given the construction methodology. As a more concrete example, the discussion in Appendix A on the investigation of the low bias in Aura-MLS is not drawn upon in the main part of the manuscript and leaves the reader with the open question on how this influences the validity of SWOOSH below 100 hPa. Here again I can't find too useful guidance given by the authors to the user.

- Along, with the previous comment, a more accurate communication of the uncertainties of the database is necessary to prevent false conclusions in studies that e.g. look at trends in these species or compare chemistry-climate models to these observations. The different definitions of uncertainty you provide confused me, not last due to the fact that the labeling of the different sigmas in the figures is not always consistent with what you write in the text. Instead of burying the discussion and definitions of the uncertainties in the Appendix, I would expect this to be a key part of the discussion within the paper and explained in a clearer more concise way. For example, you could show in addition to the applied offsets shown in figures 5 and 6, similar latitude-height distributions of the RMSS or combined standard deviations (possibly for two different time periods in the early/later part of the record). This should more clearly illustrate where the user can trust the dataset and where not (e.g. < 100hPa).

- Another major problem I have is with the communication of the data product that uses anomaly filling of data where there is basically no information from the satellite instruments available. These locations should really get a special uncertainty associated with or be flagged in an obvious way. This information is buried in the text and not brought up in the conclusions anymore where also shortcomings of the database should be summarised and highlighted.

- The manuscript states several times that the evaluations presented would illustrate the use of SWOOSH for studies of variability on different timescales. However, Section 5 provides very limited evidence for this and in my eyes is unsatisfactory if the features shown are not validated against independent instruments. I would have ex-

pected a better sanity check of the merged product using long-term observations from other measurement platforms for comparison. The authors could as a suggestion compare their new water vapor product to the Boulder FPH dataset, which would span the whole time period of SWOOSH and hence could be used to check whether the merging worked satisfactorily. Similarly, long ozonesonde records exist (see e.g. Randel and Thompson, 2010), which could be similarly used to test the QBO in the tropics. This is in particular important since the QBO is seen to have very different structures in different instruments (due to sampling, vertical resolution, etc issues see Tegtmeier et al., 2013).

- Finally, the dataset is provided in different coordinate systems and resolutions and I would assume that the uncertainty estimates should increase when moving from 10 down to 2.5 degrees latitude resolution. Also, I would expect that the higher resolution dataset would be noisier than the lower resolution one especially in the early part of the record. It would therefore be important to illustrate and discuss these characteristics and differences, and I suggest to add some figures with meridional profiles of both absolute values and uncertainties to the manuscript.

Minor and technical comments

Abstract: The abstract does not provide enough information on the characteristics of the database. I suggest to add that you have 3D and 2D climatologies, different coordinate systems (list them), height range covered (300, 100 hPa upwards to 1, 0.1, 0.001 hPa?), latitude/longitude resolutions available, and that the instruments considered are all only from NASA satellites (in contrast to e.g. Froidevaux et al. 2013 who use also Canadian Space Agency observations).

P1L13 Suggest to delete 'its use for studies of' I don't think you have shown the use of SWOOSH for the study of climate variability in this paper. This is because you haven't compared to other measurements or models.

P1L19 'climate impacts' is a misleading choice of words.

P1L25 see major comment above. I don't think you do the past work on merging datasets of the stratospheric community justice by claiming this.

P2L5 It is confusing to mingle merged total column data sets and vertically resolved data sets here.

P2L8 To my knowledge the SPARC Data Initiative does not offer a merged water vapour data set. Hegglin et al. (2014), see full citation below, should be used instead.

P2L21 A more recent comparison of these satellite instruments supporting this conclusion is given by Tegtmeier et al. (2013).

P2L25 The Kley et al. (2000) reference is outdated given that you would also like to compare Aura-MLS (from 2004 onwards) with these earlier satellite observations. Here a reference to the SPARC Data Initiative water vapour assessment (Hegglin et al., 2013) that you used earlier in the wrong context should be added. This paper shows that Aura-MLS, SAGE II, and HALOE do in fact agree much better within 10

P6L26 This implies that HALOE offers measurements down to the ground, while HALOE data providers now recommend that the data should not be used below 100 hPa. Please amend.

P6L29 Add reference to Hegglin et al. (2013).

P8L26 The coincident criteria you choose are much looser than what is generally used in validation studies (mostly within 6 hours and 400 km). Have you tried to make them stricter and how would this affect your offset uncertainty estimates? It seems scientifically not correct that by loosening your coincidence criteria (which should introduce larger biases) let's you achieve improved uncertainties on your error estimates (on the grounds that you have more profiles in the comparison that affect your standard error of the mean).

P10L5 Looking at figure 3 this implies that Aura-MLS has a high bias of up to 7

C5

P10L8-13 I don't understand your argumentation here nor do I agree with it. After all you should assume that the FPH is your truth. By choosing sat+FPH as your reference you are decreasing the percentage bias estimate which seems arbitrary.

