Response to Anonymous Referee #3 (amtd-8-C5364-2016-1)

Our responses are written in italic.

We thank the reviewer for his/her helpful comments and suggestions to improve the manuscript.

General Comments

This paper clearly and concisely describes a new instrument, AIMS-H2O, for the in situ detection of water vapor in the upper troposphere and lower stratosphere (UTLS). The physical details, calibration procedure, detection methodology, data evaluation, sources of uncertainty, and flight performance are discussed in detail. Apart from a few comments/concerns enumerated below, the paper is well ordered, and well written. I recommend publication after revision.

Specific Comments/Concerns

1. The authors' correctly state that the largest percentage differences have been observed at the lowest mixing ratios encountered in the UTLS. The critical range is typically below 10 ppmv. Suggestion: If the performance at and below 10 ppmv is a point of emphasis, it would be worthwhile showing a second calibration curve that concentrates on values centered around 5 – 10 ppmv. The inset in Figure 6 of Thornberry et al., 2013, and the in-flight calibration cycle shown in Figure 9 of the same paper are good examples. Also, there is a significant degree of noise/variability (white and non-white) in the Pt/H2 reference standard evident in Figure 6 (top panel). This is most evident at the lowest mixing ratios.

Question: Does this variability in the reference -+-1 ppmv at a mixing ratio of ~1.5 ppmv – limit the accuracy of the calibration at low mixing ratios?

The top panel of Figure 6 could be indeed misleading. The water vapor mixing ratios shown here are not from a reference measurement since this is an inflight calibration sequence but calculated from the AIMS-H2O count rates. So, the noise on the H_2O signal is basically the signal noise from AIMS-H2O, not from the Pt/H_2 reference standard. Moreover, the lowest mixing ratios in Figure 6 are short flushes with dry synthetic air at the beginning and at the end of a calibration sequence and thus not related to the H_2 reference.

From ground measurements, the combined variability (noise level) of the MBW reference instrument and the calibration standard is much lower (<0.01 ppmv at 1 ppmv and ~0.1 ppmv at 30 ppmv). This variability does not contribute significantly to the overall accuracy of the calibration.

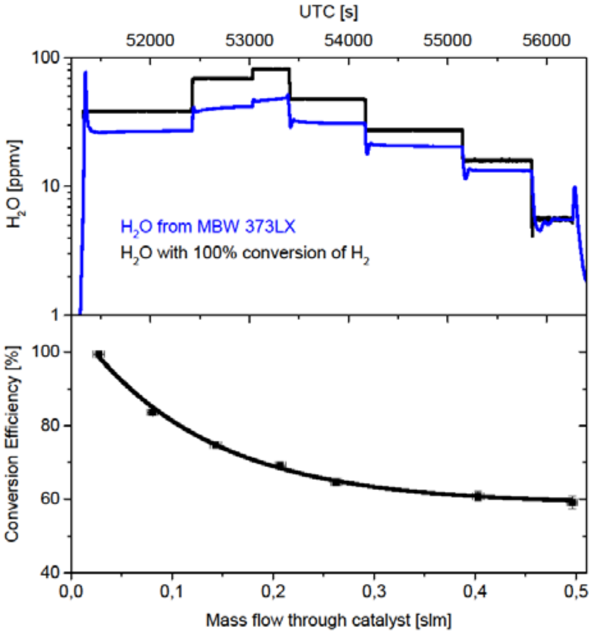
2. The accuracy of the AIMS measurement is directly dependent upon the Pt/H2 transfer standard, which in turn requires regular calibration with respect to a primary reference standard. In this case, an MBW 373-LX frost-point hygrometer.

a. The authors do an excellent job identifying and discussing sources of measurement uncertainty and bias, in particular, discovering that ion transmission through the quadrupole is dependent upon the temperature of the aircraft cabin. This is an excellent demonstration of differences between the flight and laboratory that cannot be anticipated. It also emphasizes the necessity of in-flight calibration with respect to the Pt/H2 source, as calibrations of AIMS performed on the ground are NOT valid in flight. Thus, it is absolutely essential that the accuracy of this source be established and shown.

