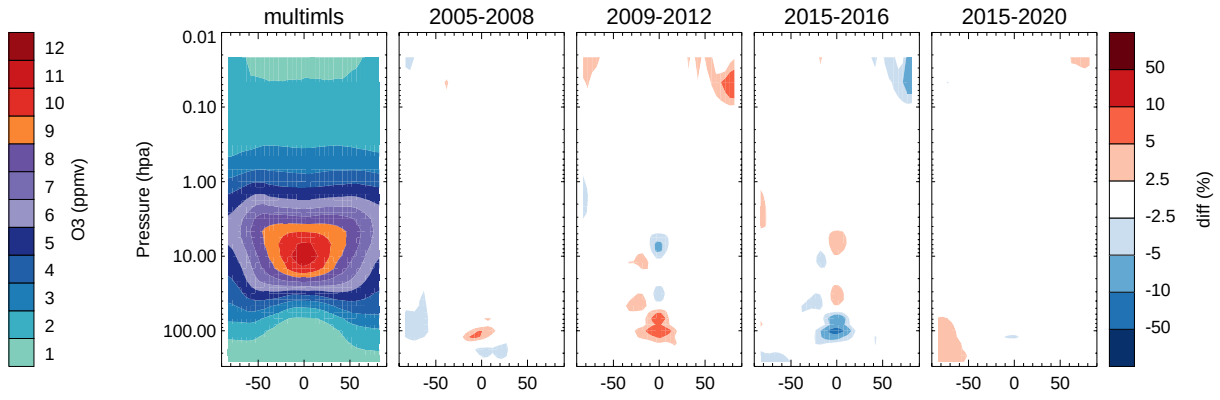
Review of
"Overview and update of the SPARC Data Initiative: Comparison of stratospheric composition measurements from satellite limb sounders"
by M. Hegglin et al.

This manuscript presents an overview an update of the SPARC Data Initiative, in particular of the assessment of all satellite limb measurements of stratospheric trace gases. The data initiative is a really important and ambitious project. It aims to evaluate the state of the available datasets comparing them against the multi-instrument mean to gain an idea of the usefulness of such datasets.

The methodology used is to compare zonal means to identify biases among the datasets. My main concern with this manuscript is that such approach is too simplistic as here described, the drawbacks of such comparisons are mentioned in the text (i.e., the possible sampling biases and the geophysical variability) but not really incorporated into the analysis.

For example, in Toohey et al (2013) some instruments can potentially have ozone sampling induced biases greater than 10% (or even 20% in some locations -see Figure 3 of Toohey paper). Yet the authors define excellent agreement as 2.5% through-out the entire zonal mean disregarding the location of such biases, I encourage the authors to consider the value of the sampling biases and to define the agreement criteria accordingly. So, for ozone below 100hPa and near the poles excellent agreement could be defined as between 10% due to the expected sampling biases.

Similarly, when comparing "cross-sections" (as in Figure 5) for different time periods, each "climatology" will be affected by different variability. For example, the figure below shows "climatologies" constructed using MLS data for different time periods.



As can be seen, some differences that will be interpreted as bias in the current analysis, are just natural variability (values up to 5 to 10% in the ozone example above). The methodology should identify areas of high variability and change the criteria agreement for such areas. They could also identify how many months are needed to decrease such variability within their current excellent agreement criteria and only include "climatologies" on such scales.

I understand that including the sampling biases and the variability into the analysis may be over-ambitious for the already ambitious project. But at least a more thorough discussion of the caveats of the agreement criteria is needed.

Comments:

P2 line 20: the vertical resolution is not 1 to 4km according to the satellite instruments section.
P2 line 28: There should be more recent citations for stratospheric dynamics and transport.

P3 line 4: Delete the in-preparation paper.

P2 line 28: Space missing between N2O5 and HNO4.

P3 line 29: add "and" between CH2O and CH3CN

Figure 2 is hard to read. Consider getting rid of the alternating blues background and filling the cells using solid colors instead of Xs. For example

SPARC data initiative	Mol1	Mol2	Mol3	Mol4	Mol5	Mol6	Mol7	Mol9	Mol10
Inst1									
Inst2									
Inst3									
Inst4									

P4 line 14: It is not clear to me what is meant by random sampling error? The sampling error is anything but random is determined by the sampling and the variability of the measured parameter. Are the "highly structured and transient features that may not be resolved by some instruments" refer on P4 line 25 what is meant by random sampling error? Please clarify.

