

We thank all the reviewers for carefully reading our manuscript and for the detailed feedback aimed at helping us to further improve the manuscript. Below we address the raised concerns in a point by point fashion. Changes are highlighted in the attached revised version.

Anonymous Referee 1

This paper presents a comparison of the SEALDH-II hygrometer with the German PTB water vapor standards. The essential aspect of SEALDH-II is that it is calibration free and this validation effort closes the traceability chain with the German water vapor standard. While I have only minor comments regarding the comparison with the German standard itself, this paper raises significant questions regarding its position within the water vapor observation community. The questions, which I outline below, need to be addressed before this paper can be published. Therefore I would evaluate this paper as accept after major revisions.

General comments:

I have only minor comments regarding the technical work itself and go into detail these below. However, the larger concern is the novelty and importance as expressed in this paper. The authors claim that SEALDH-II is a novel hygrometer, which is a calibration-free, tunable diode laser spectrometer that bridges the gap between metrological water vapor standard and field deployed hygrometers. However, last year the authors published work (Buchholz et al., 2016) on the novel Hygrometer for Airborne Investigations (HAI), which they developed in cooperation with the Research Center Jülich. The claims made in that paper read very similar than the claims made about SEALDH-II. Both are claimed to be calibration free, with some level of metrological traceability. However, the HAI paper by the same lead author is not even referenced here, which is quite odd. It is not clear what the connection is between these two instruments and which is more novel than the other. The authors should clarify the connection between these two instruments, before claiming that SEALDH-II is a novel hygrometer.

==> Thank you very much for that comment.

The instruments HAI and SEALDH-II are only similar on the first glance: The used technique (dTDLAS with a calibration-free evaluation approach) is similar, both instruments can be deployed on airborne platforms, and both are hygrometers. To grasp the differences between the instruments, we should first take look at the most important features of HAI: HAI is a multi-channel and multi-phase instrument developed for the German research aircraft HALO. The term “multi-channel” means here, that HAI measures with two different laser systems at 1.4 and 2.6 μm at two different measurement locations. The first measurement location is the “open-path” cell, which is installed on the fuselage of HALO, i.e. outside of the airplane. The second location is a “close-path” cell inside of the cabin aircraft. Additionally, HAI comprises two separate close-path cells, one for each wavelength.

The open-path cell only analyzes pure gas-phase water. By connecting the closed path cells to a forward-facing trace gas inlet, in order to sample “total water” (i.e. not only gas phase but also ice particles and droplets), the entire configuration of open and closed path cells allows “multi-phase” water measurements. In flight sections through clouds, liquid and/or frozen water particles are sampled via the inlet and then evaporated in heated sampling lines. Thus, the closed path cells determine the so called “total water”. Both instruments follow on slightly different path of the concept to ensure a permanent, highly defined, multi parameter instrument control for each individual data point during the application. Goal of this is to minimize uncertainty about the validity and correctness of each measurement point. To keep the SEALDH paper concise, we maybe didn't elaborate clearly enough on this connection between the instruments.

Indeed, the novelties of HAI and SEALDH-II are in different fields.

HAI is more complex, more versatile. Its particular novelty is the multi-phase, multi-wavelength approach and its field application and validation on one of the most modern airborne carriers. HAI's unique setup allows for the first time a highly redundant overall configuration which is enabling different versions of in-flight validations e.g. of sampling errors etc. Thus, HAI is clearly the most complex and versatile, calibration-free, airborne, multi-phase hygrometer.

HAI was until now not compared and validated at a primary standard and hence does not (yet) achieve SEALDH-II's metrological correctness/reliability/trust level. All HAI papers are more focused on field related science questions: For example, we wrote a paper about how to validate the dTDLAS relevant pressure in the open-path cell [1] which allowed us to be sure about the accuracy of HALO's gas pressure readings. However, this is not achieved for the open path gas temperature so far.

SEALDH-II on the other hand, compared to HAI, is significantly more advanced from a metrological point of view, i.e. with respect to data control, data treatment, metrological linkage, reliability level. In particular SEALDH-II is much more stringently validated at the highest primary metrological water standard. The novelty of SEALDH is thus clearly focused on the **metrological validation**.

Therefore, the full quality control from SEALDH-II still needs to be further developed, transferred and implemented in HAI. The present SEALDH manuscript deals exactly with these advantages of SEALDH II and with the results of the more stringent metrological validation experiment realized with SEALDH II. The novelty of SEALDH against HAI therefore is clearly given by the experimental rigorousness of the metrological validation. A further important novelty is that the metrological validation is realized over the full concentration and pressure range of the most relevant atmospheric regions of the UTLS and below. We added that some explanations to the manuscript, thank you for that hint!

We tried to make these differences clearer in the manuscript.

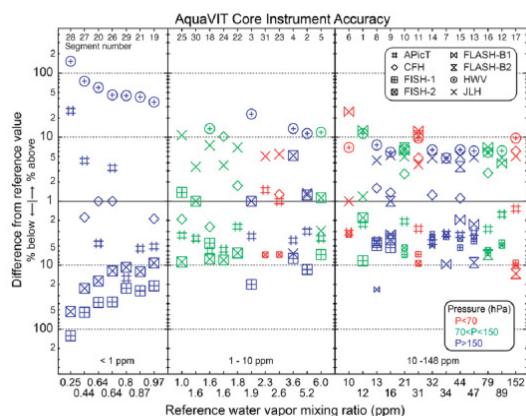
The authors also claim that this effort is the first metrologically validated humidity transfer standard. This may or may not be completely true.

==> We might have formulated our point not well enough, which could have led to some misconception by the reviewer. In essence, every metrological transfer standard which is used e.g. to compare different national primary standards with each other, or to transfer the representation for the metrological humidity scale to industry is a "metrologically validated humidity transfer standard. However in contrast to the reviewers statement, we wrote more detailed: *"With this validation, SEALDH-II is the first metrologically validated humidity transfer standard which links several scientific airborne and laboratory measurement campaigns to the international metrological water vapor scale."* We believe that this statement is true and justified, given the present manuscript and our previous papers on SEALDH. If the reviewer knows any publication about a full metrological validation of airborne laser hygrometers which contrasts or amends our statement, we would be thankful to include those in our paper and to adapt our statement.

During the AquaVIT -II campaign this community made a dedicated effort for a metrological validation of a number of instruments. The authors collected all observations, but never released the metrological reference observations. However, they presented this work at several conferences. That work may actually be the first metrological validation of several transfer standards.

==> It would have been very insightful if they had done exactly that. The peer reviewed paper and the White-paper [2], [3] of AquaVIT-1 explain in detail the deviations (an exemplary work!) but they are NOT linked back to any metrological source or device. The reason is obvious: At that time, the only available "metrological linked" devices were dew point mirror hygrometers (DPMH) which can be

traced back via a comparison with a primary generator. DPMH are often seen as the simple way of getting “metrological links”. The DPMH used during AquaVIT-1 was - to our knowledge - “only” traced back at 1 bar under static conditions. However, most of the AV-1 campaign was done at reduced pressures where the traceability was not given. What’s more, it is questionable whether the traceability can be transferred from an extractive device (the DPMH) to an 80 m³ large vessel (AIDA) which was operated in a quasi-static mode, but which – in particular at the low humidity levels and gas pressures – was far from reaching a static equilibrium as it showed significant drifts over time. It is fundamentally wrong to believe, that an instrument which is used outside its validated range (pressure, temperature, concentration,...) and operation conditions for which it is built (time response, gas matrix, impurities,...) can be used as a “metrological” standard even if it was once metrologically calibrated/validated. Doint mirror hygrometers only show correct values when they are in equilibrium (ice layer, humidity, mirror temperature). At most of the AquaVIT conditions, it also has to be considered that even short (1-2 m) sampling pipes to DPMHs required equilibration times at the low humidity levels (e.g. ppmv range) of many hours up to a day in metrological institutions. Especially instruments with a low flow (e.g. DPMHs) have great problems to reach a stable equilibrium and are prone to long term drifts. Hence, the best reference instrument for AIDA would have been a sampling-free instrument like APICT, which was – however – due to many reasons not chosen as reference. We assume that the authors of the AQUAVIT papers were aware of that and therefore both papers do NOT claim any metrological linkage. Instead, all instruments were compared to the mean value of the core instruments. AquaVIT, thus answered mainly the question about the “spread between the instruments” Due to the lack of a suitable reference, the AQUAVIT comparisons are not suitable to give answers about the level of absolute accuracy. The argument which is sometimes made, that the mean value of AquaVIT is like the “true H₂O value” is as wrong as to claim that one instrument was more accurate than another.



Referring to AquaVIT2, it is a fact that no final results and certainly no full papers have been published; neither from the simultaneous comparison exercise inside AIDA (called AV2a), which by the design of the experiment would very hardly ensure tractability, nor from the sequential metrologically motivated comparison with a PTB transfer standard generator (called AV2b). It also needs to be noted that the sequential comparison was only applied to a quite small sub group of the AV2 participants and hereby mostly to the group’s internal standards and to a much lesser extend to very few instruments used in AIDA. The AV2b comparison thus had a much reduced participation compared to the AV2a part. Even if the AV2b data once will be published, they will for sure be not comparable in rigor (setup, number of data points, concentration levels, and pressure levels) and in quality (control over system, time for validation) with the data presented here. As a consequence, AV2b is of much more indirect nature, which needs up to three calibration levels/steps to link the instrument of the relevant group to the primary standard at PTB. Second, the time planned for each individual comparison in AV2b was quite short to match the total time available for AV2a. Thus, for each of the instruments/group standards only very

few data points at 1000 hPa could be measured, with a quite short measurement time per data point. The pressure dependence could not be studied in detail.

To summarize, there are no papers and certainly no peer-reviewed full papers from AV-2b or 2a. We did work on this data and published some posters on preliminary results.

One of the drivers for the AquaVIT campaigns was the disagreement between some aircraft and balloon borne observations. Water vapor observations of less than 5 ppmv were the most important range of this disagreement. Given the uncertainty of SEALDHII, this instrument would not contribute to this concentration range. Despite the un-questioned quality of the observations presented here, this raises the question of the importance of their results, especially given the frequent reference to AquaVIT. The authors should clearly point out, whether they have achieved a metrological validation of a field deployed instrument that can measure stratospheric water vapor (100 hPa, 5 ppmv) with an uncertainty of less than 5-10%.

==> Thank you for that comment; nevertheless we are also a bit puzzled by it.

SEALDH is NOT an instrument designed especially for UT/LS application, nor is it one of the common “standard” even commercially available laser hygrometers like the ones from WVSS, Picarro, LosGatos, etc. Instead, SEALDH is designed as a metrological transfer standard which is still suitable for field applications and ensures full traceability to the primary water scale. Thus, the application range is wider than just UTLS and the requirements are also much more rigid and demanding. It is clearly stated in the abstract that we claim a “primary validation”, that covers the relevant mid to upper tropospheric and lower stratospheric (MT/UT/LS) concentration and pressure range. This was done via “15 different H₂O concentration levels between 5 and 1200 ppmv”. At each concentration level, we studied the pressure dependence at 6 different gas pressures between 65 and 950 hPa.” The introduction states (as well as the referred paper in section 2 [4]) that the instrument is built for a concentration range from 3 - 40000 ppmv; 0 -1000 hPa with a calculated uncertainty of $4.3\% \pm 3$ ppmv. The lower detection limit is defined at that point where measurement value equals uncertainty (as explained in the paper: “The calculated mixture fraction offset uncertainty of ± 3 ppmv defines the lower detection limit.”). The measurements in figure 8 show how conservative these estimations for the calculated uncertainties were since the “real” deviations are only in the 0 - 20% range (with the lowest level at 35%). SEALDHs science and metrology targets are thus much wider and of different perspective than those of the AQUAVIT 1 or 2.

The reasons why we explained the AquaVIT study in such a detail are the following: Firstly, it is the most extensive comparison exercise which was carried out on such a level (representative for the community, externally reviewed, blind submission, clear statements, exemplary work, etc.) that allows a reliable relative performance statement about the state of the art of airborne hygrometry. Secondly, as you can see in the graph above, AquaVIT covered the range up to 150 ppmv. Most instruments deployed during AquaVIT had a smaller concentration range compared to SEALDH-II because they were specifically developed for stratospheric measurements. With reference to the reviewer’s question, we have to comment that one should not compare apples with oranges. SEALDH-II is intended as a transfer standard for the troposphere AND lower stratosphere and thus designed as a wide range hygrometer. This asks for pragmatic compromises which have consequences. If we had in mind to make a comparison only for the stratospheric range, we could easily replace the 1 m long extractive cell with a fiber-coupled multipath cell having 10x - 50x more path length. This would drastically improve SEALDH-II’s sensitivity AND offset stability in the low concentration range. As the sensitivity and offset largely scales with the cell path length, this would lead for SEALDH-II to a calculated uncertainty (see [4]) of $4.3\% \pm 0.3$ ppmv (@ 10 m) and of $4.3\% \pm 0.06$ ppmv (@50 m) respectively, which seems quite well suited for a stratospheric application.

It also has to be kept in mind that we stated conservative calculated uncertainties; the “real” experimental deviations are usually smaller, as described and shown in this paper. In our experience – also as participant in AV 1 and 2 - we made the experience, that metrological humidity traceability options in the atmospheric sciences are more or less ignored, despite the availability of a fully validated metrological reference scale and infrastructure. It seems highly likely for us that an improved traceability would lead to better absolute accuracy of airborne hygrometry. SEALDH was designed to enable this link between metrology and field sciences. With the series of SEALDH papers and in particular with the one here under review, we intend to present the data that demonstrate this capability in a metrological sense and not just present another UTLS hygrometer.

So the argument of the reviewer that only improvements in stratospheric sensing would be required or relevant is a bit short sighted and ignores to a large extent our focus towards implementing general traceability for airborne hygrometry over the full range from LT to LS. Atmospheric science studies the entire atmospheric water cycle and not just its subsections in the stratosphere and as such, it would definitely be beneficiary to establish traceability with a wide range instrument such as SEALDH in order to improve the comparability between all atmospheric sub-compartments and not just between UTLS focused instruments.

We added a few lines to explain why we compare to AquaVIT.

Specific comments:

Line 18: I believe that the term ‘bridges this gap by implementing an entirely new concept’ is overselling their result. While their work is important, TDLAS technology is not new and has been around for quite a while. The authors own work on HAI show that this is not an ‘entirely new concept’. Furthermore, given the relatively large metrological uncertainties at true stratospheric water vapor concentrations, I don’t see, where a gap is being bridged.

==> We deleted the word “entirely” which was colloquial English.

Furthermore, the reviewer should be reminded that “The “new concept”” is not TDLAS. The new concept is validated traceability employed via a special, first principles TDLAS approach based on high-accuracy line data! It is a very common misconception that researchers who only “use” TDLAS think that all instrumental approaches have more or less identical properties – in particular with respect to the requirement for calibration. The statement to be “calibration-free” is also very often misused by lab focused TDLAS-groups without airborne experience as well as by commercial TDLAS instrument manufacturers. This is actually the reason why we term our approach “dTDLAS”, in contrast to alternative evaluation approaches. Thus, overselling happens at other points of the TDLAS community, certainly not at our point. TDLAS in itself only stands for “Laser absorption spectroscopy with a tunable diode laser”. “TDLAS” says nothing about the modulation approach or the data extraction and evaluation approach, so one has to be precise here.

Furthermore, SEALDH-II fortifies TDLAS with an unprecedented concept for a “complete” internal quality control mechanism, which yields a large and almost complete set of environmental/boundary parameters, the so called “housekeeping data”. This ensures that every single (!) captured raw scan and hence every reported H₂O value of the instrument has an assigned multi-parameter derived quality statement. This in combination with our first-principles calibration-free approach and our own high accuracy spectral data truly separates our instrument from any other, and allows us to achieve trustworthy and reliable measurements. It also ensures that a metrological validation (like the one in this work) in a protected laboratory environment can be transferred to harsh field environment (see [4] where the compensation of external effects are demonstrated with an environmental simulation chamber). It would be

helpful if the reviewer states any papers where a comparable or even better dTDLAS concept has been realized and was validated with a national primary standard, with our rigorouslyness.

Until then, we recommend not to generalize over an entire instrument class. But if one wants to do so, it is unavoidable to acknowledge, that instruments, which have been developed and published so far, do NOT have a data quality control concept at the level of SEALDH-II. This is often caused by the lack of a full physical model between the measurement value and the response of the instrument – the same reason why these instruments eventually have to be calibrated. The instrumental SEALDH paper [4] describes several important quality control scenarios which can be detected during SEALDH-II operation just because of the rich house-keeping data sensing any change in boundary conditions of the instrument and the measurement zone.

==> We revised the sentence.

Line 25, 'first metrologically validated': Aren't the AquaVIT-II and to some extent even the AquaVIT-I measurements metrologically validated?

==> As discussed above: AquaVIT was a major step towards assessing the quality of airborne hygrometers which we wrote and honored in the paper as well. Though, AquaVIT was not linked to the SI.

The Aquavit group could not / or didn't want to select a single reference instrument to serve as the main reference point, as there was obviously no knowledge about a traceability path of any of the participating instruments. As one of the participants in Aquavit 1 (and AV2), I remember well that there was no acceptance of the traceability concept and no discussion about a traceable reference instrument. We could not reach the conclusion that a single instrument was qualified as a reference type "gold standard". Hence, the best compromise the group could find was to take an average over the most mature "core" instruments. Without the concept of a unique primary reference and without any acceptance for traceability, it can't be stated that Aquavit1 is traceable, not even partially.

