Reply on RC2
Gerald Wetzel et al.


Response to Referee 3:

First of all we thank the referee for the effort to carefully reading the manuscript and for all comments.

General comments:

My main concern is how the combined error (Eq. 3) is calculated when comparing the two instruments. One issue is that this seems to neglect any potentially correlated error source between the two instruments. The classic example is a spectroscopic error, where if both retrievals use the same spectroscopic database this equation will overestimate the combined error, but there also may be correlated effects from non-LTE errors or other effects absent in both forward models. I understand that these errors may not be perfectly correlated between the two instruments due to differences in retrieval methods (I see different microwindows were mostly used), spectral resolution, etc., but if the dominant source of estimated systematic error between the two measurements is a potentially correlated error like a spectroscopic error it draws into question some of the conclusions made. It is not easy to correctly account for these correlations, but I would suggest at least stating what the dominant source of systematic error is for each case and analyzing if it is potentially correlated between the two instruments. Since some of the main conclusions of the paper are to recommend caution in areas where the observed differences between MIPAS-E and MIPAS-B is larger than the estimated systematic error it is critical that the estimated systematic error is interpreted correctly.

As the referee already writes, correct error assessment is a difficult task. However, we did not attribute larger VMR differences between the two MIPAS instruments to spectroscopic inaccuracies of lines when the same spectroscopic parameters for the retrieval of the target gas were used. Nor have we attributed major retrieval discrepancies to other possibly correlated errors. In case where different spectroscopic data for the target gas were used during the retrievals (like COCl2) this was already mentioned in the text. Non-LTE errors do not play a significant role in the MIPAS-B retrievals and are therefore neglected for this instrument. We are sure that we did not treat relevant correlated errors as uncorrelated. “Effects absent in both forward models” as mentioned by the referee are of course not included in the error estimation but such effects actually may lead to “unexplained relative biases” as mentioned in the manuscript text.
A full validation of every species measured by MIPAS is a monumental task, and the MIPAS-E to MIPAS-B comparisons done by the authors is one piece of that puzzle. This is fine, I don't think the authors need to include more data or analysis, but as a naive MIPAS-E data user some of the results are hard to interpret on their own. The main takeaway that I get is that I should go read the MIPAS product quality document instead (of which a version of this manuscript serves as input to). Once again, this is not a problem by itself, but the manuscript could use some further explanation on how this work fits into the larger body of MIPAS validation efforts.

The MIPAS product quality document not only includes the validation results related to MIPAS-B, but in addition ground-based, ACE-FTS, lidar, radiosonde, and ozone sonde validation results. We added some text to the corresponding sentence in the manuscript. This quality (documentation) readme file is very comprehensive (177 pages). We had a lot of discussion in the MIPAS Quality Working Group on how we can split all the results into reasonable concise publications. After all, one decision was to make a paper showing the intercomparison results between both MIPAS instruments (among other planned validation publications). Hence, this manuscript is not an overall MIPAS validation study of all atmospheric parameters and many instruments but “only“ an intercomparison study between two similar instruments. Anyhow, we already included statements concerning the behaviour of recognized differences in comparison to previously published peer-reviewed publications. However, we add some more information from the mentioned readme document at appropriate points in the manuscript and we also give comparative information from the SPARC Data Initiative for main gases.

Specific comments:

Section 2.1 l. 68: "hereinafter also referred to as MIPAS-E..."
There are some places where simply MIPAS (mostly the remainder of this section) is used to refer to MIPAS-E. I would recommend always using MIPAS-E when referring to MIPAS on ENVISAT for clarity.

We followed the reviewer’s suggestion.

Section 2.1 l. 78: "in steps of 3 km below 45 km"
What about above 45 km?

The steps are coarser above 45 km. We changed the text accordingly to give more information here.

Section 2.1 l. 94: "... an equivalent improvement in the vertical and horizontal (alongtrack) sampling"
I understand details of the sampling for the FR/OR modes of MIPAS can be found elsewhere, but the horizontal/vertical sampling of each mode should be stated in this section. Especially since the change in MOPD is stated.

We changed the text in the manuscript to give information on the nominal tangent altitudes.

Section 2.1 l. 118: "All molecules except HDO ..."
Is there a fundamental reason why HDO could not be validated as well?

HDO was not part of the ESA validation contract, since it was not clear at the time the contract was concluded whether HDO could be retrieved from MIPAS-B spectra in a correspondingly high quality.

Section 2.2:
Has there been any validation of the MIPAS-B data products separate from MIPAS-E that could be mentioned here?

MIPAS-B was also involved in validation of ILAS/ILAS-2, SMILES, GOMOS, and SCIAMACHY measurements. We included corresponding citations in the Introduction part of the manuscript.

Section 2.2 l. 125: "... MIPAS-B performance is superior, in terms of NESR ..."
Is the improvement in NESR from the averaging spectra or is there an instrumental difference that provides better NESR?

The MIPAS-B NESR for a single spectrum is slightly lower than the one for MIPAS-E. However, the main part of the NESR reduction in the MIPAS-B spectra comes from the multiple averaging.

