

Interactive comment on “The SPARC water vapour assessment II: Profile-to-profile comparisons of stratospheric and lower mesospheric water vapour data sets obtained from satellites” by S. Lossow et al.

Anonymous Referee #1

Received and published: 7 December 2018

General remarks

This is an impressive and important paper and it must have been a huge task to get all the information and data together. However, the huge dimension of the task is reflected in the length of the paper and thus it is very hard to read. It is not only the length that gives a hard reading, it is also the surprisingly difficult interpretation of "relative bias" and the listing of a very large number of single features of curves that might sometimes be simply noise. Thus there is potential to shorten the paper. I will explain that in more detail below.

C1

Otherwise, I recommend the paper for publication after consideration of the following comments.

Major issues

1) The reader easily gets lost in all the details that the paper provides. Although the conclusion section gives a kind of summary, it is still hard to get an overview for a potential user on which data to use for which purpose. The conclusion section, however, is a good start for this. I suggest to either make lists of bullet points in the last section where information is easier to be found than in the running text, or, even better, a table mentioning advantages and disadvantages for each of the datasets, with respect to both bias and drift issues. If you dare to give recommendations, these would be welcome.

2) Isn't there a lot of averaging in the calculation of biases, for instance? An example where I note this is Fig. 4. Coincidences are found and averaged over all seasons and over the whole latitude band that is available, sometimes 20°. Doesn't this averaging make the results appear much better than they actually are? A good place to discuss this is in the 2nd paragraph of section 4.1.

3) Equation 8:

I understand your point that none of the two compared datasets should be ennobled as reference. But with your definition of relative bias you get some non-intuitive results that should be mentioned. First, a relative bias undershooting -100% does not imply that x_1 is negative. Instead we have

$$x_1 = (2 + b)/(2 - b) x_2,$$

which unfortunately is a quite non-linear function. For instance we get $x_1 = x_2/3$ at a bias of -100% , $x_1 = x_2/2$ at $b = -2/3$, and $x_1 = 3x_2$ at a bias of $+100\%$. In effect, the fair comparison comes at the price of a non-linear and not easily comprehensible relative bias measure. To my impression, all the relative biases given in the paper

C2

easily mislead the reader (and perhaps the authors as well). For instance, in line 8 on page 16 you say "biases ... even exceed ... $\pm 100\%$ in some occasions", which implies that x_1 either undershoots $x_2/3$ or exceeds $3x_2$. Is this really what you have intended to say?

4) Page 16, lines 9 to 33: There is a quite detailed description of minor wiggles, local maxima and minima, of the summary bias curves (rhs of Figure 5). Are you sure that these wiggles are physically significant? That is, is there a physical argument that, for instance, the 50% percentile profile minimises at 60 hPa, or isn't this rather incidentally given the sensors that you have in your comparison data base. The non-aggregated line displays different wiggles and it seems to me that you describe noise.

5) Histograms: I doubt that showing and discussing these histograms is of any value. The reader generally wants to know whether a certain data set can be used for an application, that is, whether it has a small bias and low drift in a given altitude (and latitude etc.) region. Perhaps it is important then to have more than one dataset at hand for estimates of uncertainties. But it is not clear to me which purpose the knowledge of these histograms should serve. What, in a concrete situation as sketched, is the gain one has from knowing that the distribution of drift values over 33 data sets peaks in a certain bin? I find this discussion unnecessary and it can be removed without any intellectual loss to the paper.

Furthermore, there are technical issues with the histograms. For instance, the histogram of figure 12 contains about 60 bins which is not ideal. There are several rules for an optimal number of bins, and the result in this case is of the order 15. Consequently there is much noise in the curves and it is not clear whether a single peak as the one described in line 4 on page 22 is real or just noise. A similar comment applies to figure 7, but the curves are much smoother there.

Deleting the histograms and the corresponding text would also have the advantage of getting rid of the unnecessary speculation whether these look Gaussian or not. My

C3

impression is, they do not, but my feeling is also that this an irrelevant point. Whether these distributions are Gaussian or not is not used anywhere in the paper as the basis of an argument.

