

Interactive comment on “Evaluation of the IAGOS-Core GHG Package H₂O measurements during the DENCHAR airborne inter-comparison campaign in 2011” by Annette Filges et al.

Anonymous Referee #1

Received and published: 18 April 2018

This manuscript describes the new water vapor instrument for IAGOS commercial flight measurements, its calibration (in several ways), and some preliminary comparisons with other water vapor instruments during an aircraft-based inter-comparison campaign in 2011. The paper is, in many instances, very detailed and provides ample descriptions of the calibration and evaluation, but at times does not provide the information necessary to completely understand some of the findings. My finding is that the paper can be published in AMT if the shortcomings (noted below) are improved.

General Comments

I don't agree with the measurement "repeatability" values determined in this paper

C1

for WV mixing ratios < 100 ppm. On page 14 it is stated that an "upper limit for the measurement repeatability" ... is ... "4 ppm or 5%, whichever is greater for WV < 100 ppm". If that is the case, why are there so many data points for WV < 100 ppm in Figure 7 that have standard deviations > 5 ppm (5%)? The upper limit estimate for WV > 100 ppm is much more in line with the Figure 7 data.

The abstract is very, very detailed, and can be shortened by describing the overall work in more general terms. An example of too much detail is the inlet system description that is described again in similar (or even less) detail on page 6. Also, what were the quantitative goals of comparing the CRDS instrument to others? In other words, what levels of performance were you hoping to obtain from the CRDS in terms of precision, accuracy and stability?

The frequent interchange between using ppm and % as units of mixing ratios (mole fractions) can lead to confusion, especially when it is desired to report relative differences or uncertainties (in percent). An example is Lines 5-6 on page 10 where "relative" is required to differentiate relative differences (in %) from mole fractions (in %). Mentally converting % to ppm is not terribly difficult, but by changing all the text and graph axes from % to ppm, you would completely avoid any potential confusion. Some of your 2-panel figures show mixing ratios in % (top panel) while differences or residuals are presented in ppm (bottom panel).

In many places uncertainties and error bars are presented without any indication if they are based on 1 or 2 standard deviations of the mean. It is helpful to report uncertainties, but only if the reader knows on which statistics they are based.

Please refrain from including Figure caption information in the main body text of the paper. It makes the paper more tedious to read. This occurs in numerous places throughout the manuscript.

During the comparison flights, did all instruments sample the airstream from a common Rosemount TAT inlet, or did each have its own inlet? I believe only the FISH inlet is

C2

described as being different. If there was a common inlet for the other instruments, how was the sample stream physically split between them? If each instrument had its own inlet, how much of a difference might the different types and/or locations of inlets play in the overall comparisons? The absence of information about the sampling inlet for each instrument makes it very difficult to understand the inlet-dependent discussion on Page 16. I therefore suggest that this discussion be omitted from the paper because (a) it requires adding much greater detail about inlets and (b) inlet influences aren't necessarily a part of the evaluation of the CRDS measurement capabilities for IAGOS.

The clarity of the presentation within this manuscript would be significantly improved through editing by a native English speaker. Hopefully that can be done before revisions are submitted.

Specific Comments:

Page 1, Line 16: "North-Germany" is not a proper noun, so "North" should not be capitalized and the hyphen should be omitted, i.e., "northern Germany". This occurs several times within the manuscript, e.g., Pg 4 L19.

Pg 1, L19: What (specifically) is meant by "assessment validation"? Are there targets for the measurement precision and/or accuracy of this new IAGOS instrument that are the basis for its evaluation as described in this paper?

Pg 3, L6: "data of sufficient quality in the UTLs" - sufficient for what? This comment ties in with my previous comment about targets for "sufficient" measurement precision and/or accuracy.

Pg 3, L12-17: why include this long list of satellite-based water vapor sensors when their data are not included in this paper?

Pg 3, L18-19: "insufficient spatial resolution" - what spatial resolution is sufficient? And don't you instead mean "vertical resolution" here when discussing satellite and remote-sensing observations?

C3

Pg 4, L6: "The central problematic of all these different ... are the remaining, unexplained discrepancies ...". Are the discrepancies "remaining"? From what? Are they a problem? It depends on the science you are investigating. Please improve the clarity of this statement.

