Anonymous Referee #2

This manuscript describes the GOZCARDS data set comprised of several key stratospheric constituents. The goal of GOZCARDS is to combine data from multiple satellites into a consistent record suitable for long-term analysis. This paper describes the methodology used in constructing the data set, as well as provides analysis examples for each of the primary constituents. Continued long-term stratospheric time series analysis depends on our ability to combine data from multiple instruments, and as such documentation and validation of the methods used to construct this data set are very important to the research community. The manuscript is appropriate for publication in ACP with some notable revisions.

General Comments:
Overall the analysis and approach is good. However, I fear the length of the manuscript (27 figures, plus appendix and supplemental material) makes the paper in its current form very difficult to digest. I feel some of the analysis confuses the main point of the paper, which is the methods used to construct the data set, and the consistency between the source data sets used by GOZCARDS which directly affects the uncertainty of the final product. While I understand shortening the manuscript at this point is easier said than done, I suggest the authors consider whether some of the detailed analysis in particular can be removed, and possibly put into a separate manuscript. Though I'm not sure it would help significantly with length, in several sections the manuscript could be written more concisely. I will make a few suggestions in my specific comments, but general suggestions include writing in present active tense as much as possible, limiting qualitative descriptors, especially when the actual results follow in parentheses, and using tables as much as possible.

Response: This main criticism has been addressed by significantly reducing the size of the paper, following most of the reviewer comments/suggestions). However, we note that this paper is really the culmination of a 5-year effort to produce the datasets and to provide some results and overview of the possibilities, which are enhanced as well by other recent (or upcoming) papers analyzing ozone or HCl variations, for example. The paper has been shortened, largely following the suggestions from the reviewers and 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 / multi-year 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.

For example, Tables 1 and 2 are very informative, and easy to refer back to as needed. Can the general data screening information be put into a table? It is convenient to have all the filtering information in one place, but there are many details that are covered in the relevant publications that are not needed in the narrative. The authors might put the filtering information in a table, sticking just to the actual filtering and not including all the details as to what each filter is doing, for example: UARS MLS precision > 0, MMAF_STAT = "G", "t", "T", Quality=4, per Livesey et al (2003).

Conversely, if the authors do want to describe what each filter means (precision > 0 ensures only a negligible contribution to the retrieval from a priori, as an example), they might consider putting it in the Appendix as the SAGE filtering is done. Either way the information is presented in one place, but not in the main portion of the paper. There are other bits of information that seem to be repeated multiple times, such as the ACE_FTS data processing problems in 2010, that can be noted in the tables or in the data provenance information, rather than repeatedly in the text.

Response: Agreed, we have put the general screening in a Table (Table A1) in the Appendix but have kept some small discussion in the main text for the less well documented screening issues.

The description of the merging process seems overly complicated. For HCl and H_2O, if I understand correctly, the reference level is the average between mean HALOE values from August 2004-Nov 2005 (all that exist), mean ACE-FTS values over the same time period (all that exist), and mean AURA MLS values, but only averaged over the months that ACE-FTS has data. However, since the figures and discussion are keyed to a multi-step process, I do not recommend any large changes, just look for opportunities to simplify the text when possible. Also in the wording, keep in mind that the multi-instrument mean used as reference is not necessarily the correct answer, it is just a common reference. If AURA MLS were perfect, for example, it would still be adjusted in this procedure. In later discussion when referring to data sources that are biased low or high, I believe it is important to clearly state that they are biased high or low relative to the reference level.

Response: Since a simplification was sought here, text cuts (by close to 30%) were implemented for the 1ˢᵗ part of Sect. 3.2 (Fig. 1 description of the HCl and H_2O 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 (we have done this - also when considering incorrect 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.

Specific Comments/Wording Suggestions (many are minor and not mandatory)

p5850. General in abstract, don't need abbreviations unless you are using the abbreviation multiple times in the abstract (i.e. don't need MERRA and ESDR).