The authors provide a thorough description of the uncertainties and nonlinearities of the Pt/H2 transfer standard, e.g., its dependence upon temperature, flow rate, exposure time, and history of use, but then do not provide sufficient quantitative analysis/ evidence of its accuracy and performance with respect to the MBW.

Suggestion: The paper would benefit greatly from the inclusion of a figure showing calibration data for the H2/Pt transfer standard along with the empirical fit, an indication of the stability of its calibration over time, and a discussion of accuracy that is more than “the total accuracy of the calibration source is around 6%.”

We agree that a more quantitative description of the calibration source is beneficial for the paper since the inflight calibration is a major part of the instrument. Therefore, we added the following figure to manuscript:



The upper panel shows a time series of a ground calibration of the catalytic source versus the MBW for one ground calibration. The black line is the expected H_2O mixing ratio calculated from flow conditions and the H_2 concentration in the calibration cylinder, assuming that every H_2 molecule is transformed to H_2O. Both lines only match for the lowest H_2O concentration corresponding to the smallest flow through the catalyst. With increasing gas flow, the conversion decreases down to final values of 60% at 0.5 slm catalyst flow. The CE decreases exponentially with the flow (lower panel), a feature consistently observed for every ground calibration. If the residence time of the H_2 molecules in the catalyst would be the only parameter controlling the conversion efficiency, one would expect a linear decrease of the CE with increasing flow. Hence there must be additional mechanisms influencing the CE which could be e.g. occupation of calibration sites or an inhomogeneous efficiency of the reaction within the catalyst. The change of CE for one specific flow setting varies by less than 5% between two adjacent ground calibrations.

Question: Also, given that the stability of the catalyst is the largest source of uncertainty, how frequently is the H2/Pt source calibrated?

During a measurement campaign, the source is usually calibrated between every two flights. This allows to trace the stability of the catalyst for every flight. We added a sentence accordingly.

b. Suggestion: Despite the fact that the calibration/performance of AIMS changes in flight, an end-to-end demonstration of its success in the laboratory would also help to validate the overall approach, e.g., showing:

i. empirical determination of H2/Pt performance with respect to the MBW (as suggested above)

ii. calibration of AIMS on the ground with the H2/Pt (could be combined with the low water suggestion earlier)

iii. verifying AIMS performance through simultaneous measurement by AIMS (using H2/Pt calibration) and MBW (i.e., add a figure like Figure 8 of Thornberry et al., 2013)

AIMS-H2O is indirectly calibrated with the MBW using the H_2*/Pt calibration source as transfer standard. Hence, using the MBW as a validation for AIMS in the laboratory is a circular argument. Since the ground calibration shows no different features compared to the inflight calibration, we decided to emphasize the calibration during flight in this work. However, describing the performance of the catalytic source with respect to the MBW as suggested above might be beneficial.*

c. Question: What is the direction of the error introduced by the temperature dependence?

That depends on the direction of temperature change relative to the calibration environment. The transmission through the quadrupole (and thus the count rate) decreases with increasing temperature. So, an increase in temperature would result in an underestimation of the H_2O *mixing ratio.*

Question: I presume the direction of desorption is predominantly a high bias?

That is true.

Question: Is there evidence of desorption/hysteresis after takeoff?

We try to avoid such effect by flushing the inlet with dry synthetic air on the apron and during takeoff. We switch to ambient air as soon as the mixing ratio is below roughly 1000 ppmv (indicated by other H_2O *instruments or by temperature). When we do so, we do not observe significant desorption features.*

Question: Is there evidence of hysteresis after thick cloud encounters?

Since we sample via a backward faced inlet, we only measure gas phase water vapor independent of the cloud conditions. So, there is no hysteresis due to clouds.

Question: Is it possible to speculate why Thornberry et al., 2013, saw an effect with CO2 and AIMS does not?

Since the source of the CO_2 interference was not clarified, it is hard to speculate why there is a difference. One reason might be that we use different product ions for the quantification of H_2O mixing ratios ($H_3O^+(H_2O)$ instead of H_3O^+).

