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.

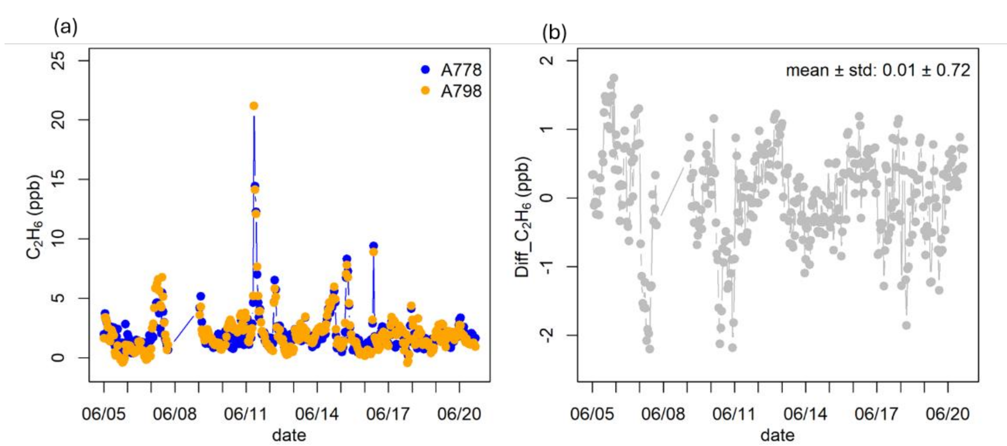


Figure S2: (a) Calibrated hourly ethane data from Aeris MIRA Ultra instruments (serial number A778 and A798) at CAO over the time period of June 2024, and (b) the difference of ethane between these two Aeris MIRA Ultra instruments.

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."

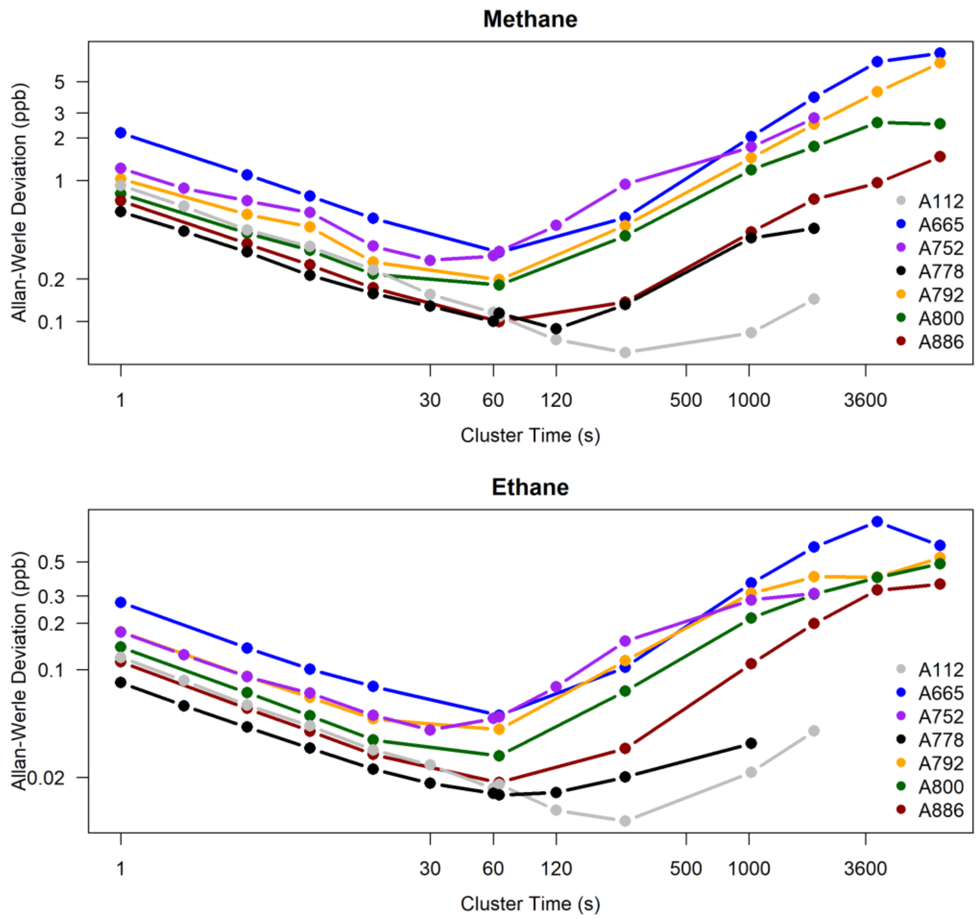
- **Why some instruments were chosen for tests/ studies:** An explanation should be provided why the authors did not perform all tests described in Table 1 for all instruments. The authors state they have 8 unique instruments used in the study, but Table 1 only describes 5 unique serial numbers (665,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



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.

