Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-33-RC1, 2018

Atmospheric Measurement Techniques Discussions

Open Access

EGU

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.

Anonymous Referee #1

Received and published: 4 April 2018

General Comments

This manuscript presents useful analyses regarding intercomparisons of H2O time series from many satellite measurements, in terms of monthly means, data spread, correlation coefficients, and drifts between the time series, as part of the SPARC WAVAS-II assessment. The methods are generally sound, although some aspects could/should be clarified, and some plots take a while to digest, as there are many curves and data sets to consider, which also makes a clear/useful summary somewhat difficult to present. Different readers or investigators may proceed differently in terms of how to use or discard certain data sets, based on these sorts of results. I find that (at least) more caveats regarding the error bars and significance levels would be useful, given that autocorrelation effects are ignored. I also find that the conclusion regarding H2O data usage is fairly bland and "politicially correct", but not very useful scientifically, since some data sets clearly exhibit more outliers or drifts than others. Such issues are not easy to deal with in terms of trying to assess the trends in H2O, ultimately, and this manuscript does not try to guide the reader in that direction, which is maybe alright for a more limited and well-defined scope in this paper. This seems to imply that everyone should try to draw their own conclusions about how to decide, which may indeed be better than trying to impose only one particular solution, and there would be a lot of extra work involved to make these sorts of assessments. This work may well be followed, in time, with more details in a related manuscript (mentioned in this work) or by other work on H2O trend assessment, and this manuscript should stand on its own as well. I include some suggestions for improvements in the specific comments below; there are also a few statements where not enough information is provided. Overall, this work will provide benefits to investigators of global stratospheric H2O; after some changes based on the suggestions below (ranging from minor to somewhat more major), it could be made (even) more suitable for publication, and I would want to see it published (my recommendation really stands somewhere between "only minor" and "some more major" revisions).

Specific Comments (some in between minor and major)

1. Impact of different vertical (and horizontal) resolutions: I realize that this is a somewhat difficult topic to deal with, but one should at least acknowledge that these differences between measurement systems can lead to differences in the time series (which are then interpolated to a fine grid); for example, a tape recorder signal will not look exactly the same to different instruments. If you could touch on this for some of the highest and lowest resolution instruments (ignoring the horizontal component), and comment on whether you think there are impacts for these results, this would enhance

the paper quality. Also, indicating the actual "canonical" vertical resolution for a typical stratospheric measurement (e.g., under the instrument names in Table 1, or by adding a column to this Table), would provide useful information that the reader does not have to try to fetch elsewhere, or remember. There are enough assumptions made already about the readers' knowledge of each instrument system (even regarding first-order information like vertical resolution).

2. Impact of known issues in certain measurements: There is not quite enough discussion, in my view, of certain known issues that could play a significant role in these intercomparisons. Each instrument team representative could (have) provide(d) more feedback on actual knowledge of instrument degradation issues or known drifts. For example, there have been issues with drifts in MIPAS ozone in the literature (which are touched on briefly in the context of some of the MIPAS ESA versions), and what about H2O? A clearer summary, up front in the drift section for example, regarding what retrieval versions have some correction and which ones do not, would be useful. There has also been some evidence for drifts in the MLS measurements versus sonde data (as presented by Hurst et al.); is this detectable in the plots or drifts you come up with here? Also, is there another part of this WAVAS assessment that attempts to consider drifts with respect to ground-based measurements, and would that not be a good cross-reference to consider, if so? (If there is not, that will be work for the future, I reckon - and these assessments undoubtedly take up a lot of time and work, I am aware of this, not trying to downplay the useful work that has been done already).

3. Some of the drifts are quite large, and each instrument's results tend to get a bit buried in the sea of curves (e.g. in Figs 2 through 4, and Figure 10; showing Figure 10 for all latitude bins could also be helpful, in the main text, even if Figs. 11 through 13 are quite nice, but somewhat less easy to grasp than Fig. 10. One also wishes for a more quantitative summary of the accuracy of expected trends based on the "combination" of knowledge shown here, even without getting into the trends themselves. The spread between the curves in Figure 10 could be such a measure, say in percent/decade

rather than ppmv/decade (just a personal preference but not critical since the vertical gradients in H_2O are not as strong as for O_3, for example). First, one might want to eliminate some of the outliers (e.g. despike the drift results using techniques you have used with the MAD for example, or some 4 or 5 sigma type of screening), and then calculate the rms spread versus pressure. One could also superpose three curves (one for each latitude bin you have considered). In fact, in an ideal world, showing this for even more latitude bins to check for consistencies or inconsistencies (and systematic effects for certain instruments) would provide an even more complete picture (such as in a latitude/pressure contour plot). As comprehensive as this work looks already, there is even more information to go after, as you imply in reference to other manuscript(s) in preparation or in press - and this is not something that is a serious flaw, as long as there is some level of consistency between the latitude bins chosen here (if not, then the reader can conclude that more work is really needed to try to make sense of all these measurements). Even a 0.5 ppmv/decade drift is fairly large, since this is about 10% and the trends in H_2O (or expectations for long-term trends) are not larger than this.