In Aquavit2, the only possible traceability path was planned via an extra, sequential, comparison of instruments (called AV2b) with a mobile traceable transfer standard from PTB which was transported to KIT. However, only a small number of instruments participated in the traceable comparison and no final data evaluation has been published yet. Therefore, the outcome of the traceable side-comparison called Aquavit2b is still open. Due to the large size of the AIDA vessel, it is also difficult to directly transfer the traceability to the AIDA experiment, which definitively limits the metrological impact of the Aquavit1 and 2 comparisons. Furthermore as mentioned above, AV2b was suboptimal with respect to the number of data points per instrument, the lack of comparison data at reduced pressure, and the short measurement time per data points which made it difficult to reach a stable equilibrium. Furthermore, mainly reference generators and not instruments were tested in AV2b, which results in a fairly long traceability chain over 3 and more levels, which significantly increases the uncertainty.

Note: It is unclear what the reviewer means with "to some extent"? Traceability is a binary property - so either "Yes, the instrument/data is traceable" or "No, it's not".

Lines 38 and 57: The tropical tropopause is highly relevant for atmospheric water vapor and may show values of less than 1 ppmv. This lower limit is a common value for some regions and seasons.

==> That's correct – even though to my knowledge averages lows are more in the range of 3-4 ppm and values below 1ppm are actually not very common. Nevertheless, we added the word "typically" to keep the statement general.

Line 46: The target accuracy for field weather stations is certainly a lot lower than 15%. Field weather stations report relative humidity and 2%-5% accuracy (in RH) are more common requirements.

==> That's was indeed confusing; we meant relative deviation and had more standard instrumentation in our mind rather than sophisticated instrumentation which is rarely used.

Note: Maybe we were talking about different things. We meant "relative accuracy", i.e. an absolute uncertainty of 5% rH at 50% rH or even 30% rH corresponds to a relative target uncertainty of 10% or even 16.5 %. So maybe we are not that far away from each other.

We revised the sentence.

Line 50: WVSS-II instruments are another variant of TDL instruments. They are commercially available instruments but probably not standardized.

==> WVSS is wavelength modulation spectroscopy based sensor which definitively has to be calibrated. We revised the sentence

Lin 55: Currently, in situ observations of water vapor are done at least up to 10 hPa and at gas temperatures of less than -90 deg C.

==> Changed

Line 79: : : 'does not facilitate a clear accuracy assessment'. This study is still highly valuable and able to characterize the status of in situ observations during that campaign. The fact that large differences, seen in earlier campaigns, were not repeated there is of great value, even though there is no direct metrological connection.

==> We agree with your sentence – it's not in contradiction to ours. If you look at the (Rollins et al., 2014) paper, you cannot draw a clear picture of the reason or cause for deviations: Were these deviations "real" e.g. because of different gas samples? Or: Was it caused by effects which modify the gas sample? Or by external driving forces like an increase of the instrument temperature? Environmental parasitic humidity or external pressure change? Calibration issues? Or one of the many other "airborne typical issues"?

Total comparisons like that are indeed valuable but they make it quite difficult and often even impossible to assess the internal causes of the deviation. They show more a "comparative state of the art", but make individual instrument assessments and assignments of systematic errors to certain problem or cause difficult. Hence, they do not show where improvements have to be done in the future. AquaVIT instead tried to distinguish between different effects (e.g. only the stationary, flat segments of the comparison where AIDA was as constant as possible). Deviations are only calculated within these "quasi-stationary" areas. This was done in order to suppress/minimize dynamic effects. The evaluation and comparison of the dynamic sections in AQUAVIT could still be done; to our knowledge, there is no intention to do so.

Line 87-88: Differences of a factor of two stimulated AquaVIT. Disagreements of 10% were largely considered within the individual instrument uncertainties.

==> Let's take a look at the AquaVIT paper: Uncertainties in the range above 5 ppmv: APicT <5%; CFH 4%; FISH-1 6%; FISH-2 6%; Flash 10%; HWV 5%; JLH 10%.

The results of AquaVIT (see plot above) show that two instruments deviate by up to 20% from each other. Keep in mind that the reference value is, as discussed above, was just a mean value – the "true" value is therefore still unknown. Even if we consider the mean value could be the "true" value, we would have the problem that the uncertainty statements of the AQUAVIT participants are insufficient. An uncertain-

ty is NOT the typical deviation of an instrument [5], [6]. One also could rephrase the statement and state: “AquaVIT shows that the at least some instruments underestimate their uncertainty”. However, I think this statement would be short-sighted since in the atmospheric communities “uncertainty”, “error”, “deviation”, “accuracy”, and “reproducibility” are often used interchangeably which causes this dilemma.

Each Aquavit participant had to state the performance or reliability or trustworthiness range of its instrument by giving an “offset term” and a relative “error”. Those numbers were not analyzed nor had to be justified or even mathematically deduced and were more or less estimates by the instrument PIs. As there was no “true” reference, i.e. no validated absolute value, the error statements of the AV instrument could only be coarsely checked. Thus again, a metrological evaluation of the uncertainty by referencing an instrument to a metrological standard and to derive a metrological uncertainty is a completely different almost complementary approach.

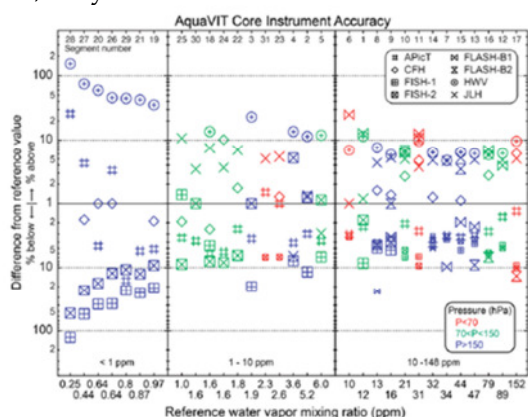
Line 90f: The goal of AquaVIT was to evaluate instruments under controlled conditions, not to rigorously evaluate each instrument’s uncertainties. No gold standard was included since no recognized standard was available for this setup.

==> Yes! We fully agree !

This is exactly the reason why our primary SEALDH validation is an indeed novel and highly relevant for the field! Contrary to a reviewer statement made earlier.

Line 128: Systematic differences of 20% and more were seen during AquaVIT at the lowest mixing ratios, i.e. below 3 ppmv. The authors should point out that SEALDH-II would not help addressing this concentration range.

==> I guess this question is referring to your third comment and hopefully already answered there. One more comment: The “span of up to 20%” describes the following graph. Keep in mind that these instruments are usually flown by themselves. If instrument A and instrument B report two measurement values, they could deviate 20% from each other (in the lower range even more).



Line 131: I doubt that this is the ‘first comparison’ with a metrological standard. Water vapor has been measured for a long time and a lot of validation efforts have happened, not all published. The AquaVIT-II activities, in which the authors have played an important role, is just one example.

==> The reviewer needs to be precise here: We do NOT claim the “first comparison with a metrological standard” in general, that would not be true. Instead, we present in this paper the “first comparison of an airborne hygrometer (SEALDH-II) with a metrological standard for the atmospheric relevant gas pressure (65 – 950 hPa) and H₂O concentration range (5 – 1200 ppmv)”. We believe that a detailed statement like this is true. In addition, we think it can be stated here, that AquaVIT-I was by no means

linked to the SI as discussed above and thus has nothing comparable to offer. For AquaVIT-II we don't understand the *diffuse, undocumented* "doubts" casted by the reviewer. We certainly don't know of any peer-reviewed final publication originating from Aquavit-2 which could challenge our claims, neither A) from the simultaneous comparison exercise in AIDA, which by the design of the experiment would be very hard to make traceable (but as said there is no full paper on AV-2), nor B) from the sequential comparison with a PTB transfer standard generator. The sequential comparison was furthermore applied only to a quite small sub group of the AV-2 participants and hereby mostly to the group's internal standards (not the instruments used in AIDA for AV-2a). This comparison part, termed AV-2b, did not have the wide participation like the AIDA part. In addition, it also did not cover the large validation range (in pressure and concentration nor did it invest sufficient amount of time as our work presented here (23 days! Permanent operation)). Even if the AV-2b data once will be published, they will for sure be not comparable in rigorousness and quality with the data presented here. One reason for this is the more indirect nature of the comparison, which needs up to three calibration levels/steps to link the instrument of the relevant groups to the primary standard at PTB. PTB just recently modified its primary standard to provide different gas pressure levels; we can for sure say that no such a comparison was linked to PTB (Germany) in the past. So again, there are no papers and certainly no peer-reviewed full papers from AV-2b. To our knowledge, there still needs to be a final data discussion meeting amongst all participants to be announced in order to decide about the further evaluation and use of the AV2 and the AV2b data.

We are of course very interested in learning which comparisons the reviewer has in mind, and would be grateful if the relevant citations could be forwarded to us. This includes publications of comparisons as well as any other metrological validations of airborne laser hygrometers, we would be very delighted to read, assess, discuss, and if suited add them.

Until then, we think our above statement is valid as well as justified and it's on the reviewer to provide references and detailed information if he is aware of comparable earlier work. To our best knowledge our data are unique in quality and in impact.

Line, 158, 165, 345-349: The lower limit of 3 ppmv is a significant limitation, since the <10 ppmv range is essential for stratospheric observations. At 5 ppmv an uncertainty of 3 ppmv makes the measurement effectively useless for stratospheric research. This should be discussed in greater detail.

==> We looked through the paper to make sure at any point that SEALDH-II is not presented as a stratospheric instrument as we also explained at other comments. That has never been our intention. We added some words.

Lines 207ff: There are other calibration free instruments. HAI, published by the authors is one of them. Some of the frostpoint hygrometers, which are being used on aircraft and balloons may be considered calibration free in the same sense. Other TDL instruments are equally considered calibration free under the definition of the authors.

==> We indeed understand the point which the reviewer wants to make, but - as explained above - we don't "oversell" and we have given in the paper presented here detailed technical explanations what we mean with "calibration-free" and even more important we have given experimental justification why we can claim this properties. It is certainly quite decisive not just to CLAIM that the instrument is calibration-free (which (too) many groups do) but also to proof it and to VALIDATE this property as we do here via a side by side comparison with the world's highest accuracy water scale and one of the few primary water standards. To our knowledge, none of the airborne TDLAS instruments has ever done this. For organizational reasons and HALO mission deployments, HAI was never available long enough in order to experimentally validate this property at the primary standard. Even though there were HAI papers coming out recently, none of them was focusing on the validation of the absolute accuracy, as

this exercise was done with the SEALDH instrument which wasn't in such long and frequent airborne missions. Again, we ask the reviewer to state papers from other air- or balloon-borne TDLAS instruments which prove a primary metrological validation exercise. We don't know any. Essentially there are many groups which claim that property but no one so far challenged, validated, and quantified experimentally the degree of accuracy which can be achieved.

Referring to a frost point instrument, we can make the following statement. Calibration-free would mean the following: The instrument is assembled and e.g. the temperature/pressure sensors (if not first principle methods) are calibrated. Then the "target" gas is guided through the frost point hygrometer and it has to report an absolute value of e.g. "500 ppmv". In an ideal world, this might be possible (using the Sonntag's equation [7], ideal gas law, etc.); however, in a real world it is not. This also explains that even the most sophisticated frost/dew point hygrometer have to be calibrated regularly. Reasons for that are e.g. technical problems such as: The temperature sensor has to measure the surface (!) temperature of the mirror; the constant airflow above the mirror transfers heat away. I agreed, all this could be modelled and corrected in principle – however nobody (to our knowledge) has ever published such a work. Operational problems such as: Any kind of hydrophobic substance, any dirty, micro scratches on the mirror, etc. change the ice layer behavior locally. Therefore, the "frost point" can be shifted slightly which is corrected during calibration. Physical problems: The enhancement factors (which describe the dew point shift if the gas is not ideal (Sontag equation is not accurate in this case)) cannot be calculated due to the lack of a full physical model.

Line 216, 'accuracy': JCGM (2008) recommends not using this term in a quantitative sense. The authors should explain what they refer to here.

==> Line 216 contains: "... philosophy leads to measurements which are very reliable with respect to accuracy, precision" and the comment is probably referred to the line: "measurement accuracy" defined as "closeness of agreement between a measured quantity and a true quantity value of a measurand. NOTE: The concept 'measurement accuracy' is not a quantity and is not given a numerical quantity value. A measurement is said to be more accurate when it offers a smaller measurement error"

As mentioned above, we believe that in non-metrological, application and field science oriented journals like AMT, we have to find a bridging language which can convey and explain the general idea and results of the scientific work. For this paper here, we decided to stay as often as possible within the terminology of the AMT community and to find a compromise between the common understood language and the rigorous metrological terminology. We strongly believe that our paper would not profit, probably be less understandable, and thus have less impact if we add not commonly used expressions such as trueness and closeness. Therefore, we assume that every reader will understand the usage of "accuracy". A metrologically much more rigorous discussion of the implications of our work will be subject to upcoming papers in common journals of the metrological community like "Metrologia" or the like.

Lines 220ff: The authors point out later in the manuscript, that calibration in the strict sense improves the measurements only, if the ambient conditions can be replicated during the calibration. They should elaborate on this topic and consolidate the various paragraphs throughout the manuscript.

==> Long-term stability (long = relatively longer than the drift of the instrument) of the instrument is a requirement for any calibration. We wrote: "(...) a calibration can only improve the accuracy for the relatively short time between two calibration-cycles by adding all uncertainty contributions linked to the calibration itself to the system."

This sentence refers to every calibration process. The replication of the ambient condition is necessary if the impact of the ambient conditions on the calibrated device is unknown. As in the comment for Line

216, we will give a more rigorous discussion in planned papers for the metrological community. In the interest of keeping this manuscript in an acceptable length to content ratio we can't discuss general topics like this in the given journal and paper.

Lines 264-267: Delete. These sentences contribute nothing and could be deleted.

==> We don't understand why these lines should be deleted! These lines contain: "In this paper, we present data from a permanent 23 day long (550 operation hours) operation in automatic mode. Despite a very rigorous and extensive monitoring of SEALDH-II's internal status, no malfunctions of SEALDH-II could be detected. One reason for this are the extensive internal control and error handling mechanisms introduced in SEALDH-II, which are mentioned above and described elsewhere (Buchholz et al., 2016)." This comment might link to the question above "why SEALDH-II has a new concept" which we hopefully have been clarified. Typical TDLAS instruments such as e.g. the WVSS-II (mentioned above) just give a value with a minimalistic "set" of operational data. A definite assessment whether the measurement process went well or had issues inflicting the quality is not possible. SEALDH-II embodies a new, holistic, embedded concept. Therefore, this question should now be answered and the content of the sentence should be clear: "No error" with an instrument such as SEALDH-II means: "no error" and not "maybe / probably a correct measurement"

Lines 301f, 'One has to compare : : ' No, this comparison does not have to be done. The purpose of AquaVIT was very different and a metrological standard was not available at that time. This statement should be deleted.

==> We have already revised this sentence according to the comments above. It seems that the reviewer presumes that we do not appreciate the tremendous contribution which AquaVIT brought to the hygrometer community. We wrote (line 89) "AquaVIT was a unique first step to document and improve the accuracy of airborne measurements in order to make them more comparable". If the reviewer would like to add an even more appreciative sentence, we are happy to include it!

Lines 306fff (section 4.1): Isn't the point of controlled static setups to minimize the impact of dynamic effects on the uncertainty estimation? Fundamentally the uncertainty of the SEALDH-II cannot be better than that of the THG. Therefore, the authors should quantify the impact of the THG dynamic effects on their static uncertainty estimation of SEALDH-II, if that is possible.

==> This statement is correct. The figure 2 shows exactly that: The uncertainty stated there (line 643) is "The uncertainties of every individual calibration point are stated as green numbers below every single measurement point". Figure 5 left allow double checking of this statement. The max peak-to-peak deviation is approx. 4 ppmv at 1250 ppmv equals 0.3%. The Figure 5 right is an extreme case where the dew point mirror data are oscillating. A section like that is not preferable/acceptable for a validation.

We added that for clarification.

As a site note: A few comments above, the reviewer stated "2%-5% accuracy (in RH) are more common requirements". Figure 5 right shows that a measurement which has a small dynamic change can easily oscillate a frost point mirror hygrometer in the order of 2%. Therefore, a reliable accuracy of 2% is not "easily" achievable with frost point mirror hygrometers which are often seen as the most accurate hygrometers.

Line 314: What is an 'indirect, inertia, thermal adjustment process'? The authors should find a better term for what is meant here.

==> Revised

Line 342: PHG should be Primary Humidity Generator.

==> Revised, thank you

Lines 365ff, 'It is important : : ': What does this sentence mean? Any uncertainty estimate always implies that the true uncertainty could be smaller. It could also be at the estimate. Lines 376 through 378 are somewhat contradictory. The authors place great value that the measurements presented here are the first metrological validation of SEALDH-II. How would non-metrological validations done previously provide contribute?

==> Response to the first question: We have already answered this question to some extent above: The words "uncertainty", "error", "deviation", "accuracy", and "reproducibility" are often used interchangeably in different communities: E.g. if an instrument has a typical "deviation" of 2%, the uncertainty is sometimes just defined as 2%. From a metrological point of view, such an "uncertainty definition" has to be seen critical. An uncertainty budget/consideration should include all significant factors to understand their individual contributions. An aggregated value, such as a deviation, does not facilitate further statements about the "reliability" of this so defined uncertainty. The more the deviation measurements includes other parameters such as instrument temperature, external/internal pressure, power supply fluctuations, vibrations etc. the more reliably a deviation statement can be transformed into an uncertainty - even with a black-box-like system. SEALDH-II is designed as a fully white box system to avoid in the first place these kinds of "deviation=> uncertainty" discussions.