Section 2.3 l. 200: "A bias between both instruments is considered significant if the SEM is smaller than the bias itself."
Should twice the SEM be used here instead to be at the ~95% confidence interval?

In principle this is a question of definition. All error estimations performed in previous MIPAS validation papers (which were cited in the text) refer to the 1-sigma confidence limit. That's why we decided to do the same here (also for reasons of consistency). We added a corresponding sentence to the text here to clarify that all errors refer to the 1-sigma criterion.

Section 2.3 l. 204: "Since the vertical resolution of the atmospheric parameter profiles of both instruments is of comparable magnitude, a smoothing by averaging kernels has not been applied to the observed profiles"
I assume that the error estimates for both instruments also do not include the classical "smoothing error"? If it is then this would cause the error estimates to be inflated since both instruments have a similar vertical resolution.

Yes, the assumption of the referee is correct. A smoothing error is not contained in the error estimates.

Section 3 l. 226: "Trajectory matches are based on diabatic 2-day forward and backward trajectories with a collocation criterion of 1 h and 500 km as described in section 2."
Is it possible to demonstrate how well the trajectory matching is working? If I understand correctly there are conditions where the measurement locations are collocated enough that trajectory matching is not necessary, maybe this can be used to show the effectiveness of the trajectory matching.

It is difficult to give an exact number of how well the trajectory calculation works here but 2-day trajectories are nothing special today in terms of accuracy of the analyses (see, e.g. Dee et al., Q. J. R. Meteorol. Soc. 137, 553–597, 2011, and Hoffmann et al., Atmos. Chem. Phys., 19, 3097–3124, 2019). However, we compared direct coincidence VMR profiles between MIPAS-B and MIPAS-E (already applied for several species in the cited older validation papers) and found no significance difference to the trajectory match results. This was also confirmed in a MIPAS temperature validation study by Zhang et al., Journal of Atmospheric and Solar-Terrestrial Physics, 72, 837–847, 2010.

Section 3.1 l. 245: "... although the standard deviations exceed the expected precision ...
Could this be because the trajectory matching introduces some variance into the comparisons as well? Even if the trajectory matching is perfect the collocation is still only within 500 km and 1 hr, which would contain some atmospheric variability.
Yes, a certain atmospheric variability can be a part of this effect. However this should be also visible in a similar way in the comparison of the longer-lived tracer species what is not really the case.

Section 3.1 l. 252: "A possible reason for this difference between both MIPAS sensors could be an inaccuracy in the altitude assignment..."
Presumably this error is included in the error budgets of the instruments, you should be able to quantify if this could actually be the case.

The line of sight error and the connected altitude assignment is included in the error bars. As written in Section 2.3 the primary vertical coordinate of MIPAS-E is pressure whereas for MIPAS-B it is altitude. Hence this inconsistency might cause some inaccuracies due to the interpolation to a common altitude grid. For instance, the tropopause altitude difference between both MIPAS instruments is up to 1 km. This yields also for the hygropause difference. We added this information in the manuscript text.

Section 3.2 l. 259: "FR and OR mode comparisons show different vertical shapes of the differences between MIPAS-E and MIPAS-B"
Is there a significant difference between the retrieved vertical resolution of H2O in the FR and OR modes? Particularly with the strong altitude gradient of H2O a small change in vertical resolution could cause a large observed difference. In general for every species I wonder how much of the difference between the two modes can be explained from the changing averaging kernel.

Changes in the vertical resolution from the FR mode to the OR phase may cause some different structures in the VMR profile deviations between both MIPAS instruments, especially in the case of strong vertical VMR gradients. We added some text in the manuscript. However, this possible effect is very difficult to quantify also because the MIPAS-B measurement grid was not constant over all flights and the corresponding retrieval averaging kernel has therefore also changed a bit. Furthermore, we made some test calculations in the case of temperature and found that the influence of the varying averaging kernels can virtually always be neglected here (Zhang et al., Journal of Atmospheric and Solar-Terrestrial Physics, 72, 837–847, 2010).

Section 3.6:
Are there any estimates of how much error could be introduced due to an imperfect photochemical correction? I'm wondering if there could be some effect where the balloon flights tend to occur at a similar time each day and so you don't average over an ensemble of random SZA differences, but I'm just throwing things out there.

The balloon flights were not performed at a similar time each day. The situation of the trajectory matches was also not similar. The number of trajectory matches is dependent on the wind speed (e.g. if there was no wind in the region of the balloon within 2 days we wouldn't get any matches off the balloon). Studies of the simulated ratios of nitrogen compounds like NO2 have shown that chemical models are generally more accurate in terms of (relative) ratios compared to absolute amounts of nitrogen species (see e.g. Wetzel et al., J. Geophys. Res., 107, D16, 10.1029/2001JD000916, 2002). Furthermore, taking the photochemical correction into account improved the agreement between the measured data from both instruments.

Technical corrections:

Section 3.1 l. 250: Is MIPAS here referring to MIPAS-B?
Yes. We changed this in the text.