Section 2: This entire section is inconsistent. Some instrument subsections have information about its vertical resolution, some have information about the spectral range and spectral resolution, some have information about which retrieval they are using, some explain the measurement concept (i.e. GOMOS), some have FOV information, or tangent heights. Please be consistent, if you feel the need to explain on detail for an instrument, then explain it for the other, regardless if it is given on the tables.

P5 line 22: Please add emission after nadir. They are other type of measurements using the nadir view.

Figure 3: What is wrong with SCIAMACHY and OMPS around 45W and 60S? Is this related with the South Atlantic Anomaly, if those please mention it here or in the SCIAMACHY / OMPS sections. This figure is also missing the sampling patterns of LIMS, MAESTRO and SAGEI. If MAESTRO sampling is

missing because is the same as ACE-FTS, please make sure to specify that in the ACE-FTS subplot title or in the caption.

P6 line 17: delete space before 70S

P7 line 2: "was about 3 to 5km" contradicts the statement made in the introduction "1 to 4 km"

P7 line 15: if POAMII and POAMIII tracks are identical why show them on Figure 3. (Just curious what happen to POAM1)

Section 2.6. I thought there were different retrieval versions for this instrument. Which one are you using and why?

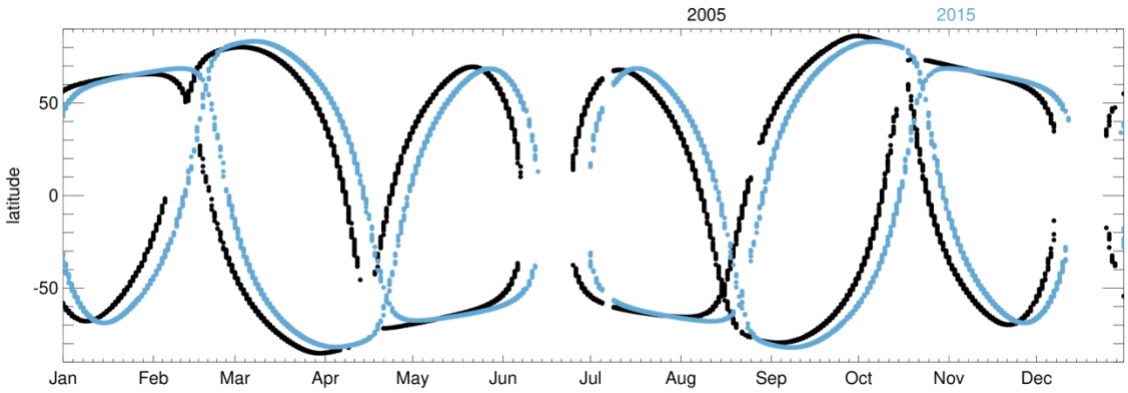
P8 line 14: This is the same principle of sun occultation which was not explained previously. Why explain it here? or better yet Why not explain the sun occultation and other methods the first time they are introduced...

P8 line18: altitude resolution of GOMOS is 0.5 to 1.7 versus 3 to 5 UARS MLS are you should you do need to use the averaging kernels. Could you prove that the comparisons do not improve when comparing so dissimilar resolutions?

P8 line 26: I thought it was 0.025 cm-1 (see for example https://amt.copernicus.org/articles/2/337/2009/amt-2-337-2009.pdf)

P9 line 3: Please justify why are you using this retrieval and not any of the many others…

P9 line 23: "the latitude scanning (see figure 3) is the same each year". That is not true, it changes slowly, the pattern may be the same but not the time of the measurements, see below:



P9 line 25: Methodology discussions should be moved to section 3

P10 line 5: Again, I am assuming that their sampling patterns are identical. In that case, Figure 3 should be labeled differently so that the reader know that their sampling pattern is the same..

P10 line 8 Methodology discussions should be moved to section 3

P10 line22: The instrument was not damaged. Please change to: Unfortunately, during launch, most of the aperture was obstructed by a plastic film used for insulation that became detached during the ascent to orbit. (or something similar) https://doi.org/10.1117/12.623574

P10 line 24 Coverage is 65S – 82N https://doi.org/10.1029/2007JD008824

P11 line 11: Coverage is between 38S-65N in north looking days and 65S-38N on south looking days so, as shown in figure 3, I could be argued that the coverage is actually 65S-65N. What does nominal mean in this instance?

