

Atmos. Meas. Tech. Discuss., referee comment RC1  
<https://doi.org/10.5194/amt-2022-114-RC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Referee report on amt-2022-114

Anonymous Referee #1

---

Referee comment on "Long-term validation of MIPAS ESA operational products using MIPAS-B measurements" by Gerald Wetzel et al., Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2022-114-RC1>, 2022

---

Review of "Long-term validation of MIPAS ESA operational products using MIPAS-B measurements", by G. Wetzel et al.

General comments:

This manuscript describes a comprehensive validation effort, using balloon-borne profiles from MIPAS-B, to study the biases and associated variability and uncertainties in the retrievals of a large number of VMR profiles from MIPAS aboard ENVISAT (MIPAS-E); some trajectory-based studies are used to provide additional "coincidences" between the balloon and satellite profiles, and more statistical analyses. This work covers the upper troposphere to the mid- to upper stratosphere. While there is no discussion of any systematic temporal changes (or trends), in part because the time period covered (2002-2012) is not quite long enough to study this well enough based on a few balloon flights, there are some noted differences between the two separate time periods when MIPAS-E was observing in different modes (the original full spectral resolution mode, FR, and the post-2004 optimized resolution mode, OR). One of the main conclusions is that the harder to measure species (for both instruments) lead to poorer overall agreement than for the species with stronger signals (and I assume that this is probably not too unexpected).

The comparisons are presented in fairly simple ways in a consistent fashion, which makes the large number of plots easier to digest (however, there is an issue with some of the font sizes, see comments later on). The summary Table is a good way to provide top-level conclusions, even if this can be somewhat oversimplified and difficult to do with a broad brush when differences change somewhat rapidly with altitude. Adding some suggested explanations for the larger differences (especially when outside the combined estimated

error bars) could be useful, if possible and if not completely speculative. A few more comments regarding other relevant work (in particular, satellite-to-satellite intercomparison results from the SPARC Data Initiative) would be recommended and welcome, as this could reinforce the impression that MIPAS-E might have a real bias (or not), at least in a few specific cases (the same could be done versus ACE-FTS data in particular, since Raspollini et al. have already discussed some of those comparisons, although using a multi-satellite approach as done by the SPARC DI would be viewed as more comprehensive, even if multi-satellite means have potential issues as well, if some measurements are clearly less desirable than others). One should not forget that MIPAS-B is not necessarily "perfect data" either, so untangling a real bias versus just a relative bias can be difficult. It would also help if estimated systematic error bars for the MIPAS-E results were included in one of the Tables, since these values are provided for MIPAS-B, and the combined uncertainties are used (so estimates of error bars for MIPAS-E exist as well). Using a lower to mid-stratospheric range might be good enough for this, or one could consider separating this into two Tables - for two regions where the error bars might be significantly different; I am open to either approach, as long as more information is provided regarding the 'typical' satellite error bars (in tabular form). Otherwise, the manuscript is written in a fairly easy to follow manner, and I have no major objections or issues.

After a few improvements, which do add up to almost (but not quite) a major revision (see below for more details), I would recommend that this work proceed to publication in AMT, since this topic is well-suited for the AMT Journal (and there is also not much discussion in this manuscript regarding composition changes or processes in the stratosphere, for example). More specific (and also some very minor editorial-type) comments follow.

Specific comments:

L200-201, here, why is a 2-sigma type of criterion not used, namely assign the term "significant difference" only when twice the SEM is smaller than the bias itself?? This is more in agreement with what most scientific studies would argue "significance" applies to (and if you disagree, please give some argument on this topic in your reply and in the text). The main impact might be in the Table of overall conclusions, where you discuss what may be "significant" (or unexplained) differences. Whether many of the plots should be changed is something else to think about - I am not necessarily arguing for this (but please be very specific regarding the meaning of the error bars given in these plots, 1-sigma or two-sigma, it seems that you list and show one sigma results...yes?).

L203, it would not be out of the question that unexplained errors in MIPAS-B could also be invoked to better "explain" relative differences between the two retrievals, at least in some cases possibly; also, unusual atmospheric variability could be partly responsible for a lack of "perfect coincidence" (also, trajectories and associated results are not "perfect" either). I just think that assigning "all" the "blame" for significant (enough) differences to MIPAS-E is not the only possible solution; perhaps you could admit to this without it invalidating the usefulness of MIPAS-B or these results, as I am not suggesting this at all either... "Unexplained relative biases" might well be a more reasonable way of wording this, for example. It would also be useful to mention what the estimated systematic errors for

MIPAS-E are, in Table 1, since this could also give the reader some feeling for which retrieval might be expected to be more accurate, if this is sometimes possible to say. However, certain factors like spectroscopic uncertainties (for example) would likely affect both retrievals in the same way, and if this sort of error was a dominant source of error, then neither instrument would be expected to be significantly more accurate (for absolute measurements) than the other... Just showing error bars for MIPAS-B is not really justified, in my view, and since such error bars do exist for MIPAS-E, why not give the reader some feeling for this as well? Are there enough issues in terms of the different satellite retrievals that this becomes a difficult problem to formulate? My issue here is that you have used some estimates, so why not provide at least a first-order example in Table 1, or a similar Table? Tables do oversimplify things, especially if there is a fair amount of altitude dependence in the estimated uncertainties (error bars), but having something would be better than nothing. Please clarify, inasmuch as possible.

