We would like to thank the referee for the review and comments to improve the quality of this work. Following these comments, we have made major modifications and point-by-point responses. Please find our response below (in blue).

General Comments.

- There is a clear need for deriving and applying water vapor corrections for GHGs and trace gases, especially in a field setting, over time. This approach seems like a viable approach for field deployment.
- The clearest benefit to the community is in communicating the design of the apparatus and experiment and in evaluating its efficacy. However, much of the paper reads as an assessment of a small number of instruments with a limited number of experiments. It seems to me that these assessments provide an incomplete picture of the instruments tested, and sometimes leave more questions open than they answer. Focusing the paper on the performance of the apparatus (as opposed to the instruments under test) would seem to be of greater benefit to the community, and better reflect the purpose of the paper as indicated by the title.
- In the assessment of the apparatus, I see a few gaps that I think merit additional testing and discussion:
    - A better assessment of the performance of the water vapor generator would be helpful: accuracy, linearity, repeatability / stability over time and over ambient temperature, etc.

We agree that we have not addressed the accuracy of the water vapor generator in that we have not measured the water vapor mole fractions with an independent instrument. We unfortunately do not have access to a chilled mirror hygrometer or similar. For our application, the exact water vapor mole fraction isn't crucial as – we just want to correct the trace gas measurements for the effects of water vapor, – but for other applications the accuracy would absolutely be important. We thank the reviewer for pointing out this oversight on our part and we have corrected the text in Lines 382-388 to acknowledge this limitation of our study.

We're not sure what the reviewer means by linearity.

For our application, the key considerations are the repeatability and temporal stability of the water vapor calibration curves. In Table 1 and 2 and the original Figures 3-10, we included a total of 8 tests for A778, 10 tests for A798, 5 tests for CFKADS2025, 4 tests for CFKADS2218, and 5 tests for CKADS2025. The Aeris tests spanned 3-4 weeks for each instrument, whereas the Picarro tests spanned only 1-2 days for CFKADS2218 and CKADS2025, but approximately 3 weeks for CFKADS2025.

To more fully assess the repeatability of the results, we analyzed the saturation test results in this context for A798 (figure below). One test was performed on 17 March 2026, followed by five consecutive tests on 14 April 2026. At high water vapor concentrations, the resulting CH_4 correction differed by 4.7 ppb (the March 17 correction minus the mean of the April 14 corrections at 20,000 ppm H_2O). The differences were smaller at lower water vapor concentrations, with differences of 0.1, 0.5, and 1.5 ppb CH_4 at 1500, 5000 and 10000 ppm H_2O, respectively. Thus, the repeatability is better than the WMO target compatibility of 2 ppb for H_2O less than 10000 ppm, but we do note that the full calibration curve is not adequately described by a second order polynomial for this instrument and the residuals are large for very dry conditions (below 2000 ppm H_2O).

We performed the same analysis for the low water vapor tests listed in Table 1, which were conducted on 16-18 March. At 1500 ppm H_2O, which was within the measured range of all four tests, the fitted CH_4 correction due to water vapor ranged from approximately -2.0 to 0.6 ppb. This plot more clearly represents the repeatability of the calibration curves, so we have added this plot in the supplement (Figure S6) and the analysis (Lines 264-271) to more explicitly characterize the repeatability of the water vapor correction.

