The authors thank David G. Johnson for constructive comments and suggestions for revision. In the following our replies are given in blue, we have revised the manuscript accordingly.

Replies to review #1 comments:

General comments and overall impression:

The authors have summarized a large body of work in comparing profiles of HDO, H_2O, and the ratio made by 3 instruments (SMR, MIPAS, and ACE-FTS), with two retrieval versions each for MIPAS and ACE-FTS. I feel strongly that this work is worth publishing, but in my opinion (and others may reasonably disagree) it is also incomplete. While I provided many comments below, my intent is not to be overly critical, but to suggest additional content so as to increase the return from the effort that the authors have already put into this manuscript.

General comment #1: I believe that the primary reason that one compares measurements made by multiple instruments is to validate a priori estimates of measurement accuracy. This validation is imperfect, and the comparison cannot be used to estimate accuracy since the difference between measured values does not tell you if any one value is correct. However, although I may have missed it, I did not see a discussion of the retrieval accuracy for any of the three instruments, only precision, so I do not know if the differences that were discussed were consistent with estimates of individual biases. And in the end, there were so many reasons for differences in the profiles (differences in spectral databases, calibration errors, time/space coverage, and so on), that I am not sure that I have learned anything about the absolute accuracy of the various measurements of H_2O, HDO, and delta D.

General response #1: Our work is part of a larger assessment. These comparisons aim to provide a contemporary picture of the typical differences in the observational database of δD and to draw some conclusions what these differences may mean for scientific analysis. It is not a validation of a single data set, aiming to validate its a priori estimates of accuracy. This aim is now much more clearly stated. Arguably, the database for δD is sparse, relative to that of H_2O. This leaves some room to address issues of individual data sets. We tried this to the best of the data we had available. Clearly multiple aspects play a role for the differences among the data sets, there is no simple attribution to a single cause.

I would ask the authors to:

General comment #2: Provide a quantitative discussion of sources of systematic error for HDO, H_2O, and delta D (some errors that affect individual profiles will cancel in the ratio, others will not) for each instrument. Perhaps provide an error budget in a table for each molecule and instrument. Or, if the paper is already struggling with length, maybe put it into the supplement and at least include a rolled up estimate that can be compared to the observed differences between instruments.

General response #2: There are several sources for the systematic errors for each instrument and it is difficult to state the exact contribution for each of them. It is complicated to get the full picture, but we have tried to include the known issues effecting the retrievals e.g. spectral database, calibration or sampling error (for the zonal and seasonal means) where applicable. In the revised version we expanded on that based on new retrieval tests for MIPAS and SMR.

A δD error budget as such only exists for MIPAS. It is a common problem for many comparisons that such information do not exist for all data sets. For the MIPAS v5 data set the total error of stratospheric δD has been estimated to be about 100 - 150 per mille (Steinwagner at al., 2007). However, this is based on one retrieval in the tropics only. It provides a measure of accuracy, but it may not be applicable for latitudes outside the tropics. Nevertheless, this information has been added to the data set description.

General comment #3: Try to say something conclusive about what was learned from the comparisons about the quality of the data. Plausible qualitative explanations are provided for differences observed in various regions (at the end of the Conclusion, for example), but I still don't know which profile to believe.

General response #3: As stated above the aim of this study was to provide a picture of the typical differences in the observational database of δD.

All data sets have their strengths and weaknesses. Which data set is the most reliable can depend on a number of aspects, like application, time, altitude or latitude of interest, for example. From the data we have, we would generally rate the ACE-FTS data sets as the best. For them the least number of systematic errors have been identified, yet that does not mean that they are perfect. Version 3.5 is an improvement over version 2.2 and has the benefit of a larger altitude range and time period covered. Behind that we would rate the MIPAS data sets. Here primarily the differences in the vertical resolution of the HDO and H_2O data used for the calculation of δD cause concern. A rating between the two MIPAS data sets is difficult, there is nothing much between them. The largest issues occur for the SMR data set, relating primarily to the characterisation of the sideband filtering and the issues in the spectroscopy. Despite that, this data set shows good consistency in latitudinal cross sections, for example.

General comment #4: An additional personal preference: I feel that the comparisons are further complicated by showing two retrieval versions each for MIPAS and ACE-FTS. I would prefer sticking with the latest release (not beta or test) version. A simple comparison between versions for each instrument might be called for if there is an extensive publication record for the older version and it is necessary to show the difference, but it would simplify things to show only 3 comparisons (3 pairs selected from a set of 3 data sets) instead of 10 (10 pairs selected from a set of 5 data sets).

General response #4: None of the data sets are beta or test versions. The ACE-FTS v2.2 and MIPAS v5 data sets has been used in previous studies (e.g. Randel et al., 2012; Lossow et al., 2011). The ACE-FTS v3.5 data set covers a longer observational period compared to v2.2. Also, the microwindows have been optimised allowing HDO and H_2O information to be retrieved at higher altitudes, as described in Sect. 2. The MIPAS v20 data set is based on an improved calibration of the spectra provided by the European Space Agency (ESA).

Within the WAVAS assessment we had an open data policy. Everyone was invited, to get a coverage of the observational database from satellites as complete as possible. The different instrument teams decided which data set versions that should be included in the evaluations. In accordance to that, the decision was made to include two MIPAS and ACE-FTS data sets to provide an overview of the differences between the well-validated older versions compared to the newer versions.

