Reply on RC1:

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: Evaluating the overall quality of the preprint

Overall, the preprint is well-written and easy to read. A technical study assessing the long term, tower-based capabilities of an Aeris MIRA instrument for methane and ethane measurements is a valuable contribution to scientific literature. Except for several areas that need more clarity, instrument characterizations appear to be scientifically sound and of relevance for future field studies. My primary concerns in this preprint are the insufficient field assessment for ethane measurements and why some instruments did not undergo all or any of the laboratory tests. Additionally, there are several areas that need more clarity in the paper. In short, I think this paper is quite interesting and captures the potential for MIRA-based methane measurements in long-term monitoring, but more work needs to be done to assess the potential capability of long-term MIRA-based ethane measurements.

Thank you for these comments and we appreciate the time you took to review this paper. The primary concerns in the general comment (field assessment of ethane measurements and not completing all laboratory tests for all instruments) were addressed in the following specific comments.

Specific Comments: Addressing individual scientific questions/issues

The introduction is easy to read with appropriate background information included. The introduction (line 61-62) and the title of the paper suggest that a large part of the novelty of this study is the laboratory and field assessment of MIRA Ultra for a long-term, tower-based network of both dry methane and ethane. The author understands that the collocated measurements of methane and ethane are important for distinguishing thermogenic methane emissions from total methane emissions. The preprint, however, does not adequately demonstrate confident tower-based ethane measurements.

Thank you for the comment. We are confident that the bias of the ethane signals within the network of Aeris CH_4/C_2H_6 systems we deployed surrounding the Denver-Jules Basin is sufficiently low on time scales relevant to inversions of tower data. We primarily base this confidence on the 10-day test with 4 instruments (Test 4), but we acknowledge that the original text did not adequately describe that the test was essentially a field deployment. This test used the same setup as the field deployment, with four completely independent Aeris systems, including field calibration cylinders. Additionally, the concurrent laboratory test had a large range of methane and ethane sampled because of a probable natural gas leak within the building. The laboratory HVAC system is not very stable and the room temperature during the test varied by about 13 °C which is similar to the temperature variability of the sheds in the field. The sentences "Due to a natural gas leak within the building, the range of methane and ethane was similar those typically measured downwind of oil and gas fields, providing a more thorough test than would background levels. The sampled methane and ethane varied from 2030 - 2378 ppb CH_4 and 0.3 - 13.0 ppb C_2H_6 in the laboratory during this 10-day (15-25 November 2024) test.

Room temperatures in the laboratory varied by 13 °C throughout the test." were added to Line 221-227 (Section 2.5.1). Therefore, the concurrent laboratory test result reflects the instruments' performance in the field. To emphasize that this test was undertaken in conditions similar to the field, with the systems set up exactly as in the deployment, the concurrent laboratory test was renamed as the pre-deployment test and Table 1 was modified to reflect this change. We also note that the intra-network bias for both methane and ethane is the relevant quantity for our intended application of methane and ethane emissions determination via inversion. We evaluated the intra-network ethane bias by comparing the Aeris instruments to the mean (through the 10-day test in the original manuscript and the additional test described below). While we tie our ethane results to the NOAA internal scale using the calibration cylinders, that tie is secondary to internal agreement for our application.

We unfortunately do not have other continuous (measurement frequency ~<5 s) instruments with ethane measurements available for a long-term comparison in the field. We did have two Aeris instruments co-located at CAO for about one month (June 2024). These instruments have different noise levels (based on Allan-Werle deviation tests), different water responses, and they were calibrated with separate cylinders. Although it would be preferable to compare the result to a known instrument, the comparison between the two Aeris instruments is quite independent (with results as shown below). The mean difference of ethane was well within the goal of 0.3 ppb C_2H_6. The instruments were calibrated every three hours and showed a relatively higher standard deviation (i.e., if we were to re-do this knowing what we now know, we would do hourly low cylinder calibrations). The figure was added to the supplementary material. We added a sentence to the methods (Line 263 on page 10): "Additionally, we deployed two Aeris instruments with separate calibration cylinders at CAO for about one month in June 2024." We added a sentence to the results: "The mean difference of ethane for two independent Aeris systems deployed at CAO for June 2024 was 0.01 ppb C_2H_6 (Figure S2)." was added to Line 417.
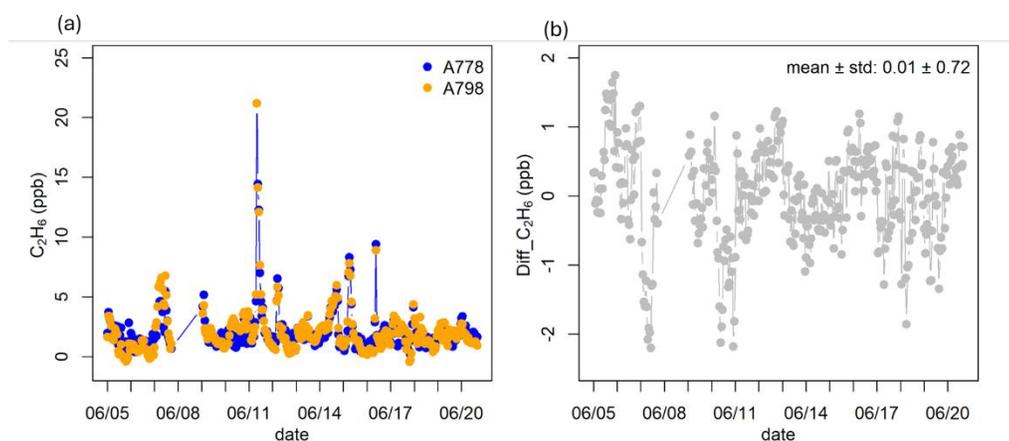

Figure S2: (a) Calibrated hourly ethane data from Aeris MIRA Ultra instruments (serial number A778 and A798) at CAO over the time period of June 2024, and (b) the difference of ethane between these two Aeris MIRA Ultra instruments.

I recommend modifying language in the intro, discussion (line 514), and/or title, so they are more reflective of the results and analysis performed.

The sentence (Line 61-62) was rewritten as "In this paper, we present the first systematic laboratory assessment of multiple Aeris MIRA Ultra instruments for CH_4 and C_2H_6 measurements and a long-term, tower-based comparison of CH_4 mole fractions with National Oceanic and Atmospheric Administration (NOAA) Picarro measurements."

The sentence (Line 513-515) was rewritten as "The system shows the potential to quantify the ratio between anthropogenic and biogenic methane sources, for regions with mean enhancements of greater than 30 ppb CH_4 and 3 ppb C_2H_6, by providing continuous ethane measurements, assuming the ethane to methane ratio of the sources is known."

Authors perform technical and practical analysis, addressing field-based concerns such as time synchronization (which is problematic with the Aeris instrument) and cold/warm start delays. The preprint can benefit with some clarity on their methodology/results in the following areas:

- **Section 2.2: bias and precision goals:** The primary take away in this section is the reference precision goal, however, it is unclear how the authors derive 3ppb as the methane goal. < 10% of the typical enhancement should be 4ppm (based on line 488). Overall, this section can be written more clearly as well as concisely.

To clarify Section 2.2, we rewrote it as "We set bias (defined as the long-term mean deviation from the true value) and precision (defined as the standard deviation of hourly differences) guidelines for CH_4 and C_2H_6. In general, network compatibility goals are the maximum instrument biases that can be accepted without adversely affecting model interpretation of gradients (GAW Report No. 292, 2022), whereas the total uncertainty contains both bias and random noise. Summertime tower CH_4 enhancements in the Permian Basin averaged around 60 ppb, while wintertime enhancements averaged around 200 ppb (Monteiro et al., 2022). Considering the lower emissions with CH_4 enhancements around 40 ppb in the Denver-Julesburg Basin, a desire for bias less than 10% of typical network enhancements any time of the year, and instrument capabilities, we adopted a more conservative bias goal of 3 ppb CH_4. The average C_2H_6 to CH_4 ratio is about 5%-10% including all biogenic and thermometric methane sources at the Denver-Julesburg Basin area (Daley et al., 2025). Considering also instrument capabilities, we set the corresponding bias guideline for C_2H_6 at 0.3 ppb. We set precision guidelines for the hourly differences (3 ppb CH_4 and 0.3 ppb C_2H_6) to limit the deviations that may impact emissions from inversion modelling on shorter timescales, e.g., weekly. Future instrument improvements allowing further reductions in bias and noise would be beneficial, particularly for regions with lower emissions."

- **Why some instruments were chosen for tests/ studies:** An explanation should be provided why the authors did not perform all tests described in Table 1 for all instruments. The authors state they have 8 unique instruments used in the study, but Table 1 only describes 5 unique serial numbers (665, 792, 800, 886, and 778). If some instruments were upgraded by the manufacturer

and returned with the same serial number, those instruments can be referenced as A792.v1 and A792.v2 or another shorthand to make it easier to follow.

