

## **Anonymous Referee #1**

This paper describes the construction of an important and comprehensive long-term dataset for key trace gases in the stratosphere. It is important to have a peer-reviewed and citable description of datasets like this GOZCARDS dataset. The manuscript serves this purpose very well, and I support publication in ACP.

My main criticism concerns the length of the manuscript. Over 100 journal pages, almost 30 figures, plus appendix and supplement are a tall order. I wish the authors would have made the effort to be more concise. At this point it is probably difficult to make the paper much shorter, but I have some suggestions:

*Response: This main criticism has been addressed by significantly reducing the size of the paper, following most of the reviewer comments/suggestions; we certainly appreciate all the reviewer efforts regarding suggestions to help address the length issue.*

*We decided to present the merging methodology in this paper but also included some original analyses (after having released the data for community analyses) in order to make it more relevant for ACP. We also include useful details such as data screening procedures, not always described well enough in past papers, in order to assist anyone potentially wanting to duplicate this dataset (and for completeness) – but now relegate most of this to the Appendix. We also believe that ACP papers (and this paper in particular) should not deal primarily with methodology - so we were not inclined to remove everything of interest regarding early sample results using GOZCARDS. One can get criticized the other way sometimes and get suggestions to try publishing in AMT instead, if one deals only or mainly with methodology (or techniques or validation). While this paper is still somewhat long because of the dual goal (methodology and some results + the description of a multi-species effort), we feel that at this point, a 23% reduction in the main text (and our cutting out 3.5 Figures out of the original 27) without eliminating most newer discussions/results is a reasonable compromise.*

### **1 Major comments**

GOZCARDS HCl sample results: Given recently published results, e.g. Mahieu et al. 2014, I feel that it is not necessary to discuss the HCl variations at various levels in such detail. I think Figures 7 and 8 already present all the relevant information. Figures 9 and 10, to me, are not needed (a look at Fig. 7 gives almost the same information). The discussion on pages 5865 last paragraph to 5868 first paragraph is also not necessary. I think it provides no really new or essential information, and is not necessary for the main topic of the paper, which is description of the overall dataset. Moving the important citations to the early part of section 3.3 would be enough, in my opinion.

*Response: We have made significant cuts here. However, showing that the HCl behavior is turning around in the later years is useful and new (and in our view, informative for ACP audience, and as a brief update to the Mahieu et al. work, which mainly looked at columns - and stopped in 2011). We believe that it is most useful to show (keep) the 2 left panels of Fig. 10 to show the non-linearity and connection to surface total chlorine changes. So we took out two thirds of Figs. 9 & 10 (combined) [we are just left with the new Fig. 9] and shortened the text significantly (by about 45% - see the new discussion before Sect. 4).*

GOZCARDS H<sub>2</sub>O sample results, pages 5873 to 5876: This reads almost like a review paper. Again, I think many readers will be familiar with the overarching issues, and would appreciate a briefer and more concise discussion. Figures 15 and 17 already give a very good description of long and shorter term water vapour variations, even without much text. Is Figure 16 really needed? Are all the (confusing and overall similar) thin and dashed lines needed in a dataset description paper?

*Response: Agreed, we have significantly shortened this H<sub>2</sub>O discussion (by almost 40% for section 4.3) and we simplified Fig. 16 (now with no thin lines), without taking it out altogether, as one can refer to this as a useful diagnostic for testing model variability.*

GOZCARDS Ozone sample results: Since GOZCARDS is primarily a stratospheric ozone profile dataset, I wonder if the long comparison and discussion with the Ziemke and Chandra 2012 column dataset is really that helpful. As the authors show, this dataset seems to definitely have a problem, and generally TOMS and OMI are not recommended for long-term trend analyses any more. It seems to me that this part of the current paper dates back a few years. There are several more recent analyses of ozone profile trends (Tummon et al. ACPD 2014, Harris et al. ACPD 2015, WMO 2014). I feel it would be much better to tie in the GOZCARDS data with these publications, and drop the comparison with the Ziemke and Chandra column data. At the same time, the discussion could be shortened, and number of figures in this part could be reduced.

*Response: We think that cutting out the column ozone discussion completely would remove a significant analysis as we have not really seen published work on this subject, despite what some may view or characterize as “common knowledge” regarding TOMS/OMI column data time series analyses. This ozone discussion may “date back” a bit, but it is new evidence nevertheless (many papers discuss “old data”), and the ozone topic is the primary reason to tie this paper to the SP<sup>2</sup>N special issue (on ozone only). The reviewer states “as the authors show”, and we did feel we needed to show something on this topic, as we did not see such a mention (to the ZC12 work) in the WMO Report, for example. Unless specific published references already exist to clearly make our point and/or to imply that other column series are superior to TOMS/OMI, we plan to retain the flavor of this ozone-related “sample result”. For brevity, however, we remove Fig. 23 and Fig. 25 and retain only Fig. 24 (Fig. 22 in the update); this preserves our point regarding comparison issues for GOZCARDS stratospheric column relative changes versus those provided by ZC12, but more succinctly. Overall, the Sect. 5.3 text has been cut by 37%.*