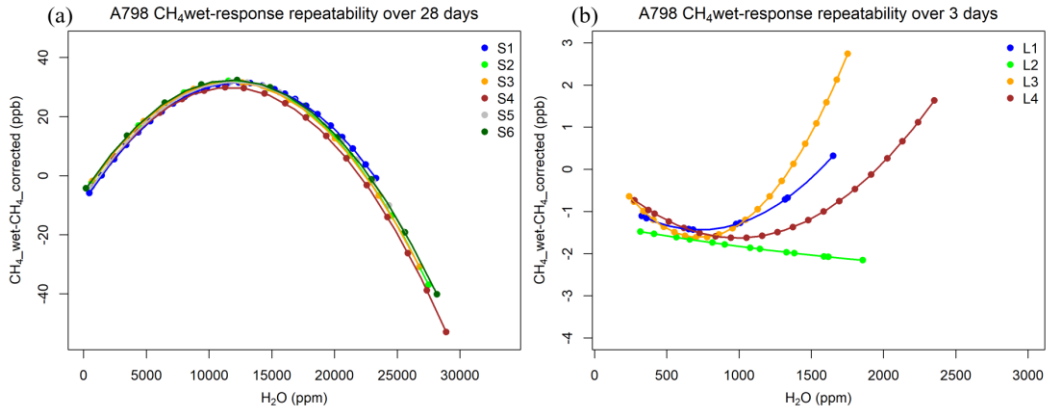


Figure S6. (a) CH_4 wet-response repeatability of the Aeris MIRA serial number A798 analyzer over 28 days for the saturation tests. The y axis shows the CH_4 correction, defined as the difference between the wet and corrected CH_4 mole fractions (CH_4_wet – CH_4_corrected). Different colors indicate saturation tests S1-S6. Lines represent the fitted water-vapor response functions for each range. (b) CH_4 wet-response repeatability of the Aeris MIRA serial number A798 analyzer over 3 days for the low water vapor tests.

Thank you for also bringing up the stability of the results over temperature. If the water generator is brought to a field site, it is likely that the temperature is not as well controlled as our laboratory. Of course, the ambient temperature itself determines the

saturation water vapor level, so performing the testing in an unheated shed in the winter, for example, would not be ideal. This reviewer comment prompted us to examine whether temperature variability could account for the observed short-term H_2O variability during the calibration steps for developing the calibration curve. We compared the standard deviation of the H_2O during the calibration steps to the standard deviation of the enclosure inlet temperature (T2, see below). The correction coefficient (R^2) was 0.006 for the 313 windows from the 3-5 min tests. Thus, variation in enclosure inlet temperature explains essentially none of the observed short-term H_2O variability, which may be caused by Aeris H_2O measurement noise, MFC flow-control variability, and/or analog setpoint-voltage stability. We added this figure to the supplement (Figure S5) and added the description text to Lines 257-262.

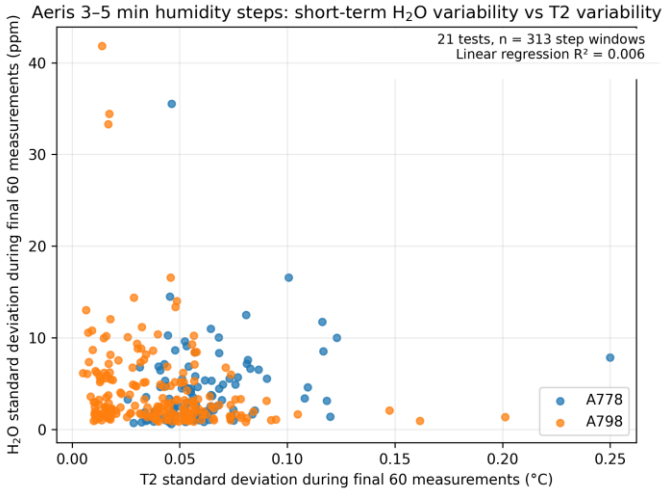


Figure S5. Standard deviation of the H_2O during the calibration steps to the standard deviation of the enclosure inlet temperature (T2) of 21 tests (313 windows) for Aeris MIRA analyzers (serial number A778 and A798). The correction coefficient (R^2) was 0.006 for these 3-5 min tests.

- The authors use the instruments themselves as water vapor monitors. Did they ever measure the same water vapor sample with two or more instruments to compare the measurements of water vapor?

The reviewer is corrected to point out that we do not have any head-to-head comparisons of our system to an existing humidity measurement (beyond that of the instruments to be calibrated). Using a sensor, such as a chilled mirror hygrometer inline would be a potentially valuable addition, particularly if the system is to be used for other applications. We added this point to the discussion (Lines 382-388). However, for our application, the key point is the value of water reported by the instrument, since, in practice, that is what we will be using in the field to correct for the remaining water vapor after using a Nafion dryer to remove most of the water. The actual value of the

water and its calibration is not crucial (unless the calibration of the water reported by the instrument changes dramatically in the time between calibrations or during calibrations of during individual steps of the calibration).