It's worth noting that instruments A792 (used in 4 tests) and A665 (used in 3 tests and field deployment) were the most tested instruments and 792 was the sole instrument involved in determining *Uncertainty due to instrument noise, cylinder calibration, and ethane cylinder assignment uncertainty,* meanwhile A665 and A792 showed the greatest sensitivity to water vapor (lines 275-278) and the most unrealist deviations (lines 382-384). During the field study why would the authors choose to use A778 (which was involved in no other Table 1 tests) and A665 that had the most unrealistic deviations and the greatest sensitivity to water vapor? Given that your results often show instrument dependent characteristics, how do you justify using test results from other instruments to perform the field assessments.

We agree that the original text indicating 8 unique instruments was confusing. Thank you for pointing this out. We changed the text to read (Line 74-75), "As the manufacturer upgraded the configurations for this instrument throughout 2024, we assessed and used two versions of the Aeris MIRA Ultra, including up to four instruments concurrently." In fact, we did do initial testing of 8 Version 1 instruments, but 4 of them failed those initial tests in that they were extremely noisy. We returned these Version 1 instruments to the manufacturer without further testing. The instruments were then upgraded by the manufacturer and returned under the same serial number. Version 1 is no longer commercially available, so this study focused more on Version 2 (as mentioned at Line 79-Line 80). We describe further our reasoning for the various tests below.

Serial number A778 (Version 1) used for cylinder calibration and ethane cylinder assignment uncertainty. We chose this instrument because it exhibited the least noise of the available instruments at the time. Rather than a test of the instrument itself, the goal was to obtain the most accurate calibrations for the field cylinders possible. To clarify, the sentence "The goal of this test was to obtain the most accurate calibrations for the field cylinders possible, so we chose the serial number A778 as it exhibited the least noise of the available instruments at the time." was added to Line 154-156.

We agree with the reviewer that using a single instrument to assess instrument noise introduces limitations. For the uncertainty due to instrument noise test, it was expensive in terms of both calibration gas and time to sample a cylinder for multiple days. We selected instrument A792 to test the instrument noise because its performance represents a high-end estimate among the deployed instruments. It was mentioned at Line 145-Line 147 that "Instrument A792 was selected for this test because its Allan-Werle deviation is a high-end estimate of the group (i.e., neither the lowest nor the highest; see Section 3.2)." We added another three Aeris MIRA Ultra instruments' Allan-Werle deviation test results (with two hours of sampling each) and updated Table 1 accordingly.

Methane

Ethane
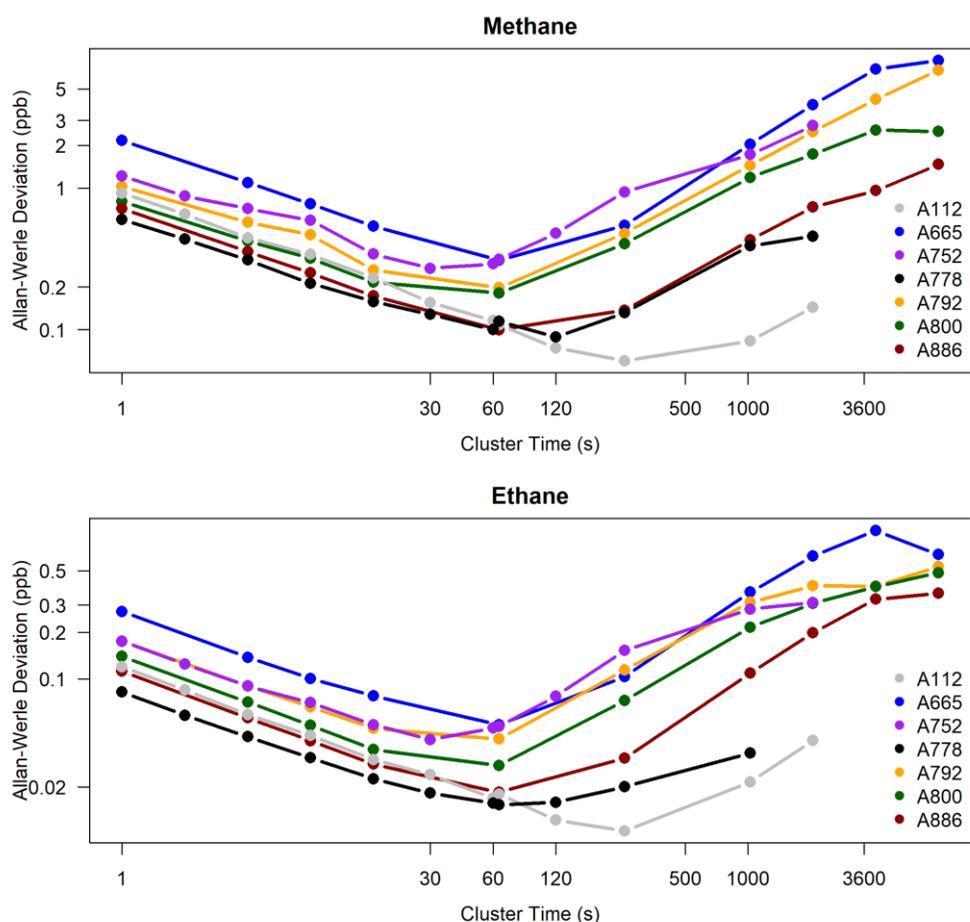
Given that CAO is the only tower equipped with a co-located Picarro instrument, and that the serial number A665 showed the largest water vapor sensitivity and the most unrealistic deviations among the instruments, we co-located the serial number A665 with the Picarro to rigorously assess the field performance of the Aeris MIRA Ultra methane measurements. The other instruments, which exhibited more stable behavior, should perform better than A665. The sentences "Since the serial number A665 showed the largest water vapor sensitivity and the most unrealistic deviations among the instruments, we co-located it with the Picarro to rigorously assess the field performance of the Aeris MIRA Ultra methane measurements. The other instruments, which exhibited more stable behavior, should perform better than A665." were added to Line 250.

As stated in the Discussion section, we do recommend Allan-Werle Deviation tests (sampling a cylinder for at least two hours) and water vapor testing for each instrument. The sentence "For future applications, it would be advantageous to quantify instrument-specific noise and water vapor related uncertainty for each individual instrument, particularly if concurrent testing of multiple instruments is not practical." was added to Line 485.

- **Calibration cylinder usage and description:** Line 160-162 is confusing. It reads as though there are 17 cylinders and each calibration cycle includes 4 min for each cylinder, however, the

paper says this is a 16-minute process. Does that mean only 4 cylinders are used? This needs to be clarified. Why was the test repeated for 8-16 hours? That's a large range.

The sentence was rewritten as "Each calibration cycle included 4 min for each of three calibrated cylinders and one unknown cylinder (totaling 16 min), with the remainder of the half hour (14 min) sampling room air".

The large range (8 to 16 hours) was primarily for practical purposes, as the tests were completed during the workday (8 hours) and overnight (16 hours) in order to facilitate calibrating all of the field cylinders in a timely manner.

In test three, why did you include all calibration cycles (line 166), when the cylinders did not stabilize for the first few hours? Shouldn't the 'NOAA C2H6' column also include the cylinder assignment error?
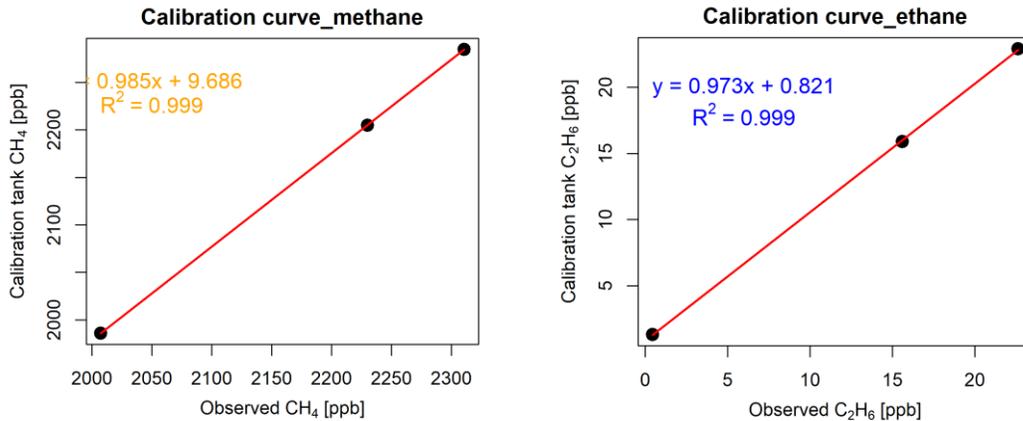
It was mentioned at Line 165-Line 166 that "The initial instability occurred for all cylinders and the cause is unknown." And we also added that "Because the cause and period of instability remain unknown and no clear threshold for exclusion could be established, we included all the calibration cycles in Test 3 for transparency and reproducibility. Excluding a consistent number of cycles as warm-up for each cylinder did not appreciably affect the final calibrated cylinder values." to Line 166.

Regarding the NOAA C_2H_6 column, the ethane mole fractions used in our analysis were those assigned by the NOAA CCL for the tertiary standards. NOAA does not provide an additional "cylinder assignment error" for these values, so no such uncertainty term could be included.