6) Section 5.4: While it is certainly necessary to compare the methods of drift determination between the current paper and that of Khosrawi et al., I doubt whether it is useful to spend so many lines of text to it and to dwell on so many details. Looking at the corresponding figures, it is evident that the two methods yield essentially consistent results. There are only a handful of exceptions with drift estimates exceeding 2σ . Instead of quoting all the boring numbers (e.g. σ is not much larger than 2, namely 2.01) it would be more useful to think about the possible reasons why the two methods differ for certain data set combinations more than for others. A more fundamental question is whether such differences are generally expected at all and whether a significant deviation is a surprise or not. To my opinion, this section should be rewritten or drastically shortened.

Minor issues

It is a problem that a paper by Walker and Stiller is quite often quoted, which is still in preparation. The reader has no possibility to consult it and does not know when and in which Journal this will eventually become possible. Are there perhaps other "grey" sources of information that may at least partly replace the "in preparation" paper?

Page 3, lines 6-14: it would be nice to have an indication of the relative contribution of each pathway to WV transport into the tropical stratosphere. The sentence "about 3.5 ppmv to 4.0 ppmv of water vapour enter the stratosphere" is not understandable. I expect to read something like "per year a total mass of x kilogram of water vapour enters the stratosphere".

Page 6, line 16: I am surprised that mixing ratios between -20 and zero are not excluded. Please explain.

C4

Page 7, lines 16 ff: I have only a vague feeling why comparing A to B might give coincidences different from comparing B to A. Perhaps this can be explained a bit clearer. I assume that the "lower half of the comparison matrix" contains just the coincidences of A to B. What then do you mean with "we used those results for the upper half of the matrix"? What does this upper half contain if not the B to A coincidences and in which way can the A to B results be used for the B to A comparisons? And finally, what are the "lower boundaries of the comparisons"?

Page 8, line 1: in which respect is the coverage of the hygropause limited? I don't understand what you mean.

Eq. 1: I wonder why the vectors x_{high} and $x_{apriori}$ have the same number of components although x_{high} refers to high resolution (i.e. more values) while $x_{apriori}$ refers to low resolution (less values).

Page 9, line 25: the following is not an equation, please replace "equation" with "quantity".

Eq. 5: Why do you need the factor $4 \ln(2)$ under the \exp ?

Page 10, beginning of section 3.3: I suggest to write "time period" instead of "time" and "latitude band" instead of "latitude". I was puzzled by assuming that t is a point in time and ϕ is a certain latitude and then reading in Eq. 6 that there can be more than one, namely $n_c(t, \phi, z)$ measurements for this point. Later it gets clear, but it is better if it is clear right from the beginning.

Section 3.4: you say $f(\dots)$ is the regressed bias time series. Before, we had $\bar{b}(\dots)$ as the bias between two data sets. Unfortunately, it is not very clear to me what the relation is between $f(\dots)$ and $\bar{b}(\dots)$. Probably there is one and it would be good to know it.

Page 13, line 16: "comparison results ... are not considered any further". Is this really meant or rather "... are not used for other calculations"? To me, it sounds useless to

C5

compute something that is not considered further.

Page 13, line 27: figures (plural).

Page 14, line 8: "this has not further been pursued".

Page 16, line 11/12: This is not surprising, as this variation is controlled by a smaller and smaller number of outliers.

Page 17, line 11: "within a small altitude range". Actually, this may be a small range in pressure, but certainly quite a large distance in altitude. Please correct.

Fig. 7, right panel: Note that 1% bins are of unequal size in terms of actual differences between x_1 and x_2 . This results from the non-linear character of your bias definition, cf. comment above.

Page 18, line 25: "in the right columns". line 27: "contributing to" and "results are based on the aggregation".

Figure 10: Please provide for the sake of completeness the mathematical definition of σ . How is it computed for the present application? This should fit after eq. 9.

Figure 11: I wonder why the smaller (blue lines) drifts are statistically significant whereas the larger (grey) ones are not.

Page 24, line 25: is this really from top to bottom? The figure (15) says the opposite. Please check.

Page 27, line 9: please rewrite the sentence. The impression of drifting data sets is surely not what is intended here.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-380, 2018.

C6