- The apparatus relies on the fact that the Nafion membranes do not allow significant quantities of the analyte gas to pass through the membrane. As far as I'm aware, the literature evidence shows small effects on CO_2 and CH_4 when there is gas on both sides of the membrane. Is there any literature on immersing the Nafion in liquid?

As for previous literature, this concept has been used a couple times that we found. Dembowski et al. (2021) used this concept (with mass flow controllers and a submerged Nafion tube) to study the role of relative humidity in determining which aluminum mineral forms under various conditions. The NOAA system in Rella et al. (2013) used a gas-permeable membrane-based water calibration in a water bath, so not a Nafion but a similar idea. We addressed the permeation concerns with results shown in the next comment. In summary, the submerged Nafion did not show any effects of permeation, but we did see a small effect for a Nafion with gas on both sides inline after the water generator. Based on this, we do not believe permeation in our water generator to be a significant concern.

References

Dembowski, M., Loring, J. S., Bowden, M. E., Reynolds, J. G., Graham, T. R., Rosso, K. M., and Pearce, C. I.: The controlling role of atmosphere in dawsonite versus gibbsite precipitation from tetrahedral aluminate species, Dalton Trans., 50, 13438–13446, https://doi.org/10.1039/d1dt02081a, 2021.

Rella, C. W., Chen, H., Andrews, A. E., Filges, A., Gerbig, C., Hatakka, J., Karion, A., Miles, N. L., Richardson, S. J., Steinbacher, M., Sweeney, C., Wastine, B., and Zellweger, C.: High accuracy measurements of dry mole fractions of carbon dioxide and methane in humid air, Atmospheric Meas. Tech., 6, 837–860, https://doi.org/10.5194/amt-6-837-2013, 2013.

- It would be very helpful if the authors performed an experiment where they humidified the calibration standard with the apparatus and then dried it, to demonstrate the accuracy of the humidifier apparatus. Without this test, there could be a bias in the humidity correction depending on the dissolved gas concentrations in the water.

Thank you for raising this issue, prompting us to do additional testing. As shown in the figure below, a calibrated reference cylinder (424.4 ppm CO_2 and 1977.2 ppb CH_4) was measured with Picarro analyzer CFKADS2025 under three configurations: directly

after the humidity generator, after the humidity generator followed by a downstream Nafion dryer, and through the Nafion dryer with the humidity generator bypassed. The external downstream Nafion dryer was a 144-inch Perma Pure Nafion tube (MD-070-144S-2), with counterflow provided by a zero cylinder (0.4 ppm CO_2, 17.6 ppb CH_4, 14 ppb CO). In (a) we see the water vapor, as measured by the Picarro during these three tests. In (b), the Generator only line indicates the CO_2 changing by up to 0.175 ppm for water vapor of 1.4%. After drying with the Nafion, there is a possible permeation effect from the external Nafion of up to 0.075 ppm, essentially the same as when only the Nafion was used (no water generator). Since the Generator plus Nafion line is essentially the same as the Nafion only line, this indicates that the Generator itself did not affect the CO_2 of the air sample. In (c), there is possibly a slight water vapor dependence for the Generator only test, but no evidence of permeation effects caused by the Generator or the external Nafion. In (d), we see a large water vapor effect for CO, and again no evidence of permeation effects. We have added this figure to the supplement (Figure S8) and text to the Results section (Lines 320-332).