4. Your final statements (in the Abstract and in the Conclusions) about being able to consider "all data sets... when data set specific characteristics (e.g. a drift) and restrictions (e.g. temporal and spatial coverage) are taken into account" seem to be too much of a "politically correct" stretch, even though I realize that this is often done to please every team making up an assessment-type paper. What is really required (besides a lot of work) to try to actually take such effects into account and really assess trends in H_2O (as is done for ozone)? This is certainly missing from this work - precisely because this is a lot easier said than done. I will not try too hard to force a different consensus view or statement, but it is something to reconsider, I would argue, in terms of what the best message for readers really is, as a scientific statement about the uncertainties and possibilities, given the large spreads or at least, the existence of several outliers. How is one supposed to "consider" known drifts, or the large spread in (some of) these results? Either they all get characterized better versus ground-based

data (if and where possible), or versus some "cleaner" average satellite time series - or some other solution, if a more satisfying recommendation can be pursued. At the very least, I am asking for some thought about this and an attempt to make a more useful statement, even if this may just be a suggestion in order to arrive at a hopefully robust consensus about the state of the trends in H2O. I am not asking that all that work be done for this manuscript, just for a better suggestion than "use everything but consider everything", which is basically not providing any useful recommendation in terms of specifics. If it really is too difficult to arrive at a better (consensus) statement, saying that this was attempted is still better than just leaving the vague conclusion in as it stands now; hopefully this stimulates a bit more discussion, at the very least (again, without requiring a lot of detailed work). This may have been a point of discussion already among co-authors.

5. It is stated that the error bars (and drift) estimates do not take into consideration in the regressions (see the statement near the top of page 6). I assume that this is also the case for the drift estimation (calculated via a simple linear trend model applied to the time series of differences). Since a lot of the results depend on the significance level, the underestimated set of error bars, which is typically the result of the neglect of autocorrelation effects, will imply that somewhat erroneous conclusions are arrived at whenever you discuss significance levels. This could often be a non-negligible effect indeed. The trend estimates are not likely to change very much, and this is more of an impact on the error bars themselves, which means that you would have more slant lines across more boxes in Figures 11-13, and Figure 10 would also be affected (mainly for the panel on the right). I realize that this is also a lot of work to try to do rigorously, so any rough estimate of the impact of this (for an example, not for all the series) would be a useful addition to the work already done, mainly as a comment, not necessarily in terms of changing the Figures themselves. One could comment, for example, that it is likely that most (or almost all) of the boxes would end up being non-statistically-significant (this is apparently already the case actually). This does not invalidate the fact that there are outliers in an often systematic way, and that the drifts

do show relative inter-instrument effects and certain tendencies, even if not statistically significant. So I still think that the flavor of (and interest in) the results can be preserved, despite the fact that there is a lack of full rigor in the treatment of statistical significance and some of the conclusions. Some sort of statement that goes slightly beyond merely stating that this effect is completely neglected would be useful for readers, since this is a mathematically known issue (that often gets ignored). If you can prove that this is an insignificant omission, please show this with further analyses (although I personally do not believe that this is the case, without more investigation). If this is completely ignored, however, even if one states that it is being ignored, the reader will get the (incorrect) impression that the results of significance are to be taken at face value, which is really not the case. A more cautionary note is what I am mainly after here, since these issues are too often ignored altogether, but not a complete rework using different statistical methods (e.g., a bootstrap method could also be applied for error estimates).

6. You sometimes state that noisy measurements are a cause for poorer results, for example in the correlations. You have also used the different impact of clouds as an explanation of some differences. I am not convinced that these explanations are well enough proven, at least by what I see in this manuscript. When you have monthly means, most of the results will have very small standard errors in the means, and noise itself becomes much less of an issue. This can be demonstrated for each instrument, and some will be more "noisy" than others, this is still true, depending on the number of data points (and the actual single-profile noise). I would urge you to more carefully consider those comments and convince yourselves of certain cases where these statements are made. If you are not convinced, or convincing, maybe you should invoke unknown systematic effects as well, as another option that could also play a role (when one wants to try to provide a range of options for some differences without a more complete investigation, which could take a while for the multitude of data sets you are considering here, I agree). Alternatively, if you do have a few good examples, you could add supplementary material to show differences of time series

with the noise values (for example), or something else relating to cloud studies (or another reference, possibly). I also tend to think that sampling will play a bigger role than one might think in some of the differences (for example in the issues mentioned in the paragraph before Section 4.4), even if you already do mention some examples of this issue. I just want to avoid statements that sound like careful work has gone into the conclusion, with little proof given; while there can be (is) some level of trust regarding work not shown in the paper, I am not entirely convinced that all the possible explanations have been vetted enough. One can try to investigate and comment about a few of the more obvious cases with poorest results, for example, at certain pressures or latitudes. It is also somewhat surprising to see some of the differences between the various MIPAS retrievals, in terms of how some of the larger differences can come about. Again, the main flavor of the results will most likely not change, and it would be a large undertaking to try to resolve "everything", so I (and others) will need to read this as the way things are, for now, not that one shouldn't expect better agreement in the end, given enough time, etc... but there are still a number of unresolved issues. Some of the writing gives a lot of description of differences, with maybe not enough "potential explanations", and we can infer that some things are not understood yet, which is not completely unexpected either. This does not make this work unworthy of publication, in my view, although one would prefer to see a lot of these uncertainties reduced, or somewhat better explained.