==> Response to the second question: We described AquaVIT as the largest, most representative, non-metrological laboratory comparison exercise. From our point of view, AquaVIT is unique in terms of quality level and the conclusions which could be drawn from it. Therefore, we did not describe other airborne or laboratory campaigns since a detailed analysis would lengthen this paper unnecessarily and would not add value since the overall statements remains. If the reviewer is interested to read more about other water vapor related comparisons: here are some entry points [8]–[15]

Do the authors imply that non-metrological validations are equally useful or even better suited to address the uncertainty issue at low pressures and low mixing ratios? As shown in this manuscript, the uncertainty of SEALDH-II at true stratospheric values (low pressure and low mixing ratios) is too large to be scientifically relevant.

==> First part of the question: A non-metrological validation has by definition no direct link the SI units. Therefore, it is well suited for comparing instruments relatively to each other but as soon as the absolute value is of strong interest, a metrological link should be established, to provide comparability between different instruments. This is a general recommendation and does explicitly not depend on any pressure and/or mixing ratio range.

==> Second part: SEALDH-II is – as mentioned frequently – not designed as a dedicated stratospheric instrument. It could be adopted for LS conditions but so far it isn't ; we clarified this in this rebuttal several times as well as in the paper. SEALDH-II could, as explained above, set up with a longer optical path length to serve as a standard for stratospheric values. But currently, SEALDH-II is set up for tropospheric concentrations up to 40000 ppmv and with its uncertainty of $4.3\% \pm 3$ ppmv not well suited for single digits ppmv measurements. The measurement principle, however, could be easily adapted to LS conditions.

Lines 428f: Why the authors would want to suppress this systematic pressure dependence? In instrument comparisons and atmospheric measurements the systematic biases are often the determining factors.

==> We don't suppress. If we did, we would only show the "corrected" i.e. calibrated data. As we have a clear physical explanation for the pressure dependence, which is an inherent deficit of the VOIGT line

shape, we certainly know that by expansion of our physical measurement model – i.e. inclusion of a higher order line shape profile – we could "remove" this pressure deviation at the cost of higher computational efforts and an enlargement of the number of necessary fit parameters. Recent developments in higher order line shape models (such Speed-dependent-Voigt (SDV) [16] or Hartmann-Tran-Profile (HTP) [17]) are most suited for such an improvement, which we have planned for future updated versions of our fitting model.

Line 667f: What do the authors want to say here? The sentence as is doesn't make sense.

==> Maybe we talk about different lines: "The different dynamic characteristics of SEALDH-II (fast response time) and THG (quite slow response) lead in a direct comparison to artificial noise." We added here a sentence.

Figures 6-8: The abscissa should be shown as the Log of P. This makes it easier to relate the altitude and emphasizes the lower pressures, where water vapor is more challenging.

==> Thank you for that comment. SEALDH-II is a wide range (3 - 40000 ppmv) hygrometer. Therefore, the entire pressure range is equally important. This comment is probably inspired by the fact that the reviewer assumed SEALDH-II might be a purely stratospheric instrument. We clarified that (see multiple comments above)

The authors use the term 'calibration-free' excessively and should reduce it to the necessary amount. The term is defined in a dedicated section and does not need to be repeated subsequently. The authors do not seem to be completely familiar with the water vapor observation community. Stratospheric water vapor is also observed on large and small balloons reaching all the way into the middle stratosphere. These measurements use a variety of techniques, none of which are referenced, but should be referenced. Water vapor is also measured using remote sensing (Raman and DIAL lidar), which are technologies comparable to their own. In particular DIAL measurements are considered calibration free and traceable measurements.

==> Thank you for this comment. SEALDH-II is not a stratospheric hygrometer; therefore, we didn't focus on this single atmospheric region. The target of this paper is a "Pressure dependent absolute validation from 5 – 1200 ppmv at a metrological humidity generator" and not to write a review about different water measurement techniques even up to the middle stratosphere. As an entry point for such a review for different measurement methods: [18]

We would be interested to learn more about the "default" traceability of DIAL measurements. The technique certainly requires traceable absorption coefficients, which are very difficult to get, due to the lack of traceable spectral data. We would be delighted to receive some references about this specific topic.

Lines 61-64: These statements are much too broad and even incorrect. The vast majority of water vapor observations has been quite sufficient for validation studies of models. The limiting factor in model validation is usually the availability and coverage of these observations, not their quality. The authors should change this statement.

==> The statement was: "These and other impacts complicate reliable, accurate, long-term stable H₂O measurements" followed by a brief outline why water vapor measurements remain a quite difficult in-situ measurement in the field, even if they are nearly always needed in atmospheric science. Up to now, the lack of sufficient accuracy may have limited important scientific interpretations (Krämer et al., 2009; Peter et al., 2006; Scherer et al., 63 2008; Sherwood et al., 2014)."

==> We added the comment about availability and coverage – thank you for that.

We would like to note that we can't follow the reviewer's opinion that "quality" is not a limiting factor in atmospheric science. This is definitively too general. I remind e.g. of a paper about observations of atmospheric super saturations of 300% in the UTLS, well beyond the homogenous nucleation threshold! It is not unlikely that these were caused by significant offsets of non-traceable instruments, which then led to this extreme numbers. Other examples could open discussions on sampling effects in gas lines to extractive water sensors. Here, our 4-fold redundant HAI sensor could show in the future a quantification of such effects. We are convinced that data quality and quantification of disturbances is actually often a topic in atmospheric H₂O detection.

Technical comments:

Line 52: standard (singular)

=> thank you

Line 60: delete 'a quite'

=> thank you

Line 60: measurements (plural)

=> thank you

Line 73: Better: The latter is particularly important for investigations in heterogeneous regions in the lower troposphere as well as for investigations in clouds.

=> thank you

Line 101: :: : inside the aircraft: ::

=> thank you

Line 281: Replace ')(' with ' , '

=> thank you

Line 359: Delete '(primary standard = calibration-free)', which is a meaningless repetition here. Also delete 'calibration-free' in the same line, which is again a repetition.

=> We revised to emphasis that two calibration-free and one calibrated device are compared with each other.

Line 155, 389, 413: What is the meaning of 'holistic' in this paper? Better to delete this term.

=> We introduces this term for a situation where a "state of the art detection principle" is embedded in a sophisticatedly controlled/supervised environment. I guess the comment above "why is SEALDH-II a new concept" and this comment are linked. We emphasized that better in the text.

- [1] B. Buchholz, A. Afchine, and V. Ebert, "Rapid, optical measurement of the atmospheric pressure on a fast research aircraft using open-path TDLAS," *Atmospheric Measurement Techniques*, vol. 7, pp. 3653–3666, (2014), doi:10.5194/amt-7-3653-2014.
- [2] D. W. Fahey, H. Saathoff, C. Schiller, V. Ebert, T. Peter, N. Amarouche, L. M. Avallone, R. Bauer, L. E. Christensen, G. Durr, C. Dyroff, R. Herman, S. Hunsmann, S. Khaykin, P. Mackrodt, J. B. Smith, N. Spelten, R. F. Troy, S. Wagner, and F. G. Wienhold, "The AquaVIT-1 intercomparison of atmospheric water vapor measurement techniques," *Atmospheric Measurement Techniques*, vol. 7, pp. 3159–3251, (2014), doi:10.5194/amtd-7-3159-2014.
- [3] D. Fahey and R. Gao, "Summary of the AquaVIT Water Vapor Intercomparison: Static Experiments," source: https://aquavit.icg.kfa-juelich.de/WhitePaper/AquaVITWhitePaper_Final_23Oct2009_6MB.pdf (last accessed: May 2014), (2009).

- [4] B. Buchholz, S. Kallweit, and V. Ebert, "SEALDH-II—An Autonomous, Holistically Controlled, First Principles TDLAS Hygrometer for Field and Airborne Applications: Design–Setup–Accuracy/Stability Stress Test," *Sensors*, vol. 17, no. 1, p. 68, (2016), doi:10.3390/s17010068.
- [5] Joint Committee for Guides in Metrology (JCGM), "Evaluation of measurement data - An introduction to the 'Guide to the expression of uncertainty in measurement' and related documents," *BIPM: Bureau International des Poids et Mesures*, www.bipm.org, (2009).
- [6] JCGM 2008, "JCGM 200 : 2008 International vocabulary of metrology — Basic and general concepts and associated terms (VIM) Vocabulaire international de métrologie — Concepts fondamentaux et généraux et termes associés (VIM)," *International Organization for Standardization*, vol. 3, no. Vim, p. 104, (2008), doi:10.1016/0263-2241(85)90006-5.
- [7] D. Sonntag, "Important new Values of the Physical Constants of 1968, Vapour Pressure Formulations based on the ITS-90, and Psychrometer Formulae," *Meteorologische Zeitschrift*, vol. 40, no. 5, pp. 340–344, (1990).
- [8] M. Heinonen, "A comparison of humidity standards at seven European national standards laboratories," *Metrologia*, vol. 39, no. 3, pp. 303–308, (2002), doi:10.1088/0026-1394/39/3/7.
- [9] J. M. Livingston, B. Schmid, p. b. Russell, J. Redemann, J. R. Podolske, and G. S. Diskin, "Comparison of Water Vapor Measurements by Airborne Sun Photometer and Diode Laser Hygrometer on the NASA DC-8," *Journal of Atmospheric and Oceanic Technology*, vol. 25 (10), pp. 1733–1743, (2008).
- [10] S. J. Abel, R. J. Cotton, P. a. Barrett, and a. K. Vance, "A comparison of ice water content measurement techniques on the FAAM BAe-146 aircraft," *Atmospheric Measurement Techniques Discussions*, vol. 7, no. 5, pp. 4815–4857, (2014), doi:10.5194/amtd-7-4815-2014.
- [11] M. Helten, H. G. J. Smit, and D. Kley, "In-flight comparison of MOZAIC and POLINAT water vapor measurements," *Journal of geophysical research*, vol. 104 (D21), pp. 26087–26096, (1999).
- [12] S. J. Abel, R. J. Cotton, P. A. Barrett, and A. K. Vance, "A comparison of ice water content measurement techniques on the FAAM BAe-146 aircraft," *Atmospheric Measurement Techniques*, vol. 7, no. 9, pp. 3007–3022, (2014), doi:10.5194/amt-7-3007-2014.
- [13] R. A. Ferrare, S. H. Melfi, D. N. Whiteman, K. D. Evans, F. J. Schmidlin, and D. O. Starr, "A comparison of water vapor measurements made by Raman lidar and radiosondes," *Journal of Atmospheric & Oceanic Technology*, vol. 17, pp. 1177–1195, (1995), doi:10.1175/1520-0426(1995)012<1177:ACOWVM>2.0.CO;2.
- [14] A. K. Vance, S. J. Abel, R. J. Cotton, and A. M. Woolley, "Performance of WVSS-II hygrometers on the FAAM research aircraft," *Atmospheric Measurement Techniques*, vol. 8, no. 3, pp. 1617–1625, (2015), doi:10.5194/amt-8-1617-2015.
- [15] H. Vömel, V. Yushkov, S. Khaykin, L. Korshunov, E. Kyrö, and R. Kivi, "Intercomparisons of Stratospheric Water Vapor Sensors: FLASH-B and NOAA/CMDL Frost-Point Hygrometer," *Journal of Atmospheric and Oceanic Technology*, vol. 24, no. 6, pp. 941–952, (2007), doi:10.1175/JTECH2007.1.
- [16] C. D. Boone, K. A. Walker, and P. F. Bernath, "Speed-dependent Voigt profile for water vapor in infrared remote sensing applications," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 105, no. 3, pp. 525–532, (2007), doi:10.1016/j.jqsrt.2006.11.015.
- [17] J. Tennyson, P. F. Bernath, A. Campargue, A. G. Cs??sz??r, L. Daumont, R. R. Gamache, J. T. Hodges, D. Lisak, O. V. Naumenko, L. S. Rothman, H. Tran, J. M. Hartmann, N. F. Zobov, J. Buldyreva, C. D. Boone, M. D. De Vizia, L. Gianfrani, R. McPheat, D. Weidmann, J. Murray, N. H. Ngo, and O. L. Polyansky, "Recommended isolated-line profile for representing high-resolution spectroscopic transitions (IUPAC technical report)," *Pure and Applied Chemistry*, vol. 86, no. 12, pp. 1931–1943, (2014), doi:10.1515/pac-2014-0208.
- [18] Manfred Wendisch and J.-L. Brenguier, *Airborne Measurements for Environmental Research: Methods*

and Instruments. WILEY-VCH Verlag GmbH & Co. KGaA, (2013).

We thank all the reviewers for carefully reading our manuscript and for the detailed feedback aimed at helping us to further improve the manuscript. Below we address the raised concerns in a point by point fashion. Changes are highlighted in the attached revised version.

Anonymous Referee

The paper addresses the relevant scientific questions on how to measure atmospheric water vapour more accurately. The experiments are thoroughly conducted and the results well discussed, as they seem to tackle the real measurement issues. The text gives enough details and clarifications, so that it is fairly easy to follow, although it would benefit from shortening it a bit.

My main comment is about the argument that the instrument is calibration-free. The authors do discuss this in page 6, however I believe more careful wording would be needed. Namely, the instrument does indeed measure the water vapour concentration without relating it to the quantity of the same kind (humidity). This could arguably be called an absolute measurement, where water vapour concentration is indirectly measured through quantities of different kind by using an improved physical model. However, that is in essence true also for any other instrument type, e.g. gravimetric hygrometer through mass, chilled mirror hygrometer through temperature, an impedence-based hygrometer through impedence etc. Even though the authors do fairly discuss what they mean by calibration-free, it should still be noted, that in order to obtain the water vapour concentration indirectly, the instrument has to measure different parameters directly (temperature, pressure etc.), which eventually requires a calibration of the individual instruments.

==> Thank you very much for that comment.

In the atmospheric science communities, there are several words commonly used: Calibrated, calibration-free, self-calibration, in-situ-calibrated; outside of these communities as well: first-principles method, primary method etc.

In Metrology a “calibration” is defined by (JCGM 2008, 2008): “calibration (...) in a first step, establishes a relation between the measured values of a quantity with measurement uncertainties provided by a measurement standard (...), in a second step, this information is used to establish a relation for obtaining a measurement result from an indication (of the device to be calibrated)”. This metrological explanation is pretty close to those calibrations “typically” used for laser- based instruments [1] and also often described in relevant individual papers such as [2]–[5].

The reason for choosing “**calibration-free**” is to unambiguously emphasize that SEALDH-II does not use / rely on any kind of such a classical calibration process based upon a water vapor reference.

The word “**self-calibration**” is used for systems which calibrate themselves during operation using an internal reference (e.g. [6]). Typical examples are temperature sensors, which use the Curie effect or phase change transitions as temperature reference points inside of the sensor. These kind of sensors do not need an *external* calibration process; however, their accuracy depends directly on the quality of the “built-in reference” and the internal calibration cycle which is done with sensor dependent strategies. “**In-situ**” or “**online**” calibrated systems follow usually measurement cycles following a pattern like “calibration – measurement – calibration etc.” to cyclically link the absolute accuracy of the instrument to a reference analyzer/generator which is during the calibration process connected to the instrument to be calibrated.

From a metrological point of view, the term “calibration-free” is not as broad as the terms “first principle method” or “primary method”. The word “absolute” is commonly seen as the opposite of “relative”. E.g. gas pressure transmitters are sold as relative and absolute versions. If we named a calibrated instrument not as an absolute device, we would confuse many readers significantly since this view “relative to an absolute reference” seems not to be commonly used. Thus, we strongly believe that an optimum, com-

monly understandable word choice for our instrument characteristics which provides and ensures the desired clarity and quick comprehensibility of the idea is “calibration-free”.

When we talk about “calibration-free” in the context of SEALDH-II, we do not refer to a special feature of this individual instrument realization; we refer to a feature of the entire instrument family i.e. the evaluation principle itself.

Let’s think about a typical scenario of a calibrated instrument: The application requires from the instrument sufficient long-term stability (the expression “long-term” indicates here: “longer than the time span between two calibration cycles”). The correlation between measurement “value” and the real “physical quantity” has to be deterministic. Both properties together allow an absolute calibration of an instrument. Subclasses are instruments which have an offset drift; they still can be calibrated but will only provide a relative measurement rather than an absolute measurement.

Measurement principles which allow a calibration-free evaluation have in common, that they do not rely on/include any “instrumental specific” adjustment parameter and that they derive/calculate the final measurement result based on a first-principles based approach which relies on a physical model of the instrument and the measurement process. In our case, the instrumental response relies on a physical model of the light absorption induced by a spectrally sufficiently resolved ro-vibrational absorption transition of the water molecule. One of the most essential parameters is hence the line strength of the chosen water transition. This is a “molecular property”, which does not have a spatial or temporal variation, which could modify the instrument response. The molecular parameters of the water transitions are in that sense ideal “transfer standards” to link instrument and measurements to the SI

The reviewer made the argument, that we still use calibrated devices inside of SEALDH-II. This is true, but these calibrations are only needed because we need absolute values for L , p , T , etc. These calibrations are not used to remove non-understood instrumental deviations/drifts outside the physical model of the measurement process, i.e. to calibrate any instrumental specific deviations. As an example: SEALDH-II needs the length of the optical path. If one wanted to get rid of this “measurement”, we could do it directly: Length is defined by the speed of light and time. Time is defined by a molecular transition (energy). It would be highly impractical to set up an atomic clock inside of SEALDH-II, but it’s not impossible, just impractical for the desired airborne application. In general, other national primary standard also do not rely on a complete set of “only primary principles”, even if this is possible; it would be confusing to say “the primary standard is calibrated”. Therefore, we believe people outside of Metrology will understand “traceable, based on first principle” as “calibration-free” and vice versa.