I noticed that your NOAA tertiary standards range from 1985.9-2284.7ppb methane and 1.3 and 22.9 ppb ethane. In test four, the maximum methane values exceeded 2350 ppb and the minimum ethane values were below 0.5 ppb (line 222). Do you generally trust the instrument's response outside the calibrated range, and have you checked how often your field measurements fall outside of the range?

It was mentioned at Line 171-Line 174 that "Other factors such as equilibration after gas switching, nonlinearity of the true calibration curve, and changes in the calibration slope on time scales shorter than the time between full calibration cycles (tested in Section 3.6.1) are assumed to be small and not included in the uncertainty estimate." Based on our laboratory analysis (Test 3), all the calibration curves were well described by linear regressions with R^2 values above 0.99 (an example of methane and ethane calibration curves is shown below). This strong linearity supports the assumption that the instrument maintains a linear response slightly outside the calibration range. Therefore, we trust the instrument's response outside the calibration range, but it is true that field cylinder assignment errors can lead to systematic errors outside the calibration range. We added "All the calibration curves showed a strong linearity with R^2 values above 0.99. Extrapolation errors outside of the field calibration range because of non-linearity are thus not likely to be significant. In the field, with only two cylinders utilized to determine slope, field cylinder assignment errors lead to systematically increasing errors outside the range of the field cylinders. For example, if the low cylinder (1 ppb C_2H_6) at a field site is assigned correctly, but the high cylinder at 24 ppb is assigned a value 0.2 ppb C_2H_6 from its true value, the error of an atmospheric sample at 36 ppb is 0.3 ppb." was added to Line 354 (Section 3.4). In practice, peaks

outside the calibration range do occur, but these are generally isolated and primarily occur in stable conditions at night.

Calibration curve_methane

$0.985x + 9.686$
$R^2 = 0.999$

(y-axis: Calibration tank CH_4 [ppb]; x-axis: Observed CH_4 [ppb])

Calibration curve_ethane

$y = 0.973x + 0.821$
$R^2 = 0.999$

(y-axis: Calibration tank C_2H_6 [ppb]; x-axis: Observed C_2H_6 [ppb])

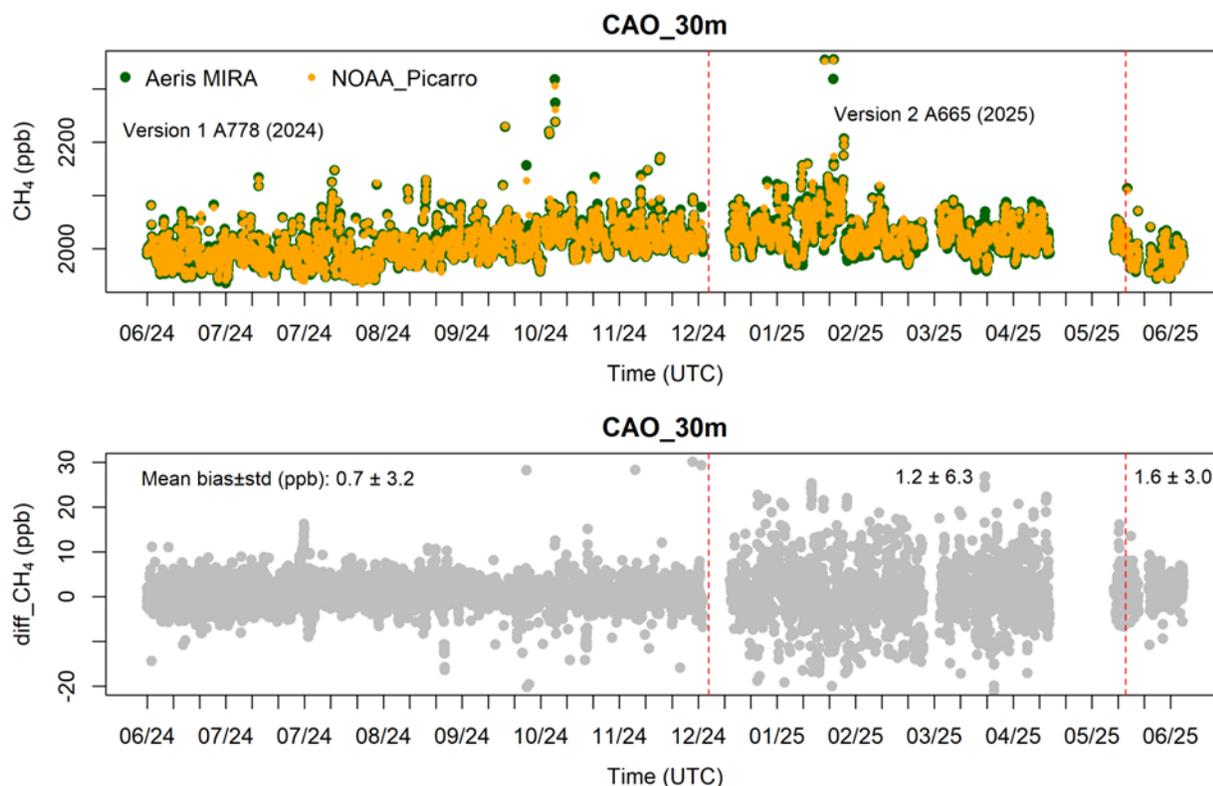- **Water Vapor corrections:** What does a perfect water correction mean (135-136)?

The sentence was rewritten as "If the water correction was correctly described by a stable relationship in the manufacturer software, the reported CH_4 and C_2H_6 dry mole fractions would not depend on the water vapor level of the resulting air."

- **Field Deployment/Design:** Is a 20-minute lag time common for these sorts of measurements? I don't think it is correct to assume large variations in Aeris or Picarro measurements are from mismatched timing alone (lines 258-260). The author needs to add more support on why and how they chose to eliminate time series data points when either the Picarro or Aeris standard deviation exceeds the fiftieth percentiles. What is the local time period you are running these standard deviations? Can you site studies that employ a comparable methodology- it just seems a bit arbitrary?

For our application, the overall impact is minor as we use hourly averaged data for the flux calculations. This timing shift was taken into account when we archived the data for all the sites. We have shown that this works for these types of measurements in previous studies (Richardson et al., 2012 and 2017). The lag time in the tubing could be improved by using an extra pump to increase the flow rate, but instead we use a single pump to simplify the field operation and maintenance.

Times with large standard deviations were screened out to minimize noise caused by mismatches in air sampling. It is reasonable to exclude highly variable hours for the comparison (Levin et al., 2020; Richardson et al., 2017; Miles et al., 2018), but we agree that 50th percentile was needlessly aggressive. The methane threshold of Aeris' hourly standard deviation or NOAA Picarro's standard deviation greater than the 50th percentile was about 3.5 ppb. Based on the study of Richardson et al (2017) describing the tower measurement network in/around Indianapolis, we relaxed the threshold to the in-situ hourly standard deviation of methane larger than 7 ppb to include more data. The sentences (Line 258-Line 260) were rewritten as "To minimize noise that might be caused by mismatches in timing, the hours

with large atmospheric variability were removed (Levin et al., 2020; Richardson et al., 2017; Miles et al., 2018). We used a threshold of 7 ppb CH_4 for the standard deviation within each hour above which the data for that hour were excluded (Richardson et al., 2017)." Figure 9 was updated as shown below. The standard deviations are a bit larger, but the overall message is the same.

References

Levin, I., Karstens, U., Eritt, M., Maier, F., Arnold, S., Rzesanke, D., Hammer, S., Ramonet, M., Vítková, G., Conil, S., Heliasz, M., Kubistin, D., and Lindauer, M.: A dedicated flask sampling strategy developed for Integrated Carbon Observation System (ICOS) stations based on CO_2 and CO measurements and Stochastic Time-Inverted Lagrangian Transport (STILT) footprint modelling, Atmos. Chem. Phys., 20, 11161–11180, https://doi.org/10.5194/acp-20-11161-2020, 2020.

Richardson, S. J., Miles, N. L., Davis, K. J., Crosson, E. R., Rella, C. W., and Andrews, A. E.: Field Testing of Cavity Ring-Down Spectroscopy Analyzers Measuring Carbon Dioxide and Water Vapor, https://doi.org/10.1175/JTECH-D-11-00063.1, 2012.

Richardson, S. J., Miles, N. L., Davis, K. J., Lauvaux, T., Martins, D. K., Turnbull, J. C., McKain, K., Sweeney, C., and Cambaliza, M. O. L.: Tower measurement network of in-situ CO_2, CH_4, and CO in support of the Indianapolis FLUX (INFLUX) Experiment, Elementa: Science of the Anthropocene, 5, 59, https://doi.org/10.1525/elementa.140, 2017.

Miles, N. L., Martins, D. K., Richardson, S. J., Rella, C. W., Arata, C., Lauvaux, T., Davis, K. J., Barkley, Z. R., McKain, K., and Sweeney, C.: Calibration and field testing of cavity ring-down laser spectrometers measuring CH_4, CO_2, and $\delta^{13}CH_4$ deployed on towers in the Marcellus Shale region, Atmos. Meas. Tech., 11, 1273–1295, https://doi.org/10.5194/amt-11-1273-2018, 2018.