Minor Comments

- pg. 2, L3, change ratio to ratios.

- pg. 2, L8, change :when data" to "if data", although my specific comments are not that positive regarding this sort of statement.

- pg. 3, L18; I would suggest something shorter/better like "with water vapour abundances recovering after 2004-2005".

- pg. 4, Eq. (1), interesting use of "z" for pressure, rather than "p", but this is just a personal preference, nothing to really change.

- pg. 5, lines 8 and 11, there is a change in tense (from "was calculated" to "are discarded"); I would recommend using the same tense in general, inasmuch as possible (e.g., "were discarded").

- pg. 5, L15, maybe change "refined" to "relaxed", or "less stringent criterion".

- pg. 6, L11, typo in "alltogether" [altogether].

- pg. 6, L12, change "ratio" to "ratios".

- pg. 6, L22, delete "the" before "the".

- pg. 6, L26, change "was 3.6" to "removed 3.6".

- pg. 6, L29, add a comma after "measurements".

- pg. 7, L1, change "result" to "results".

- pg. 9, L18, are compared qualitatively.

- pg. 10, L26, delete "and" before "2011".

- pg. 11, L14, available for these altitude and latitude regions.

- pg. 11, L18, change "as those" to "than those".

- pg. 11, L33, change "in order" to "on order".

- pg. 12, L14-15, in the other data sets, anomalies up to only 0.4-0.8...

- pg. 12, L17, anti-correlated with the time series

- pg. 12, L23, add a comma after V5H.

- pg. 13, L1, are found during 2004-2008, when SAGE II...

- pg. 13, L9, change "regions" to "region".

- pg. 13, L17, change slightly decreasing to decreasing slightly.

- pg. 13, L21, 22, there is a lot of repetition of the same references to the drops in H_2O.

- pg. 14, L17, change "coefficient" to "coefficients", and one line lower, change "is" to 'are".

- pg. 14, L20, change "in form" to "in the form" or to "as".

- pg. 14, L30, change "NOM than" to "NOM as".

- pg. 14, L31, most data sets between 1 and 30 hPa.

- pg. 15, L9, change "varies" to "vary".

- pg. 15, L19, "the number of months of overlap between time series is given..."

- pg. 15, L25, the number of overlap months is not that high...

- pg. 16, L2, the number of overlap months is rather low (same for L5/6, and L10, and L20 on pg. 17).

- pg. 16, L7, overview of the temporal...

- pg. 16, L11, An example of a negative correlation, despite a high number of overlapping months, is for the correlation ...

- Also on this page, it is not very surprising when ACE-FTS data versions or MIPAS versions correlate well... it is the opposite hat is more surprising.

- pg. 16, L23, which "both quantities" are you talking about here, please clarify.

- pg. 17, L14, change "were both data sets" to "for which both data sets".

- pg. 17, L19, change "at the" to "on the" (x-axis and y-axis).

- pg. 18, L17, drifts are significant in most cases.

- pg. 18, L26, change "pattern" to "patterns" [are...].

- pg. 18, L34, Exceptions are HIRDLS... and MAESTRO...

- pg. 19, L25, change "because" to "that".

- pg. 19, L27, influenced differently by clouds.

- pg. 19, L30/31, Larger deviations in the lower mesosphere occur in the case of the MIPAS NOM data sets, which are close to their upper retrieval limit there, and thus more uncertain.

- pg. 20, Since the dehydration is only partly a seasonal oscillation,...

- pg. 20, L8, were also assessed; this indicates if the longer-term variations (trends) ...

- pg. 20, L11, change "level" to "levels".

- pg. 20, L11, The most significant drifts were found in the tropics, where there is low..., which...

- pg. 20, L13, Drifts were also calculated...

- pg. 20, L22, delete "amongst others".

- pg. 20, L32, that in addition to correcting...

- pg. 20, L33, changes in the calibration were madew within the HIRDLS mission...

- pg. 20, L34, data sets encounter large uncertainty...

- References, there are a few refs. with no doi numbers (Randel, Remsberg, von Clarmann).

- pg. 28, bottom line, and the spread are more easily compared.

- Figure 5, one could also find an rms diagnostic versus pressure and contract the time axis this way, as an additional Figure that could overlay the three latitude regions, so as to get a maybe better overview of this quantity over pressure and latitude (and some of the colors are not so easy to differentiate).

Printer-friendly version

Discussion paper

- Figure 10, last sentence, Th second number indicates the number of months for which both data sets ...

- Figure 11, line 2, change "at" to "on" (x-axis, y-axis). Line 4, upper left the overall overlap period between data sets is given. The second number indicates for how many months ...

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