P12L3 The term 'to create statistical agreement' is a misleading terminology to use here. If I understand your methodology right you simply bias-correct the mean. What bringing into 'statistical agreement' means is to do a quantile-adjustment, i.e., also correcting for the variance and variability differences in two datasets. Please change wording.

P13L1 Please specify vertical range.

P13L5-8 This would be an example figure that I would like to see evidenced in the appendix, since the question of the impact of averaging kernels on instrument or model-instrument comparisons is always asked by reviewers and users.

P13L12-15 The study by Hegglin et al. (2014) who introduce a new merging technique based on a chemistry-climate model as transfer function seem to reach the conclusion that a simple merge between two datasets such as HALOE and Aura-MLS does lead to a wrong bias-adjustment due to a potential degradation of HALOE and SAGE II observations towards the end of their lifetimes. Another study by Brinkop et al. (2016) provide additional evidence that support the conclusion of this study. The fact that your methodology is based on bias-adjustment during overlap periods of instruments may hence be a problem for the merging and should be mentioned in this manuscript.

P14L31-P15L1 That's not the full story. Equivalent latitude is expected to steepen gradients also in the subtropics and in the tropopause region. Please amend this statement.

P15L23 onwards: please improve consistency of the annotations of these different standard deviations or uncertainties in the text and the figures 8 and 9.

P17L2 This recommendation to the user of your dataset to use the combined stan-

C6

dard deviation (s), which seems equivalent to the light grey shading in figures 8 and 9 (please clarify if this interpretation is not correct), seems scientifically not justifiable to me. The figures show that s is basically constant over the whole time period, which does not reflect in any way that there are known uncertainties stemming from the pure facts that HALOE measurements have a much sparser sampling and have been strongly affected during the early 90's by the Pinatubo aerosol.

P18L6-11 An equivalent statement to this one needs to show up in the conclusion sections.

P18L24 See comment above, there are new studies that claim the drop in the merged HALOE-Aura-MLS datasets is overestimated (Hegglin et al., 2014; Brinkop et al., 2016), so this cannot be used as proof that your dataset is showing the right behavior. Please add this caveat.

P20L5 I don't see where you have explained the differences to the Froidevaux or Randel and Wu methodologies further above.

P21L3 There are indeed non-American satellite instruments as well that provide water vapor measurements from space. These are the Canadian ACE-FTS or the Swedish Odin-SMR, both of which are still in space and would be very useful for extending the water vapor record (at least as long as they can keep up in space as is true for Aura-MLS).

Appendix A: It is not clear to me how this information on the Aura-MLS low bias affects the data screening used in your study.

P22L15 delete 'the' between 'kernel to' and 'degrade'

P22L18-24 This seems to me a very surprising result and in my eyes warrants further investigation. Do you really use the full retrieval procedure that is used in an equivalent way for the retrieval of MLS L2 data? The shape of the a-priori profile seems too close to the L2 data profile so that it would be surprising that other measurement aspects

create the strong oscillations instead. Did you discuss this with the MLS folks?

P23L1 The dry bias in Aura-MLS at this altitude has also been pointed out in the study by Read et al. (2007) and Hegglin et al. (2013).

P23L29 This seems to contradict your interpretation of your own evaluation in P22L18-24.

References: Please use the correct reference (see below) for referring to GOZCARDS, the paper has been published in ACP last year already.

Figure 4: Suggest adding a relative difference plot to make this figure consistent with Figure 3. It is not good practice to cover up the axis tick marks with the legend (in both Figures 3 and 4).

Figure 10: This is not a good color scale to use. It makes HALOE appear to have an artifact in equivalent latitude with a distinct high-bias in the Southern hemisphere middle stratosphere when compared to the other panels. Or is it possibly an artifact of the use of equivalent latitude, which introduces too high values at these latitudes? Please test and comment.

References

Bodeker, G. E., et al., A vertically resolved, global, gap-free ozone database for assessing or constraining global climate model simulations, Earth Syst. Sci. Data, 5, 31–43, doi:10.5194/essd-5-31-2013, 2013.

Brinkop, S., et al., The millennium water vapour drop in chemistry–climate model simulations, Atmos. Chem. Phys., doi:10.5194/acp-16-8125-2016, 2016.

Froidevaux, L., et al., Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS): methodology and sample results with a focus on HCl, H2O, and O3, Atmos. Chem. Phys., doi:10.5194/acp-15-10471-2015, 2015.

Hegglin, M. I., et al., Variation of stratospheric water vapour trends with altitude from

merged satellite data, Nature Geoscience, doi:10.1038/NGEO2236, 2014.

Read, W. G., et al., Aura Microwave Limb Sounder upper tropospheric and lower stratospheric H2O and relative humidity with respect to ice validation, J. Geophys. Res., 112, D24S35, doi:10.1029/2007JD008752, 2007.

Interactive comment on Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-16, 2016.

C9