Pg 4, L15-18: This is the type of general statement that belongs in the abstract instead of the very detailed descriptions.

Pg 4: I find a lot of the information presented in the abstract is duplicated here in the introduction. The introduction should explain why your investigation is important and how you intend to perform it. Specific information about instruments, including the MOZAIC humidity device, belongs in a separate section describing the various instruments.

Pg 5: There is a lot of repeated information on this page. Please ensure that you don't write the same sentence more than once in the paper, with the exception of abstract and conclusions.

Pg 5, L24: What "impact on gas density and spectroscopy" is being minimized? This statement is vague and needs clarification.

Pg 6, L4-6: This description was already given in great detail, in the abstract (where it likely does not belong).

Pg 6, L13: How are there only small humidity differences between the cabin and the outside air?

Pg 6, L27-30: This would be the best place to explain that the instrument zero ("offset") is not stable (over what time scale?) and requires frequent adjustment, while the instrument gain is very stable and needs calibration only infrequently.

Pg 9, L30-32 and Figure 3: Are the reported uncertainties 1-sigma? 2-sigma? This comment applies to the entire paper, wherever uncertainties are given (text and Figures).

C4

Pg 10, L1: "0.0024 ± 0.0021" is "not significant"? What does "significant" mean here? It can't possibly mean "not statistically different from zero" as the reader might expect.

Pg 10, L12-14, L20: The differences between calibration curves presented in ppm in Figures 3b and 4b and expressed in ppm and % (rel) in the text may cause confusion here. If you express all mixing ratios, differences and residuals in ppm it avoids any confusion. For the Figures 3b and 4b it would add value to include a right axis of relative differences/residuals. Given the near-linearity of the differences/residuals against mixing ratios, I think the relative data (different symbol/color) would be quite constant over the entire range of mixing ratios.

Pg 10, L26-28: Do you have enough evidence to make this statement based on a conjecture that the instrument "was not calibrated well enough" in 2009? This seems like a hand-waving statement with little factual backing. Is there further evidence that allows you to select the 2013 calibration as being correct?

Pg 10, L33: I have to assume that the mixing ratio ranges compared with each other instrument were chosen based on the best measurement ranges of the other instruments. Is this the case? If so, please inform the reader of this earlier in the paper. I see the range information in Table 1, but the table is not mentioned until page 13. An earlier statement about the choices of comparison ranges for different instruments and a mention of Table 1 are needed, otherwise it appears the comparison ranges were chosen randomly.

Pg 11, L5-13 and Figure 5: The residuals in Figure 5b (presented in ppm) are discussed in the text in relative (%) terms. Adding a right axis for relative residuals to Figure 5b (using different symbol/color) would strengthen the discussion of relative residuals in the text.

Pg 11, L30-31: It is odd that the residuals at ~250 ppm are negative (Figure 6b) while at the lowest mixing ratio they are positive. With an offset of 16 ppm one would expect the residuals at both of the lowest mixing ratios to be of the same sign. Any idea why

C5

they are negative at ~250 ppm and positive at the lowest mixing ratio?

Pg 12, L1: If "the lowest measurement was made at around 300 ppm", why are there data points below 300 ppm in Figure 6?

Pg 12, L6-7: "which was tried to exclude" is an example of where a native English speaker could help clean up the grammar in this manuscript.

Pg 12, L10: It's not "zero air" if it contains 2 ppm CH₄. You might want to use a different term.

Pg 12, L20: Of course, "the pressure sensor is sensitive to water vapor", as it is to any gas-phase constituent. I think you are trying to convey that the pressure sensor responds in a non-linear way to increasing amounts of water vapor, creating a water vapor-dependent bias in the pressure readings. In L25, what does "assumed that the pressure changes linearly" mean here? If the pressure changes linearly with changes in water vapor you don't have a problem with water vapor-dependent biases. I don't understand this assumption.

Pg 12, L12-25: Don't the water vapor-dependent bias in the pressure readings also affect the in-flight data of the CRDS? Here you dwell on how this affects the CO₂ dilution calibration method, but don't discuss the effects on the actual in-flight measurements. If this also affects the flight data, why only discuss it here in the "Calibration by CO₂ Dilution" section?