L238, it would be good to add just a sentence or so on the main differences between the current manuscript and the Raspollini et al. (2020) document, since these seem to deal with largely the same results. In fact, stating how ACE-FTS comparisons have enhanced these comparisons could be illuminating, especially when the MIPAS-B/MIPAS-E differences are (significantly) larger than one might have expected. On the same topic, I find that you should add at least a few sentences, when appropriate, regarding the results of the SPARC Data Initiative, for some of the species, especially when MIPAS-E biases (with respect to the satellite instrument mean) appear to follow the same tendency that is found here (although it is also interesting if they do not follow this tendency); I realize that the satellite intercomparisons can also be subject to discussion regarding where the "real truth" might be, as it is not necessarily found by showing a multi-instrument climatological mean. Nevertheless, I believe that it is a problem not to mention that document at all (or, actually, the more recent update by Hegglin et al., 2021) and give that work some credit in terms of at least relative bias identification for MIPAS-E; these sorts of studies rely on many more profiles and therefore, in principle, biases can be more robustly identified (although they are also relative biases, and exact knowledge of truth is always a difficult question). On this topic (SPARC DI), I recommend that at least a sentence or so be considered for each of a few of the species mentioned in this manuscript (H<sub>2</sub>O, N<sub>2</sub>O, CH<sub>4</sub>, HNO<sub>3</sub>, and NO<sub>2</sub> are the main ones - while MIPAS-E ozone, in particular, is not seen to have significant issues), if it seems relevant/appropriate - but doing a bit more homework on this issue and adding some additional relevant text would be a change for the better.

L252, it would also be an improvement if you carried out a "gedanken" experiment, in order to at least roughly estimate what altitude uncertainty might be required to lead to such temperature differences (is it 100 m or more than 1 km, say?).

L267-270, see the comment for L252 also, is your thinking regarding H<sub>2</sub>O pure speculation or would there be a reasonable change in altitudes that could account for the observed relative biases in H<sub>2</sub>O (how large a change in *z*, if this is something one can explore "on the back of the envelope", without running full retrieval tests?). If this is just pure speculation, it is probably best to remove the text, I would say. If there are changes that can account for both the T and H<sub>2</sub>O differences, that might start to be more believable.

L289, "on the order of". Also, how do the SPARC DI results compare to these biases in MIPAS-E, i.e. is MIPAS-E on the high side versus other satellite data as well? If so, this might help your argument; if not, it may be more difficult to decide what to conclude - but adding a brief comment on this topic could well be useful.

L300, change "stated" to "mentioned". See my comment above for SPARC DI relevance, please check for CH<sub>4</sub> and N<sub>2</sub>O as well.

L305, how large is the NO<sub>2</sub> photochemical correction compared to the differences (pre-correction) between the two data sets? That is, does the correction actually improve the level of agreement? Again, the SPARC DI results might help the interpretation here (worth a try).

Table 3: Please be a little more consistent in terms of the comments when mentioning whether differences are estimated to be significant or not (this gets back to the 1-sigma versus 2-sigma question as well); in particular, there is a mention for H<sub>2</sub>O regarding differences [generally] being within the combined systematic errors, but why not be more specific also for CH<sub>4</sub> and N<sub>2</sub>O?

Figure 1: This Figure could have larger fonts for the readers to be able to read the y-axis label (altitude) and the x-axis as well (one can remove the words "Volume Mixing Ratio" and just write "H<sub>2</sub>O (ppmv)" or "H<sub>2</sub>O / ppmv"). The larger labels in the plots do allow one to understand which species is shown, but the x-axis and y-axis labels could still be improved; at the same time, this would also allow for a larger font size in the numbers shown along the axes.

Figures 2 and similar: I should have mentioned this in the quick review, but the font size for the listed differences in these Figures is probably too small for readers to see well enough on a printed page (without using a zoom feature on the electronic version, even if this is the most likely use of published material these days). It would be good to reduce the unnecessary text and enable larger font sizes for the main comments; also, some things can be abbreviated and some can better be described in the legend(s) instead of inside the various annotated plots.

L389-390, do the differences between the OR and FR time periods suggest anything regarding the validity of the MIPAS-E data sets (absolute values and scatter or precision)? For example, is the OR mode (in some cases at least) maybe less robust or accurate than the FR mode, or is this too difficult to really ascertain?

Very minor (editorial-type) comments:

L24, change "where" to "when".

L32, I suggest: "This includes an assessment of the data agreement between both sensors, taking into account the combined errors from both instruments."

L36, "a 5-20% level of agreement between the retrieved... For C<sub>2</sub>H<sub>2</sub>,...larger differences (within 20-50%) appear in this altitude range."

L43, "... operated between 2002 and 2012."

L52, ...logistical requirement that the satellite...

L75, solar time of 10:00...

L77, During each orbit, approximately...

L89, ...investigations, it was decided...

L91, back in operation

L97, "...was steadily increased..." [since this did happen]

L99-100, "... anomaly occurred, resulting in the loss..."

L141, comparable to or slightly better than

L142, overview of

L150, consistent with

L155, retrievals [plural might be better here]

L250, change MIPAS to MIPAS-B for extra clarity.

L257, add a comma before "we carefully looked at".

L281, the statistical agreement between the two data sets...

L321-322, suggesting the need for a more careful use...

L342, Deviations for CFC-11...are somewhat larger, up to ...

L346, is also clearly seen if one considers previous comparisons...

L350, is only available for CCl4 profiles

L361, positive bias for MIPAS-E...

L369, which is at the limit of the combined systematic errors.

L375, There is general agreement between both instruments between ...

L387, add a comma before "exceeding"

L393, available for COCl2.

L406, negative bias in MIPAS-E ...

L422, acts as a precursor for the stratospheric aerosol layer

L426, The agreement between the VMR profiles

L448, a somewhat poorer agreement

L456, "on the quality of the MIPAS satellite data."