Response: Agreed, we deleted the terms "ESDR" and "MERRA" from the Abstract.

L9-L10: I'm not sure about the word debiased, it implies the full bias is removed, but in fact only the relative bias is removed. What about "To merge ozone records we first adjust each time series by its mean offset with respect to SAGE II during periods of measurement overlap to remove relative bias, and then average. For other species we use a multi-instrument mean computed during overlap periods as the reference."

Response: Agree to the "relative to SAGE II" modification (simple), but not on the 2nd simplification, which uses more words but does not state accurately enough what is done (we prefer to leave the details for the main text rather than oversimplify). Also, our slight revision is shorter than the above suggestion.

L17-18: and lower stratospheric Aura MLS data

Response: Agreed, done.

L20: On 6-8 year time scales [in general work to consolidate space by limiting the use of parentheses].

Response: Agreed, done.

L25-p5851 L2: this could be shortened: "Short-term tendencies of lower stratospheric and column HCl vary, with increases from 2005-2010 at northern mid-latitudes and deep tropics, but decreases after 2011 in northern mid-latitudes and increases at southern mid-latitudes."

Response: Agreed, done (even slightly more succinctly).

p5851 L17 I think you can end the sentence at reference and the meaning is clear [saving little bits of space]. L17: remove next sentence, this is confusing and not needed in the abstract.

Response: Agreed, done.

p5852 L10: of independent ground-based data (as in analysis of independent...)
Response: Agreed, done.
L19-20: A slow recovery of the ozone layer towards pre-1985 levels is expected...
Response: Agreed, done.
L22: High quality ozone and other constituent datasets are needed to document past variability and to constrain... [suggestion to reduce word count].
Response: Agreed, done.

p5853 L8: here is an example of how one might consolidate the words a bit: ... on a common latitude/pressure grid, using the following high quality satellite limb sounding instruments:[move list starting on line 15, and rest of paragraph, up.] Then change last sentence to "All source data sets still have shortcomings or imperfections, but in creating the GOZCARDS Earth Systems Data Record (ESDR) we maintain the integrity of the original data and do not arbitrarily

disregard data, nor do we attempt to fill in spatial or temporal gaps in the record." The sentence on ways to fill in data is beyond scope of paper and is not needed.

Response: Agreed, done.

P5854 L7: see Sect. 5.1.1 L8: profiles in two narrow latitude bands every day.

Response: Agreed, done.

L11-14: consolidate here [you've just introduced the source files and grid, so no need to refer back to these ideas]: Next we merge the GOZCARDS source data files by computing average biases between different source files during periods of overlap, and then adjusting each source file to a common reference to remove the relative biases. Non-zero biases ...

Response: Agreed, consolidation was done.

L16: such as differences in radiance measurement system or wavelength range, the retrieval algorithm, or in the vertical, temporal or spatial resolution. [no need to refer to Level 1 or Level 2, especially if you have to then define what you mean]

Response: Agreed, done.

L17: remove sentence starting "A useful reference" and start next sentence as Toohey et al. (2013) studied ...

Response: Agreed, done.

L21: '... vs. averages of output sampled at sub-orbital track locations'

Response: Agreed, done.

P5855 L4-5: remove clause "these analyses ... discussed later"

Response: Agreed, done.

L8: SI2N initiative (....) [remove 'which stands for'] L18: replace 'main' with 'source' [since source data sets already defined]

Response: Agreed, done.

L19-20: Section 6 briefly mentions GOZCARDS N2O, HNO3 and temperatures...

Response: Agreed, done.

P5856-5857 L5: As mentioned before, the filtering done according to various data teams can be summarized either in a table or in the appendix. **The special filtering for ACE-FTS is appropriate for the text body.** Also there is no discussion of SAGE here, or reference to the appendix that I saw. I suggest moving this information out of the main text body, but if it is not moved, SAGE should be mentioned.

