Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-33-AC2, 2018

Atmospheric Measurement Techniques Discussions

Interactive comment on "The SPARC water vapour assessment II: Comparison of stratospheric and lower mesospheric water vapour time series observed from satellites" *by* Farahnaz Khosrawi et al.

Farahnaz Khosrawi et al.

farahnaz.khosrawi@kit.edu

Received and published: 25 June 2018

We thank Hugh Pumphrey for the constructive, helpful criticism and the suggestion for revision. We have revised the manuscript accordingly.

General comments
This is a useful summary paper, which, for the most part, presents a large and complex body of information in a digestible manner. Most of the items that make it difficult for the reader are related to the large number of datasets from a single instrument

(MIPAS). If it were possible to do anything to separate the inter-MIPAS comparisons from the (more important) comparisons between different instruments, then I would like to see this done. But I recognise that this might be too large a change to be made. Like referee 1, I am conscious that this paper does not provide any kind of guidance to the reader as to which data sets are the most useful. I imagine that this is deliberate and is done to avoid annoying any of the data providers. I nevertheless feel that some sort of opinion as to which datasets are the most useful for which purposes would not be out of place.
We can understand that the huge number of MIPAS data sets is somewhat overwhelming. However, since these data sets exists, they also have a right to be assessed. These data stem from 4 different processors and there are a lot of differences between the data sets as shown in our paper as well as in the other WAVAS papers. The intention of WAVAS is to provide a full assessment of "all" available stratospheric data sets. Plenty of time series analyses and assessments using less data sets can be found elsewhere (e.g. Hegglin et al, 2013; Hegglin et al., 2014; Khosrawi et al., 2016; Weigel et al., 2016; Noel et al., 2018; Lossow et al., 2018). We included now an paragraph in the manuscript on the differences between the MIPAS data sets (see our answer below to the specific comment on Figures 2-4). We agree that it would be good to give some guidance on which data set to use for further studies. However, this is quite difficult to judge since this decision depends on the scientific application and on which altitude region or latitude region the study is focused on. For trend analyses the longest data sets with the highest spatial and temporal coverage have of course for such studies a clear advantage. We changed the last paragraph of the conclusion as follows and hope that this will give at least some guidance: *Nevertheless, although the water vapour data sets have been thoroughly assessed in this study it is difficult or rather impossible to judge on which data set is the best one to use for future modelling and observational studies. This simply can only be answered with respect to the*

specific science application the data set should be used for. For future studies on e.g. water vapour trends we can state that the data sets that provide the longest measurement record with a high spatial and temporal coverage have an advantage over the ones which provide only observations in specific latitude bands and/or altitude regions. For data sets that have a drift relative to other data sets as e.g. SMR 489 GHz, the drift has to be taken into account and data sets that are simply too short (less than one year) as e.g. ILAS-II and SMILES cannot be used for trend studies at all.

Specific comments

- *Page 4 line 10: The authors note that the data from UARS MLS are not considered. I do not think these data would add much as there are less than 18 months worth. But the authors have included the ILAS-II and SMILES data, which cover even shorter time periods, so I think they should explain why they are including ILAS-II and SMILES, but not including UARS MLS. (Disclaimer: I am responsible for the UARS MLS water vapour data.)*
  The aim of WAVAS II was to include all data sets that performed observations in the period from 2000 to 2014 (or extended to 2016 as it is done in some other WAVAS II papers). To our knowledge UARS/MLS H_2O measurements ceased in 1993 and that only measurements from the other trace gases are available until 2001. An assessment of the pre-2000 data sets was done within the first WAVAS project and can be found in the SPARC WAVAS Report published in 2000.
- *Figure 1: The labelling of the colour bar is rather cluttered; it might be preferable to label only 2,3,4,5,6,7, and 8 ppmv.*
  We agree and changed the labeling of Figure 1 as suggested.
- *Page 8 line 14: I would remove the words "(contour time series)" as the data are*

presented as an image âĂŤ no contours have been drawn.
It is correct that we have not drawn any contour lines. However, the definition of a contour plot is as follows: "A contour plot is a graphical technique for representing a 3-dimensional surface by plotting constant z slices, called contours, on a 2-dimensional format. That is, given a value for z, lines are drawn for connecting the (x,y) coordinates where that z value occurs." This exactly what we are doing, but instead of using lines we use filling of the contours. Thus, it correctly should read "filled contour". However, this is detail is not very useful and somehow we need to distinguish our time series plots from each other and thus we would prefer to keep the header using the term "contour".

