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 impedance-based hygrometer through impedance 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 <u>Metrology</u> 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, commonly 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 vise 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 frame work 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, weather 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 explicitely.
==> 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 <u>chemical</u> 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 confidential 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 accordantly 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.