Response to the Editor

Dear Dr. Janssen,

We sincerely thank you for your decision and for the attention to detail. We have carefully revised the manuscript according to your requests and implemented all suggested technical and formatting improvements to ensure consistency with AMT standards. Below we provide a detailed point-by-point response.

Our responses are in blue, with specific changes to the text highlighted in *blue italics*.

All line numbers in this document correspond to the line numbers in the updated version of the manuscript.

With best regards,
Di WANG
On behalf of all authors

Dear Mr Wang, dear authors,

I would like to thank you again for submitting to AMT and I also thank you for responding to the referee's and my requests. I appreciate the great detail that you have dedicated to each of your responses that helped to clarify most of the concerns. Nevertheless, there remains one issue which has not been addressed satisfactorily. This requires that you go once again through a revision of your manuscript.

The comparison of your bag measurements with independent methods is a substantial part of the paper. However, as already indicated in the last referee report

"It stands out from the comparison of corrected data with independent observations (in situ for near-ground sampling, IASI satellite data for higher altitudes) that the applied correction model often fails to significantly eliminate the observed offsets between raw bag data and the respective reference observations, especially in the case of high-altitude sampling. In many cases, correction-induced shifts of data were even in the wrong direction, i.e. away from reference observations, which makes the entire correction approach questionable or at least insufficient. This holds even for near-ground samples, where the environmental conditions should be known best and exposition of samples to potentially adverse conditions should be shortest between collection and analysis."

it remains unclear whether the correction has a significant impact on comparisons with other platforms, whether in situ or satellite. What readers want to know is whether it affects the comparison between the bag data and the other instruments. Does it improve or worsen the comparison? For example, the data in Fig. A3 are too scattered to actually 'see' the effect of the correction. Clear evidence is needed to show whether the diffusion correction improves the comparison with other observations, such as a statistical analysis based on the correlation coefficient or residuals showing how the agreement between the bag data and the validation platform changes with the correction. The results of this analysis should be discussed critically in light of the referee's remark cited above. The same applies to the comparison with satellite data. Please check all your statements regarding this comparison. The current discussion is inappropriate, particularly the statement in the abstract that 'the corrected observations match the Picarro in situ observations and IASI satellite retrievals'. According to panel c of figure 10, where both the uncorrected and corrected data are substantially higher than the observations, the statement is both incorrect and misleading, as the graph suggests that the correction has no significant effect on the agreement between the bag and the IASI data.

We sincerely thank you for this constructive follow-up. We fully understand that the key remaining issue concerns whether the diffusion correction significantly improves the agreement between the air-bag measurements and the other dataset (Picarro and IASI).

In this revised version, we have (1) performed a quantitative statistical analysis (correlation coefficients and mean absolute errors), (2) clarified that the comparison with IASI is qualitative only, and (3) note the difference in the upper troposphere, and critically discussed the sources of discrepancy and the limitations of satellite representativeness.

For the comparison between the ground-level bag measurements and the in-situ Picarro observations, the correction was indeed small for both δ^2H and d-excess, which is expected because the air inside and outside the bags had nearly identical humidity and isotopic composition. As described in Section 2.1 (Eq. 8), the diffusion effect depends on the humidity and isotopic difference between the inside and outside of the bags.

We have accordingly revised the text to provide clearer quantitative comparisons and explanations (lines 614-626):

"At near-ground level, both the raw and corrected δ^2H and d-excess values from the bag measurements show close agreement with the in-situ Picarro observations (Fig. 10a and b), indicating minimal diffusion effects under nearly identical humidity and isotopic conditions between the inside and outside of the bags. Most data points are distributed along the 1:1 line (Fig. A3), while d-excess shows slightly larger scatter. After applying the correction, the agreement improved: the correlation coefficient between the bag and Picarro δ^2H values increased from 0.90 to 0.99, and for d-excess from 0.41 to 0.72. The mean absolute error (MAE) between the corrected bag and Picarro measurements is 3.0‰ for δ^2H and 1.4‰ for d-excess. These discrepancies are within the uncertainty range derived from laboratory diffusion experiments and model–experiment mismatches, which were comprehensively incorporated in the error

estimation (Sections 3.2 and 3.4): 0.5 ‰ for $\delta^{18}O$, 4.1 ‰ for δ^2H, and 2.9 ‰ for d-excess. The differences also could be due to the fact that the Picarro in-situ observations are period-averaged, whereas the bag samples represent instantaneous values.”