Suggestion: The paper would benefit by a table, similar to Table 2 of Thornberry et al., 2013, that summarizes and quantifies the different sources of uncertainty.

We decided to discuss the sources of uncertainty in the text rather than in a table because the different contributions are not independent or can act in various ways. Thus, it is not possible to just add up all contributions in order to obtain the overall uncertainty of the measurement.

3. The use of dual methods of determining the water vapor mixing ratio is clever, particularly as the methods complement each other with different strengths in different mixing ratio ranges. (And it appears that differences between these two methods aided the discovery of the temperature dependence.) The requisite agreement between the two methods also provides a means of corroborating the accuracy of the measurement.

Suggestion: Would it be possible to show explicitly/quantitatively how well these methods agree in flight with a time series plot and/or a regression analysis?

Both methods agree very well (within +- 5%) for mixing ratios between 15 and 70 ppmv. We added that information to the manuscript. For lower and higher mixing ratios, there is a systematic deviation between both methods. Method 2 remains more sensitive towards higher mixing ratios, therefore the upper measurement limit is higher when including this evaluation. Using only method 1 for the high mixing ratios, the variability in the count rate produces unrealistic high scatter in the H_2O signal due to the low sensitivity of the instrument. On the other hand, the production of $H_3O^+H_2O$ at mixing ratios below 15 ppmv is almost independent of the count rate on the (in that case abundant) educt ion O_2^+. For that reason, we use method 1 in the low ppmv regime (discussed in section 6.1).

4. Question: When calibrating in flight I presume the calibration flow is sufficient to halt any sample flow from the atmosphere? Suggestion: If so, you might add a sentence or two about this.

That is true. The excess flow is drained trough the inlet. We added a sentence.

5. Figure 10 shows the good broad-scale linear agreement between the two instruments. Question: Was the linear fit done on the raw data or on the log of the data? In other words, do the slope and intercept shown on the figure mean that AIMS = 0.99961xSHARC+0.001, or that log(AIMS) = 0.99961xlog(SHARC)+0.01?

The fit in Figure 10 was indeed a fit of the log-log Data. Therefore, the equation in section 7 was wrong. For a fit of the linear data, we obtain AIMS = 1.007xSHARC+1.66.We changed the numbers in the Figure and in the text to the linear fit of the linear data.

a. The biggest concern here is that log-log plots hide detail that can matter when evaluating the agreement between a pair of instruments. Indeed there are some significant trends/differences evident at both the high and low ends of the scatter plot that suggest that the agreement at any given time during a flight can be significantly greater than the implied agreement of _1% and 0.01 ppmv. (See also the data shown in the time series of Figure 9.) The combination of Figures 2 and 3 of Rollins et al., demonstrates the degree to which significant systematic differences between pairs of instruments can be buried in a log-log plot and composite fit.

We agree that a log-log plot can hide details about the agreement of different instruments. However, the scope of this plot is to show that both instruments agree reasonable over the whole measurement range. A thorough quantitative analysis of the agreement between the different water vapor instruments onboard HALO during ML-CIRRUS (AIMS-H2O, FISH, HAI, SHARC) is in preparation but beyond the scope of this paper.

Suggestion: A figure like that of Rollins et al., 2014, Figure 3, would provide a more detailed exposition of the agreement between SHARC and AIMS.

We agree with the referee. A comparable figure is in preparation for an inter-instrument comparison paper.

b. Additionally, it is informative to quantify the uncertainty of the fit. Certain linear fitting algorithms have the capability of estimating uncertainty in the fit parameters that arises due to uncertainties in both the x and y datasets. This might constrain the number of significant figures you report for the fitted slope.

In the linear fit, the numbers of significant figures are reduced according the uncertainty of the fit parameters.

c. Question: What is the stated accuracy of SHARC? Suggestion: This number should be cited, since you conclude that the disagreement between AIMS and SHARC during the first segment of the 7 April 2015 flight is within their combined uncertainty.

The accuracy of the SHARC instrument is 10%.This is added to section 7.