- *Page 8 line 27: Again (and in several subsequent places, including in the supplement), remove the word "contour" as figures 1, 5, and S1-S3 contain no contours.*
  As stated in our answer above, these are nevertheless contour plots, but without explicitly plotting contour lines. Thus, we would rather keep the word "contour" in the text and figures to differentiate these figures from the other time series plot where we consider the time series on specific pressure levels.
- *Figure 2-4: I do rather wish that the various teams involved with MIPAS would agree on one best product. Half of the products shown in these figures are from this one instrument. I understand that the instrument has various operating modes which are not directly comparable, so a single product may not be practical. But 13 different products are very confusing for the reader and the data user. It might have been preferable to first form some sort of combined or approved MIPAS dataset (or, at most, one for each operating mode) to be compared to other instruments. I do not imagine that the authors will want to re-design the entire paper along these lines. But for the purposes of these figures it might be better to show only one MIPAS dataset (and possibly, only one ACE-FTS data set âĂŤ*

why do we need V2.2 if V3.5 is supposed to be an improvement?).
We can really understand this point of criticism since at a first glance including all 13 MIPAS data sets looks a bit like an overkill. To simply pick one data set (or a selection of data sets) from MIPAS is not possible due to the differences between these data sets (due to usage of different microwindows, different retrieval choices etc). We have to apologize here that we completely missed out to motivate in the manuscript why we want/need to include all MIPAS data sets in this comparison. Therefore, we included in Section 2 a similar paragraph as the one given in Nedoluha et al. (2017) on the differences of the MIPAS data sets: *This especially holds for MIPAS where 13 data sets have been included in this comparison. The MIPAS measurements are processed by four different processing centers: (1) the University of Bologna (Dinelli et al., 2010), (2) the European Space Agency (ESA; Raspollini et al., 2013), (3) IMK/IAA (von Clarmann et al., 2009; Stiller et al. 2012), and (4) Oxford (Payne et al., 2007). The four processors differ in several respects, such as their choices of spectral ranges (so called micro-windows), the vertical grid on which the retrievals are performed (pressure or geometric altitude), the choice of regularization (and related to this, the vertical resolution), the choice of spectroscopic database, the sophistication of the radiative transfer (in particular, whether or not non-LTE emissions are considered), and whether or not any attempt is made to account for horizontal inhomogeneities, and the a priori and the assumed p-T profile. Indeed, the temperature used might be a large source of error for species retieved in LTE regions. Some of the different processing schemes also make use of different level-1b data versions (here V5 and V7) based on different ESA calibrations. The spread of results seen for MIPAS indicates how specific choices within a retrieval approach may influence the retrieval results.* Selecting one specific MIPAS data set (the best one, obviously) might rather be an outcome of this study but not an input. Re-

garding the two ACE-FTS versions that are included in this assessment: We had an open data set policy to represent a database as complete as possible. All data sets were allowed to participate. The ACE-FTS team wished to include both data sets, v2.2 (well validated) and v3.5 (not really validated, covering a longer time period).

- *Page 10-12: Many of the features of the data described here are rather hard to see in figures 2-4, on account of the large number of lines. I am not sure what to suggest (other than not showing all the MIPAS data!).*
  We agree that with such a high number of data sets the features we described here becomes hard to see. However, this is the drawback of performing a multi-dataset assessment. For the sake of completeness it is important to have all data sets included. Nevertheless, by just zooming into the figures we managed to see these features despite the high number of instruments. Nowadays, many scientist anyway read papers rather on the computer screen than printing them out. Separating the MIPAS data sets from the other data sets is no solution since then we would not be able to include MIPAS into the comparison (since picking one data set is also no option as discussed above). Comparisons of water vapour time series using less data sets are published elsewhere (e.g. Hegglin et al., 2013; Hegglin et al., 2014; Weigel et al., 2016; Khosrawi et al., 2016; Noel et al. (2018); Lossow et al., 2018). Further, there actually has been some optimisation in the plotting sequence of the time series with the aim to benefit the sparse data sets. Furthermore, the time series analyses shown Figs. 2-4 provides only a qualitative assessment. From these figures we learn more on the characteristics of the data sets when we look at the outliers instead of on the data sets that agree well with each other. Further, the more important results in this study are the assessment of the correlation and drifts where we overcome the problem of the huge amount

of data sets by using the matrix plots and giving quantitative estimates of the differences.