For higher altitudes, we can only rely on satellite data IASI, but as described in the Methods section, the IASI dataset is not strictly comparable with the local high-resolution observations. The MUSICA retrievals from the IASI satellite instrument provides water vapor isotope data at three altitude levels: 1-3 km in the lower troposphere, 4-7 km in the mid-troposphere, and 8-12 km in the upper troposphere. However, these measurements represent a vertical average over layers determined by the averaging kernels. While applying the averaging kernels to smooth our observed profiles would allow a more quantitative comparison, the multi-level data with averaging kernels for 2020 are not publicly available and were still inaccessible at the time of our delayed response submission. Therefore, we averaged our observations over the corresponding altitude ranges; as a result, the comparison remains mainly qualitative. The effect of the averaging kernels on altitude could contributes to the differences between IASI and our observations in the upper troposphere, which also highlights the importance of our high-resolution vertical measurements.

In addition, even if the IASI or our observational profiles were adjusted according to the averaging kernels, three other factors would still further limit their comparability, which we now explicitly discuss below:

(1) Temporal mismatch: IASI provides only two overpasses per day (local overpass around 09:30 am and 09:30 pm), whereas our observations were made during daytime. Diurnal variability can introduce significant differences (Herman et al., 2014; Lacour et al., 2018). Moreover, IASI data are not available for every observation date, even at locations close to our sampling site.

(2) Spatial mismatch: The IASI footprint (~12 km at nadir) represents a coarse spatial average, whereas our measurements are local and instantaneous. Therefore, the satellite retrievals cannot fully represent the fine-scale variability at our sampling site.

(3) Most important—Cloud-sampling limitation: IASI cannot observe through or below thick clouds, such as anvil clouds. Consequently, it mainly samples the environment outside convective detrainment regions. Since convective detrainment isotopically enriches vapor (Kuang et al., 2003; Moyer et al., 1996; Smith et al., 2006; Vries et al., 2021; Webster and Heymsfield, 2003), the air masses observed by IASI—being farther from detrainment—are expected to be more depleted. This sampling limitation could therefore contribute to the lower δ^2H values in the upper troposphere (around and above cloud top) retrieved by IASI compared with our in-situ observations.

The diffusion of air bags should result in systematic bias. But IASI data sometimes agree well with our measurements, sometimes show higher values, and sometimes lower values (Fig. 10). These fluctuations arise from representativeness differences inherent to the satellite retrievals, not from diffusion-related biases. Therefore, such variations cannot be corrected using the diffusion model, which accounts for isotopic fractionation during air-bag storage but not for spatiotemporal (both horizontal and vertical) discrepancies between measurement platforms, nor for the observational limitations of satellites regarding cloud processes.

We have now included the uncertainty ranges of the IASI retrievals in Fig. 10, which just account for retrieval—fit noise and atmospheric temperature a priori constraint. The uncertainty range is already large. In the 4000–7000 m range, it almost entirely covers our observation range, demonstrating good consistency within errors. However, in the upper troposphere (7000–12000 m), the IASI data still appear lower than our observations. This discrepancy likely arises primarily from cloud-sampling limitations, as well as from temporal and spatial representativeness mismatches, rather than from diffusion-related biases. We explicitly acknowledge this difference in the revised Discussion and thoroughly analyze all potential causes (lines 627-676):

"To compare corrected measurements with observations at higher altitudes, we refer to the IASI satellite dataset. We acknowledge that this comparison is qualitative, as satellite data represent vertical averages over layers defined by averaging kernels, and differ in measurement footprints (both horizontal and vertical) and spatio-temporal sampling disparities (Shi et al., 2020). A more quantitative analysis could be facilitated if an averaging kernel is used to smooth the observed profiles (Herman et al., 2014), but even after such adjustment, substantial differences would likely persist. IASI provides only two overpasses per day (around 09:30 LT AM and PM), whereas our drone-based measurements were conducted locally during daytime. The IASI footprint (~12 km at nadir) provides coarse spatial averaging, whereas our measurements are local and instantaneous, limiting its ability to capture fine-scale variability at the sampling site. Importantly, IASI cannot observe through or below thick clouds such as anvils, and thus mainly samples air outside convective detrainment regions. Because convective detrainment isotopically enriches vapor (Kuang et al., 2003; Moyer et al., 1996; Smith et al., 2006; Vries et al., 2021; Webster and Heymsfield, 2003), the air masses observed by IASI—being farther from detrainment—are expected to be more depleted. This cloud-sampling limitation may therefore contribute to the lower δ^2H values retrieved by IASI in the upper troposphere compared with our in-situ observations. Such fundamental mismatches in vertical weighting, temporal coverage, and spatial representativeness are noted in previous satellite–in-situ intercomparisons, which found that satellites capture large-scale isotopic gradients but smooth out fine-scale vertical and diurnal variations (Herman et al., 2014; Lacour et al., 2018). These complementary characteristics highlight the value of combining satellite and high-resolution in-situ observations: satellites provide synoptic coverage, while local observations resolve small-scale processes that shape the isotopic structure of the atmosphere. Overall, our comparison with IASI should thus be regarded as a qualitative consistency check confirming large-scale and temporal coherence rather than a one-to-one correspondence.