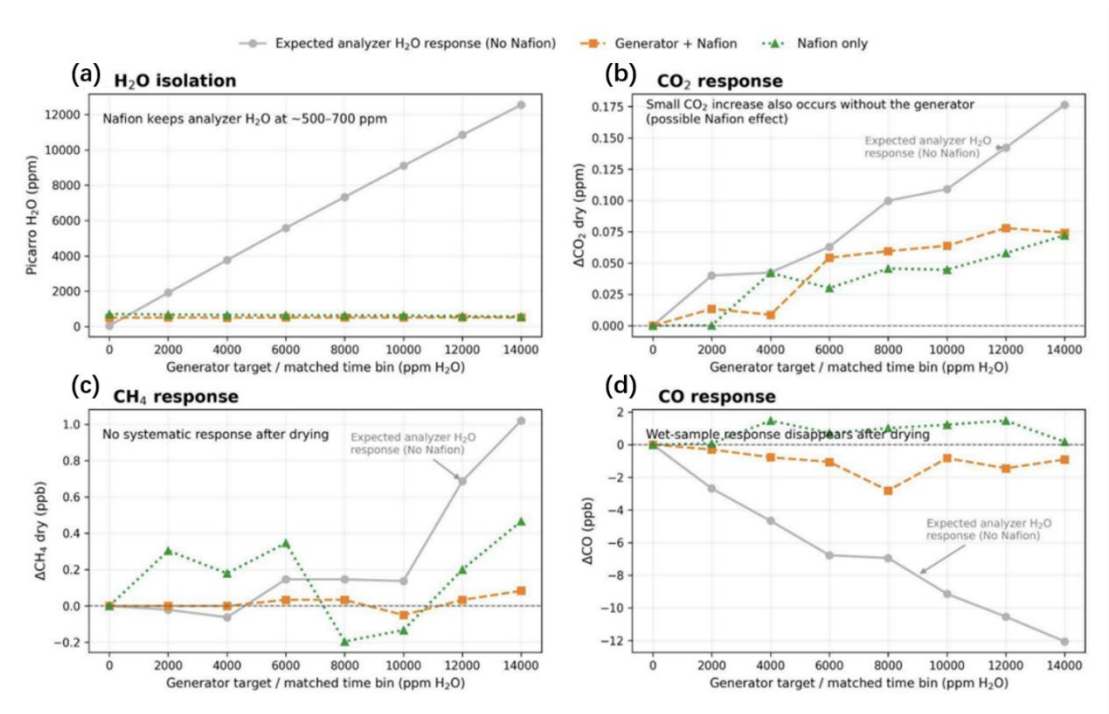


Figure S8. (a) water vapor mole fraction measured by Picarro (CFKADS2025), for three tests (calibration cylinder gas through the water generator (Generator only), calibration cylinder gas through the water generator and an external Nafion (Generator + Nafion), and calibration cylinder gas through an external Nafion (Nafion only). (b) CO_2 mole fraction (corrected to the dry mole fraction using the manufacturer correction). (c) CH_4 mole fraction (corrected to the dry mole fraction using the manufacturer correction). (d) CO mole fraction.

- The authors are inconsistent in their choice of units for water vapor concentration. Please select one and use it consistently.

Thank you for pointing this out. We have changed the text and figures to consistently use ppm H_2O.

Specific Comments

L44: please provide reference for WMO/GAW compatibility standards.

Thank you for the comment. The reference (WMO, 2025) has been added to Line 44.

References

WMO (2025): Measurement guidelines for the observations of carbon dioxide, methane, and nitrous oxide at global atmosphere watch stations, GAW Report No. 318, World Meteorological Organization, Geneva, Switzerland.

S2.1: Apparatus. This is an interesting design. It looks as though there is a possibility of condensation occurring in the wet side output lines – they may need to be heated for better robustness. I would like to see more testing just of the setup itself, with just one of the instruments and gases. Do you get the same answer when you scan the apparatus up in concentration and then back down? Does the step size matter? Do the steps need to be in order of increasing or decreasing concentration, or can they be randomized? Is there benefit to keeping the flow through water chamber fixed, or was it varied? Can an effect be seen in the resulting H_2O and or analyte concentrations?

We thank the reviewer for these thoughtful suggestions regarding the robustness and operation of the P-WAVES apparatus. We agree that the response of the system to different water-vapor concentration sequences and step durations is important for evaluating its practical operation. We found that 3-5 min step produced very similar results, indicating that this duration was sufficient for the system and analyzer response to stabilize for the conditions tested. We also found that increasing water-vapor concentrations was preferable to decreasing concentrations. This behavior is likely related to the relatively slow removal of water from the Nafion tubing and other wetted surfaces ("stickiness" of water). When the water-vapor concentration was decreased, a longer equilibration period was required to reach the new water-vapor level. Consequently, concentration steps in the decreasing direction would require longer durations than 3-5 min steps used in our standard procedure.