Pg 13, L26: It is strange to have a section entitled "Summary" in the middle of this paper. How about "Calibration Summary" instead?

Pg 13, L30: "Note that both offsets, or rather their uncertainties, are likely not reliable." So, are the "offsets" (y-intercepts) likely not reliable, or their uncertainties, or both?

Pg 14, L8-9: What is "a total accuracy" and what does "and down to 0.3 ppm for the lower mixing ratios" mean? How low is "lower"?

C6

Pg 14, L14-15: What are the "stable atmospheric conditions" mentioned here? This sentence needs a comma after "analysers".

Pg 14, L13: Why use the term "repeatability" instead of precision? From the subsequent description it appears that you are looking at variations in the 0.4 Hz measurements from the 60-second averages. I would call this "precision" and not repeatability, since repeatability can imply stability in results obtained at different times, such as re-sampling the same air mass five minutes later and seeing how "repeatable" the measurements are.

Pg 14, L18-19: Isn't the 60-second average calculated from the 2.5-second measurements? Then why do you calculate the standard deviation of their differences when the standard deviation of the 60-second average directly provides a direct statistical measure of variability in the shorter-term measurements?

Pg 14, L24: "results based on a larger number of data are highlighted". Aren't most 60-second averages based on 24 measurements made at 0.4 Hz?

Pg 14, L26-27: "upper limit" ... "4 ppm or 5%, whichever is greater for WV < 100 ppm". In Figure 7, why are there so many data points for WV < 80 ppm that have standard deviations > 4 ppm (5%)? The upper limit estimate for WV > 100 ppm is much more in line with the Figure 7 data: most of the data points for WV > 100 ppm lie below the 30 ppm standard deviation line and to the right of the 5% diagonal. This cannot be said about the WV < 100 ppm data, many of which lie above the 4 ppm standard deviation line and to the left of the 5% diagonal. Please explain the connection between this Figure and the numbers you have extracted from it that appear to underestimate the precision values for measurements at WV < 100 ppm.

Pg 15, L9-16: No need to describe the different symbols since that is done in both the Figure 8 caption and legend. What is more important here is a description of what the Figure shows about response time of the CRDS. I think that showing only a short period of Figure 9 would better support this discussion of response times <10 seconds.

C7

Specifically, the flight segment from 12.35 to 12.55 UTC on 1 June 2011 when water vapor increases and falls an order of magnitude in about 3 minutes.

Pg 15, L22-25: This discussion of laboratory tests of response time is not very informative or conclusive because the reader has no idea how much water vapor is in the "wet and dry air". Also, "recovery time to 99% of a challenge" is not understandable.

Pg 15, L27-31 and Pg 16 L1-6: Please remove Figure 11 symbol descriptions and the explanation of data binning since all this information appears in the Figure caption.

Pg 16, L1-2 and Figure 11 caption: "Absolute" differences steers my thinking towards absolute values, which are definitely not what Figure 10 shows. "Absolute" is unnecessary in both instances because the differences are clearly in units of ppm, i.e., not relative differences.

Pg 16, L4-6 and Figure 11: If the differences are binned by CR-2 water vapor in 1 ppm bins for CR-2 < 10 ppm, why are there multiple blue diamonds and multiple blue triangles in the lowest CR-2 mixing ratio bin (and some other bins)? Aren't the averages computed from all segments of all flights? If not, why are the average differences in each bin determined for subsets of the 4 flights?

Figure 11: The log-scale Y-axis is problematic for average differences < 1 ppm and for the error bars that appear asymmetric around the mean. Why don't you instead plot average relative (%) differences using a linear-scale Y axis? Average differences < 1 ppm will then be displayed correctly and the error bars will be symmetric around the mean values. I think the relative values will be much more like flat lines with CRDS-CR2 values above zero and the other instrument pairs below zero.

Pg 16, L11-33: This section is difficult to understand because there are no descriptions of the various inlets for the different instruments. All that is known at this point is that the FISH inlet is different from the CRDS inlet, although there is now a hint that the other instruments do not share the CRDS inlet. Is it really important or necessary to

C8

include this section of the paper when it requires adding more details of the various instrument inlets? This section seems to cover a different topic from the main thrust of this paper and may even be worthy of a separate paper in the future.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-36, 2018.