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 (-ln transmission) 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.)?

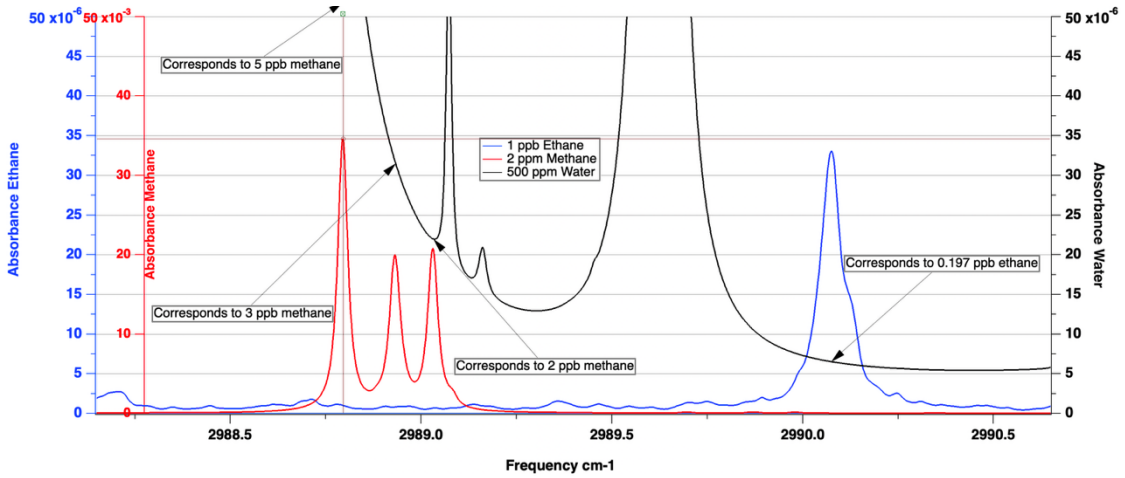

Figure Caption: Absorbance simulations for a 15 m pathlength, sampling cell pressure of 240 mb and a temperature of 42ºC. HITRAN 2020 absorption parameters of methane and water were employed in these simulations with Harrison ethane absorption parameters (not incorrect HITRAN values).

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.

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. After zeroing the results in Fig. 3, do the results shown reflect the retrieved values using the AERIS software? This needs clarification. 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. Can corrections to this software be implemented to minimize these differences? How are the water vapor concentrations being measured (dew point hydrometer)?

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%.

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

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.

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? 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 CH_4 and C_2H_6 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?

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.

Regarding Figure 9: Why is the mean CH_4 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. 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. 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 CH_4? 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 spectra provided the AERIS software would allow this.

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

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.

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.

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.

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?**

Line 25: " With appropriate engineering and calibration, the Aeris MIRAUltra 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.**

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 50 Basin)." – **This is true,**

but to truly disaggregate CH_4 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).

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.**

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".

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.

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.

Line 121: the typical C_2H_6/CH_4 ratio in the DJB is in the range of 5 to 10% and not the other way around as in the text. Please correct.