In Figure 9, I would recommend re-evaluating the Aeris data in early May and right after the version 2 switch that looks like lines - I suspect this is unreal data. The x-axis should likely be date (DD/YY) or something more description. It looks like Picarro data has more missing datapoints, particularly during sharp peaks where the Picarro indicates a point at the maximum and the Aeris data is tracking multiple points along the enhancement peak.

Thank you for pointing out this possible point of confusion for readers. The instrument experienced a technical issue (related to the calibrating setting in a configuration file) between May 10 and June 4. The plot type used both lines and points. The figure was updated by removing the lines to connect points and the x-axis representing MM/YY.

- **Allan Deviation:** The allan test appears to be run for 40 minutes for V2 of A665, A800, and A886 and for 5 hours for A792. I am not convinced that A792 is the most representative instrument beyond 5 minutes, and I worry this study is putting too much emphasis on results from a single instrument, when there is a large variation between individual instrument performance and sensitivities.

As stated in the figure caption, we used two hours of data for most of the instruments and 50 hours of data for A792. When calculating the Allan-Werle deviation, the time series is broken into segments, the longest of which is 40 min for the instruments tested for 2 hours and 300 min for A792. Then the deviation amongst the mean of those segments is calculated. The caption was correct, but we clarified that the A792 test was 50 h. We also added another three Aeris MIRA Ultra instruments' Allan-Werle deviation test results and clarified accordingly in the caption. In terms of the 50 h test, it is expensive in terms of both calibration gas and time to sample a cylinder for such long time. We selected the instrument A792 to test the instrument noise because its performance represents a high-end estimate among the deployed instruments. It was modified at Line 145-Line 147 that "Instrument A792 was selected for this test because its Allan-Werle deviation indicated relatively high variability compared to the other instruments (see Section 3.2) so that the results represent the upper end of the expected variability range."

Overall, I think the information, tables and figures are relevant for the main text, yet I am not sure it is worth including in such detail that standard deviation reduces as averaging time increases. Lines 232-234, 385-394, and Figure 8, are not particularly novel or necessary to include in the main text.

Line 385-394 and Figure 8 were moved to the supplementary material. The sentence from Line 232-234 was removed. We also clarified that we were checking for a reduction in the noise based on the square root of the number of observations, which would indicate random noise. The noise decreases, but not as fast as would be expected for random noise.

The discussion addressed some crucial concerns about field deployment of the MIRA instrument for long-term methane and ethane measurements. Using your logic in lines 486-496, the bias threshold (section 2.2) for each tower network should be dependent on the expected methane enhancement (e.g. Indianapolis should strive for an uncertainty threshold below 0.5ppb methane) and thus make this sort of analysis not practical outside of a large metro area or are with significant O&G operations. You addressed this issue similarly in lines 489-492. My main concern with your discussion is that you say "the [Aeris MIRA Ultra measurement] system shows promise for distinguishing among multiple methane sources by providing continuous ethane measurements, depending on the magnitudes of methane and ethane emissions". While showing promise is vague and the second part of your sentence creates a wide caveat, I think this is too strong of a statement for the lack of ethane results during the field deployment.

We agree that the ability to quantify regional emissions with these instruments does depend on the magnitude of the enhancements. The sentence (Line 513-515) was rewritten as "The system shows the potential to quantify the ratio between anthropogenic and biogenic methane sources, for regions with mean enhancements of greater than 30 ppb CH_4 and 3 ppb C_2H_6, by providing continuous ethane measurements, assuming the ethane to methane ratio of the sources is known." We are confident, based on the 10-day pre-deployment test (reworded from laboratory test to emphasize that the conditions were not particularly ideal and the systems were completely independent), that the bias of the methane/ethane signals within the DJ Basin network is sufficiently low on time scales relevant to inversions of tower data. As stated above, we also added an in-situ to in-situ comparison (using two completely independent Aeris instruments deployed at CAO) to the supplement, indicating mean bias of 0.01 ppb. The Aeris systems are quite noisy, particularly for ethane, so shorter time frames are problematic unless some improvements are made to the Aeris instruments themselves.

It's a good point that we implied a requirement of 0.5 ppb CH_4 uncertainty threshold for Indianapolis. The 5 ppb quoted in the text came from a site that is rural which isn't the most relevant for this situation, so we changed the text to refer to "11-21 ppb for downwind urban sites, depending on the site" at Line 490. Our CH_4 compatibility for INFLUX using Picarros is about 1.0 ppb (Richardson et al., 2017).

We also reworded Section 2.2 where we defined the compatibility goals based on the previous comments of this reviewer and comments of Reviewer 2.

Technical Corrections: typos, etc.

- Line 19-31: Myhre et al., 2013 is an outdated source. I would recommend referencing the latest IPCC report.

The latest IPCC, 2021 report was added to Line 31 and the section of references.

- "also" used twice in the last sentence of section 2.2 (line 121-123)

One "also" was removed.

- Subscripts for methane and ethane in table 2 display in the midline of text and uncertainty terms (like Ut) do not have subscripts ("Ut")

The subscripts for methane and ethane and U_t were modified in the table.

- Remove or replace "obviously" line 191

"obviously" was removed.

- Figure 2: "Latitude" y axis label is off center; caption and legend should clarify what the Oil & Gas data points are indicating. I would assume they are active sites during the time of study but I'm not sure. On the tower illustrations it would be helpful if you added the approximate location of the picaro and the Aeris from the inlet line.

Thank you for pointing out these issues. Figure 2 and the caption were updated. The Oil & Gas data points are active sites during the time of study (https://ecmc.state.co.us/dashboard.html#/dashboard) and we added this reference. We replaced the map with one with cities to give more context for the locations. The sentence "The Picarro and the Aeris instruments were located in a sea container next to the tower." was added to Line 250.

- Figure 5 and 6: subplots e and f should have the same x-label; Additionally, descriptions of subplots c and d can be worded clearer

Figure 5, Figure 6 were updated. The caption for (c-d) was rewritten as "(c) difference between linear interpolation values using 5 h calibration cycles and the observed values. (d) difference between linear interpolation values using 1 h calibration cycles and the observed values."

- Figure 7: This figure would look more appealing if the zero horizontal line were aligned between subplot A and B. Thus, plot A y axis would range from 0+/-x and plot B y axis would range 0+/-z.

Figure 7 was updated.

- Table 3: what does "typical" mean. In the top half of the table ()* means noise and in the bottom half of the table () means precisions. If that is not correct, please change symbols to be clearer.

"Typical" here means the system was working properly. "Typical" was removed from the table. Your description is correct.

Reply on RC2:

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).

This study represents a good first step in attempts to characterize the AERIS instrument performance for methane and ethane concentration measurements. The ultimate goal of this study is to assess the suitability for a suite of 4 AERIS instruments to be employed in a regional tower based flux study over the Denver-Julesburg Basin (DJB) as shown in Fig. 2. **This reviewer recommends publication of the methane results after major revisions, as discussed below. The ethane results, however, are more problematic and will require additional data and validations. Instead I would bill the AERIS ethane measurements as a work in progress requiring further validation.** This study rightfully places a large emphasis on the interference due to water vapor, and more specifically on how differences in water vapor between calibration and measurement cycles can result in biases. The 10-day laboratory comparisons of the retrieved AERIS methane concentrations from all 4 tested instruments with a Picarro instrument (Fig. 7) presents convincing evidence for the long term AERIS accuracy. The uncertainty due to errors in the water vapor correction between measurement and calibration in Section 3.1.2 is reasonable and in rough agreement with my simulations below based upon the known specific wavelengths employed. Here I am assuming a sample cell pathlength of 15-m (should approximate the AERIS pathlength to within 25%), Voigt line shapes employing the stated sampling cell pressure and temperature. The red profile simulates the absorbance for 2 ppm methane, the blue profile simulates the absorbance for 1 ppb ethane, and the black profile simulates the absorbance for 500 ppm water vapor (the upper limit claimed for the difference in water vapor between sample and calibration). The shown methane and ethane absorbances are for 0 water and the actual spectra will reside on top of the water feature. The indicated water interferences are based on the ratio of peak absorbances (assuming that's what the AERIS software uses). If the AERIS software uses the integrated absorbances, these water vapor interferences may be slightly different. Assuming the former, this simulation shows a water interference of 0.197 ppb for ethane, which is not too different than your 0.15 ppb value. The interference on the red methane very much depends on which of the 3 lines that are employed by the AERIS software. The average interference using all 3 lines is approximately 3 ppb, which matches your upper range. The above simulations reveal a number of important features. First the absorbance (-lntransmission) for 1 ppb ethane is approximately 3 orders of magnitude weaker than ambient methane levels around 2 ppm. Secondly, the water vapor correction will dramatically be affected by the Nafion scrubber performance. Even small changes in this performance will adversely affect the measurement accuracy. Thirdly, as can be seen, the water vapor tail around the ethane line is relatively flat but the features around the 3 methane lines are steeply sloping. Hence, the water vapor interference on methane will be more dramatic than ethane, as borne out by your Fig 3 plots. **This reviewer thus strongly recommends that water vapor concentrations need to be retrieved along with the methane and**