Different example: Let’s think for a moment how our general approach contrasts the sometimes made argument that frost/dew point mirror hygrometer didn’t need a calibration: Here, “calibration-free” would mean the following: After mechanically and electronically setting up an instrument, the “target” gas is guided through the frost point hygrometer and it would report an absolute value of e.g. “500 ppmv”. In an ideal world, this might be possible (using the Sonntags equation [7], ideal gas law, etc.); however, in a real world it is not. This also explains that even the most sophisticated frost/dew point hygrometer have to be calibrated regularly.

Reasons for that are e.g.:

Technical problems such as: The temperature sensor has to measure the surface (!) temperature of the mirror; the constant airflow above the mirror induces heat transfer losses; or the contamination of the condensation mirror with other gas constituents induces offsets by shifting the ice/dew formation; or the temperature sensor used for the mirror surface temperature suffers drifts caused by aging processes. We agree that all this could be modelled and corrected in principle – however nobody (to our knowledge) has ever published such a work. The model would have to be so general, that exact knowledge of the geometry (= length) of the chamber would be sufficient to calculate the correction function based on first principles.

Operational problems such as: Any kind of hydrophobic substance, any dirt, micro scratches on the mirror surface, etc. locally modify the ice/dew deposition behavior. The “frost/dew point” can be shifted slightly from the ideal situation which is corrected for via a calibration process. Therefore, the calibration-free dew point mirror hygrometer would need a broad set of additional measurement devices to analyze, detect, and quantify such influencing effects and to adopt the right frost/dew model.

Further, physical problems need to be solved: The enhancement factors (which relate the dew point shift if the gas is not ideal (Sontag equation is not accurate in this case)) cannot be calculated due to the lack of a full physical model. Condensation/freezing models are currently not entirely isolated from material properties. There seems to be no fundamental problem but the current models cannot cope with the complexity at the required very high accuracy of a research type frost point mirror. Vice versa, if dew/frost point mirrors were operated in a “relaxed” sensitivity regime, we would expect, that the small shifts induced by the problems mentioned above become insignificant, hence could be ignored, thus making a “reduced”-sensitivity dew point hygrometer also potentially calibration-free. The question what calibration-free means and how it is achieved therefore always has to be answered in the framework of the performance requirements for a the desired application: Is the physical model of the measurement process and the realization of the corresponding instrument understood well enough so that the measurement process can be evaluated accurately enough without relying on a comparison with an external reference process (i.e. via a calibration to a reference), then a physical model could allow a calibration-free first-principles evaluation. If the accuracy requirements are too high for a given instrument/evaluation configuration then the instrument will require calibration via a comparison with an external/internal reference.

The design process of SEALDH-II was exactly governed to answer these questions: I.e. the question to be answered was e.g. what “enabling” accuracies of the spectral parameters, the temperature, pressure and length measurements and other controlling critical effects are needed and have to be embodied in the instrument in order to be able to realize a first principles TDLAS based measurement of the water vapor mixture fraction with a total accuracy sufficient for the airborne applications.

To summarize: Papers should have “clear” and commonly understandable statements for the target communities, which are here the applied atmospheric sciences. From this point of view, “calibration-free” is more widely comprehensible. “Traceable via a first principle approach” would be more precise but is much less common and understandable outside of the Metrology community, since the exact definitions of “traceability” and “first principle” is still to be disseminated more commonly.

It could be further discussed, though, whether the principle gives a potential to serve as a primary standard. They (the primary standards) do employ the absolute measurement in this sense, but they also need to be generally accepted (or chosen by convention, according to VIM). A similar situation is with chilled mirror hygrometer, which is not treated as a primary standard, but is nevertheless typically used in conjunction with it (or the SPRTs with fixed points for instance). And regularly calibrated against it.

==> We also thought about that, but eventually decided not to discuss that in an atmospheric sciences journal as this would be out of the scope of AMT. A discussion like this would fit better in "Metrologia" or similar more metrologically oriented journals. Due to this, we intend to provide a more metrologically oriented discussion of our results in a future paper.

The way from a “traceable via first principle” to a “national primary standard” is long-term process. Since the readers of AMT are dominated by the atmospheric sciences communities, we decided against an out-of-scope discussion focusing on general Metrology. Therefore, we also didn’t explain e.g. if and how SEALDH-II could get a CMC entry.

In this respect also a more evaluation of the long-term drift would need to be conducted before a new metrological classification could be discussed, despite the argument of the offset compensation.

==> We fully agree with this statement. But this "long-term" evaluation needs time i.e. several months to 1-2 years to be long-term. Data in this paper show already, that the long-term stability seems higher than the uncertainty of the dew/frost point hygrometer. Therefore, a comparison like that has to be directly done at the national primary standard, which is pretty occupied and busy for service calibrations at PTB. But: We are currently working on that; e.g. the primary standard was recently upgraded with a setup to facilitate validations at different gas pressures.

For this reason I would suggest to avoid the notion of calibration-free standard, but rather to stress out an alternative advantages of the SEALDTH-II and of its evaluation.

==> We hopefully convinced the reviewer, that the word choice is a good compromise between metrology and atmospheric science and that it is the best fit for the atmospheric community, which is the one we are focusing on with this paper.

Specific comments:

- Page 1, line 23: SEALDH is not the first metrologically validated humidity standard; consider rephrasing

==> We wrote "With this validation, SEALDH-II is the first metrologically validated humidity transfer standard which links several scientific airborne and laboratory measurement campaigns to the international metrological water vapor scale"

It emphasizes on the linkage between airborne, laboratory and Metrology. We added words to make it even clearer. To our knowledge, SEALDH is the first "airborne, metrologically validated humidity transfer standard". If there are previous publications which demonstrating these properties, we ask the reviewer to please provide a reference to the publication.

- Page 2, line 34: Water vapour measurement is often needed: : : The word measurement or similar is missing

==> Revised

- Page 2, line 46: consider deleting words "such as"; giving the reference is enough

==> Revised

Page 2, line 61: falsification is a strong word; consider revising

==> Revised

- Page 3, line 86: instead of "entirely transferred to", "represented by" would sound more appropriate (or similar)

==> Revised

- Page 3, line 100: Are you talking about desorption? If so, put it more explicitly.

==> We mentioned that in the sentence before: "...can lead to signal creep due to slow adsorption and desorption processes,..."

- Page 4, lines 120 to 124: Please consider revising in the light of general comment above.

==> We revised the last sentence

- Page 4, line 128: Why is it called Selective Extractive: : It seems to me that Selective would be enough (selection usually means extraction).

==> SEALDH refers to the entire instrument family. "Selective" stands here for the fact that SEALDH-II is gas species selective, i.e. we are refereeing to a very high chemical selectivity as a consequence of the high spectral resolution of the employed diode lasers. During the evaluation processing, we can distinguish between water vapor and its other phases as well as other species such as CO₂, CH₄, aircraft fuel vapor, etc. which might be in the sampled air. (Btw, this is also a major difference between a "single information based" evaluation (ice layer on frost point mirror) and a full model based spectroscopic evaluation. (See for details about SEALDH-II's evaluation [7])

"Extractive" stands for "taking" = extracting a gas sample in order to analyze it "inside" the instrument. In contrast to "open path"- TDLAS, where the gas sample remains where it is (i.e. in the atmosphere) and the light, used for sensing, is "brought" to the gas sample.

- Page 5, line 140: Can you provide any reference for White-type cell?

==> Sure, we published the full "technical" description in [7]. The general reference is [8]; our specialized version is: [9]. We added both references in the paper as well.

- Page 5, line 146 and 147: Is the uncertainty expanded (k=2)? Please add a comment.

Instead of linear uncertainty it would be better a linear part of the uncertainty or similar (the same goes for the rest of the text).

==> The answer is yes; details in [7]. We have had the experience in other papers that readers outside of Metrology got confused since they are not used to different uncertainty models and confidence intervals. In order to avoid this confusion (and discussions with non-metrology reviewers), we deliberately adopted our performance parameters to coincide with the common positions in the atmospheric sciences community. A reviewer from a different paper commented on a similar topic that if one relies on HITRAN, he should not use metrological confidence intervals since the entire HITRAN database does not have common metrological standards for uncertainty calculations (HITRAN only deals with coarse uncertainty "classes" or "corridors"). Therefore, he told us that a "(k=2) note" would claim a knowledge which we did not have.

=> Thank you; we revised the wording

- Page 5: line 151: Authors are advised to replace units, such as ccm and SLM with the SI units through the entire paper.

==> We changed ccm to ml. (For all mass flow calculations, it is important to distinguish between "liter per minute" and "standard liter per minute" (normalized to standard pressure))

- Page 5, line 161: Section 2.1 is actually 2.2. The same goes for the sections 3 and 4.

Also avoid calibration-free wording.

==> see above

- Page 6, line 180: variables are not constants; consider rewording: : where kB is Boltzmann constant,: : S(T) is already explained in the previous page

- the second half of the page 6: please see the general comments above

==> Good point; we revised.

- line 229: please consider replacing the word recirculation. Are you talking about back-flow due to partial pressure gradients?

==> Recirculation is defined by local flow field gradients which point more than 90° away from the main flow direction; usually caused by vortex structures. Even if the main flow is forward facing, parts of the flow can go backwards. In general: The higher the flow, the lower the proportional backward facing flow.

- Line 234: THG seems to include both the generator and the reference instrument; I think it's better to keep them separate (here and in the rest of the text) in order not to

confuse the two purposes. Or simply use setup, where appropriate.

==> Yes that is true: In line 257, we introduced the abbreviation: "We termed this entire setup 'traceable humidity generator', THG, and will name it as such throughout the text." Issues which are linked to the entire setup are addressed by "THG", others by the individual components. We fully agree, that the overuse of abbreviations is tempting, however in this case, it should help to distinguish between the different parts rather than writing: "... the setup, consisting of frost point mirror, pressure control unit, water vapor source, heat control, co-flow unit, does the following..."

- Line 274 to 277: Have you considered the effect of the water vapour equations used (pure saturation pressure and the enhancement factor) at two different pressures to the deviation in response?

==> Short answer is "yes". The uncertainty of the national primary standard is strongly affected by the uncertainty of the Sonntag's equation as well as the factors named above. The uncertainty of the dew point mirror hygrometer (see figure 2) includes all of these considerations as well. We briefly documented the working principle of the primary standard elsewhere [10]; this publication also contains several references about the principle and the different issues when operating it.

- Line 297 and elsewhere: precision would better be replaced by resolution

==> The spectroscopic community defines precision differently than other communities. We used the term accordingly to [11], [12] which seems to be the community standard.

- Line 331: linear -> linear part

==> Revised

- Line 372: Consider replacing the ::one single performance statement:: with the assessment of weather the uncertainty is within the expected/estimated value.

==> Indeed, this sentence could be confusing. We revised it.

- Line 406: water scale would better be replaced by dew-point scale or similar

==> Revised

- Conclusion: Please add a discussion of the long-term drift evaluation.

==> We would like to do that but the measurements in this paper do not allow a reliable statement of the long-term drift. The data suggest that the long-term drift seems to be very small but this dataset does not allow quantifying it on a high accuracy level.

- Figure 4: The variable u (mi) is not explained

==> Thank you, we added that.

- [1] R. J. Muecke, B. Scheumann, F. Slemr, and P. W. Werle, "Calibration procedures for tunable diode laser spectrometers," *Proc. SPIE 2112, Tunable Diode Laser Spectroscopy, Lidar, and DIAL Techniques for Environmental and Industrial Measurement*, vol. 2112, pp. 87–98, (1994), doi:10.1117/12.177289.
- [2] M. Helten, H. G. J. Smit, W. Sträter, D. Kley, P. Nedelec, M. Zöger, and R. Busen, "Calibration and performance of automatic compact instrumentation for the measurement of relative humidity from passenger aircraft," *Journal of Geophysical Research: Atmospheres*, vol. 103, no. D19, pp. 25643–25652, (1998), doi:10.1029/98JD00536.
- [3] J. Podolske, G. Sachse, and G. Diskin, "Calibration and data retrieval algorithms for the NASA Langley/Ames Diode Laser Hygrometer for the NASA Transport and Chemical Evolution Over the Pacific (TRACE-P) mission," *Journal of Geophysical Research*, vol. 108, no. D20, p. 8792, (2003), doi:10.1029/2002JD003156.
- [4] D. Tátrai, Z. Bozóki, H. Smit, C. Rolf, N. Spelten, M. Krämer, a. Filges, C. Gerbig, G. Gulyás, and G. Szabó, "Dual-channel photoacoustic hygrometer for airborne measurements: background, calibration, laboratory and in-flight intercomparison tests," *Atmospheric Measurement Techniques*, vol. 8, no. 1, pp. 33–42, (2015), doi:10.5194/amt-8-33-2015.
- [5] M. Zöger, A. Afchine, N. Eicke, M.-T. Gerhards, E. Klein, D. S. McKenna, U. Mörschel, U. Schmidt,

- V. Tan, F. Tuitjer, T. Woyke, and C. Schiller, "Fast in situ stratospheric hygrometers: A new family of balloon-borne and airborne Lyman photofragment fluorescence hygrometers," *Journal of Geophysical Research*, vol. 104, no. D1, pp. 1807–1816, (1999), doi:10.1029/1998JD100025.
- [6] "What Are the Differences between Self-Calibration and External Calibration?" [Online]. Available: <http://digital.ni.com/public.nsf/allkb/305DB655C3C5C7628625710A005198F7>.
- [7] B. Buchholz, S. Kallweit, and V. Ebert, "SEALDH-II—An Autonomous, Holistically Controlled, First Principles TDLAS Hygrometer for Field and Airborne Applications: Design–Setup–Accuracy/Stability Stress Test," *Sensors*, vol. 17, no. 1, p. 68, (2016), doi:10.3390/s17010068.
- [8] J. White, "Very long optical paths in air," *Journal of the Optical Society of America (1917-1983)*, vol. 66, no. 5, pp. 411–416, (1976), doi:10.1364/JOSA.66.000411.
- [9] B. Kühnreich, M. Höh, S. Wagner, and V. Ebert, "Direct single-mode fibre-coupled miniature White cell for laser absorption spectroscopy," *Review of Scientific Instruments*, vol. 87, no. 2, pp. 0–8, (2016), doi:10.1063/1.4941748.
- [10] B. Buchholz, N. Böse, and V. Ebert, "Absolute validation of a diode laser hygrometer via intercomparison with the German national primary water vapor standard," *Applied Physics B*, vol. 116, no. 4, pp. 883–899, (2014), doi:10.1007/s00340-014-5775-4.
- [11] D. W. Allan, "Statistics of atomic frequency standards," *Proceedings of the IEEE*, vol. 54, no. 2, pp. 221–230, (1966), doi:10.1109/PROC.1966.4634.
- [12] P. Werle, R. Mücke, and F. Slemr, "The limits of signal averaging in atmospheric trace-gas monitoring by tunable diode-laser absorption spectroscopy (TDLAS)," *Applied Physics B*, vol. 57, no. 2, pp. 131–139, (1993), doi:10.1007/BF00425997.

**SEALDH-II – a calibration-free transfer standard for airborne water vapor measurements:
Pressure dependent absolute validation from 5 – 1200 ppmv
at a metrological humidity generator**

Bernhard Buchholz^{1,2,4}, Volker Ebert^{1,2,3}

¹ Physikalisch-Technische Bundesanstalt Braunschweig, Germany

² Physikalisch Chemisches Institut, Universität Heidelberg, Germany

³ Center of Smart Interfaces, Technische Universität Darmstadt, Germany

⁴ currently at Department of Civil and Environmental Engineering, Princeton University, USA.

Corresponding author: volker.ebert@ptb.de

Abstract

Highly accurate water vapor measurements are indispensable for understanding a variety of scientific questions as well as industrial processes. While in metrology water vapor concentrations can be defined, generated and measured with relative uncertainties in the single percentage range, field deployable airborne instruments deviate even under quasi-static laboratory conditions up to 10-20%. The novel SEALDH-II hygrometer, a calibration-free, tuneable diode laser spectrometer, bridges this gap by implementing an ~~entirely~~-new holistic concept to achieve higher accuracy levels in the field. Here we present the absolute validation of SEALDH-II at a traceable humidity generator during 23 days of permanent operation at 15 different H₂O concentration levels between 5 and 1200 ppmv. At each concentration level, we studied the pressure dependence at 6 different gas pressures between 65 and 950 hPa. Further, we describe the setup for this metrological validation, the challenges to overcome when assessing water vapor measurements on a high accuracy level, as well as the comparison results. With this validation, SEALDH-II is the first airborne, metrologically validated humidity transfer standard which links several scientific airborne and laboratory measurement campaigns to the international metrological water vapor scale.

1. Introduction

Water vapor affects, like no other substance, nearly all atmospheric processes (Ludlam, 1980; Möller et al., 2011; Ravishankara, 2012). Water vapor represents not only a large direct feedback to global warming when forming clouds, but also plays a major role in atmospheric chemistry (Held and Soden, 2000; Houghton, 2009; Kiehl and Trenberth, 1997). Changes in the water distribution, as vapor or in condensed phases (e.g. in clouds), have a large impact on the radiation balance of the atmosphere. This justifies that water vapor is often mentioned as the most important greenhouse gas and one of the most important parameters in climate research (Ludlam, 1980; Maycock et al., 2011). Water vapor measurements ~~is~~are often needed for other in-situ atmospheric analyzers to correct for their water vapor cross-interference. The high (spatial and

temporal) variability of atmospheric water vapor, its large dynamic range (typically 3 – 40 000 ppmv), and its broad spectroscopic fingerprint typically require complex multi-dimensional calibrations, in particular for spectroscopic sensors. These calibrations often embrace the water vapor content of the gas flow to be analyzed as one of the key calibration parameters even if the instrument (e.g. for CO₂), is not intended to measure water vapor at all.