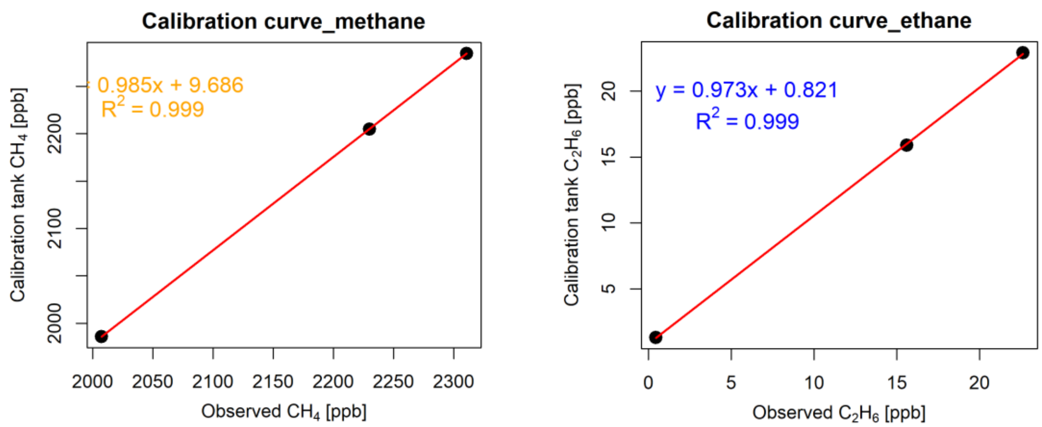
In test three, why did you include all calibration cycles (line 166), when the cylinders did not stabilize for the first few hours? Shouldn't the 'NOAA C2H6' column also include the cylinder assignment error?

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.



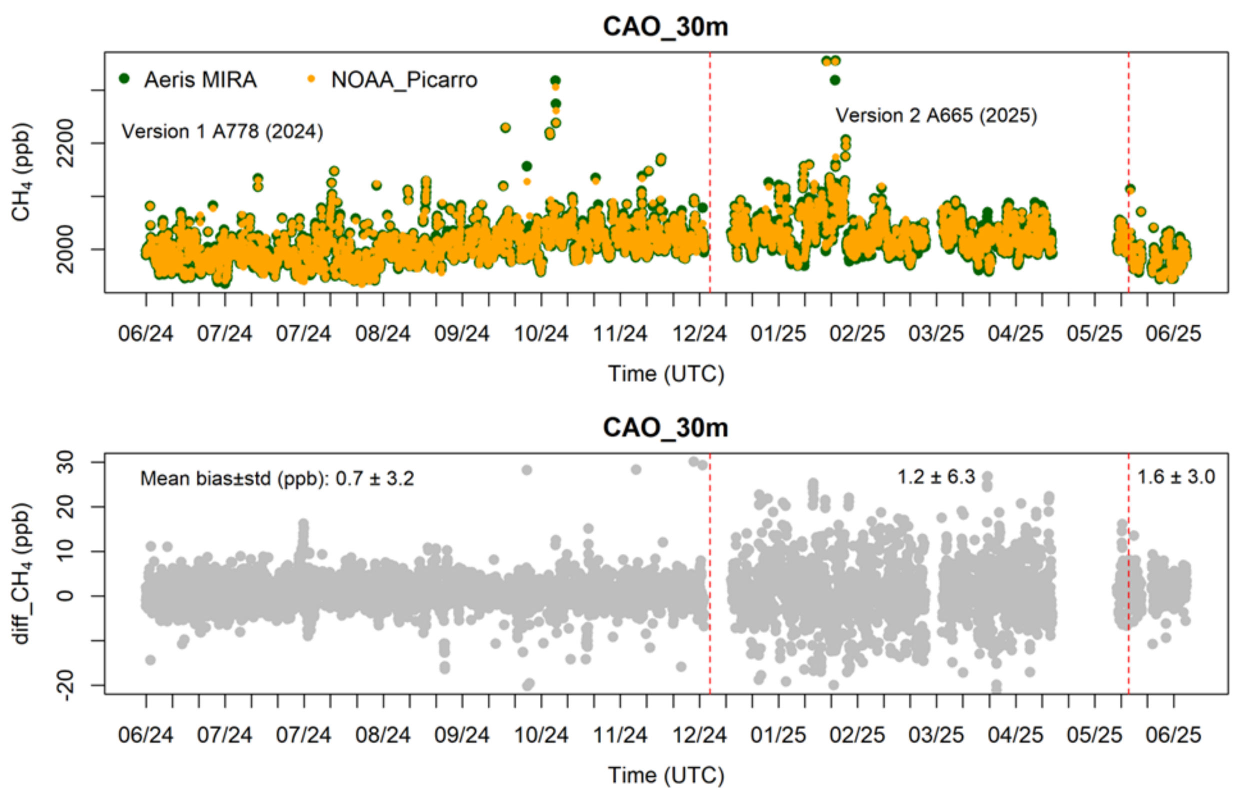
- **Water Vapor corrections:** What does a perfect water correction mean (135-136)?

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 50^{th} 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.