Response: Agreed; as mentioned earlier, most of the data screening is now described in Table A1 (Appendix).

P5857 ACE-FTS: this discussion was a little confusing, and could be simplified. "When analyzing ACE-FTS data we found it necessary to remove occasional large outlier values that

could significantly impact the monthly zonal means. Our outlier screening procedure ... for each year of data."

Response: Agreed, done.

L12: What do the authors mean by independent zonal means from the ACE-FTS team? Is this a data set that the ACE-FTS filtered using their own method and it compares well? I'm not clear on how the zonal means supplied by the team are independent.

Response: Yes, the ACE-FTS team used a slightly different method; we added a brief clarification for this sentence.

L17: remove 'in the tropics' because the limits are extended in all latitude bands L17: remove parentheses around good.

Response: Agreed, done, and with slightly more consolidation.

L21-22: such data sparseness can increase trend uncertainties, for example.[I don't follow the rest of this sentence, comparisons with AURA gave confidence in trend results from ACE-FTS?]

Response: To remove confusion and length, and as this is not critical information, we have largely cut this out.

L24: add this last filter to the table or appendix, as it is standard.

Response: Agreed, this is now in Table A1 (Appendix).

L28: have been set to -999 in the GOZCARDS files.

Response: Agreed, done.

Section 3.1: This section is a bit confusing. First extra screening is covered, then the band deterioration on MLS, then validation studies, then more screening, more on the band deterioration and then more validation studies. Suggest adding "Froidevaux et al. (2008) found anomalously high MLS HCl vs. aircraft data at 147 hPa at low latitudes; these values are not included in the GOZCARDS HCl data set." to the screening info after the HALOE aerosol screening. Then summary of validation work. Then 'Finally, AURA MLS HCl data are not recommended for trend analysis at pressures < 10 hPa (reference?) even if monthly ... measurements. Aura MLS switched to backup band 14 after primary HCl band 13 showed signs of rapid degradation in early 2006. For pressures 10 hPa and above, the long-term band 14 data are considered robust, but drift in band 14 HCl in the upper stratosphere led us to not include AURA MLS HCl upper stratospheric source data in the GOZCARDS record.

Response: Agreed, we have rearranged and shortened this section.

Section 3.2: [the first two sentences don't add new information, but it would be helpful in Table 2 to note that the AURA MLS is included only from 10 hPa and below] 'Although not included in the final merged product, we do use the 2004-2005 absolute Aura HCl measurements at pressures less than 10 hPa to compute the offsets for the ACE-FTS and HALOE source files in a consistent manner at all pressure levels. Figure 1 ...

Response: Agreed, done.

P5860 L14: remove phrase in () L18: remove "too"

Response: Agreed, done.

L29: the average of the months when data from both instruments exist (i.e. at ACE-FTS temporal sampling).

Response: Agreed, done.

P5861 L4: we weight L11: thus removing any relative bias

Response: Agreed, done.

L15-16: checked against independent data sets from two institutions? This needs more detail, or to be removed. Did the authors compare to two other merged data sets?

Response: No, this was an independent computational check within our team (among co-authors at different places), but in the interest of shortening the text, we have removed this detail.

L29: is biased low relative to the reference mean and needs to be increased by the offset value.

Response: Agreed, done.

P5862 L7: detailed examples of upper and lower stratospheric offsets

Response: Agreed, done.

L 18-20: This seems a general data information statement, and should be moved up to the ACE-FTS screening section, if it needs to be mentioned explicitly.

Response: Agreed, done.

P5863 L17: ... merged values is given by the range of available ...

Response: Agreed, done.

L18: remove sentence 'We have made such a calculation....'

Response: Agreed, done.

L23: in relation to the reference values

Response: We have changed to: "biased significantly more than others, in a relative sense."

P5864 L4 ... comparisons; although not an official product, users can readily...

Response: Agreed, done.

L6-16: this seems a more general description of the GOZCARDS files. Should some or all of this information be covered elsewhere, before getting into individual constituent discussion? Maybe at the end of the data screening/binning section.