In particular for field weather stations, water vapor analyzers often are seen as part of the standard instrumentation in atmospheric research. This seems reasonable due to several reasons: slow H₂O concentration change over hours, the typical mid-range humidity levels (approx. above 5000 ppmv), no significant gas pressure or temperature change, target accuracy often only in the on the order of 5-15% relative deviation, and the absence of “non-typical atmospheric components” such as soot or hydrophobic substances. Water vapor measurements under these conditions can be performed by a variety of different devices (Wiederhold, 1997): Capacitive polymer sensors e.g. such as (Salasmaa and Kostamo, 1986) are frequently deployed in low cost (field) applications. Small-scale produced, commercially availableStandardized spectral absorption devices such as e.g. (Petersen et al., 2010) are often used in research campaigns. Dew-point mirror hygrometers (DPM) are known for their high accuracy. However, this is only true if they are regularly calibrated at ~~a~~-high accuracy (transfer-) standards in specialized hygrometry laboratories such as in metrology institutes (Heinonen et al., 2012).

As soon as hygrometers have to be deployed in harsh environments (e.g. on airborne platforms), this situation changes entirely: The ambient gas pressure (~~70-10~~ – 1000 hPa) and gas temperature (~~-80-90~~ – 40°C) ranges are large and both values change rapidly, the required H₂O measurement range is set by the ambient atmosphere (typically 3 – 40000 ppmv), mechanical stress and vibrations occur, and the sampled air contains additional substances from condensed water (ice, droplets), particles, or even aircraft fuel vapor (e.g. on ground). These and other impacts complicate reliable, accurate, long-term stable H₂O measurements and briefly outline why water vapor measurements remain ~~a quite~~ difficult in-situ measurements in the field, even if they are nearly always needed in atmospheric science. Usually, the availability and coverage of observations limit model validation studies in the first place but also Up to now, the lack of sufficient accuracy may have limited important scientific interpretations (Krämer et al., 2009; Peter et al., 2006; Scherer et al., 2008; Sherwood et al., 2014).

Over the last decades, numerous hygrometers were developed and deployed on aircraft (Buck, 1985; Busen and Buck, 1995; Cerni, 1994; Desjardins et al., 1989; Diskin et al., 2002; Durrý et al., 2008; Ebert et al., 2000; Gurlit et al., 2005; Hansford et al., 2006; Helten et al., 1998; Hunsmann et al., 2008; Karpechko et al., 2014; Kley and Stone, 1978; May, 1998; Meyer et al., 2015; Ohtaki and Matsui, 1982; Roths and Busen, 1996; Salasmaa and Kostamo, 1986; Schiff et al., 1994; Silver and Hovde, 1994a, 1994b; Thornberry et al., 2014; Webster et al., 2004; Zöger et al., 1999a, 1999b) (non-exhaustive list), but those often show results which are not sufficient for validation ~~or falsification~~ of atmospheric models in terms of the required absolute accuracy, precision, temporal resolution, long-term stability, comparability, etc. These problems can be grouped into two major categories: accuracy linked problems and time response linked problems. The latter is particularly important for investigations in heterogeneous regions in the lower troposphere as well as for

~~investigations in clouds. The latter one is in particular important for investigations in strongly, spatially structured regions in the lower troposphere as well as for investigations in clouds.~~ In these regions, even two on average agreeing instruments with different response times yield local, large, relative deviations on the order of up to 30% (Smit et al., 2014). In contrast to time response studies, accuracy linked problems in flight are difficult to isolate since they are always covered by the spatial variability (which leads to temporal variability for moving aircraft) of atmospheric H₂O distribution. Comparing hygrometer in flight, such as, for example in (Rollins et al., 2014), does not facilitate a clear accuracy assessment.

Therefore in 2007, an international intercomparison exercise named “AquaVIT” (Fahey et al., 2014) was carried out to compare airborne hygrometers under quasi-static, laboratory-like conditions for upper tropospheric and stratospheric humidity levels. AquaVIT (Fahey et al., 2014) encompassed 22 instruments from 17 international research groups. The instruments were categorized in well-validated, often deployed “core” instruments (APiC, FISH, FLASH, HWV, JLH, CFH) and “younger” non-core instruments. AquaVIT revealed in the important 1 to 150 ppmv H₂O range, that -even under quasi-static conditions- the deviation between the core instrument’s readings and their averaged group mean was on the order of ±10 %. This result fits to the typical interpretation problems of flight data where instruments often deviate from each other by up to 10%, which is not covered by the respective uncertainties of the individual instruments. AquaVIT was a unique first step to document and improve the accuracy of airborne measurements in order to make them more comparable. However, no instrument could claim after AquaVIT that its accuracy is higher than any other AquaVIT instrument, since no “gold standard” was part of the campaign, i.e., a metrological transfer standard (JCGM 2008, 2008; Joint Committee for Guides in Metrology (JCGM), 2009) traced back to the SI units. There is no physical argument for the average being better than the measured value of a single instrument. Instead, many arguments speak for systematic deviations of airborne hygrometers: Most hygrometers have to be calibrated. Even for a perfect instrument, the accuracy issue is ~~entirely transferred to~~ represented by the calibration source and its gas handling system, which in this case leads to two major concerns: First, one has to guarantee that the calibration source is accurate and stable under field conditions, i.e., when using it before or after a flight on the ground. This can be challenging especially for the transportation of the source with all its sensitive electronics/mechanics and for the deviating ambient operation temperature from the ambient validation temperature (hangar vs. laboratory). Even more prone to deviations are calibration sources installed inside ~~of an the~~ aircraft due to changing ambient conditions such as cabin temperature, cabin pressure, orientation angle of instrument (important, if liquids are used for heating or cooling). Secondly, the gas stream with a highly defined amount of water vapor has to be conveyed into the instrument. Especially for water vapor, which is a strongly polar molecule, this gas transport can become a critical step. Changing from high to low concentrations or even just changing the gas pressure or pipe temperature can lead to signal creep due to slow adsorption and desorption processes, which can take long to equilibrate. In metrology, this issue is solved by a long validation/calibration time (hours up to weeks, depending on the H₂O concentration level), a generator without any connectors/fittings (everything is welded) and piping made out of electro-polished, stainless steel to ensure that the equilibrium is established before the actual calibration process is started. However,

this calibration approach is difficult to deploy and maintain for aircraft/field operations due to the strong atmospheric variations in gas pressure and H₂O concentrations, which usually leads to a multi-dimensional calibration pattern (H₂O concentration, gas pressure, sometimes also gas temperature) in a short amount of calibration time (hours). Highly sensitive, frequently flown hygrometers like (Zöger et al., 1999a) are by their physical principle, not as long-term stable as it would be necessary to take advantage of a long calibration session. Besides the time issue to reach a H₂O equilibrium between source and instrument, most calibration principles for water vapor are influenced by further issues. A prominent example is the saturation of air in dilution/saturation based water vapor generators: gas temperature and pressure defines the saturation level (described e.g. by Sonntag's Equation (Rollins et al., 2014)), however, it is well-known that e.g. 100.0% saturation is not easily achievable. This might be one of the impact factors for a systematic offset during calibrations in the field. The metrology community solves this for high humidity levels with large, multi-step saturation chambers which decrease the temperature step-wise to force the water vapor to condense in every following step. These few examples of typical field-related problems show, that there is a reasonable doubt that deviations in field situations are norm-distributed. Hence, the mean during AquaVIT might be biased, i.e. not the correct H₂O value.

The instruments by themselves might actually be more accurate than AquaVIT showed, but deficiencies of the different calibration procedures (with their different calibration sources etc.) might mask this. To summarize, AquaVIT documented a span of up to 20% relative deviation between the world's best airborne hygrometers – but AquaVIT could not assess absolute deviations nor explain them, since a link to a metrological H₂O primary standard (i.e., the definition of the international water vapor scale) was missing. Therefore, we present in this paper the first comparison of an airborne hygrometer (SEALDH-II) with a metrological standard for the atmospheric relevant gas pressure (65 – 950 hPa) and H₂O concentration range (5 – 1200 ppmv). We will discuss the validation setup, procedure, and results. Based on this validation, SEALDH-II is by definition the first airborne transfer standard for water vapor [which links laboratory and field campaigns directly to metrological standards](#).

2. SEALDH-II

2.1. System description

This paper focuses on the metrological accuracy validation of the Selective Extractive Airborne Laser Diode Hygrometer (SEALDH-II). SEALDH-II is the airborne successor of the proof-of-concept spectrometer (SEALDH-I) study published in (Buchholz et al., 2014), which showed the possibility and the achievable accuracy level for calibration-free dTDLAS hygrometry. [The publication \(Buchholz et al., 2014\) demonstrates this for the 600 ppmv to 20000 ppmv range at standard ambient pressure\). The instruments SEALDH-I, SEALDH-II and also HAI \(Buchholz et al., 2017\) are all three built with the design philosophy that every single reported value of the instrument should have a "related boundary/operation condition](#)

150 snap shot” allowing to exclude the possibility of any instrumental malfunction during the measurement.
151 SEALDH-II is from this perspective the most “holistic” approach (capturing much more boundary
152 condition data (Buchholz et al., 2016)), while HAI can serve as a multi-channel, multi-phase hygrometer for
153 a broader variety of scientific questions.

154
155 ~~The publication (Buchholz et al., 2014) demonstrates this for the 600 ppmv to 20000 ppmv range at standard~~
156 ~~ambient pressure).~~ SEALDH-II integrates numerous different principles, concepts, modules, and novel
157 parts, which contribute to or enable the results shown in this paper. SEALDH-II’s high internal complexity
158 does not allow a full, detailed discussion of the entire instrument in this paper; for more details the reader is
159 referred to (Buchholz et al., 2016). The following brief description covers the most important technical
160 aspects of the instrument from a user’s point of view:

161 SEALDH-II is a compact (19” rack 4 U (=17.8 cm)) closed-path, absolute, directly Tunable Diode Laser
162 Absorption Spectroscopy (dTDLAS) hygrometer operating at 1.37 μm . With its compact dimensions and the
163 moderate weight (24 kg), it is well suited for space- and weight-limited airborne applications. The internal
164 optical measurement cell is a miniaturized White-type cell with an optical path length of 1.5 m (Kühnreich
165 et al., 2016; White, 1976). It is connected to the airplane’s gas inlet via an internal gas handling system
166 comprising a temperature exchanger, multiple temperature sensors, a flow regulator, and two gas pressure
167 sensors.

168 Approximately 80 different instrument parameters are controlled, measured, or corrected by SEALDH-II at
169 any time to provide a holistic view on the spectrometer status. This extensive set of monitoring data ensures
170 reliable and well-characterized measurement data at any time. The knowledge about the instruments status
171 strongly facilitates metrological uncertainties calculations. SEALDH-II’s calculated linear part of the
172 measurement uncertainty is 4.3%, with an additional offset uncertainty of ± 3 ppmv (further details in
173 (Buchholz et al., 2016)). The precision of SEALDH-II was determined via the Allan-variance approach and
174 yielded 0.19 ppmv ($0.17 \text{ ppmv} \cdot \text{m} \cdot \text{Hz}^{-1/2}$) at 7 Hz repetition rate and an ideal precision of 0.056 ppmv (0.125
175 $\text{ppmv} \cdot \text{m} \cdot \text{Hz}^{-1/2}$) at 0.4 Hz. In general, SEALDH-II’s time response is limited by the gas flow through the
176 White-type multi-pass measurement cell with a volume of 300 ~~cc~~ cm³. With the assumption of a bulk flow
177 of 7 SLM at 200 hPa through the cell, the gas exchange time is 0.5 seconds.

178 SEALDH-II’s measurement range covers 3 – 40000 ppmv. The calculated mixture fraction offset uncertainty
179 of ± 3 ppmv defines the lower detection limit. This offset uncertainty by itself is entirely driven by the
180 capability of detecting and minimizing parasitic water vapor absorption. The concept, working principle,
181 and its limits are described in (Buchholz and Ebert, 2014). The upper limit of 40000 ppmv is defined by the
182 lowest internal instrument temperature, which has to always be higher than the dew point temperature to
183 avoid any internal condensation. From a spectroscopic perspective, SEALDH-II could handle
184 concentrations up to approx. 100000 ppmv before spectroscopic problems like saturation limit the accuracy
185 and increase the relative uncertainty beyond 4.3%.

2.1. Calibration-free evaluation approach

SEALDH-II's data treatment works differently from nearly all other published TDLAS spectrometers. Typically, instruments are setup in a way that they measure the absorbance or a derivative measurand of absorbance, and link it to the H₂O concentration. This correlation together with a few assumptions about long-term stability, cross interference, gas temperature dependence, gas pressure dependence is enough to calibrate a system (Muecke et al., 1994). Contrarily, a calibration-free approach requires a fully featured physical model describing the absorption process entirely. The following description is a brief overview; for more details see e.g. (Buchholz et al., 2014, 2016; Ebert and Wolfrum, 1994; Schulz et al., 2007).

In a very simplified way, our physical absorption model uses the *extended* Lambert-Beer equation (Equation 1) which describes the relationship between the initial light intensity $I_0(\lambda)$ before the absorption path (typically being in the few mW-range) and the transmitted light intensity $I(\lambda)$.

$$\text{Equation 1: } I(\lambda) = E(t) + I_0(\lambda) \cdot Tr(t) \cdot \exp[-S(T) \cdot g(\lambda - \lambda_0) \cdot N \cdot L]$$

The parameter $S(T)$ describes the line strength of the selected molecular transition. In SEALDH-II's case, the spectroscopic multi-line fit takes into account 19 transition lines in the vicinity of the target line at 1370 nm (energy levels: 110 – 211, rotation-vibrational combination band). The other parameters are the line shape function $g(\lambda - \lambda_0)$, the absorber number density N , the optical path length L and corrections for light-type background radiation $E(t)$ and broadband transmission losses $Tr(t)$.

Equation 1 can be enhanced with the ideal gas law to calculate the H₂O volume mixing ratio c :

$$\text{Equation 2: } c = -\frac{k_B \cdot T}{S(T) \cdot L \cdot p} \int \ln \left(\frac{I(v) - E(t)}{I_0(v) \cdot Tr(t)} \right) \frac{dv}{dt} dt$$

The additional **variables-parameters** in Equation 2 are: constant entities like the Boltzmann constant k_B ; the optical path length L ; molecular constants like the line strength $S(T)$ of the selected molecular transition; the dynamic laser tuning coefficient $\frac{dv}{dt}$, which is a constant laser property; continuously measured entities such as gas pressure (p), gas temperature (T) and photo detector signal of the transmitted light intensity $I(v)$ as well as the initial light intensity $I_0(v)$, which is retrieved during the evaluation process from the transmitted light intensity $I(v)$.

Equation 2 facilitates an evaluation of the measured spectra without any instrument calibration at any kind of water vapor reference (Buchholz et al., 2014; Ebert and Wolfrum, 1994; Schulz et al., 2007) purely based on first principles. Our concept of a fully calibration-free data evaluation approach (this excludes also any referencing of the instrument to a water standard in order to correct for instrument drift, offsets, temperature dependence, pressure dependence, etc.) is crucial for the assessment of the results described in this publication. It should be noted that the term "calibration-free" is frequently used in different communities with dissimilar meanings. We understand this term according to the following quote (JCGM 2008, 2008): "calibration (...) in a first step, establishes a relation between the measured values of a quantity with measurement uncertainties provided by a measurement standard (...), in a second step, [calibration] uses this information to establish a relation for obtaining a measurement result from an indication (of the

device to be calibrated)". Calibration-free in this sense means, that SEALDH-II does not use any information from "calibration-, comparison-, test-, adjustment-" runs with respect to a higher accuracy "water vapor standard" to correct or improve any response function of the instrument. SEALDH-II uses as described in (Buchholz et al., 2016) only spectroscopic parameters and the 80 supplementary parameters as measurement input to calculate the final H₂O concentration. The fundamental difference between a calibration approach and this stringent concept is that only effects which are part of our physical model are taken into account for the final H₂O concentration calculation. All other effects like gas pressure or temperature dependencies, which cannot be corrected with a well-defined physical explanation, remain in our final results even if this has the consequence of slightly uncorrected results deviations. This strict philosophy leads to measurements which are very reliable with respect to accuracy, precision and the instrument's over-all performance. The down-side is a relatively computer-intensive, sophisticated evaluation. As SEALDH-II stores all the raw spectra, one could – if needed for whatever reason – also calibrate the instrument by referencing it to a high accuracy water vapor standard and transfer the better accuracy e.g. of a metrological standard onto the instrument. Every calibration-free instrument can be calibrated since pre-requirements for a calibration are just a subset of the requirements for a calibration-free instrument. However, a calibration can only improve the accuracy for the relatively short time between two calibration-cycles by adding all uncertainty contributions linked to the calibration itself to the system. This is unpleasant or even intolerable for certain applications and backs our decision to develop a calibration-free instrument to enable a first principles, long-term stable, maintenance-free and autonomous hygrometer for field use e.g. at remote sites or aircraft deployments.