- *Figure 5: The black dots are very difficult to see, especially against the darker end of the colour scale. Potential solutions include joining the dots with a line and/or using a colour (red?) which does not form part of the colour scale.*
  Thanks a lot for the suggestion. We increased the size of the dots and changed the colour from black to red.

- *Page 14 lines 1-15: One of the most striking features of the figure is the change in 2012 caused by the end of the Envisat mission and hence of the myriad MIPAS datasets. It strikes me that the use of the max-min difference to quantify spread means that this plot mostly tells you about where the noisiest dataset is at its noisiest. I have to question whether this is the most useful measure of either atmospheric variability or overall data quality.*
  We have tested several methods to calculate the spread and derived qualitatively the same results. We prefer the spread calculation using the max-min differences since it makes the spread calculation most comprehensible and shows most clearly that the largest spread between the data sets is found where the largest variability in H_2O is found, in agreement with what was found in Lossow et al. (2017). Further, it should be noted that a pre-screening has been performed to remove outliers and to get reasonable estimates. It is correct that a striking feature is the change in 2012 due to the end of the Envisat mission. However, keeping these two years in the figure is worth since it quite clearly shows that with a few data sets the spread is decreasing, but the characteristic features (largest spread found in the areas of largest variability) are not that pronounced any longer. Thus, showing that a few data sets are not sufficient to get a good statistic.

- *Figures 7-9: These figures are an interesting way of showing a large amount of summary information in a clear way, and in a small space. Something that caused me a bit of confusion was the way that the numbers in the upper triangle do not always align with those in the lower triangle. This is because different levels have different datasets available. It might be worth inserting blank rows into one or other triangle in each pane so that the two triangles have the same numbering scheme.*
  We agree, but we have to keep these gaps to save space, because otherwise the boxes get even smaller as they already are.

- *Figures 11-13: In addition to the suggestion I make regarding figures 7-9, figures 11-13 have text on them which is VERY small. It is commonly recommended that text on a figure should be no smaller than the figure caption text in the final typeset version of the article. There is clearly a bit of leeway on this recommendation, but the text on this figure is so tiny that it is very annoying for the reader, especially for middle-aged readers who are still cross that they need reading glasses. I am not sure what to suggest here, because simply making the text bigger will not work: in some cases it is already impinging on the diagonal lines.*
  We agree that the numbers in the boxes are really hard to read. However, we really tried to find a solution for this problem, but could not come up with a better idea. Nevertheless, the numbers are additional information and the most important information in this figure is the drift given by the colours and the colour bar as well as if the drift is significant or not by the green boxes and a slash. For reading the numbers one in fact has to use the pdf and zoom in. However, to make it easier for the readers who prefer a paper version, we shortened the caption of Fig. 11 so that the size of these figures is now a bit increased and added these three figures to the supplement where we can provide them in a even larger size than in the manuscript.

- *Page 21: dedication. I too have good memories of working briefly with Jo Urban, and was saddened to hear of his passing at such a young age.*

  Technical corrections

- *Page 2 line 1: "allowed considering the time period" reads rather oddly. Maybe write "allowed us to consider the time period" or "allowed the consideration of the time period".*
  We have changed this phrase as suggested.

- *Page 3 line 17: "One drop (also known as the millennium drop) . . . " The "also" does not read right as you have not first given another name by which the drop is known. Maybe write "One drop (sometimes known as the millennium drop) . . . ".*
  We changed the sentence as follows: *One drop (sometimes denoted as the millennium drop) occurred in 2000.....*

- *Page 11 Line 25: remove comma after "Both"*
  Done.

- *Page 14 line 30: replace "than" with "as"*
  Done.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-33, 2018.