Section 2.17: I do not believe TES is used at all in the manuscript. Please delete this section or include examples using TES data.

P11 line 18: the apodized or unapodized detail was not mention for MIPAS.

P12 line10: The data preparation and handling of the dataset is a really important step of the whole endeavor. Please provide a brief description here, or in each instrument section. Also, the SPARC data initiative does not have information on the new datasets included in this manuscript.

P12 line19: The conversion from altitude to pressure levels as well as the conversion from number density to volume mixing ratio will introduce an uncertainty. According to 10.5194/amt-9-2497-2016 this bias could be up to 5% in the upper stratosphere (Figure 8/section 6 of that paper).

P12 line16: is not clear to me what is meant by hybrid log linear interpolation, do you mean that you linearly interpolated on log(pre) instead of on pre. A brief explanation will suffice.

The terminology on Table 3 is barely used through-out the rest of the text. I recommend changing the blue-red color bar of Figures 5, 6, 8 etc, to a blue-white-red color bar with the white covering from -2.5 to 2.5. That way the reader could easily identify where the excellent agreement is found.

P12 line 24: The authors seem to use climatology file when they meant zonal monthly mean files. It is confusing.

Figure 4: Again not climatology file, zonal monthly mean file, space missing between and-LST_MEAN. MIPAS was not available in 2018. Do you mean 2008?

P12 line 29: Toohey et al 2013 is not the only study about sampling biases for limb instruments, there are more, for example: 10.5194/amt-7-1891-2014, 10.5194/acp-16-11521-2016, 10.5194/acp-18-4187-2018, 10.5194/amt-12-2129-2019

P13 line 26: How do the authors define if measurements from a given instrument are deemed unrealistic. I think the MIM will be more robust if the authors defined quantitively which measurements to include or exclude.

P14 line 17: Since the second summary is not going to show here, please delete this paragraph.

P16 Line15: Why are the author evaluating different retrievals for OMPS? If they are going to do that, they should include all the retrievals from MIPAS and OSIRIS, etc.

P16 line 23: Wang et al is published already. And it shows how to get rid of the altitude registration problem. Such correction should be implemented for SAGEIII/ISS. As well as for SAGEII since that product is also affected.

P17 line 27: I thought kernels were not considered at all through the whole study. Further I do not see what retrieving in log space has to do with the kernels. The averaging kernels are concentration dependent disregarding if they are retrieved in log space or not. Retrieving in log space only ensures that the retrieval will be positive.

P19 line 1: instead of the current figure 5 why did the authors do not show the figure 5 and 6 equivalent for CH4. It will presumably show the same results as Figure 7 but I will be consistent with the previous to sections.

Figure 7 what is the shaded region is this the excellent agreement region or the very good please specify in the caption.

Figure 8: the color bar units are missing. Is the level in the differences really 2 or 2.5.

P20 line11: Not “climatologies”, zonal monthly mean from SMR, MIPAS, …

P20 line18: Again, be consistent in Figure 5 and 6 you only show the MIM, not the individual instruments zonal means. Now, looking at those, why are the authors including SMR into the MIM when there is a clear high bias around 20hPa in the tropics and a clear artifact around 60N.

P21L16: Were some instruments not included for O3, H2O in figure 14 due to differences in time period, Why is this important for CO but not for H2O, O3, etc?

P22 line 3: Why was ozone at higher altitudes not corrected using a chemical box model? There is a strong diurnal variation in the US-LM affecting those concentrations.

P22 line 25: Why does figure 11 does not include the MIM? Also, this figure should include panels showing the difference versus the MIM to be consistent. How are the authors evaluating the good agreement (P22 line27) or the agrees well (P22 line31) or the very good agreement without it.

What does the “s” stands for in ACE-FTS s10am etc?

Figure 12: What is the “s” or the “ss”

P24 line 1: sorry if this is obvious and I am just not understanding it. How come if you compute anomalies as monthly – MYM (multy year mean) might display a diurnal cycle while doing (monthly – MYM) / MYM *100. does not?

P24 line 11: “possibly due to its higher vertical resolution” This statement could be proven by applying for example the MIPAS averaging kernels to HIRDLS and repeating the comparison.

P24 line 17: Again, please be consistent, in previous sections there has always been mention of the previous evaluation.

P24 line32 SMR data is not used at all in Figure 13, i.e., the only comparison shown in the manuscript.