We did not systematically randomize the concentration sequence because the same water-retention behavior makes randomized sequences difficult to implement efficiently: a transition from a high to a low water-vapor concentration can require substantially longer equilibration times. Thus, for routine measurements, we used an increasing sequence of water-vapor concentrations, which provided more rapid and

reproducible equilibration. A randomized sequence could be used in principle, but longer equilibration periods would be required to ensure that the previous water-vapor condition did not influence the subsequent measurement.

During the experiments, the flow was kept constant at 160 sccm. We chose a fixed flow rate to maintain consistent operating conditions among concentration steps and experiments. The system could potentially accommodate a higher flow rate, but we did not systematically investigate the effect of flow rate on the generated humidity or analyzer response in this study.

Regarding condensation in the wet-side output lines, we recognize that condensation could potentially occur if the gas temperature falls below its dew point, particularly at high water-vapor concentrations. The present experiments were conducted without active heating of the output lines, and our results did not indicate an obvious condensation-related artifact under the conditions tested. Nevertheless, we agree that heating the wet-side lines could improve the robustness of the system. Particularly for operation at high humidity, large temperature gradients, or longer sampling lines. We have added this consideration to the discussion of the apparatus limitations (Lines 376-381).

L132: it would be helpful to describe how the offset was included in Eq. 1

We have clarified how the offset correction was incorporated into Eq. 1. The text "The offset was determined from the difference between the known values of the cylinder for each species and the uncorrected (wet) mole fractions measured by the analyzers at the lowest water vapor level." was added to Lines 164-165.

L153-L155: the selection of RS is not a particularly helpful statistical choice. For example, for a highly linear dependence (which would presumably be a desirable outcome), the RS of the quadratic term will tend to infinity. Perhaps just estimating the error in the correction from each term at a couple different water vapor concentration levels would be more clear and helpful, or reporting the std dev of each coefficient (not normalized by the mean)

Thank you for the comment. We agree that the relative stability (RS) of the quadratic term can become very large when the underlying response is highly linear. We removed RS from the revised manuscript and instead report the standard deviation of each fitted coefficient across repeated measurements. We have updated Table 1 and 2, and revised the corresponding text.

Figure 3: (also applies to Fig 5,6,7): Putting the wet/dry ratio and the residuals in the same panel is very unclear. Please split these into two panels, where the top panel is the wet/dry ratio and fit, and the bottom panel is the residual of the fit. The symbols in

general are quite small and difficult to read.

Thank you for the comment. Figure 3 and 5-7 were updated as the reviewer suggested.

Figure 3: it would be nice to show the time series associated with one of these plots.

Thank you for the comment. In the revised manuscript, the updated Fig.2 (a) shows the time series plot of one of the saturation tests performed with the Aeris MIRA A778. We believe this addition provides a clearer view of the temporal behavior of the measurements.

Fig 3: There are two sets of tests, one from 0-3% water, and one from 0-2500 ppm water. It is confusing that the polynomial fit over the wide range also includes the low range. If the authors are suggesting a piecewise fit, then I'd like to see clearly how they operate in the transition region. Otherwise, it's difficult to know what correction to use in which situation.

Thank you for raising this issue and its potential confusion. We did not intend to suggest a piecewise fit. We suggest using the low water vapor fit if the sample air is dried (partially – there are still water vapor effects) and for the best compatibility using this approach for measuring greenhouse gases and associated tracers. If, for some reason drying is not desirable or practical, the correction to a wide range of water vapor values can be used. The analysis over a wide range also shows the expected errors if the manufacturer-corrected mole fractions are utilized, so that users can assess what works best for their application.

L276: The authors note several limitations of this study. Of them, I think that the short period of time, and the small number of tests, is the most consequential, and would in principle be the easiest to remedy with minimal extra effort. For example, I would love to see repeated tests over a period of 1-4 weeks, showing e.g. hourly measurements of the wet-dry curve for one instrument and one gas, showing setup repeatability, drift over time, as water levels drop.

Thank you for the comment. As stated for the previous comment about the repeatability and drift over time. We have added the analysis to Lines 264-271 and Figure S6 to the supplement.

The water reservoir contains approximately 130 ml of water and each test to saturation uses about 0.1 ml. Thus 50 tests to saturation thus use about 5% of the total volume of water so the reservoir level is essentially constant on the timescale or number of tests used here. We have added these details to the manuscript (Lines 99-101).