*Also, adding something really new to the ozone profile trends discussions at this point would likely require substantial efforts and a separate manuscript, given the assessments and discussions provided for GOZCARDS ozone profiles in WMO (2014) and in several other references – as mentioned in our ozone section. We think that briefly referencing/highlighting recent trend studies using GOZCARDS is sufficient, as there are a lot of details in those papers.*

Summary and conclusions could also be shortened. There, and in several other places, it seems to me that the manuscript resulted from adding newer references to an older previous manuscript version, and discussing them later / separately. E.g. Section 5.5.2 about diurnal ozone variations seems like a tag-on coming from more recent interest and publications on this topic. The reference to Mahieu et al. 2014 paper on page 5893 restarts the discussion again, making the paragraph unnecessarily long.

*Response: Agreed. We have shortened this summary significantly; the whole summary section was cut by 31%.*

The water vapour summary on pages 5894 and 5895 is also very long. It should be shortened and focused.

*Response: Agreed. We have shortened this summary significantly; the whole summary section was cut by 31%.*

Overall, the manuscript would benefit greatly from a thorough attempt at removing redundancies and shortening plus focusing on the main topics: Data set generation and sample results.

*Response: Agreed. We have shortened the main text significantly (by 23% - and deleted 3.5 Figures), also following another set of referee comments – but keeping in mind our introductory response (sample results are viewed as a useful contribution to ACP).*

## **2 Minor comments**

page 5851, lines 15 to 17: I find this sentence confusing and hard to understand. Maybe change to: In the upper stratosphere, however, SAGE II is not used after June 2000, when the (NCEP temperature-dependent) data conversion from a density/altitude to a mixing ratio/pressure grid exhibits a shift in the Southern Hemisphere. Instead HALOE data are used as a transfer reference from SAGE II before 2000 to the datasets starting in 2004.

*Response: Agreed, but this portion and the whole Abstract were shortened (per other referee comments), so the upper stratospheric details are now left out of the Abstract.*

page 5852, line 15: Here, and in most places, WMO 2011 should be supplemented by WMO 2014, or replaced by it. WMO 2014 was published in December 2014.

*Response: Agreed; in places, as appropriate, we have added the WMO (2014) reference (although in the first mention listed here, we removed the WMO reference for brevity).*

page 5854, line 6: Probably good to add "converted to mixing ratio versus pressure using their enclosed NCEP temperature profiles, and" after "were"

*Response: Agreed; we essentially used the above additional text.*

page 5854, line 8: Add a typical number of MLS profiles per day.

*Response: Agreed, done (about 3500 profiles).*

By the way, also related to the next paragraph: How is the standard error of the mean calculated in Figure A1(b)? From the Appendix and the Figure caption I get the impression that it is calculated by standard deviation divided by the square root of the number of profiles. This assumes that all the profiles are uncorrelated. This may be true for the few and far spread occultation profiles, but for the many close emission profiles it is probably not true.

*Response: Yes, this is how it is calculated (standard definition); the correlation issue is not trivial to accurately estimate, and the main result (many more data points from emission measurements) still would lead to these relatively much smaller errors (probably within a factor of two or better). Correlation occurs for some MLS bands (radiance) more than others but mainly along-track; in particular, the MLS measurement calibration (which can introduce some correlation) occurs on a much shorter timescale (a few minutes/profiles) than the typical profile*

separation in a monthly  $10^\circ$  zonal average, which mixes in many uncorrelated data points from separate orbits.

page 5854, lines 15, 16: cross-sections (UV, visible, near IR, far IR, microwave) could also be an issue.

*Response: Agreed, but this is implicitly included in the retrievals and their assumptions; if we were to produce all the potential reasons, this list would be quite long (e.g., see the MLS validation/characterization papers). Thus, and for brevity, we did not try to list all error sources.*

page 5855, 1st paragraph: Maybe add WMO 2014 as well?

*Response: Agreed, done.*

page 5856: I am missing a mention of SAGE screening. Maybe mention the discussion in the Appendix here.

*Response: The main data screening procedures have now been collected/summarized in Table A1 (Appendix A). We found a few species-specific details to be more appropriate for the species sections, but Table A1 is mentioned in these sections also.*

page 5860+5861: I find this whole discussion lengthy and difficult to understand. Could it not be shortened? Is it not simply like this: First AURA-MLS and ACE-FTS are shifted to have the same average level in 2004 and 2005. This level is equal to their combined average level. Then HALOE (which has fewer data in 2004 and 2005) is also shifted to this level.