Response: Agreed, done.

Section 3.3 This section is very long, and seems like a good place to make some larger cuts. As a reader interested in the GOZCARDS data, I am most interested in GOZCARDS comparisons with other results. I would suggest at the very least removing Figure 10. The authors can simply state, as they have, that analysis showed the short-term drifts in the period from 2006-2011 to be

most hemispherically asymmetric. I would suggest ending this section at P5866 L27 (end sentence with GOZCARDS lower stratospheric HCl trends agree quite well.) This shows what GOZCARDS can do, which is what users need from this paper.

Response: We made some significant cuts here (Sec. 3.3 is cut by 33% per word count). 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 just left with the new Fig. 9.

P5865:

L1: Remove sentence 'The GOZCARDS HCl ..." this was covered already, and is in the Tables.
Response: Agreed, done.
L4: ... documented by satellite-based ...
Response: Agreed, done.
L12: ... for the one and a half year band 13 AURA MLS data record ...
Response: Agreed, done.

P5866: L2 less negative trends.
Response: Agreed, done.
L13: Is putting Fig 9a, 9b and 9c in parentheses required? It seems a bit odd, but it might be a journal requirement. If not, I would suggest removing the ().
Response: The Journal did this – but we have reworded this in the new version.
L12-13: remove 'For the results'
Response: Agreed, done.
L17: and both positive and negative values
Response: Agreed, done.

P5868 Section 4.1: consider writing in active voice. '... we screened HALOE H2O data for high aerosol extinction values, closely following the screening used in the Stratospheric ... '
Response: Agreed, done.

P5869 L7: (Version 20 or VPMC) could be used ...
Response: Agreed, done.
L16-17: do the authors mean the resulting UARS MLS source data span 1991-1993?
Response: Yes, we clarified this.
L20: The validation results here are confusing. Is it necessary to talk about V2.2 and V3.3 AURA MLS validation? Sticking to V3.3 might limit some of the confusion.

Response: Yes, we have limited this more to the V3.3 discussion, although not exclusively so, for proper referencing of important past validation work.

p5871 L3: Sentence #2, this sentence could be removed, as the reason given for not including SAGE H2O seems good enough, that speculation to retrieval problems after the volcano are not needed.

Response: Agreed, done.

L9-L16: Do the authors mean care should be taken because there are missing bands in the GOZCARDS? Since space is at a premium in this paper, I would consider taking this paragraph out. Most users should be aware of sampling issues when averaging without full coverage and it's already been made clear that missing values are not filled in GOZCARDS.

Response: A user came to us with this specific (sort of "blind plotting") result/issue, and thus, we feel that it is worthwhile to include a note especially for the lower altitude region (not to have users, even a minority of them, rediscover the wheel in this case). The merged dataset does not readily provide flags or latitudinal coverage information for the source datasets, and users would need to do more work to track this particular issue down (maybe they should, agreed…). We prefer to keep this cautionary note (latitudinal inhomogeneity issues in early/occultation datasets), given also the interest in H_2O at lower altitudes; however, we shortened this text significantly.

L18: remove "(mentioned earlier)"

Response: Agreed, done.

P5872 L10: with a large drift in the difference time series.

Response: Agreed, done.

L15: if the two time series

Response: Agreed, done.

L16: consider removing this sentence 'The main point here... ' [this is understood]

Response: Agreed, done.

L19: end sentence at ACE-FTS data [don't need to mention again data ending in 2010]

Response: Agreed, done.

P5873 Section 4.3 This section is well written, but it reads like a completely new paper, and possibly could be part of a new manuscript. For the purposes of this paper however it seems far too detailed, especially considering the overall length of the paper. One suggestion would be to use the first sentence of the section, then skip to Fig 15 through the end of the section.

Response: Agreed, the literature review has been removed, as it is not that directly relevant or necessary. Overall, Sect. 4.3 is almost 40% shorter now.