For most intervals, IASI satellite data closely match raw and corrected δ2H measurements for altitudes 4000–7000m. In the 8000–12000m range, although applying the diffusion correction increased the correlation coefficient between the air-bag observations and the IASI data from 0.53 to 0.60, IASI data are lower than δ^2H air-

bag observations during certain periods, particularly June and September 2020; one possible reason is unaccounted environmental influences at high altitudes. However, it is noteworthy that the IASI data closely match the observed δ^2H for all other periods in the 4000–7000 m range, they are also lower in June 2020. Moreover, diffusion within the sampling bags could, in principle, introduce systematic bias, yet the IASI values sometimes agree well, sometimes appear higher, and sometimes lower than our measurements (Fig. 10). Therefore, the differences between our in-situ observations and the IASI retrievals likely arise mainly from representativeness differences intrinsic to the satellite data, as discussed above—practically the cloud-sampling limitation in the upper troposphere—rather than from diffusion-related biases. Accordingly, such variations cannot be corrected by the diffusion model, which accounts for isotopic fractionation during air-bag storage but not for spatiotemporal or cloud-sampling discrepancies between observation platforms.

The right panel of Fig.10 (Figs. 10 b, d, and f) shows the comparison of raw and corrected vapor d-excess measurements. No d-excess dataset is available for comparison for the 4000–7000m and 8000–12000m (Fig. 10d and f). As previously noted, corrected d-excess values should be and are lower than the raw data, as expected from diffusion theory. For the 8000–12000 m observations, the correction magnitude is relatively smaller than at lower altitudes due to the shorter storage time of the air bags.

Although no completely independent dataset is available for direct validation, our model—already verified by laboratory experiments—was used to perform a detailed assessment of all potential error sources and to quantify a comprehensive and conservative uncertainty range of the corrected data."

In Section 4.2, we demonstrate that our diffusion model successfully reproduces the experimental results under a wide range of environmental conditions, confirming its robustness and its ability to quantify the influence of diffusion on water vapor isotope measurements.

As emphasized in the Conclusion, an important contribution of this study is the methodological guidance it provides:

"We recommend prioritizing the use of glass containers and air bags with the lowest permeability for collecting water vapor using portable devices. Additionally, it is essential to conduct the permeability experiments described in this article before any experimental undertaking. This involves storing water vapor with known isotopic composition in the portable collection device for an extended period and then re-measuring these values to assess or determine the device's permeability parameters." This recommendation represents a key outcome of our work, offering a practical standard for future field measurements.

Regarding the vertical profiles, although no completely independent dataset is available for direct validation, our model—already verified by laboratory experiments—was used to propagate all potential error sources and to calculate a conservative and comprehensive uncertainty range.

In the revised manuscript, we no longer emphasize detailed matching with the IASI data, but rather highlight the three main contributions of this work:

(1) methodological recommendations for future isotopic vapor sampling,

(2) development of a robust diffusion model validated under various experimental conditions

(3) quantitative uncertainty assessment for the vertical profiles.

Apart from this point, I would like to ask you to consider the following minor comments to revise your manuscript.

General technical comments:

There are several occurrences of underscores in the text and formulae, such as λ_surface or M_alt. This is somewhat non-standard and reads like computer code. Could you please remove the underscore when you use the subscript in combination with a mathematical symbol or variable such as (lambda, alpha, etc.)

Thank you for the suggestion. We have removed the underscores and replaced them with proper subscripts throughout the text and equations.

Please check that the quality of graphs is sufficient and that exes labels are uncut (see Fig. 10 for example).