ethane values to assess in real time the Nafion scrubber performance in order to effect corrections. To the degree possible, this reviewer also recommends a close examination of the software fitting algorithms and parameters employed. Specifically, are the methane results based on one, two, or all three lines, and what are the absorption parameters employed (line positions, absorption and broadening coefficients, etc.)?
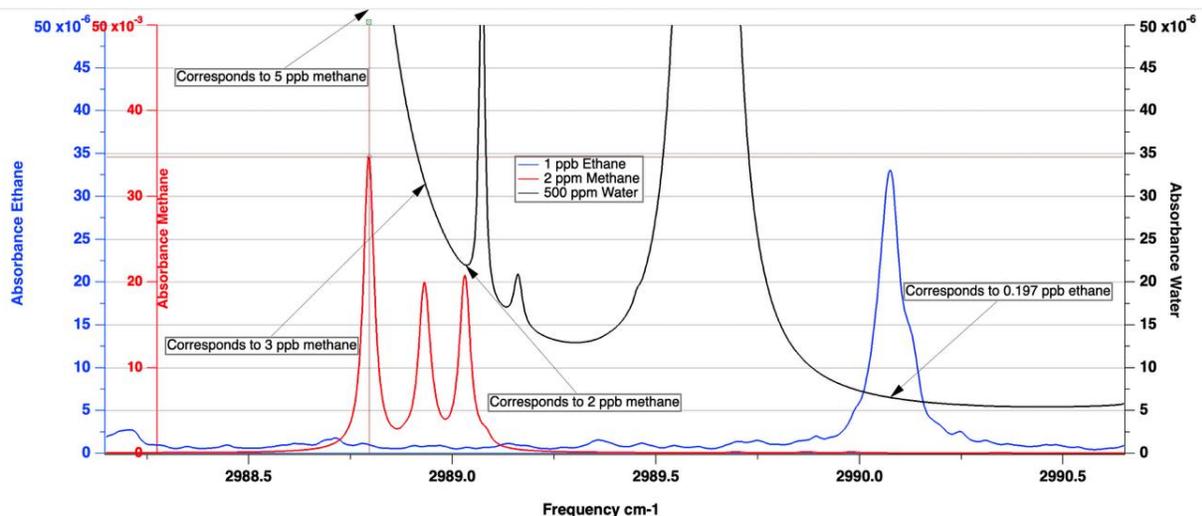

Figure Caption: Absorbance simulations for a 15 m pathlength, sampling cell pressure of 240 m band a temperature of 42oC. HITRAN 2020 absorption parameters of methane and water were employed in these simulations with Harrison ethane absorption parameters (not incorrect HITRAN values).

Thank you for the general comments and the detailed analysis. We really appreciate it. It's great to see that the water vapor effects we see are in the expected range based on your analysis. In fact, water vapor concentration was retrieved along with the methane and ethane values by the Aeris instrument. We added a phrase in Line 103 to clarify that point, "Typical differences between the dried sample and the humidified calibration gas, as measured by the Aeris instrument, were 100 - 500 ppm H_2O." We do have real-time water vapor concentrations used in this study to assess the Nafion scrubber performance and we assess the uncertainty related to the water vapor for each hour based on these values.

The software fitting algorithms are not open source, but we did get the following information from the manufacturer that "All three of the methane peaks are fitted and used to calculate the end concentration, they are weighed equally." We added the sentence, "In the wavelength range of the instrument, corresponding to frequencies of 2989.2 to 2986.4 µm, there are three methane peaks, and all are weighted equally to determine the mole fraction." to Line 73.

Specific Comments:

The results of this paper are hard to follow given the description of each test and the separation of each resulting figure. I would move each resulting figure right after the test description.

It was mentioned in the other reviewer's comment that "Overall, the preprint is well-written and easy to read." Both ways of paper organization have advantages and disadvantages, and we would like to keep the traditional paper structure (separate method and result sections). We do have thorough figure captions designed to help the reader.

In section 2.3.1 (Multi-hour cylinder tests 1 & 2), this reviewer highly recommends placing Figure 3 right after line 143. This more naturally follows the previous discussion. If possible, I would then expand the most important part of this plot (the low end with water vapor levels in the 100 -500 ppm range) as an inset to better assess the results when the Nafion drier(s) are working properly. Also, keeping in mind that these 4 instruments most likely are intended to be employed at the 4 tower sites, the authors need to explain the significant differences in the behavior of the 4 instruments, particularly at the low end.

Thank you for this comment. It's a good point to focus on the range of water vapor we see in the dried systems, although that value is typically 500-2000 ppm. 100-500 ppm is the *difference* between the dried sample gas and the humidified calibration gas. After zooming in the figure as shown below, there are limited points and the correlation between water vapor and methane/ethane is low. We feel that the important range (up to 2000 ppm (0.2%) is already visible in Fig. 3 and thus did not change the figure. It is interesting that the instruments exhibit more noise at lower water vapor levels (this can be seen in Fig. 3 in the paper), even though the line used to lock switches to methane for Version 2, according to Aeris. Note that we could not separate the impact of instrument noise and water vapor for ethane. We use an uncertainty of 0.03 ppb C_2H_6 for each 100 ppm difference of H_2O, but could be just noise. In Section 3.1.2, we stated in the original manuscript that "For ethane, it is possible that the errors (Fig. 3b) result from instrument noise rather than water vapor effects, but the assigned uncertainty is an effort to capture the range of the possible effect." For methane, the water dependence is more clear, and it does vary by instrument. In the discussion, we recommend at minimum 2-h cylinder tests to calculate the Allan-Werle deviation, water dependence tests for each instrument prior to deployment (at Line 484 and 501). If the water dependence is above the uncertainty specified here, an instrument-specific correction should be applied.

We agree with the reviewer that the water vapor sensitivity of ethane needs more validation at the low end. The sentence that "Future efforts should be focused on testing the water vapor impact on methane and ethane for the water vapor range from 100 to 2000 ppm, which is the typical water vapor range when using a Nafion drier as described here." was added to Line 486 on Page 21.
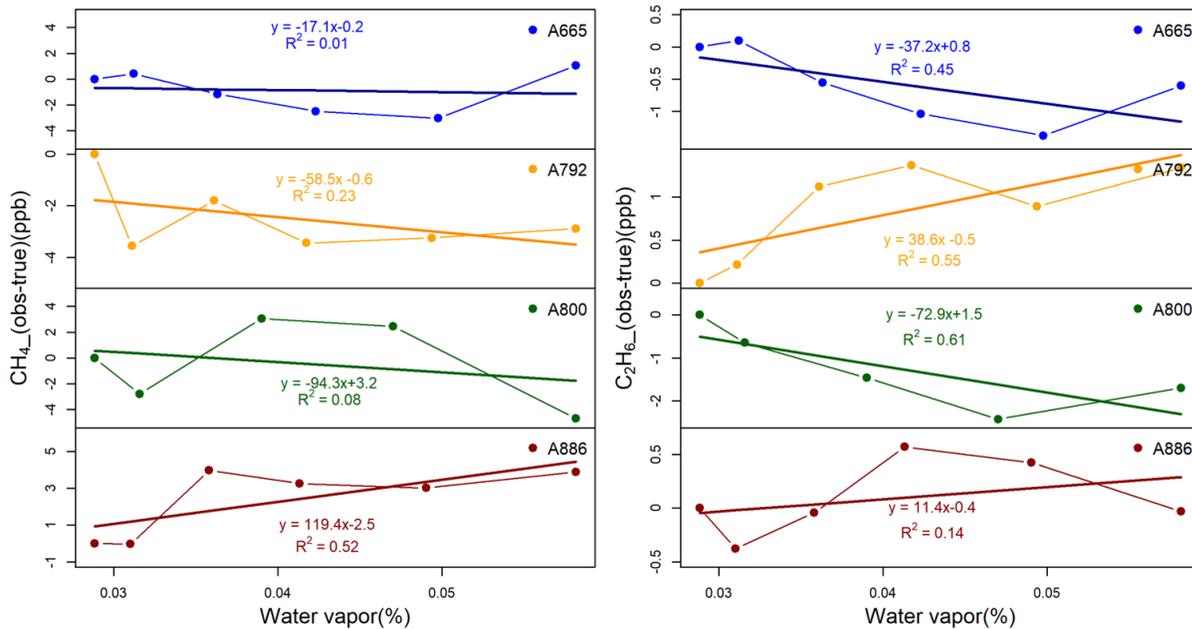
After zeroing the results in Fig. 3, do the results shown reflect the retrieved values using the AERIS software? This needs clarification.

After zeroing the results, the results shown in Fig. 3 still reflect the retrieved values using the AERIS software. We just subtracted a constant value from the data, simulating an offset calibration at the lowest water vapor value. Otherwise, it's difficult to compare the results from the separate instruments. In the original Fig. 3 caption we stated, "Note that without applying any zero or span using the Aeris software, the errors were up to 190 ppb CH_4 and 15 ppb C_2H_6." As is apparent, without calibration, the values vary wildly.

Can corrections to this software be implemented to minimize these differences?