Specific comments:

Specific comment #1: page 11, line 5, and page 34, line 15: Although this information is probably in the references (and citations in those references), it would be useful to include some discussion of the specific differences in the spectral databases used for MIPAS and ACE-FTS (and SMR, for that matter). Specifically, whose line parameters (strengths, positions, and linewidths) are used for H_2O and HDO? What are the uncertainties in the parameters for the lines used in the retrievals, and how does that affect the profiles?

Specific response #1: Since, both, ACE-FTS and MIPAS use multiple microwindows we feel that listing all lines with its parameters is a little bit too much. For ACE-FTS v2.2 the information on the considered microwindows is found in the following document: https://databace.scisat.ca/level2/ace_v2.2/ACE-SOC-0020-microwindow_list_for_v2.2_and_updates_Dec1.pdf. The corresponding document for v3.5 is located here: https://databace.scisat.ca/level2/ace_v3.5_v3.6/ACE-SOC-0027-ACE-FTS_Spectroscopy-version_3.5_Jan222016_Rev1A.pdf. Please note that a registration is required to obtain these documents. For MIPAS the microwindows are listed in Tab. 1 of Steinwagner et al. (2007).

In the revised version we have quantified the effect of the different spectral databases used in the ACE-FTS and MIPAS retrievals. The uncertainty of the different spectroscopic parameters is typically a few percent. We have performed tests with the SMR retrieval to quantify the effect of these uncertainties, assuming a 5% uncertainty in the line broadening parameter, a 10% uncertainty temperature dependence exponent and a 2% uncertainty line intensity.

Specific comment #2: page 16, line 9-10: This would probably be obvious to most people, but it would have been helpful to me to clarify to me here that by "all available data" you meant (I assume) the full data sets for each instrument as described in section 2. As distinct from the subsets used for profile comparisons as listed in table 2.

Specific response #2: With “all available data” is meant that the complete data sets are considered. The text has been changed to make that clearer.

Specific comment #3: page 30, lines 13-31: This was a section that I thought needed to be more quantitative. Sideband leakage is specified, but the bias this may cause in H_2O is not quantified. Likewise, bias due to spectroscopic parameters is mentioned, but the parameters are not identified, the uncertainty in the parameters is not specified, and the effect on retrievals is not quantified. Having this additional information is very useful when trying to make sense of the difference between SMR and other sensors.

Specific response #3: We have expanded this section in the revised version. Besides the retrieval tests regarding the uncertainty of spectroscopic parameter in the SMR retrieval, as mentioned in the specific response #1, we have also performed retrieval test focusing on the sideband leakage.

Specific comment #4: page 32, line 18: when the authors specified "homogeneous coverage in latitude and time", I was confused. A sun-sync orbit covers all latitudes but just 2 times (ascending and descending) at each latitude. Does time mean season, not time of day?

Specific response #4: The sampling differences were discussed in the context of seasonal comparisons. In that sense time means season. The MIPAS observations cover almost all latitudes twice a day. Over the course of a season this coverage is rather homogenous (the longitudes vary). In contrast, the ACE-FTS observations focus on high and middle latitudes and therefore have most of the observations in that region. On a given day only two latitudes are covered (typically one in the Northern Hemisphere, one in the Southern Hemisphere). During a season the covered latitudes vary.

Specific comment #5: Figure 1, lower left panel (H2O bias): This figure confused me. Looking at around 30 hPa, we get SMR-MIPAS←-1.4 ppmV, and SMR-ACE←-0.7 ppmV. That would suggest that MIPAS-ACE←+0.7ppmV. But the figure shows more like +/-0.2 ppmV (depending on the exact algorithm pair). However, Figure 7 shows MIPAS-ACE much closer to +0.7 ppmV, even though this is not a direct profile-profile comparison. Does this suggest that, for H_2O anyway, the direct comparison of ACE and MIPAS is invalid due to insufficient data and poor coincidence?

Specific response #5: Essentially, these results cannot be combined in a commutative manner.

The profile-to-profile comparisons (Fig. 1) among the different data sets consider different time periods and latitude bands (especially the comparisons between the ACE-FTS and MIPAS data sets, see Tab. 2). Therefore, it is already in this case not possible to deduce a specific comparison result from the combination of other comparisons.

The seasonal comparisons (Fig. 2-7) consider the complete data sets and extend the results from the profile-to-profile comparisons shown in Fig.1. This is especially relevant for the comparisons between the MIPAS and ACE-FTS data sets, which have a limited overlap time.

Given, these different conditions the comparison results cannot easily be combined or transferred from one figure to another.

Specific comment #6: Page 6 of 55, line 9: should this read "climatological comparisons"?

Specific response #6: The text has been changed to "climatological comparisons".

References

Lossow, S., et al., Comparison of HDO measurements from Envisat/MIPAS with observations by Odin/SMR and SCISAT/ACE-FTS, Atmospheric Measurement Techniques, 4, 1855 – 1874, doi:10.5194/amt-4-1855-2011, 2011.

Randel, W. J., E. Moyer, M. Park, E. Jensen, P. Bernath, K. Walker, and C. Boone, Global variations of HDO and HDO/H_2O ratios in the upper troposphere and lower stratosphere derived from ACE-FTS satellite measurements, Journal of Geophysical Research, 117(D16), D06,303, doi:10.1029/2011JD016632, 2012.

Steinwagner, J., M. Milz, T. von Clarmann, N. Glatthor, U. Grabowski, M. Höpfner, G. P. Stiller, and T. Röckmann, HDO measurements with MIPAS, Atmospheric Chemistry & Physics, 7, 2601 – 2615, doi:10.5194/acp-7-2601-2007, 2007.