Thank you for your reminder. We have checked all figures to ensure sufficient quality and that all axis labels are fully visible.

Detailed comments:

l 42 "The corrected observations match the Picarro in-situ observations and IASI satellite retrievals." -> This statement is not only wrong (see above), but it is also misleading.

We have deleted this sentence, we no longer emphasize detailed matching with the IASI data, but rather highlight the three main contributions of this work:

(1) methodological recommendations for future isotopic vapor sampling,

(2) development of a robust diffusion model validated under various experimental conditions

(3) quantitative uncertainty assessment for the vertical profiles.

l 119 "However, because of the short residence time at high altitudes, the reduced diffusivity at low temperatures, and the flexibility of the bags maintaining equal internal and external pressure, our current approach does not explicitly resolve potential temperature- or pressure-dependent changes in bag permeability. Under our experimental conditions, these effects were instead addressed indirectly through conservative uncertainty estimates. Future applications should take into account the specific experimental conditions, and future developments may extend the model to explicitly incorporate the dependence of bag permeability on partial pressure and temperature, thereby improving its applicability under a wider range of atmospheric

conditions."
"However, because" sounds strange.
You could instead write something along the following lines "Although we minimised certain diffusion-related biases, such as pressure differences and long storage times, temperature and pressure effects on bag permeability were only addressed indirectly via conservative uncertainty estimates. Future applications should consider the specific experimental conditions and extend the model to explicitly incorporate the dependence of bag permeability on partial pressure and temperature. This would improve the model's applicability to a wider range of atmospheric conditions."

Thank you for the suggestion. We have revised the paragraph accordingly as follows (lines 118-127):

"Although we minimized certain diffusion-related biases, such as pressure in different altitudes and long storage times, temperature and the pressure difference between the inside and outside of the bags on bag permeability were only addressed indirectly via conservative uncertainty estimates. This treatment is justified under our experimental conditions, which are characterized by the short residence time at high altitudes, the reduced diffusivity at low temperatures, and the flexibility of the bags maintaining equal internal and external pressure. Future applications should consider the specific experimental conditions, and future developments may extend the model to explicitly incorporate the dependence of bag permeability on partial pressure and temperature, thereby improving its applicability under a wider range of atmospheric conditions."

l 148 Here, I need to apologize, because I was obviously wrong in one of my earlier remarks on the definition of alpha. According to the "IUPAC gold book" (https://goldbook.iupac.org/terms/view/K03405), the standard definition of alpha is k_light/k_heavy. Also, your approximation is k = k_light. Please change your phrases accordingly.

We sincerely thank the editor for this clarification. In our previous revision, we mistakenly followed an earlier suggestion without carefully rechecking our own definition, which led to an incorrect expression of α. We have now corrected this and confirmed that the current formulation is fully consistent with the IUPAC definition (α = k_light / k_heavy). To facilitate checking, the relevant section is shown below, only the revised sentences are marked in light blue italics (lines 149-151):

"The flux of water into the bag, F (in $kg/m^2/s$), is expressed as:

$$F = k * (q_e - q(t)) \quad (1)$$

where q(t) represents the variation of humidity inside the air bag over time (in kg/kg), q_e denotes the environmental humidity (in kg/kg), k is water vapor conductance. This first-order formulation is derived from Fick's diffusion law(Fick, 1855).

Similarly, the flux of isotopologue, F_i, either $H_2{}^{18}O$ or HDO, moving into the bag can be described as:

$$F_i = k_i * (R_e * q_e - R(t) * q(t)) \quad (2)$$

In this equation, k_i represents the conductance specific to each isotopologue (in kg/m^2/s), R_e denotes the isotopic ratio in the environment, and R(t) is the variation of isotopic ratio within the air bag with time. Notably, the fractionation coefficient here can be denoted as:

$$\alpha=\frac{k}{k_i} \tag{3}$$

Since k represents the total conductance dominated by the lighter molecules ($H_2{}^{16}O$), and k_i is defined as the conductance of the heavier isotopologues (either $H_2{}^{18}O$ or HDO), Eq. (3) is therefore consistent with the conventional definition of fractionation factors ($\alpha=k_{light}/k_{heavy}$).”

l 278 I suggest to write "As lambda is related to the surface area of the sampling bag, " ...

We have modified as suggested.

l 342 delete ", preventing unintended air ingress.", as this has already been said in the sentence before

Thank you. We have deleted as suggested.

l 388 "The method of uncertainty estimation" -> "Method of uncertainty estimation"

We have replaced the sentence as recommended.