Yes. It was mentioned at Line 275-276 in the original manuscript that "The Aeris instruments, using the manufacturer-supplied water vapor correction, showed substantial sensitivity to water vapor for measurements of methane and this sensitivity was not consistent between individual instruments." The water vapor corrections to the post-processing software could be implemented to minimize these differences among instruments. At Line 483-485 in the original manuscript, we stated that, "Ideally the water vapor response would be tested multiple times and an instrument-specific water vapor correction applied if necessary, and repeated tests performed on a regular schedule as it is possible that the true water vapor correction of each instrument changes in time." Repeated tests, however, have indicated variability in the methane error to water vapor slope, complicating the implementation of instrument-specific water vapor corrections. For example, subsequent measurements of the slope for A112 varies from none to +0.9 ppm CH_4/100 ppm H_2O.

We added to Line 485-490 that "Future efforts should be focused on testing the water vapor impact on methane and ethane for the water vapor range from 100 to 2000 ppm, which is the typical water vapor

range when using a Nafion drier as described here. For future applications, it would be advantageous to quantify instrument-specific noise and water vapor related uncertainty for each individual instrument, and essential if multiple instruments cannot be tested concurrently prior deployment or following any software upgrades."

How are the water vapor concentrations being measured (dew point hydrometer)?

We used the Aeris instrument to measure water vapor concentration along with the methane and ethane values. We updated Section 2.1 (Line 103) to indicate, "Typical differences between the dried sample and the humidified calibration gas, **as measured by the Aeris instrument**, were 100 - 500 ppm H_2O. "

Right after line 151, I would move the Allan deviation plots of Fig. 4 here to go with the discussion. In all cases where you use the term "Allan deviation", please replace with "Allan-Werle deviation since Petter Werle's pioneering work first employed this approach to atmospheric measurements and this modified term has been adopted by the atmospheric community. Although the authors are correct in stating that the exact calibration values are not important here, only the stability of the sampled mixture, you then cannot go onto use the Y-axis Allan-deviation concentrations in further discussions of the instrument concentration precisions. The figure caption here indicates methane and ethane concentrations of about 1980 ppb and 1 ppb, respectively. Perhaps you can modify your concentration statement to indicate approximate values for the retrieved precisions to within x%.

"Allan deviation" was replaced with "Allan-Werle deviation" in the paper.

The sentence at Line 302-306 was rewritten as "In the calculated Allan-Werle deviations of the instruments, there are significant differences in performance amongst the instruments for both CH_4 and C_2H_6, with the observed Allan-Werle deviation at 1 Hz ranging from 0.7 ppb to 2.1 ppb for methane and from 0.1 ppb to 0.3 ppb for ethane across the four Version 2 instruments (Fig. 4). The Allan-Werle deviation was minimized (0.1 - 0.3 ppb CH_4 and 0.02 - 0.04 ppb C_2H_6, depending on instrument) for averaging times 30 s to 2 min, the optimal averaging time for calibration cycles."

"For all the instruments, the retrieved Allan-Werle Deviation (at 1s) is about 0.03-0.1% CH_4 and 10-30% C_2H_6." was added to Line 314 in the figure caption.

In keeping with my recommended reorganization, I would then move Fig 5 and Fig. 6 right after the discussion of Test 2 on line 151. The determination of the methane bias in both the lab and the field using simultaneous Picarro methane measurements provides a nice convincing determination. However, in the case of ethane (starting on line 230), is it valid to use average of all 4 instruments as an indicator of bias? I understand that you are trying to assess differences.

It would be ideal to have a continuous (measurement frequency ~<5 s) instrument with ethane measurements as a reference to comparing the ethane results, but we did not have one available. We do note that the comparison among the four Aeris instruments is quite independent, using completely

separate plumbing and calibration cylinders. It is the difference amongst the measurements that is important for the atmospheric inversion. We added a sentence to the methods (Line 232), "For the primary intended application of the network of determining regional emissions, intra-network compatibility is the primary concern as only enhancements are utilized, justifying comparison to the mean."

However, the actual retrieved ethane emission flux in your tower network will also depend upon the actual absolute bias amongst the 4 instruments.

The ethane emission estimated from the tower network would not depend upon the absolute bias in the 4 instruments, compared to a WMO scale, for example. When doing an atmospheric inversion, we subtract the background value (measured at the upwind site) and use the enhancement for the flux calculation, so it's only the difference amongst the instruments that is important for quantifying the regional flux.

It's unfortunate that you could not have utilized the NOAA flask ethane sampling measurements at 478 m AGL in providing a direct bias determination in the field for select time periods. Can any of these data be utilized with some type of height correction for such validation?

It is unfortunate that logistical concerns prevented us from sampling 478 m AGL with the Aeris. To evaluate the feasibility of comparing flask ethane data at 478 m AGL to the Aeris data at 30 m AGL, we gathered methane flask data (at 478m AGL) and compared it to the Aeris CH_4 data (which we know agrees well with the NOAA Picarro based on Fig. 8). We do not currently have access to multi-level Picarro data. The gradient between 30 m and 478 m AGL was 15.1 ppb CH_4, so there are local fluxes within the tower footprint. This suggests that the ethane comparison at 30 m and 478 m would be difficult to interpret.

As indicated also in our responses to Reviewer 1, we added a figure in the supplement showing the differences between two independent Aeris systems deployed for June 2024 at CAO. The bias for that time period was 0.01 ppb C_2H_6, well below our bias goal. To the main text we added, "The mean difference of ethane for two independent Aeris systems deployed at CAO for June 2024 was 0.01 ppb C_2H_6 (Fig. S2)." to Line 417.
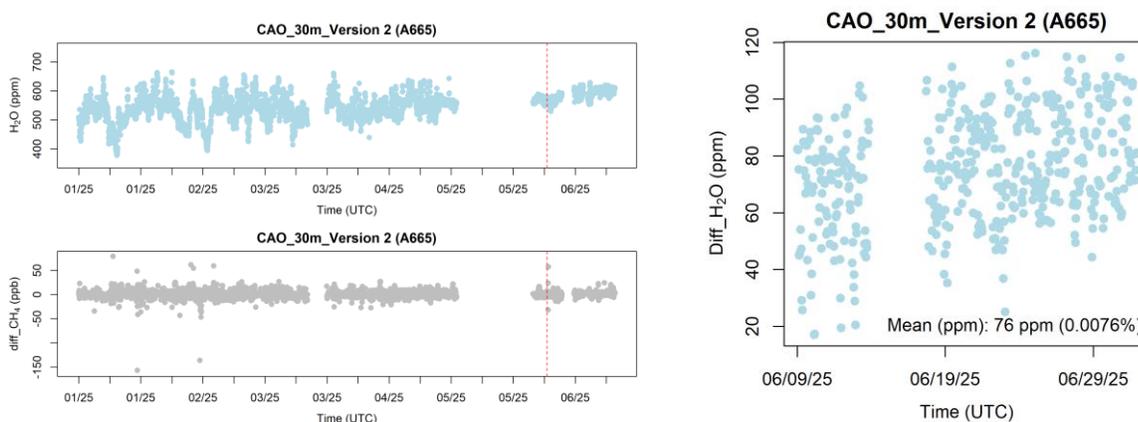
References

Vimont, I., Montzka, S., Crotwell, M., Andrews, A., Baier, B., Hall, B., Handley, P., Higgs, J., Kofler, J., Legard, T., McKain, K., Miller, J., Moglia, E., Mund, J., Neff, D., Newberger, T., Petron, G., Sweeney, C., Turnbull, J., Wolter, S., & NOAA Global Monitoring Laboratory. (2022). Atmospheric Dry Air Mole Fractions of from the NOAA GML Surface and Aircraft Vertical Profile Network. [Data set]. NOAA GML. Version 2025-12-08. https://doi.org/

Line 258: Is it valid to ignore data with "large" atmospheric variability as detected by the Picarro? I am aware that this presents challenges in terms of precise timing and residence time issues, but large ambient swings could provide important additional tests of the AERIS instrument with potential large swings in water vapor. I would recommend trying to reanalyze these time periods if possible.

It was mentioned at Line 254-259 that "The NOAA Picarro line was flushed with an additional pump with a high flow rate such that the sample air takes only several seconds (8 - 14 s) to reach the instrument and the data were reported as 2-min means every 15 min. The Aeris instrument analyzed air from 30 m AGL using a dedicated sampling line. The flow rate for the Aeris was controlled at 110 sccm, using only the pump for the instrument. With this flow rate, the air took about 20 min to travel from the inlet to the Aeris MIRA Ultra. The timing difference was accounted for in the comparison. To minimize noise that might be caused by mismatches in timing, the hours with large atmospheric variability were removed." Note the response to Reviewer 1, the sentences (Line 258-Line 260) were rewritten as "To minimize noise that might be caused by mismatches in timing, the hours with large atmospheric variability were removed (Levin et al., 2020; Richardson et al., 2017; Miles et al., 2018). We used a threshold of 7 ppb CH_4 for the standard deviation within each hour above which the data for that hour were excluded (Richardson et al., 2017)." Figure 9 was updated as shown below. The standard deviations are a bit larger, but the overall message is the same.