241 **3. SEALDH-II validation facility**

242 **3.1. Setup**

243 ~~Figure 1~~~~Figure 1~~~~Figure 1~~~~Figure 1~~ right shows the validation setup. As a well-defined and highly stable H₂O vapor source, we use a commercial Thunder scientific model (TSM) 3900, similar to (Thunder-Scientific, 2016). This source saturates pre-dried air at an elevated gas pressure in an internally ice covered chamber. The gas pressure in the chamber and the chamber's wall temperature are precisely controlled and highly stable and thus define the absolute water vapor concentration via the Sonntag equation (Sonntag, 1990). After passing through the saturator, the gas expands to a pressure suitable for the subsequent hygrometer. The pressure difference between the saturation chamber pressure and the subsequent step give this principle its name "two pressure generator". The stable H₂O concentration range of the TSM is 1 – 1300 ppmv for these specific deployment conditions. This generator provides a stable flow of approximately 4 – 5 SLM. Roughly 0.5 SLM are distributed to a frost/dew point hygrometer, D/FPH, (MBW 373) (MBW Calibration Ltd., 2010). SEALDH-II is fed with approx. 3.5 SLM, while 0.5 SLM are fed to an

outlet. This setup ensures that the dew point mirror hygrometer (DPH)¹ operates close to the ambient pressure, where its metrological primary calibration is valid, and that the gas flow is sufficiently high in any part of the system to avoid recirculation of air. The vacuum pump is used to vary the gas pressure in SEALDH-II's cell with a minimized feedback on the flow through the D/FPH and the TSM. This significantly reduces the time for achieving a stable equilibrium after any gas pressure change in SEALDH-II's chamber. SEALDH-II's internal electronic flow regulator limits the mass flow at higher gas pressures and gradually opens towards lower pressures (vacuum pumps usually convey a constant volume flow i.e., the mass flow is pressure dependent). We termed this entire setup "traceable humidity generator", THG, and will name it as such throughout the text.

3.1. Accuracy of THG

The humidity of the gas flow is set by the TSM generator but the absolute H₂O values are traceably determined with the dew point mirror hygrometer (D/FPH). The D/FPH, with its primary calibration, thus guarantees the absolute accuracy in this setup. The D/FPH is not affected by the pressure changes in SEALDH-II's measurement cell and operates at standard ambient gas pressure and gas temperature where its calibration is most accurate. The D/FPH was calibrated (Figure 2Figure 2Figure 2Figure 2) at the German national standard for mid-range humidity (green, 600 – 8000 ppmv) as well as at the German national standard for low-range humidity (blue, for lower values 0.1 – 500 ppmv). The two national standards work on different principles: The two pressure principle (Buchholz et al., 2014) currently supplies the lower uncertainties (green, "±"-values in Figure 2Figure 2Figure 2Figure 2). Uncertainties are somewhat higher for the coulometric generator (Mackrodt, 2012) in the lower humidity range (blue). The "Δ"-values in Figure 2Figure 2Figure 2Figure 2 show the deviations between the readings of the D/FPH and the "true" values of the national primary standards.

4. SEALDH-II validation procedure

4.1. Mid-term multi-week permanent operation of SEALDH-II

One part of the validation was a permanent operation of SEALDH-II over a time scale much longer than the usual air or ground based scientific campaigns. In this paper, we present data from a permanent 23 day long (550 operation hours) operation in automatic mode. Despite a very rigorous and extensive monitoring of SEALDH-II's internal status, no malfunctions of SEALDH-II could be detected. One reason for this are the extensive internal control and error handling mechanisms introduced in SEALDH-II, which are mentioned above and described elsewhere (Buchholz et al., 2016). Figure 3Figure 3Figure 2Figure 3 shows an overview of the entire validation. The multi-week validation exercise comprises 15 different H₂O

¹ The used dew point mirror hygrometer can measure far below 0°C; therefore, it is a dew point mirror above > 0°C and a frost point mirror as soon as there is ice on the mirror surface. We will use both DPH and D/FPH abbreviations interchangeably.

concentration levels between 2 and 1200 ppmv. At each concentration level, the gas pressure was varied in six steps (from 65 to 950 hPa) over a range which is particularly interesting for instruments on airborne platforms operating from troposphere to lower stratosphere where SEALDH-II's uncertainty ($4.3\% \pm 3$ ppmv) is suitable. ~~Figure 3~~~~Figure 3~~~~Figure 3~~ (top) shows the comparison between SEALDH-II (black line) and the THG setup (red). ~~Figure 3~~~~Figure 3~~~~Figure 3~~ (bottom) shows the gas pressure (blue) and the gas temperature (green) in SEALDH-II measurement cell. The gas temperature increase in the second week was caused by a failure of the laboratory air conditioner that led to a higher room temperature and thus higher instrument temperature. ~~Figure 4~~~~Figure 4~~~~Figure 4~~ shows the 200 hPa section of the validation in Figure 3. To avoid any dynamic effects from time lags, hysteresis of the gas setup, or the instruments themselves, every measurement at a given concentration/pressure combination lasted at least 60 min. The data from the THG (red) show that there is nearly no feedback of a gas pressure change in SEALDH-II's measurement cell towards the D/FPH, respectively the entire THG. The bottom subplot in ~~Figure 4~~~~Figure 4~~~~Figure 4~~ shows the relative deviation between the THG and SEALDH-II. This deviation is correlated to the absolute gas pressure level and can be explained by deficiencies of the Voigt lines shape used to fit SEALDH-II's spectra (Buchholz et al., 2014, 2016)~~(Buchholz et al., 2014)~~~~(Buchholz et al., 2016)~~. The Voigt profile, a convolution of Gaussian (for temperature broadening) and Lorentzian (pressure broadening) profiles used for SEALDH-II's evaluation, does not include effects such as Dicke Narrowing, which become significant at lower gas pressures. Neglecting these effects cause systematic, but long-term stable and fully predictable deviations from the reference value in the range from sub percent at atmospheric gas pressures to less than 5 % at the lowest gas pressures described here. We have chosen not to implement any higher order line shape (HOLS) models as the spectral reference data needed are not available at sufficient accuracy. Further, HOLs would force us to increase the number of free fitting parameters, which would destabilize our fitting procedure, and lead to reduced accuracy/reliability (i.e., higher uncertainty) as well as significantly increased computational efforts. This is especially important for flight operation where temporal H₂O fluctuations (spatial fluctuations result in temporal fluctuations for a moving device) occur with gradients up to 1000 ppmv/s.

These well understood, systematic pressure dependent deviations will be visible in each further result plot of this paper. The impact and methods of compensation are already discussed in (Buchholz et al., 2014). The interested reader is referred to this publication for a more detailed analysis and description.

SEALDH-II's primary target areas of operations are harsh field environments. Stability and predictability is to be balanced with potential, extra levels of accuracy which might not be required or reliably achievable for the intended application. Higher order line shape models are therefore deliberately traded for a stable, reliable, and unified fitting process under all atmospheric conditions. This approach leads to systematic, predictable deviations in the typical airborne accessible atmospheric gas pressure range (125 – 900 hPa) of less than 3%. One has to compare these results for assessment to the non-systematic deviations of 20% revealed during the mentioned AquaVIT comparison campaign (Fahey et al., 2014). Hence, for field/airborne purposes, the 3% seems to be fully acceptable – especially in highly H₂O structured environments.

~~This comparison with AquaVIT should just provide a frame to embed the 3%. The H₂O concentration range of Aquavit (-0 – 150 ppmv) versus this validation- range (5 -1200 ppmv), and -the instruments configuration at AquaVIT (mainly (upper) stratospheric hygrometers) versus SEALDH-II as a wide range instrument (3 – 40000 ppmv) do not allow a direct comparison. but sadly~~ Sadly, there is no other reliable (representative for the community, externally reviewed, blind submission, etc.) comparison exercise such as AquaVIT for higher concentration ranges.

~~Hence, for field/airborne purposes, the 3% seems to be fully acceptable especially in highly H₂O structured environments.~~

4.1. Assessment of SEALDH-II's mid-term accuracy: Dynamic effects

Besides the pressure dependence discussed above, SEALDH-II's accuracy assessment is exacerbated by the differences in the temporal behavior between the THG's dew/frost point mirror hygrometer (D/FPH) and SEALDH-II: ~~Figure 5~~Figure 5Figure 5Figure 5 (left) shows an enlarged 45 min. long section of measured comparison data. SEALDH-II (black) shows a fairly large water vapor variation compared to the THG (red). The precision of SEALDH-II (see chapter 2) is 0.056 ppmv at 0.4 Hz (which was validated at a H₂O concentration of 600 ppmv (Buchholz et al., 2016)) yielding a signal to noise ratio of 10700. Therefore, SEALDH-II can very precisely detect variations in the H₂O concentration. Contrarily, the working principle of a D/FPH requires an equilibrated ice/dew layer on the mirror. ~~As an~~Caused by the indirect, inertia,inertial thermal adjustment process, the response time of a dew/frost point mirror hygrometer has certain limitations due to this principle (the dew/frost point temperature measurement is eventually used to calculate the final H₂O concentration), whereas the optical measurement principle of SEALDH-II is only limited by the gas transport, i.e., the flow (exchange rate) through the measurement cell. The effect of those different response times is clearly visible from 06:00 to 06:08 o'clock in ~~Figure 5~~Figure 5Figure 5Figure 5. The gas pressure of SEALDH-II's measurement cell (blue), which is correlated to the gas pressure in the THG's ice chamber, shows an increase of 7 hPa – caused by the regulation cycle of the THG's generator (internal saturation chamber gas pressure change). The response in the THG frost point measurement (green, red) shows a significant time delay compared to SEALDH-II, which detects changes approx. 20 seconds faster. This signal delay is also clearly visible between 06:32 to 06:40 o'clock, where the water vapor variations detected by SEALDH-II are also visible in the smoothed signals of the THG. ~~Figure 5~~Figure 5Figure 5Figure 5 right shows such a variation in detail (5 min). The delay between the THG and SEALDH-II is here also approximately 20 seconds. If we assume that SEALDH-II measures (due to its high precision) the true water vapor fluctuations, the relative deviation can be interpreted as overshooting and undershooting of the D/FPH's controlling cycle, which is a commonly known response behavior of slow regulation feedback loops to fast input signal changes. The different time responses lead to "artificial" noise in the concentration differences between SEALDH-II and THG. Theoretically, one could characterize this behavior and then try to correct/shift the data to minimize this artificial noise. However, a D/FPH is

360 fundamentally insufficient for a dynamic characterization of a fast response hygrometer such as SEALDH-
361 II. Thus, the better strategy is to keep the entire system as stable as possible and calculate mean values by
362 using the inherent assumption that under- and overshoots of the DPM affect the mean statically and
363 equally. With this assumption, the artificial noise can be seen in the first order as Gaussian distributed noise
364 within each pressure step (~~Figure 4Figure 4Figure 4Figure 4~~) of at least 60 min. The error induced by this
365 should be far smaller than the above discussed uncertainties of the THG (and SEALDH-II).
366

367 5. Results

368 The results of this validation exercise are categorized in three sections according to the following conditions
369 in atmospheric regions: mid-tropospheric range: 1200 – 600 ppmv (~~Figure 6Figure 6Figure 6Figure 6~~), upper
370 tropospheric range: 600 – 20 ppmv (~~Figure 7Figure 7Figure 7Figure 7~~), and lower stratospheric range: 20 – 5
371 ppmv (~~Figure 8Figure 8Figure 8Figure 8~~). This categorization is also justified by the relative influence of
372 SEALDH-II's calculated offset uncertainty of ± 3 ppmv (Buchholz and Ebert, 2014): At 1200 ppmv, its
373 relative contribution of 0.25% is negligible compared to the 4.3% linear part of the uncertainty of SEALDH-
374 II. At 5 ppmv, the relative contribution of the offset uncertainty is 60% and thus dominates the linear part of
375 the uncertainty. Before assessing the following data, it should be emphasized again that SEALDH-II's
376 spectroscopic first-principles evaluation was designed to rely on accurate spectral data instead of a
377 calibration. SEALDH-II was never calibrated or referenced to any kind of reference humidity generator or
378 sensor.

379 5.1. The 1200 – 600 ppmv range

380 ~~Figure 6Figure 6Figure 6Figure 6~~ shows the summary of the pressure dependent validations in the 1200 –
381 600 ppmv range. Each of the 48 data points represents the mean over one pressure measurement section of
382 at least 60 min (see ~~Figure 4Figure 4Figure 4Figure 4~~). A cubic polynomial curve fitted to the 600 ppmv
383 results (blue) serves as an internal quasi-reference to connect with the following graphs. The 600 ppmv data
384 (grey) are generated via a supplementary comparison at a different generator: The German national
385 primary mid-humidity ~~primary~~ generator (PHG). This primary generator data at 600 ppmv indicate a
386 deviation between PHG and THG of about 0.35 %, which is compatible with the uncertainties of the THG
387 (see chapter 3.1) and the PHG (0.4%) (Buchholz et al., 2014). The PHG comparison data also allow a
388 consistency check between the absolute values of (see ~~Figure 2Figure 2Figure 2Figure 2~~) the PHG (~~primary~~
389 ~~standard~~—calibration-free), the THG (DPM calibrated) and SEALDH-II (calibration-free).

390 5.2. The 600 – 20 ppmv range

391 In this range, the linear part of the uncertainty (4.3%) and the offset uncertainty (± 3 ppmv) have both a
392 significant contribution. ~~Figure 7Figure 7Figure 7Figure 7~~ shows a clear trend: The lower the concentration,

the higher the deviation. We believe this is being caused by SEALDH-II's offset variation and will be discussed in the 20 – 5 ppmv range.

5.3. The 20 – 5 ppmv range

The results in this range (~~Figure 8~~~~Figure 8~~~~Figure 8~~~~Figure 8~~) are dominated by the offset uncertainty. It is important to mention at this point, that the ± 3 ppmv uncertainties are calculated based on assumptions, design innovations, and several independent, synchronous measurements which are automatically done while the instrument is in operation mode (see publication (Buchholz et al., 2016; Buchholz and Ebert, 2014)). Hence, the calculated uncertainties resemble an upper uncertainty threshold; the real deviation could be lower than 3 ppmv. A clear assessment is fairly difficult since at low concentrations (i.e., low optical densities) several other effects occur together such as, e.g., optical interference effects like fringes caused by the very long coherence length of the used laser. However, ~~Figure 9~~~~Figure 9~~~~Figure 9~~~~Figure 9~~ (left) allows a rough assessment of the offset instability. This plot shows all the data below 200 ppmv, grouped by the gas pressure in the measurement cell. If one ignores the 65 hPa and 125 hPa measurements, which are clearly affected by higher order line shape effects (see above), the other measurements fit fairly well in a ± 1 ppmv envelope function (grey). In other words, SEALDH-II's combined offset "fluctuations" are below 1 ppmv H₂O. All validation measurements done with SEALDH-II during the last years consistently demonstrated a small offset variability so that the observed offset error is around 0.6 ppmv — i.e., only 20% of the calculated ± 3 ppmv.

5.4. General evaluation

~~Figure 9~~~~Figure 9~~~~Figure 9~~~~Figure 9~~ presents a summary of all 90 analyzed concentration/pressure-pairs during the 23 days of validation. The calculated uncertainties (linear 4.3% and offset ± 3 ppmv) of SEALDH-II are plotted in purple. This uncertainty calculation doesn't include line shape deficiencies and is therefore only valid for a pressure range where the Voigt profile can be used to represent all major broadening effects of absorption lines (Dicke, 1953; Maddaloni et al., 2010). This is the case above 250 hPa. The results at 950, 750, 500, 250 hPa show that the maximum deviations, derived from these measurements, can be described ~~with one single performance statement~~~~by~~: linear +2.5%, offset -0.6 ppmv.

To prevent further interpretations, it should be noted that this result doesn't change the statement about SEALDH-II's uncertainties, since these are calculated and not based on any validation/calibration process. This is a significantly different approach: The holistic control/overview is one of the most important and essential differences between calibration-free instruments such as SEALDH-II and other classical spectroscopic instruments which rely on sensor calibration. SEALDH-II can guarantee correctness of measurement values within its uncertainties because any effect which causes deviations has to be included in the evaluation model – otherwise it is not possible to correct for it.

As mentioned before, any calibration-free instrument can be calibrated too (see e.g. (Buchholz et al., 2013)). However by doing so, one must accept to a certain extent loss of control over the system, especially in

environments which are different from the calibration environment. For example, if a calibration was used to remove an instrumental offset, one has to ensure that this offset is long-term stable, which is usually quite difficult, as - shown by the example of parasitic water offsets in fiber coupled diode laser hygrometers (Buchholz and Ebert, 2014). Another option is to choose the recalibration frequency high enough; i.e., minimizing the drift amplitude by minimizing the time between two calibrations. This, however, reduces the usable measurement time and leads to considerable investment of time and money into the calibration process. For the case of SEALDH-II, a calibration of the pressure dependence – of course tempting and easy to do – would directly “improve” SEALDH-II’s laboratory overall performance level from $\pm 4.3\% \pm 3$ ppmv to $\pm 0.35\% \pm 0.3$ ppmv. At first glance, this “accuracy” would then be an improvement by a factor of 55 compared to the mentioned results of AquaVIT (Fahey et al., 2014). However, it is extremely difficult – if not impossible – to guarantee this performance and the validity of the calibration under harsh field conditions; instead SEALDH-II would “suffer” from the same typical calibration associated problems in stability and in predictability. Eventually, the calibration-free evaluation would define the trusted values and the “improvement”, achieved by the calibration, would have to be used very carefully and might disappear eventually.