Fig 3 "α_δ 18O" should be "α18O" and "α_δ2H" should be "α2H"

Thank you for pointing out the missing updates. We have corrected them accordingly.

l 556 "determined the ranges for four errors to evaluate uncertainty" -> "investigated four error sources to evaluate the uncertainty"

We have replaced the sentence as recommended.

l 559 Consider to replace "We represent the contributions of uncertainty sources to vertical vapor δ18O and d-excess measurements at different altitudes using probability density function plots (Fig. 8)." -> "We represent the contribution of each source of uncertainty to vertical vapor δ18O and d-excess measurements at different altitudes using probability density function plots (Fig. 8)."

We have replaced the sentence as recommended.

l 599 Common confidence intervals are 68 %, 95.5 % or 99.7 %, which corresponds to 1, 2 or 3 sigma, respectively, if a standard normal probability distribution function (Gaussian) is assumed. Maybe you could give your result once for 95% and once for the 98% confidence interval.

Thank you. We have added the corresponding 95% confidence intervals (~0.9‰ for $\delta^{18}O$ and ~6.9‰ for d-excess) as suggested.

l 594 "these ranges is" -> "this range is"

We have revised the grammar.

l ~604, figure caption: delete "(legend: in situ measuremeents)"

Thank you. We have deleted as suggested.

Fig. 8, If the mismatch is represented as a pdf, it is a delta-function with infinite amplitude so that the integral is 1. Please correct in the picture by drawing a vertical line which goes to the upper axis of the graph. Also the unit of the ordinate axis must be 1/‰.

Thank you for your suggestion. We have modified Fig. 8 accordingly by adding a vertical line to represent the mismatch as a delta function and corrected the ordinate unit to 1/‰.

l 609 change "2.99" to "3" or to "3.0", because I don't think that your result is significant to 2 digits after the decimal point.

We have adjusted the decimal place of the error value.

l 627 "After correction, the a mean absolute error is 1.37 ‰. ." -> "After correction, the mean absolute error is 1.4 ‰."

Thank you. We have adjusted the decimal place of the error value and removed the extra "a".

ll 630 I strongly recommend replacing "As previously noted, raw d-excess are higher than corrected data due to kinetic fractionation. After correction, d-excess decreases." by "As previously noted, corrected d-excess values are lower than the raw data."

Thank you. We have replaced the sentence as recommended.

l 654 check sentence: ", and offers" instead of ", offers" ?

Thank you. We have modified as suggested.

Fig A3. Panel b) shows the wrong data. While δ18 should be presented, the data belong to measurements of δ2H. Please display the correct data.

Thank you for your comment. The axis label was mistakenly written as $\delta^{18}O$. We have now corrected the label to $\delta^{2}H$.

References:

Herman, R., Cherry, J., Young, J., Welker, J., Noone, D., Kulawik, S., Worden, J., 2014. Aircraft validation of Aura Tropospheric Emission Spectrometer retrievals of HDO/H 2 O. Atmospheric Measurement Techniques 7, 3127-3138.

Kuang, Z., Toon, G.C., Wennberg, P.O., Yung, Y.L., 2003. Measured HDO/H2O ratios across the tropical tropopause. Geophysical Research Letters 30.

Lacour, J.-L., Risi, C., Worden, J., Clerbaux, C., Coheur, P.-F., 2018. Importance of depth and intensity of convection on the isotopic composition of water vapor as seen from IASI and TES δD observations. Earth and planetary science letters 481, 387-394.

Moyer, E.J., Irion, F.W., Yung, Y.L., Gunson, M.R., 1996. ATMOS stratospheric deuterated water and implications for troposphere-stratosphere transport. Geophysical Research Letters 23, 2385-2388.
Smith, J.A., Ackerman, A.S., Jensen, E.J., Toon, O.B., 2006. Role of deep convection in establishing the isotopic composition of water vapor in the tropical transition layer. Geophysical Research Letters 33, 0_46-40_47.
Vries, A.D., Aemisegger, F., Pfahl, S., Wernli, H., 2021. Stable water isotope signals in tropical ice clouds in the West African monsoon simulated with a regional convection-permitting model.
Webster, C.R., Heymsfield, A.J., 2003. Water isotope ratios D/H, O-18/O-16, O-17/O-16 in and out of clouds map dehydration pathways. Science 302, 1742-1745.