We used the Nafion in the system to dry the sample air. The Nafions react slowly to water vapor changes, providing a relatively stable value. For our current application (calibration every one hour), the difference of water vapor between the sample air and calibration cylinder remains small (below 100 ppm at CAO), and this water vapor difference caused uncertainty was calculated and documented in the data file. Below are the water vapor time series plot (from January - June 2025), the difference between the Aeris and NOAA Picarro CH_4 for that time period, and the difference of water vapor between the sample air and the calibration cylinder (June 2025). The water vapor was relatively stable and did not show large swings in water vapor corresponding to the large variation of methane.

References

Levin, I., Karstens, U., Eritt, M., Maier, F., Arnold, S., Rzesanke, D., Hammer, S., Ramonet, M., Vítková, G., Conil, S., Heliasz, M., Kubistin, D., and Lindauer, M.: A dedicated flask sampling strategy developed

for Integrated Carbon Observation System (ICOS) stations based on CO_2 and CO measurements and Stochastic Time-Inverted Lagrangian Transport (STILT) footprint modelling, Atmos. Chem. Phys., 20, 11161–11180, https://doi.org/10.5194/acp-20-11161-2020, 2020.

Richardson, S. J., Miles, N. L., Davis, K. J., Lauvaux, T., Martins, D. K., Turnbull, J. C., McKain, K., Sweeney, C., and Cambaliza, M. O. L.: Tower measurement network of in-situ CO_2, CH_4, and CO in support of the Indianapolis FLUX (INFLUX) Experiment, Elementa: Science of the Anthropocene, 5, 59, https://doi.org/10.1525/elementa.140, 2017.

Miles, N. L., Martins, D. K., Richardson, S. J., Rella, C. W., Arata, C., Lauvaux, T., Davis, K. J., Barkley, Z. R., McKain, K., and Sweeney, C.: Calibration and field testing of cavity ring-down laser spectrometers measuring CH_4, CO_2, and $\delta^{13}CH_4$ deployed on towers in the Marcellus Shale region, Atmos. Meas. Tech., 11, 1273–1295, https://doi.org/10.5194/amt-11-1273-2018, 2018.

The long term laboratory tests of Fig. 7 are interesting but hard to see differences between the minute and hourly averaged data. Is it possible to provide one series of plots with 1 minute averages and another series with hourly averages?

Below are the plots with 1-minute averages and with hourly averages. These plots did not provide new information. We would like to keep the original version.
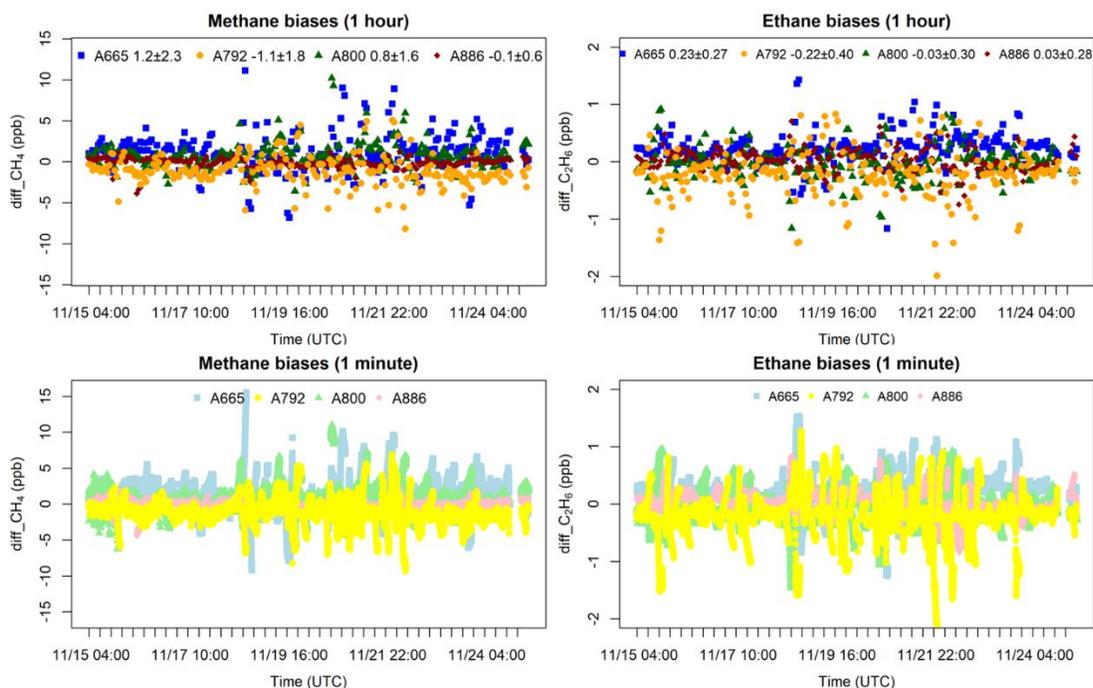
By comparing the two figures, it seems like the large deviations in methane and ethane occur at similar times. Can you comment on this? Can this be associated with large changes in for example in the laboratory temperature or pressure handling systems? The large CH4 and C2H6 deviations of 11 ppb and 2 ppb, respectively, in two of the four instruments for hourly measurements is a little disconcerting.

Tower based flux measurements using these 4 individual instruments could result in flux errors. Can the authors comment on this? Can this be due to the excessively long times (3 hours) between calibrations?

We agree that the large deviations are disconcerting! We have looked at the pressure and temperatures recorded by the instrument, including the temperature inside the optical core and the temperature inside the instrument enclosure near the gas sampling port. We did not find any relationships with the methane and ethane deviations.

Yes, the instrument is noisy with short-term fluctuations, and these large deviations are disconcerting. This is why we have Figure 8 (now in the supplement) to show if this deviation decreases when increasing the averaging time as would be expected for random noise. We always average the measurements when using the observations for a regional inversion. After averaging, the biases are less than 10% of the total enhancement used for the flux calculation.

The instrument noise is reduced by increasing the calibration frequency. To make an acceptable uncertainty due to the instrument noise, we decided to calibrate the instrument every one hour (Fig. 5(e) and Fig. 6(e)). Even more frequent calibrations, as you suggest, would further reduce the noise but the cylinders are expensive and time-consuming to deploy. For the atmospheric inversion, we are concerned primarily with bias, not noise. Our time averaging reduces the noise to an acceptably low level.

Regarding Fig. 8: the caption should be changed to read "the same as those shown in Figure 7". However, I am not so sure this figure adds anything since averaging longer than 1 hour between

calibrations may be too long, as wind conditions over the DJB can change much faster than that. This could wash out any true changes in the emission fluxes.

We think you meant the caption of Figure 6 should be changed to read "as those shown in Figure 5" which is a good point. The caption of Figure 6 was modified as "As in Figure 5, but for ethane". These figures put into context the degree of noise exhibited by the Aeris and explain the ramifications of extending the time scale between calibrations. Line 385-394 and Figure 8 were moved to the supplementary material, as suggested by Reviewer 1.

For regional inversions (e.g., Barkley et al. (2023), we use hourly wind fields to determine footprints and hourly averaged mole fraction enhancements, and we then average over all afternoon hours and then typically over many days to estimate regional emissions. Calibrating every 15 min would be more pleasing in terms of reducing the large deviations (and negative ethane results described in the manuscript), but not helpful for our purpose since we average the data (reducing the noise) before reporting regional fluxes.

Regarding Figure 9: Why is the mean CH4 bias relative to the Picarro instrument worse for the MIRA Ultra version 2 when changing the calibration period from every 5 to 1 hour? Is this significant since very few measurements were acquired in the last segment? The authors should include in the figure caption that these results are only for one AERIS instrument (A665) compared to the Picarro.

As you noticed, in the updated Fig. 8, the mean bias was 1.2 ppb CH_4 for Version 2 with offset calibrations every 5 hours, but slightly worse (1.6 ppb) for Version 2 with offset calibrations every 1 hour. It is possible that the bias changed slightly because of differences in the drying of the sample (the winter data is more dry (<1000 ppm H_2O) than summer (<2000 ppm H_2O)), but as mentioned by the reviewer the last segment contains relatively few measurements so it's difficult to say. The standard deviation is improved as expected (from 6.3 ppb to 3.0 ppb).

The instrument information was added to the figure. Figure 9 (now Figure 8) was updated.

These comparisons, particularly the figure (b) difference plots provide nice confirmatory data for the AERIS methane results in the field. Were the other 3 instruments compared as well? If not, in the future this should be undertaken to insure consistency amongst the 4 measurement site instruments.