*Response: No, this is not the way we adjust the third dataset and this will not provide the same answer as our procedure, as we wish to give equal weight to each of the three datasets (also note that the number of points changes with latitude and HALOE can often have more points than ACE-FTS, especially in the tropics). Shifting the 3<sup>rd</sup> dataset to the average of the first two is not the same as weighting them all three equally to get an average reference that depends on all three series. We describe our iterative procedure, which does use this equal weighting philosophy (since the first 2 series give a temporary merged result, then weighted by 2/3 with the 3<sup>rd</sup> series weighted by 1/3 for the last adjustments). Since a simplification was sought here, text cuts (by close to 30%) were implemented for the 1<sup>st</sup> part of Sect. 3.2 (Fig. 1 description of the HCl and H<sub>2</sub>O merging process), with little reduction in the main message. However, an oversimplification would be dangerous, for what is an essential description of the process. The Appendix has a mathematical explanation that should help, and this process can be carefully tested on sample datasets (and we have done this - also when considering the simplifications offered by the reviewers). The iterative process is what we decided upon as the best path, and it does take a few sentences to explain this. Note that this explanation paragraph is not a major portion of the full paper (or even the HCl section), but we did shorten this section. We also added some clarification words in the caption to Figure 1, which we hope can help.*

page 5865, line 23: With the addition of the Mahieu et al. 2014 references, and one or two additional sentences, section 3.3 could end right here. The rest of the section, until page 5868 line 15, is not needed, in my opinion.

*Response: We made some significant cuts here. However, showing that the HCl behavior is turning around in the later years is useful and new (and in our view, informative for ACP audience, and as a brief update to the Mahieu et al. work, which mainly looked at columns - and*

stopped in 2011). We believe that it is most useful to show (keep) the 2 left panels of Fig. 10 to show the non-linearity and connection to surface total chlorine changes. Thus, we took out two thirds of Figs. 9 and 10 (combined) and shortened the text accordingly; we are left with the new Fig. 9.

page 5868, around line 25: Is it certain that these low H<sub>2</sub>O values are wrong? Is there a reference for that? Could H<sub>2</sub>O from the gas phase have gone into Pinatubo aerosols, which consist of a sulfuric acid + water mixture? I have not done the numbers, so maybe this just shows my ignorance.

*Response: The water vapor averages are lower than usual and other tropical measurements from UARS MLS (for 22 hPa) do not typically dip lower than 3 ppmv (including 3 std. devs. from the monthly means, which are above 3.5 ppmv). The Pinatubo-related stratospheric SO<sub>2</sub> amounts are quite small compared to the 0.5 to 1 ppmv needed for these water vapor differences in 1992. In particular, Read et al. (GRL, June 1993) indicate that enhanced lower stratospheric SO<sub>2</sub> profile values (measured by UARS MLS) were only of order 10 ppbv in Sep. 1991. Thus, conversion to sulfuric acid aerosols would not be expected to deplete water vapor substantially on a large scale, even if shortly after the eruption it may have had some (more localized) impact. However, we added the following sentence: "While this method may exclude some good data points, the lowest values (< 3 ppmv) do get screened out; such outliers are not corroborated by 22 hPa UARS MLS data (with most values > 3 ppmv).*

page 5873, line 15 to page 5876, line 6: This is very lengthy, half a water vapour review paper. I think it should be shortened substantially. A few key messages, no literature review.

*Response: Agreed, we eliminated most references and shortened this significantly (section 4.3 is now almost 40% shorter), also in response to comments from another referee.*

page 5877, lines 1 to 7: I am missing clarity here. There are two aspects: 1.) SAGE retrieval needs density profiles. For V6.2 these came from NCEP operational, for V7 these came from MERRA. The effect of the different sources on SAGE ozone (number density vs. altitude) is very small 2.) Conversion to mixing ratio vs. pressure needs temperature/ pressure profiles. There, a large artefact arises in the SH upper stratosphere when NCEP is used instead of MERRA. If GOZCARDS would use MERRA here (and not the NCEP that comes with SAGE V6.2) the problem would be much smaller. I think this needs to be made clearer.

*Response: Agreed, we have reworded the relevant sentences (before Sect. 5.1.1).*

page 5878, line 20: Give some references for SAGE-HALOE comparisons, e.g. E. Remsberg et al., and others.

*Response: While the data versions in the past have often differed slightly, we do not know or find which Remsberg reference the referee might be thinking of; but we did add the SPARC (1998 Report) reference, providing overall agreement with our own results (shown in the Supplementary material).*

page 5879, line 27: I don't think SAGE ozone drifts. What drifts is NCEP / the conversion to mixing ratio versus pressure. Please correct.

*Response: Agreed, we clarified this further by rearranging the sentences and rewording (see discussion of Fig. 17, before Sect. 5.2).*

page 5881, line 17 to page 5887 line 21: Two very long subsections (review papers) that present lots of references, but very little really new information. Please shorten substantially.

*Response: Agreed, we shortened this substantially (also per comment of another referee). These combined two ozone sections are now shorter by about 28% (5.2.2 is shorter by nearly 20% and 5.3 by 37%). Also, two ozone Figures were deleted from the “sample results” section (5.3).*