6. Conclusion and Outlook

The SEALDH-II instrument; a novel, compact, airborne, calibration-free hygrometer which implements a holistic, first-principles directly tuneable diode laser absorption spectroscopy (dTDLAS) approach was stringently validated at a traceable water vapor generator at the German national metrology institute (PTB). The pressure dependent validation covered a H₂O range from 5 to 1200 ppmv and a pressure range from 65 hPa to 950 hPa. In total, 90 different H₂O concentration/pressure levels were studied within 23 days of permanent validation experiments. Compared to other comparisons of airborne hygrometers - such as those studied in the non-metrological AquaVIT campaign (Fahey et al., 2014), where a selection of the best “core” instruments still showed an accuracy scatter of at least $\pm 10\%$ without an absolute reference value - our validation exercise used a traceable reference value derived from instruments directly linked to the international ~~water-sealed~~[dew-point scale for water vapor](#). This allowed a direct assessment of SEALDH-II’s absolute performance with a relative accuracy level in the sub percent range. Under these conditions, SEALDH-II showed an excellent absolute agreement within its uncertainties which are 4.3% of the measured value plus an offset of ± 3 ppmv (valid at 1013 hPa). SEALDH-II showed at lower gas pressures - as expected - a stable, systematic, pressure dependent offset to the traceable reference, which is caused by the line shape deficiencies of the Voigt line shape: e.g. at 950 hPa, the systematic deviation of the calibration-free evaluated results could be described by (linear +0.9%, offset -0.5 ppmv), while at 250 hPa the systematic deviations could be described by (linear +2.5%, offset -0.6 ppmv). If we suppress this systematic pressure dependence, the purely statistical deviation is described by linear scatter of $\pm 0.35\%$ and an offset uncertainty of ± 0.3 ppmv.

463 Due to its extensive internal monitoring and correction infrastructure, SEALDH-II is very resilient against a
464 broad range of external disturbances and has an output signal temperature coefficient of only 0.026%/K,
465 which has already been validated earlier (Buchholz et al., 2016). Therefore, these results can be directly
466 transferred into harsh field environments. With this metrological, mid and upper atmosphere focused
467 validation presented here, we believe SEALDH-II to be the first directly deployable, metrologically
468 validated, airborne transfer standard for atmospheric water vapor. Having already been deployed in
469 several airborne and laboratory measurement campaigns, SEALDH-II thus directly links for the first time,
470 scientific campaign results to the international metrological water vapor scale.

471 **Data availability**

472 *The underlying data for the results shown in this paper are raw spectra (time vs. photo current), which are compressed*
473 *to be compatible with the instruments data storage. In the compressed state the total amount is approximately 6GB of*
474 *binary data. Uncompressed data size is approx. 60 GB. We are happy to share these data on request.*

475

476 **Author Contributions**

477 *Bernhard Buchholz and Volker Ebert conceived and designed the experiments. Bernhard Buchholz performed the*
478 *experiments; Bernhard Buchholz and Volker Ebert analyzed the data and wrote the paper.*

479

480 **Conflicts of Interest**

481 *The authors declare no conflict of interest*

482

483 **Acknowledgements:**

484 *Parts of this work were embedded in the EMPIR (European Metrology Program for Innovation and Research) projects*
485 *METEOMET- 1 and METEOMET-2. The authors want to thank Norbert Böse and Sonja Pratzler (PTB Germany)*
486 *for operating the German primary national water standard and the traceable humidity generator. Last but not least,*
487 *the authors thank James McSpiritt (Princeton University) for the various discussions about reliable sensor designs*
488 *and Mark Zondlo (Princeton University) for sharing his broad knowledge about atmospheric water vapor*
489 *measurements.*

490

491 7. References

- 492 Buchholz, B., Afchine, A., Klein, A., Schiller, C. and Krämer, M.: HAI , a new airborne, absolute, twin dual-
 493 channel, multi-phase TDLAS-hygrometer: background, design, setup, and first flight data, *Atmos. Meas.*
 494 *Tech.*, 5194, <http://dx.doi.org/10.5194/amt-10-35-2017>, 2017.
- 495 Buchholz, B., Böse, N. and Ebert, V.: Absolute validation of a diode laser hygrometer via intercomparison
 496 with the German national primary water vapor standard, *Applied Physics B*, 116(4), 883–899,
 497 <http://dx.doi.org/10.1007/s00340-014-5775-4>, 2014.
- 498 Buchholz, B. and Ebert, V.: Offsets in fiber-coupled diode laser hygrometers caused by parasitic absorption
 499 effects and their prevention, *Measurement Science and Technology*, 25(7), 75501,
 500 <http://dx.doi.org/10.1088/0957-0233/25/7/075501>, 2014.
- 501 Buchholz, B., Kallweit, S. and Ebert, V.: SEALDH-II—An Autonomous, Holistically Controlled, First
 502 Principles TDLAS Hygrometer for Field and Airborne Applications: Design–Setup–Accuracy/Stability
 503 Stress Test, *Sensors*, 17(1), 68, <http://dx.doi.org/10.3390/s17010068>, 2016.
- 504 Buchholz, B., Kühnreich, B., Smit, H. G. J. and Ebert, V.: Validation of an extractive, airborne, compact TDL
 505 spectrometer for atmospheric humidity sensing by blind intercomparison, *Applied Physics B*, 110(2), 249–
 506 262, <http://dx.doi.org/10.1007/s00340-012-5143-1>, 2013.
- 507 Buck, A.: The Lyman-alpha absorption hygrometer, in *Moisture and Humidity Symposium Washington*,
 508 DC, Research Triangle Park, NC, pp. 411–436., 1985.
- 509 Busen, R. and Buck, A. L.: A High-Performance Hygrometer for Aircraft Use: Description, Installation, and
 510 Flight Data, *Journal of Atmospheric and Oceanic Technology*, 12, 73–84, [http://dx.doi.org/10.1175/1520-0426\(1995\)012<0073:AHFHFA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0426(1995)012<0073:AHFHFA>2.0.CO;2), 1995.
- 512 Cerni, T. A.: An Infrared Hygrometer for Atmospheric Research and Routine Monitoring, *Journal of*
 513 *Atmospheric and Oceanic Technology*, 11, 445–462, [http://dx.doi.org/10.1175/1520-0426\(1994\)011<0445:AIHFAR>2.0.CO;2](http://dx.doi.org/10.1175/1520-0426(1994)011<0445:AIHFAR>2.0.CO;2), 1994.
- 515 Desjardins, R., MacPherson, J., Schuepp, P. and Karanja, F.: An evaluation of aircraft flux measurements of
 516 CO₂, water vapor and sensible heat, *Boundary-Layer Meteorology*, 47(1), 55–69,
 517 <http://dx.doi.org/10.1007/BF00122322>, 1989.
- 518 Dicke, R.: The effect of collisions upon the Doppler width of spectral lines, *Physical Review*, 89(2), 472–473,
 519 <http://dx.doi.org/10.1103/PhysRev.89.472>, 1953.
- 520 Diskin, G. S., Podolske, J. R., Sachse, G. W. and Slate, T. A.: Open-path airborne tunable diode laser
 521 hygrometer, in *Proc. SPIE 4817, Diode Lasers and Applications in Atmospheric Sensing*, vol. 4817, pp. 196–
 522 204., 2002.
- 523 Durry, G., Amarouche, N., Joly, L. and Liu, X.: Laser diode spectroscopy of H₂O at 2.63 μ m for atmospheric
 524 applications, *Applied Physics B*, 90(3–4), 573–580, <http://dx.doi.org/10.1007/s00340-007-2884-3>, 2008.
- 525 Ebert, V., Fernholz, T. and Pitz, H.: In-situ monitoring of water vapour and gas temperature in a coal fired
 526 power-plant using near-infrared diode lasers, in *Laser Applications to Chemical and Environmental*
 527 *Analysis*, pp. 4–6, Optical Society of America. [online] Available from:
 528 <http://www.opticsinfobase.org/abstract.cfm?id=142068> (Accessed 15 February 2012), 2000.
- 529 Ebert, V. and Wolfrum, J.: Absorption spectroscopy, in *OPTICAL MEASUREMENTS-Techniques and*
 530 *Applications*, ed. F. Mayinger, pp. 273–312, Springer., 1994.
- 531 Fahey, D. W., Saathoff, H., Schiller, C., Ebert, V., Peter, T., Amarouche, N., Avallone, L. M., Bauer, R.,
 532 Christensen, L. E., Durry, G., Dyroff, C., Herman, R., Hunsmann, S., Khaykin, S., Mackrodt, P., Smith, J. B.,
 533 Spelten, N., Troy, R. F., Wagner, S. and Wienhold, F. G.: The AquaVIT-1 intercomparison of atmospheric
 534 water vapor measurement techniques, *Atmospheric Measurement Techniques*, 7, 3159–3251,
 535 <http://dx.doi.org/10.5194/amt-7-3159-2014>, 2014.
- 536 Gurlit, W., Zimmermann, R., Giesemann, C., Fernholz, T., Ebert, V., Wolfrum, J., Platt, U. U. and Burrows, J.

537 P.: Lightweight diode laser spectrometer “CHILD” for balloon- borne measurements of water vapor and
538 methane, *Applied Optics*, 44(1), 91–102, <http://dx.doi.org/10.1364/AO.44.000091>, 2005.

539 Hansford, G. M., Freshwater, R. A., Eden, L., Turnbull, K. F. V., Hadaway, D. E., Ostanin, V. P. and Jones, R.
540 L.: Lightweight dew-/frost-point hygrometer based on a surface-acoustic-wave sensor for balloon-borne
541 atmospheric water vapor profile sounding, *Review of scientific instruments*, 77, 014502–014502,
542 <http://dx.doi.org/10.1063/1.2140275>, 2006.

543 Heinonen, M., Anagnostou, M., Bell, S., Stevens, M., Benyon, R., Bergerud, R. A., Bojkovski, J., Bosma, R.,
544 Nielsen, J., Böse, N., Cromwell, P., Kartal Dogan, A., Aytekin, S., Uytun, A., Fernicola, V., Flakiewicz, K.,
545 Blanquart, B., Hudoklin, D., Jacobson, P., Kentved, A., Lóio, I., Mamontov, G., Masarykova, A., Mitter, H.,
546 Mnguni, R., Otych, J., Steiner, A., Szilágyi Zsófia, N. and Zvizdic, D.: Investigation of the equivalence of
547 national dew-point temperature realizations in the -50 °C to +20 °C range, *International Journal of*
548 *Thermophysics*, 33(8–9), 1422–1437, <http://dx.doi.org/10.1007/s10765-011-0950-x>, 2012.

549 Held, I. and Soden, B.: Water vapor feedback and global warming, *Annual review of energy and the*
550 *environment*, 25(1), 441–475, <http://dx.doi.org/10.1146/annurev.energy.25.1.441>, 2000.

551 Helten, M., Smit, H. G. J., Sträter, W., Kley, D., Nedelec, P., Zöger, M. and Busen, R.: Calibration and
552 performance of automatic compact instrumentation for the measurement of relative humidity from
553 passenger aircraft, *Journal of Geophysical Research: Atmospheres*, 103(D19), 25643–25652,
554 <http://dx.doi.org/10.1029/98JD00536>, 1998.

555 Houghton, J.: *Global warming: The complete briefing*, Cambridge University Press. [online] Available from:
556 <http://britastro.org/jbaa/pdf/114-6shanklin.pdf> (Accessed 14 May 2014), 2009.

557 Hunsmann, S., Wunderle, K., Wagner, S., Rascher, U., Schurr, U. and Ebert, V.: Absolute, high resolution
558 water transpiration rate measurements on single plant leaves via tunable diode laser absorption
559 spectroscopy (TDLAS) at 1.37 μm , *Applied Physics B*, 92(3), 393–401, <http://dx.doi.org/10.1007/s00340-008->
560 3095-2, 2008.

561 JCGM 200: JCGM 200: 2008 International vocabulary of metrology — Basic and general concepts and
562 associated terms (VIM) *Vocabulaire international de métrologie — Concepts fondamentaux et généraux et*
563 *termes associés (VIM)*, International Organization for Standardization, 3(Vim), 104,
564 [http://dx.doi.org/10.1016/0263-2241\(85\)90006-5](http://dx.doi.org/10.1016/0263-2241(85)90006-5), 2008.

565 Joint Committee for Guides in Metrology (JCGM): Evaluation of measurement data - An introduction to the
566 “Guide to the expression of uncertainty in measurement” and related documents, BIPM: Bureau
567 International des Poids et Mesures, www.bipm.org, 2009.

568 Karpechko, A. Y., Perlwitz, J. and Manzini, E.: *Journal of Geophysical Research: Atmospheres*, , 1–16,
569 <http://dx.doi.org/10.1002/2013JD021350>, 2014.

570 Kiehl, J. T. and Trenberth, K. E.: Earth’s Annual Global Mean Energy Budget, *Bulletin of the American*
571 *Meteorological Society*, 78(2), 197–208, <http://dx.doi.org/10.1175/1520->
572 0477(1997)078<0197:EAGMEB>2.0.CO;2, 1997.

573 Kley, D. and Stone, E.: Measurement of water vapor in the stratosphere by photodissociation with Ly α
574 (1216 Å) light, *Review of Scientific Instruments*, 49(6), 691, <http://dx.doi.org/10.1063/1.1135596>, 1978.

575 Krämer, M., Schiller, C., Afchine, A., Bauer, R., Gensch, I., Mangold, A., Schlicht, S., Spelten, N., Sitnikov,
576 N., Borrmann, S., Reus, M. de and Spichtinger, P.: Ice supersaturations and cirrus cloud crystal numbers,
577 *Atmospheric and Oceanic Optics*, 9, 3505–3522, <http://dx.doi.org/10.5194/acp-9-3505-2009>, 2009.

578 Kühnreich, B., Höh, M., Wagner, S. and Ebert, V.: Direct single-mode fibre-coupled miniature White cell for
579 laser absorption spectroscopy, *Review of Scientific Instruments*, 87(2), 0–8,
580 <http://dx.doi.org/10.1063/1.4941748>, 2016.

581 Ludlam, F.: Clouds and storms: The behavior and effect of water in the atmosphere. [online] Available
582 from: <http://agris.fao.org/agris-search/search.do?recordID=US8025686> (Accessed 24 May 2014), 1980.

583 Mackrodt, P.: A New Attempt on a Coulometric Trace Humidity Generator, *International Journal of*
584 *Thermophysics*, 33(8–9), 1520–1535, <http://dx.doi.org/10.1007/s10765-012-1348-0>, 2012.

585 Maddaloni, P., Malara, P. and Natale, P. De: Simulation of Dicke-narrowed molecular spectra recorded by
586 off-axis high-finesse optical cavities, *Molecular Physics*, 108(6), 749–755,
587 <http://dx.doi.org/10.1080/00268971003601571>, 2010.

588 May, R. D.: Open-path, near-infrared tunable diode laser spectrometer for atmospheric measurements of
589 H₂O, *Journal of Geophysical Research*, 103(D15), 19161–19172, <http://dx.doi.org/10.1029/98JD01678>, 1998.

590 Maycock, A. C., Shine, K. P. and Joshi, M. M.: The temperature response to stratospheric water vapour
591 changes, *Quarterly Journal of the Royal Meteorological Society*, 137, 1070–1082,
592 <http://dx.doi.org/10.1002/qj.822>, 2011.

593 MBW Calibration Ltd.: MBW 373HX, 2010.

594 Meyer, J., Rolf, C., Schiller, C., Rohs, S., Spelten, N., Afchine, A., Zöger, M., Sitnikov, N., Thornberry, T. D.,
595 Rollins, A. W., Bozoki, Z., Tatrai, D., Ebert, V., Kühnreich, B., Mackrodt, P., Möhler, O., Saathoff, H.,
596 Rosenlof, K. H. and Krämer, M.: Two decades of water vapor measurements with the FISH fluorescence
597 hygrometer: A review, *Atmospheric Chemistry and Physics*, 15(14), 8521–8538,
598 <http://dx.doi.org/10.5194/acp-15-8521-2015>, 2015.

599 Möller, D., Feichter, J. and Herrmann, H.: Von Wolken, Nebel und Niederschlag, in *Chemie über den*
600 *Wolken... und darunter*, edited by R. Zellner, pp. 236–240, WILEY-VCH Verlag GmbH & Co. KGaA,
601 Weinheim, 2011.

602 Muecke, R. J., Scheumann, B., Slemr, F. and Werle, P. W.: Calibration procedures for tunable diode laser
603 spectrometers, *Proc. SPIE 2112, Tunable Diode Laser Spectroscopy, Lidar, and DIAL Techniques for*
604 *Environmental and Industrial Measurement*, 2112, 87–98, <http://dx.doi.org/10.1117/12.177289>, 1994.

605 Ohtaki, E. and Matsui, T.: Infrared device for simultaneous measurement of fluctuations of atmospheric
606 carbon dioxide and water vapor, *Boundary-Layer Meteorology*, 24(1), 109–119,
607 <http://dx.doi.org/10.1007/BF00121803>, 1982.

608 Peter, T., Marcolli, C., Spichtinger, P., Corti, T., Baker, M. B. and Koop, T.: When dry air is too humid,
609 *Science*, 314(5804), 1399–1402, <http://dx.doi.org/10.1126/science.1135199>, 2006.

610 Petersen, R., Crouce, L., Feltz, W., Olson, E. and Helms, D.: WVSS-II moisture observations: A tool for
611 validating and monitoring satellite moisture data, *EUMETSAT Meteorological Satellite Conference*, 22, 67–
612 77 [online] Available from:
613 http://www.eumetsat.int/Home/Main/AboutEUMETSAT/Publications/ConferenceandWorkshopProceedings/2010/groups/cps/documents/document/pdf_conf_p57_s1_03_petersen_v.pdf (Accessed 20 February 2017),
614 2010.

616 Ravishankara, A. R.: Water Vapor in the Lower Stratosphere, *Science*, 337(6096), 809–810,
617 <http://dx.doi.org/10.1126/science.1227004>, 2012.

618 Rollins, A., Thornberry, T., Gao, R. S., Smith, J. B., Sayres, D. S., Sargent, M. R., Schiller, C., Krämer, M.,
619 Spelten, N., Hurst, D. F., Jordan, A. F., Hall, E. G., Vömel, H., Diskin, G. S., Podolske, J. R., Christensen, L.
620 E., Rosenlof, K. H., Jensen, E. J. and Fahey, D. W.: Evaluation of UT/LS hygrometer accuracy by
621 intercomparison during the NASA MACPEX mission, *Journal of Geophysical Research: Atmospheres*, 119,
622 <http://dx.doi.org/10.1002/2013JD020817>, 2014.