P5876 L14: remove commas around "is needed"

Response: Agreed, done.

P5877 L12 were discarded L24: It is known
Response: Agreed, done.

P5880 L 11: remove 'as described in Sect 5.1'
Response: Agreed, done.
L13: remove sentence 'The monthly means ... ' I think the construction of monthly means after screening is applied follows from section 5.1 and doesn't need to be restated.
Response: Agreed, done.

P5882
L9-11: Isn't the sentence starting 'Biases' at the end of line 9 repeating what was said in the line above? It might be possible to shorten the diurnal section, just noting that the data are simply not co-located in time (sunset to 1:30am is a large time range). Solving this issue really requires model output which isn't feasible for this work (at least this version).
Response: Yes, the redundant sentence was removed and some trimming of the rest was implemented although the point is that the relative bias adjustments include diurnal effects and one does not really need model output (with their own uncertainties) to do such adjustments, thus taking care of diurnal sampling effects, at least to first order. All datasets are basically adjusted to the SAGE II average local times (sunset and sunrise combined). Also, MLS (and ACE-FTS to a large extent) do not have significantly different (local time) sampling patterns year after year (see also the stability issue below).
L21: Only at 3.2 hPa and above, correct?
Response: Yes, we clarified this.
L18-27: This paragraph is a little confusing. Why is stability of measurements coming up now (assuming the authors are not still considering diurnal ozone variations). Is this not a consideration for all constituents?
Response: This was meant to focus on the diurnally-changing region. We clarified this, but also removed the last two sentences for brevity - as the stable sampling patterns of various instruments should be known.

P5886 L3 - end of section: The Ziemke Chandra results have always been anomalous, and it is probably not worthwhile pursuing this comparison. Updated column ozone comparisons could be made with SBUV MOD (which do not agree with ZC12); SBUV is considered the better "trend-quality" data set. However, given space constraints, it is not clear that a column ozone analysis is needed. The authors could easily end the section after the review of the GOZCARDS profile comparisons that have already been done.
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 is new evidence nevertheless and it is the primary reason to tie this paper to the SI^2N special issue (which is on ozone, not on other species). 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%.

P5888 L23: We now briefly mention the N_2O, HNO_3 and temperature GOZCARDS records that were part of the delivery...
Response: Agreed, done.

P5889 L6 end sentence after data record [again don't need to repeat this information about ACE-FTS if it is mentioned in the general data screening section and/or Tables].
Response: Agreed, done.
L11-14: wording is confusing. Suggest 'Until then the GOZCARDS N_2O record will include AURA MLS N_2O though the end of 2012 only, to avoid the discontinuities resulting from the shift to 190 GHz band N_2O in the current V3.3 data.
Response: Yes, we have reworded and shortened this discussion.
 L19-20: Is this sentence referring to the first validation studies mentioned? If so, combine these sentences. 'Validation results for the first few years of AURA MLS and ACE-FTS N_2O shows agreement mostly within 5% in the stratosphere (Lambert et al., 2007; Strong et al., 2008)'.
Response: We have shortened this brief validation discussion.
L25: is illustrated in Fig. S15, showing generally highly correlated fields and insignificant drifts.
Response: Agreed, done.

P 5890 L5 remove parentheses.
Response: Agreed, done.
L19: can remove 'the quality of'
Response: Agreed, done.

P5891 L2: and 2012 agree mostly within 10-15%
Response: Yes, modified slightly.
L15-18: remove parentheses.
Response: Agreed, done.

Summary and Conclusions: This section is also quite long, and will have to be changed to reflect changes made, particularly in the summary sections for each constituent. In the interest of space, a higher level summary, focusing primarily on the data set methodology, would suffice.
Response: We have shortened this summary section significantly (a total cut of 31%); however, we believe that sample results have their place in such a summary for ACP (and for many people who have or take time to read mostly Abstract and/or Summary).