P25 line 1: please include the equivalent of Figure 5 and 6 for HO2. Showing the MIM and the difference versus the MIM for SMILES, MLS and SMR before discussing the November 2009 and February comparison.

Figure 13: are you comparing the MIM for 2 instruments. If you are, then just shown the MIM and then the differences. Were the SMILES selected according to their local time or all local times were included in this comparison?

P25 Summary evaluations. Is the reader supposed to be using the definitions on Table 3 with Figure 14, 15 and 16? If it is, the color bar should reflect the values determined in such table 3.

P26 line1: Why is there no HO2 1-sigma multi-instrument summary as for the other molecules? There are other molecules with only 2 or 3 instruments as for SF6, HF, NO, HCl.

Why is there no BrO 1-sigma summary?

Figure 15 is not consistent with Figure 14, I understand that the authors are redoing the Figure from the SPARC DI but for this paper it will be much better is they kept a simple layout as in Figure 14. That is, remove the boxes and the chemistry explanation, etc.

P26 line 11 The acronyms (CSA, ESA, JAXA, etc) are not are not defined.

P26 line 16 These manuscript only updated the trace evaluation, as written it implies that it also evaluated the aerosols observations.

P26 line 26: The doi zenovo link those not exist. As in, google cannot find it (neither the zenovo webpage). You have to go into the webpage and search for Hegglin to get the dataset. Please clarify

P27 line 5: Why generally? It always produces larger sample sizes.

P27 line 9: Toohey et al 2013 only shows the sampling biases for Ozone and water vapor. As mention here it seems that it investigated the sampling biases for all molecules included in the SPARC data initiative.

Table 1- References column: Why are some years in brackets and some don't? Cloud top is not an appropriate vertical range, please provide the lowest possible measured altitude in clear sky and then either on the text or as a tablenote specify that in the presence of clouds is from the cloud top. Why do HALOE, UARS and SMR have pressure ranges as well as altitude ranges. Please be consistent.

Table 2: Why does MIPAS have two vertical resolutions? Due to the change in spectral resolution? Please clarify.

What happened to the ascending LT of meas for SCIAMACHY?

How come TES has a LT of meas but not MLS and HIRDLS.

For SAGEIII/ISS - Wang et al has already been published.

Table A1 I think this should be part of the main text. Also, I think the variable type is wrong:

N2O, N2O_NR and N2O_STD are 3d, [lat, pre, months]

AVE-dom, ave-lat, lst's are 2D. [lat, months]

But all are described as GEO2D.

Further, looking at some files, for example the OMPS-SASK or OSIRIS have other variables:
OZONE_CONCENTRATION_STANDARD_ERROR, OZONE_VMR_STANDARD_ERROR
OZONE_COUNT_IN_BIN

Table A2. Why do some molecules used different versions? For example, MIPAS (1) uses v21 v20
Table A3. Why is v4.0 in SAGEIII/M3M in brackets?
Table A4. Occultation needs to be bold in the Stellar or lunar occultation section.
Table A5. LIMS reference, is there no published reference. This dataset is from 1979? Perhaps this was meant to be Remsberg 2009?
MIPAS additional comments: Meas mode switched in 2004 from high spectral to low spectral resolution. Not from high to high
SAGEIII/ISS reference: Wang et al has been published.
HALOE: the vertical range states up to 80km, please provide the minimum altitude

Table A6. MIPAS: Meas mode switched in 2004 from high spectral to low spectral resolution.
SAGEIII/ISS cite 10.1029/2020JD033803.
Is there an upper vertical range limit for UARS MLS ?

Table A7 The additional comment for MIPAS in the previous tables is that the Meas mode switched in 2004 from high to low …, in here it says Change in spectral resolution in 2005. Please be consistent. And also the years do not change.

Table A8 should include https://doi.org/10.1029/2007JD008723 for MLS.

Table A9: In the previous tables the authors have given the overall period for MIPAS. I Actually prefer the format in A9. But be consistent.

Table A10. No references for SAGEIII and SCIAMACHY, at least provide the one describing the instrument.
The day/night in MIPAS imply that you are taking the day night difference? If it those, this detail should be in the additional comments.

Table A11: I do not understand the additional comment for UARS MLS, please clarify
Please use the vertical range MIPAS layout for all cloud top's in these tables.

Table A12: Why is there a (UTLS) in the AURA MLS row. The resolution according to the v3 quality document varies from 4 at 10hPa to 10 at 0.046 hPa. Please clarify.