6. Question: Despite being within the uncertainty limits, is the fact that the RHi is consistently 90% during the dense homogenous cloud segment cause for concern? Could this sustained difference from thermodynamic equilibrium suggest the presence of a systematic error, i.e., calibration error of the water measurements, or (maybe more likely) bias in the temperature measurement?

The mentioned sequence is indeed a very interesting one. In contrast to most of the cloud sequences where RHi is very variable and usually scatters around saturation (as in the first sequence in the figure), the RHi signal is rather smooth and clearly subsaturated for both instruments. Since both AIMS-H2O and SHARC match very well in that sequence, it seems unlikely that both experience an identical bias (different measurement principle, different location of the inlet). So, one could either speculate about a temperature bias (maybe due to icing of the rosemount housing of the temperature sensor?) or a meteorological feature where the cloud is in a evaporating status.

A plot of RHi as a function of cloud IWC, as well as a plot of RHi as a function of temperature would be interesting. These figures aren't necessary for the publication, but might provide evidence for the discrepancies, i.e., 1) the difference between SHARC & AIMS during the heterogeneous cloud encounter; and 2) the difference between the measured RHi and the expected value of 100%. (The figure in the supplement illustrates how different measurement errors can impact observed RHi.)

Those plots will surely be part of the water vapor intercomparison paper. To 1): There seems to be o systematic bias between AIMS and SHARC during cloud encounters. However, for both instruments, there seems to be a slight low bias of RHi which is almost independent of the IWC. This might be a hint to a possible temperature bias.

Technical Corrections

Thank you for the corrections. They are applied wherever the respective sentence was not reformulated.

1. page 13526, line 5, suggested wording: "…the role of H2O in the UTLS energy budget…"

2. page 13527, line 14, insert "the": "In Part 2… the setup…"

3. page 13527, line 21, suggested wording: "…greenhouse gas and engages in a positive feedback with changes in long-lived…"

4. page 13526, line 27, change "in" to "on":…offsets on the order of...

5. page 13528, line 9, suggested wording: "…radiation budget including the effect of clouds…"

6. page 13528, line 15, remove "partly": "…but still above…"

7. page 13528, line 24, replace upper troposphere and lower stratosphere with UTLS: "…mixing ratios of the UTLS."

8. page 13529, line 4, suggested new paragraph starting with "In this work…"

9. page 13529, line 7, suggested wording: "Second, we present the in-flight calibration setup that is used to assure accurate and reliable airborne measurements, and evaluate its performance."

10. page 13530, line 14, suggested wording: "The one directly at the TGI is used to add calibration gas, and the second one is used for optional dilution of the flow with dry synthetic air."

11. page 13530, line 6, suggested wording: "The dilution flow is added…"

12. page 13536, line 1, typo: "…is shown in Fig. 4."

13. page 13538, line 21, replace "already" with "is": …(R1-R4) is completed"

14. page 13542, line 16, add comma: "…around 5 ppmv, which…"

15. page 13545, line 4, suggested wording: "AIMS-H2O flew on the DLR Falcon…"

16. page 13547, line 8, replace "sucked in" with "drawn in" or "sampled"?

17. page 13547, line 14, remove "a": "…established by regular in-flight calibration…"

18. page 13547, line 18, remove "directly" and add comma: "…15 ppmv, we use the count rate…"

19. page 13547, line 20, remove "a", and "other": "…provides better data quality. Major contributors to…"

20. page 13547, line 26, remove "a" and add commas: "…showed reasonable agreement, within _10 %, for most of the data."

21. Page 13548, line 1, suggested wording: "…proved to be well-suited to contrail and…"

Sentence is rewritten

22. Figure 2. Suggested wording, add “through”: “…inlet and passes through a pressure regulation…”

23. Figure 4. Suggested wording, change “the” to “a” and change “ist” to “is”: “A resistance of 500 MOhm is…”

24. Figure 7. The y-axis label does not match the text or the units of ES. Should it be ppmv-1?

Yes, it should be $ppmv^{-1}$*, it is corrected now.*