Yes, all four instruments were compared to a co-located calibrated Picarro during the concurrent laboratory (pre-deployment) test (Test 4). This test is as rigorous as the test at NOAA CAO, and used the same setup as the field deployment against a co-located Picarro instrument. Additionally, the concurrent laboratory had a large range of methane and ethane sampled because of a probable natural gas leak within the building. The laboratory has relatively unstable HVAC system and the room temperature during the test varied by about 13 °C which is similar to the temperature variation of the sheds in the field. As indicated also in our responses to Reviewer 1, the sentences "Due to a natural gas leak within the building, the range of methane and ethane was similar those typically measured downwind of oil and gas fields, providing a more thorough test than would background levels. The sampled methane and ethane varied from 2030 - 2378 ppb CH_4 and 0.3 - 13.0 ppb C_2H_6 in the laboratory during this 10-day (15-25 November 2024) test. Room temperatures in the laboratory varied by 13 °C throughout the test." were added to Line 221-227 (Section 2.5.1) to clarify the rigor for this test. Therefore, the concurrent laboratory test results reflect the instruments' performance in the field. The concurrent laboratory test was renamed as the pre-deployment test to emphasize that all the systems were tested exactly as deployed, with separate calibration cylinders, and in a non-idealized environment.

Despite the excellent agreement of 1.8 ppb, one should note that the standard deviations of the bias (at the 1σ level?) in the ±2.4 ppb range actually correspond to peak-to-peak deviations of approximately 4 to 5 times this. Does this larger value still reside within your target compatibility goal of 3 ppb for CH4?

The standard deviation was at 1σ level. We clarified in the caption.

These large peaks were concerning, and Figure 8 (now moved to the supplement) showed these peaks would be gone after averaging. Instead of interpreting these individual peaks, we care more about the mean biases. The mean biases are less than 3 ppb CH_4 and 0.3 ppb C_2H_6, which are less than 10% of the enhancements expected in the area and to be used for the flux calculation. The instrument is noisy and shows short-term fluctuations. It was mentioned at Line 459-461 that "The perturbations may be particularly problematic for non-continuous applications with less data available to average, such as drone-, aircraft-, and vehicle-based analyses."

Couldn't the bias standard deviation be improved by more frequent calibrations, say on the order of 15 - 30 minutes or so? If available, it would be very useful to show the ambient water vapor concentrations. Perhaps, even more useful, it would be important to attempt to derive water vapor directly from the acquired spectrum provided the AERIS software would allow this.

Yes. The noise would be improved by increasing the calibration frequency as shown in Fig. 5 (e) and Fig. 6 (e), but it is not helpful for our purpose because we use hourly wind fields and hourly-averaged mole fractions and then average the fluxes to report weekly average fluxes.

Refer to the figures and replies of previous comments "Water vapor concentration was retrieved along with the methane and ethane values by the Aeris instrument." and "We used the Nafion in the system to dry (for Version 2) the sample air. The Nafion responds slowly to water vapor changes in the sample. The water vapor was relatively stable during the measurements. For our current application (calibration every one hour), although the difference of water vapor between the sample air and calibration cylinder remains small (below 100 ppm at CAO), this water vapor difference caused uncertainty was calculated and is documented in the data files."

As no comparable field confirmatory data are shown for ethane, I would recommend modifying your target ethane compatibility results for ethane on line 432.

We clarified above that the concurrent laboratory test (renamed as the pre-deployment test) result reflects the instruments' performance in the field in terms of intra-network compatibility which is the primary concern for our application of determining regional emissions via atmospheric inversion. The measurements are connected to the NOAA internal C_2H_6 scale via the calibration cylinders.

As per my previous comment, I would increase the ambient cylinder calibration frequency to perhaps on the order of 15-30 minutes or so. As I have found, baseline noise associated with optical interference fringing can significantly contribute to instrumental noise. Hence, I would include zero air measurements along with each calibration cycle. As this fringing noise comes and goes depending upon the instrument temperature and pressure stability, I would make an effort to incorporate precise temperature and pressure sensors as close as possible to the instrument optical train, if possible. Given this, I am not convinced that increasing the averaging time as suggested on line 459 is a viable solution.

We agree with the reviewer that increasing the calibration frequency would reduce the uncertainty due to the instrument noise as shown in Fig. 5 (e) and Fig. 6 (e). The high calibration frequency, however, would not be helpful for our purpose. We use hourly averaged data for inverse flux estimates. The random errors are reduced to an acceptably low level and we retain more sampling of atmospheric conditions. Therefore, we chose to calibrate every one hour with an acceptable uncertainty.

The spectrum was locked either on methane or water vapor, so it would be lost if we chose to sample zero air.

There are temperature and pressure sensors built close to the instrument optical train. We added at Line 72-73 that "There are four temperatures sensors in various locations in the Aeris unit, including in the cell. These temperatures and the cell pressure are reported in the datasets." We did not find any associated relationships with the pressure and temperatures. Still, we agree that it seems likely that temperature variations contribute to the instrument noise. It was also mentioned at Line 504-505 in the original document that "Improving the ambient temperature control or that of the cell may also improve the instrument noise performance."

Tests should be run comparing all 4 instruments, particularly for ethane on this cycle since differences in all 4 will be important in deriving flux estimates.

We always average the data for the flux calculation. The concurrent lab test showed the four instruments met our goal for ethane (with long-term mean bias less than 0.3 ppb), and it was less than 10% of the total enhancement used for the flux calculation.

The statement on line 495 regarding large enhancements in the Denver-Julesburg basin, is not quite correct in the case of ethane. As shown in the box-and whisker plots in Daley et al. (2025), the mean and median airborne boundary layer ethane measurements over the DJB is in the 3 to 4 ppb range compared to background values in the 1 ppb range.

Ethane was removed from the statement, and the sentence was modified as "For regions with large methane signals, such as the Denver-Julesburg and Permian Basins, the uncertainty for the instrument design presented here is within an acceptable range".

Additional minor specific issues, as discussed below, should be addressed.

Line 22: "the mean total uncertainty, including both systematic errors and noise, of hourly averages was 0.8 - 3.0 ppb CH4 and 0.35 - 0.37 ppb C2H6" – **at what σ level?**

The total uncertainty was calculated by following Section 2.4. It was mentioned at Line 208-210 that "We take the square root of the squared uncertainty components to determine the quadrature sum of the total uncertainty (Eq. 1)". Therefore, it is not obtained from standard deviation of what σ level.

Line 25: "With appropriate engineering and calibration, the Aeris MIRA Ultra shows the capability to measure ethane and methane with sufficient stability to distinguish regional methane emission sources in many field settings" – **This is a very ambiguous statement not readily supported by**

the data in this paper.

The sentence was rewritten as "With appropriate engineering and calibration, the Aeris MIRA Ultra has the potential to distinguish regional methane emission sources in many field settings". We argue that the results of the pre-deployment test establish this claim, and is supported by both the long-term test for CH_4 at CAO and the co-located Aeris C_2H_6 that we added to the supplement.

Line 48: "Measuring both CH4 and C2H6 mixing ratios can provide information to disaggregate sources responsible for measured CH4 enhancements, especially in regions with co-located thermogenic and biogenic methane sources (e.g. the Denver-Julesburg Basin)." – This is true, but to truly disaggregate CH4 sources into its components, namely thermogenic, biogenic, and regional background, you actually need an additional measurement involving the other major source in the DJB (CAFOs).

We updated the sentence to read, "The system shows the potential to quantify anthropogenic and biogenic methane sources, for regions with mean enhancements of greater than about 30 ppb CH_4 and 3 ppb C_2H_6, by providing continuous ethane measurements, assuming the ethane to methane ratio of the sources is known."

Line 55: **Strictly speaking, the Aerodyne instrument is not a cavity-enhanced infrared absorption spectrometer but is an infrared absorption spectrometer. This should be changed. The term cavity-enhanced is used to indicate a specific type of infrared absorption spectrometer where the cavity is locked to the laser, which is not the case here.**

The sentence was modified.

Line 77: This sentence should be modified to be precise. The laser wavelength in all cases is swept across the absorption features of interest. Wavelength drifts are eliminated using either the water line or the methane line to keep spectra coaligned when co-averaging. Thus, strictly speaking the laser wavelength is not "locked" but the spectra are "locked".

The sentence was modified by replacing "laser" with "spectrum".

Line 97: Did you mean to state "humidifying the calibration gas"? Typically, the Nafion dryers de-humidify the air unless employed in a reverse sense. Please explain.

Yes, we did mean "humidify the calibration gas". In tower applications (urban networks like INFLUX, NorthEast Corridor, etc. and the NOAA Greenhouse Gas Reference Network), Nafions are used to both dry the air sample and humidify the calibration gas. The membrane inside the Nafions equilibrates the humidity between the outer and inner tube regardless of which has more moisture. For Version 1 we had to use the Nafions to stabilize the water vapor near the sample air level rather than drying to < 2000 ppm H_2O. It was mentioned at Line 98-99 in the original document that "using two Nafion tube dryers/water exchangers (Perma Pure LLC: MD-070-96S-2 and MD-070-144S-2) without counterflow (Fig. 1)".

Line 114: Your definition of the term "bias" as the "long-term mean" does not comport with the accepted definition as the deviation from the true value.

It was modified as "(defined as the long-term mean deviation from the true value)".

Line 121: the typical C2H6/CH4 ratio in the DJB is in the range of 5 to 10% and not the other way around as in the text. Please correct.

The sentence was rewritten as "The average C_2H_6 to CH_4 ratio is 5%-10% including all biogenic and thermometric methane sources at the Denver-Julesburg Basin area..." We rewrote this section to not be so specific.