623 Roths, J. and Busen, R.: Development of a laser in situ airborne hygrometer (LISAH) (feasibility study),
624 *Infrared physics & technology*, 37(1), 33–38, [http://dx.doi.org/10.1016/1350-4495\(95\)00103-4](http://dx.doi.org/10.1016/1350-4495(95)00103-4), 1996.

625 Salasmaa, E. and Kostamo, P.: HUMICAP® thin film humidity sensor, in *Advanced Agricultural*
626 *Instrumentation Series E: Applied Sciences*, edited by W. G. Gensler, pp. 135–147, Kluwer., 1986.

627 Scherer, M., Vömel, H., Fueglistaler, S., Oltmans, S. J. and Staehelin, J.: Trends and variability of midlatitude
628 stratospheric water vapour deduced from the re-evaluated Boulder balloon series and HALOE,
629 *Atmospheric Chemistry and Physics*, 8, 1391–1402, <http://dx.doi.org/10.5194/acp-8-1391-2008>, 2008.

630 Schiff, H. I., Mackay, G. I. and Bechara, J.: The use of tunable diode laser absorption spectroscopy for
631 atmospheric measurements, *Research on Chemical Intermediates*, 20(3), 525–556,
632 <http://dx.doi.org/10.1163/156856794X00441>, 1994.

633 Schulz, C., Dreizler, A., Ebert, V. and Wolfrum, J.: Combustion Diagnostics, in Handbook of Experimental
 634 Fluid Mechanics, edited by C. Tropea, A. L. Yarin, and J. F. Foss, pp. 1241–1316, Springer Berlin Heidelberg,
 635 Heidelberg, 2007.

636 Sherwood, S., Bony, S. and Dufresne, J.: Spread in model climate sensitivity traced to atmospheric
 637 convective mixing, *Nature*, 505(7481), 37–42, <http://dx.doi.org/10.1038/nature12829>, 2014.

638 Silver, J. A. and Hovde, D. C.: Near-infrared diode laser airborne hygrometer, Review of scientific
 639 instruments, 65(5), 1691–1694, <http://dx.doi.org/10.1063/1.1144861>, 1994a.

640 Silver, J. and Hovde, D.: Near-infrared diode laser airborne hygrometer, Review of scientific instruments,
 641 65, 5, 1691–1694 [online] Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4991817
 642 (Accessed 25 November 2013b), 1994.

643 Smit, H. G. J., Rolf, C., Kraemer, M., Petzold, A., Spelten, N., Neis, P., Maser, R., Buchholz, B., Ebert, V. and
 644 Tatrai, D.: Development and Evaluation of Novel and Compact Hygrometer for Airborne Research
 645 (DENCHAR): In-Flight Performance During AIRTOS-I / II Research Aircraft Campaigns, Geophysical
 646 Research Abstracts, 16(EGU2014-9420), 2014.

647 Sonntag, D.: Important new Values of the Physical Constants of 1968, Vapour Pressure Formulations based
 648 on the ITS-90, and Psychrometer Formulae, *Meteorologische Zeitschrift*, 40(5), 340–344, 1990.

649 Thornberry, T. D., Rollins, A. W., Gao, R. S., Watts, L. A., Ciciora, S. J., McLaughlin, R. J. and Fahey, D. W.:
 650 A two-channel, tunable diode laser-based hygrometer for measurement of water vapor and cirrus cloud ice
 651 water content in the upper troposphere and lower stratosphere, *Atmospheric Measurement Techniques*
 652 Discussions, 7(8), 8271–8309, <http://dx.doi.org/10.5194/amtd-7-8271-2014>, 2014.

653 Thunder-Scientific: Model 2500 Two-Pressure Humidity Generator, [online] Available from:
 654 www.thunderscientific.com (Accessed 12 May 2016), 2016.

655 Webster, C., Flesch, G., Mansour, K., Haberle, R. and Bauman, J.: Mars laser hygrometer, *Applied optics*,
 656 43(22), 4436–4445, <http://dx.doi.org/10.1364/AO.43.004436>, 2004.

657 White, J.: Very long optical paths in air, *Journal of the Optical Society of America* (1917-1983), 66(5), 411–
 658 416, <http://dx.doi.org/10.1364/JOSA.66.000411>, 1976.

659 Wiederhold, P. R.: Water Vapor Measurement. Methods and Instrumentation, Har/Dskt., CRC Press., 1997.

660 Zöger, M., Afchine, A., Eicke, N., Gerhards, M.-T., Klein, E., McKenna, D. S., Mörschel, U., Schmidt, U., Tan,
 661 V., Tuitjer, F., Woyke, T. and Schiller, C.: Fast in situ stratospheric hygrometers: A new family of balloon-
 662 borne and airborne Lyman photofragment fluorescence hygrometers, *Journal of Geophysical Research*,
 663 104(D1), 1807–1816, <http://dx.doi.org/10.1029/1998JD100025>, 1999a.

664 Zöger, M., Engel, A., McKenna, D. S., Schiller, C., Schmidt, U. and Woyke, T.: Balloon-borne in situ
 665 measurements of stratospheric H₂O, CH₄ and H₂ at midlatitudes, *Journal of Geophysical Research*,
 666 104(D1), 1817–1825, <http://dx.doi.org/10.1029/1998JD100024>, 1999b.

667
 668

Figures:

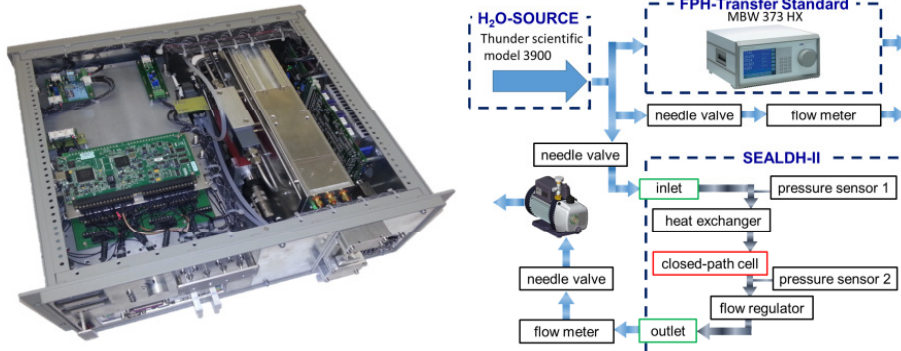


Figure 1: Left: Photo of SEALDH-II, the Selective Extractive Airborne Laser Diode Hygrometer (dimension 19" 4 U). Right: Setup for the metrological absolute accuracy validation. The combination of a H₂O source together with a traceable dew point hygrometer, DPM, is used as a transfer standard – a traceable humidity generator (THG).

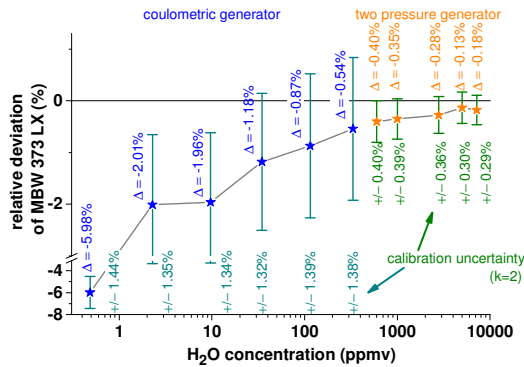


Figure 2: Calibration of the DPM (dew/frost point mirror hygrometer, MBW 373 LX, which is used as part of the THG) at the national primary water vapor standards of Germany. The standard for the higher H₂O concentration range (orange) is a “two pressure generator” (Buchholz et al., 2014); for the lower concentration range (blue) a “coulometric generator” (Mackrodt, 2012) is used as a reference. The deviations between reference and DPM are labelled with “Δ”. The uncertainties of every individual calibration point are stated as green numbers below every single measurement point.

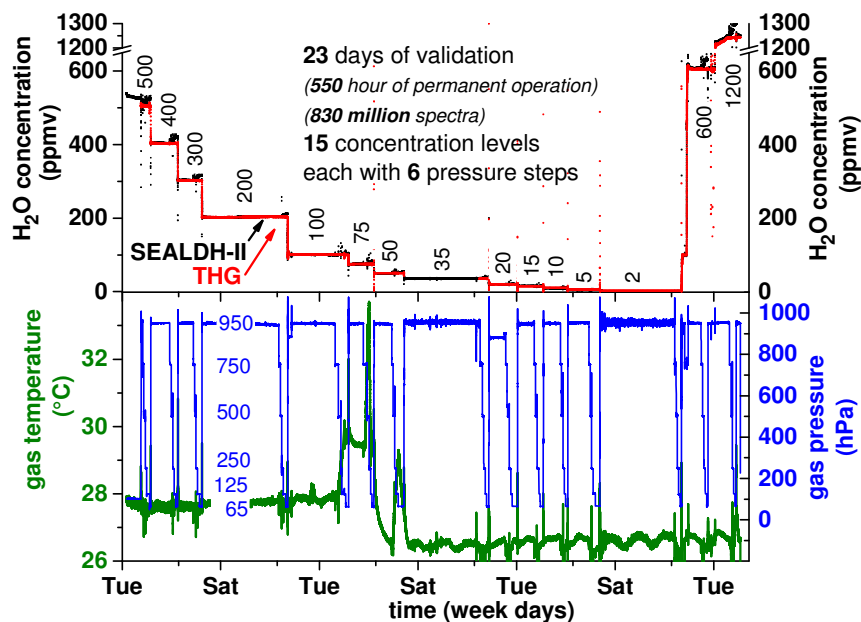


Figure 3: Overview showing all data recorded over 23 days of validation experiments. Measurements of the traceable humidity generator (THG) are shown in red, SEALDH-II data in black, gas pressure and gas temperature in SEALDH-II's measurement cell are shown in blue and green. Note: SEALDH-II operated the entire time without any malfunctions; the THG didn't save data in the 35 ppmv section; the temperature increase during the 75 ppmv section was caused by a defect of the air conditioning in the laboratory.

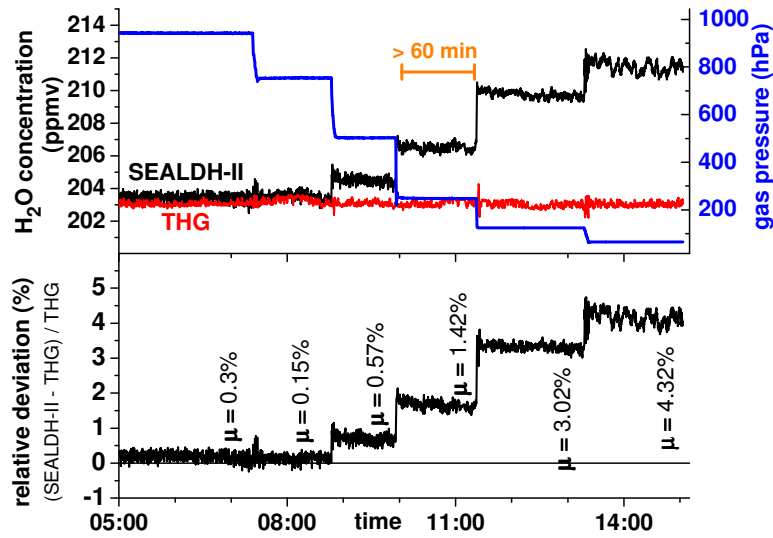


Figure 4: Detailed plot of the validation at 200 ppmv with six gas pressure steps from 50 to 950 hPa. Each individual pressure level was maintained for at least 60 minutes in order to avoid any dynamic or hysteresis effects and to facilitate clear accuracy assessments. The μ -values defines the averaged relative deviation on every gas pressure level.

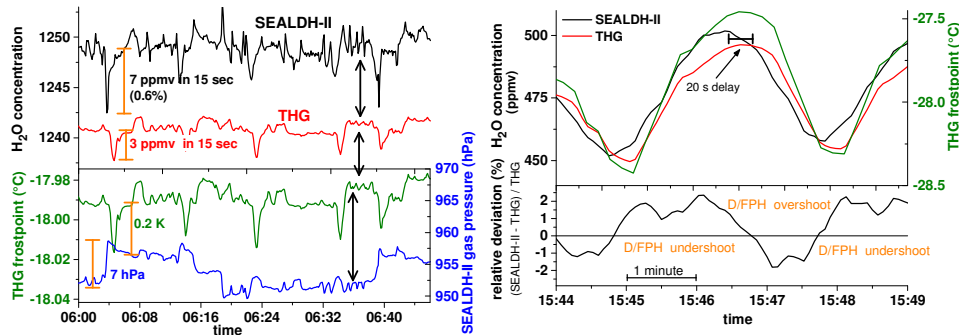


Figure 5: Short term H_2O fluctuations in the generated water vapor flow measured by SEALDH-II and the dew/frost point mirror hygrometer (D/FPH) of the traceable humidity generator (THG). The different dynamic characteristics of SEALDH-II (fast response time) and THG (quite slow response) lead in a direct comparison to artificial noise. Oscillating behaviors like in the right figure occur when the THG is not equilibrated. We did not use these data segments for the accuracy assessments.

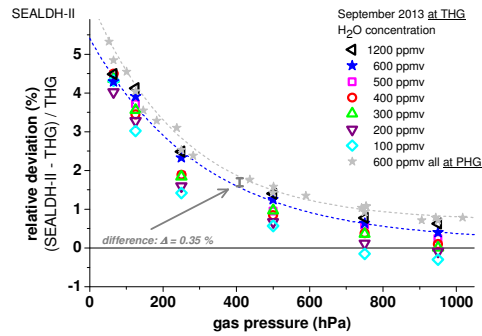


Figure 6: Gas pressure dependent comparison between SEALDH-II and THG over a H₂O concentration range from 600 to 1200 ppmv and a pressure range from 50 to 950 hPa. The 600 ppmv values (in grey) are measured directly at the national primary humidity generator (PHG) of Germany; all other H₂O concentration values are measured at and compared to the traceable humidity generator (THG). All SEALDH-II spectra were evaluated with a calibration-free first principles evaluation based on absolute spectral parameters. No initial or repetitive calibration of SEALDH-II with respect to any “water reference” source was used.

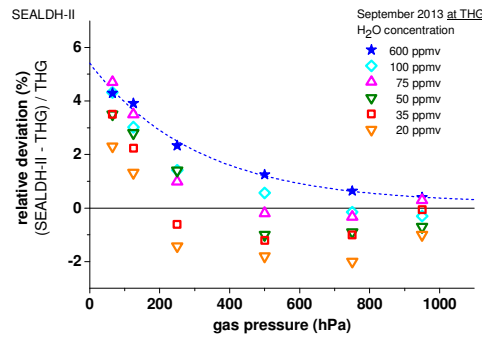


Figure 7: Comparison results as in Figure 6 but for the 200 – 600 ppmv range.

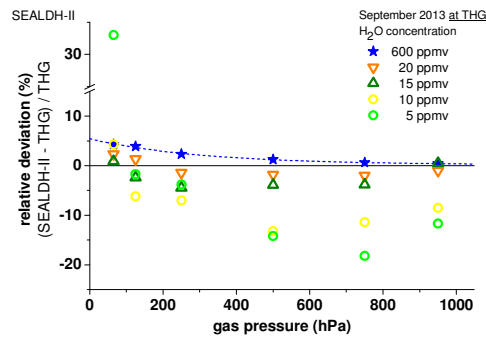


Figure 8: Comparison results as in Figure 6 and Figure 7 but for the 5 – 20 ppmv range. All spectra are determined with a calibration-free first principles evaluation concept. The major contribution to the higher fluctuations at lower concentrations is the accuracy of the offset determination (details see text).

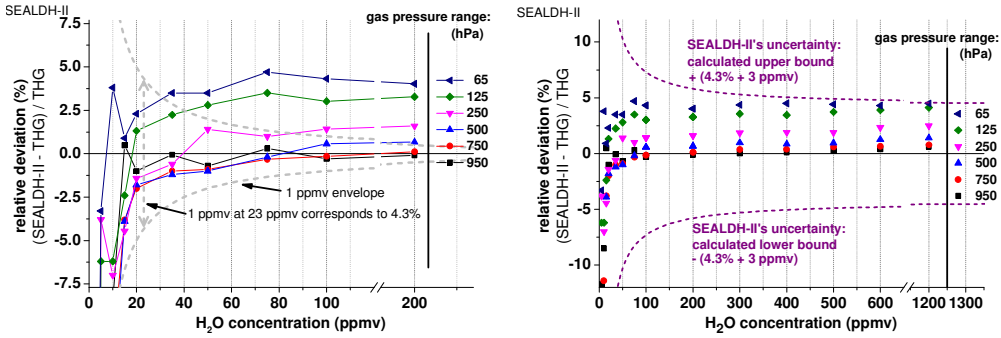


Figure 9: Direct comparison of SEALDH-II versus THG for H_2O concentrations between 5 and 200 ppmv and gas pressures from 65 to 950 hPa. Both figures show the relative deviations between SEALDH-II and THG grouped and color-coded by gas-pressure. Left plot: relative deviations of SEALDH-II versus THG below 200 ppmv; the grey line indicates the computed relative effect in SEALDH-II's performance caused by ± 1 ppmv offset fluctuation. This line facilitates a visual comparison between an offset impact and the 4.3% linear part of the uncertainty of SEALDH-II. Right plot: relative deviations for all measured data in the same concentration range. Also shown is SEALDH-II's total uncertainty of $4.3\% \pm 3$ ppmv (calculated for 1013 hPa) as a dashed line.