Dear editor,

We would like to thank you and the referees for their time and constructive feedback on our manscript. We have adjusted the manuscript accordignly and believe the changes have significanly improved the clarity and quality of it. This document contains a detailed, point-by-point response to all the comments provided to us. Changes in the manuscript are indicated where appropriate.

Reply to Reviewer number 1, Anonymous review

Dear Anonymous Referee,

Thank you for your thorough review of our manuscript. Hereby I present the replies to all your comments (in green).

General comments

- Perhaps the title can be a little more specific on the type of sensor being tested. Many in situ methane sensors have been tested so it might be nice to add some more details on which one, to emphasise the novelty of this work.
    - We changed the title to: Evaluating the performance of a cost effective in situ methane sensor for UAS-based systems and its ability to quantify facility-scale emissions." In this case, the title is less generic and incorporates the general comment of this reviewer and anonymous reviewer number 3.

- The authors have based their description of different UAV flux techniques in the introduction on a small number of references with limited newer citations. Lots of work has been done on testing different near-field flux quantification methods in recent years. In addition, perhaps some more recent UAV flux results could be included in the discussion section for uncertainty comparisons.
    - Additional, and newer references were added to the introduction

- The linearity test is not really a linearity test. This section presents the results from a simple calibration, which is fine as a calibration. However, to evaluate linearity, one must evaluate the goodness of the linear fit. This may be quantified using the root-mean-squared error of the fit, for example. Without any indication of the quality of the linear fit, the linearity remains unknown. An identity line (or one-to-one) fit does not necessarily mean that an instrument is linear, as the data could average out to provide a gain factor of unity over the full sampled range, but different parts of the sampled range may deviate from the linear fit.
    - Thank you for your comment, we added a plot of residuals vs. measured concentrations to Appendix B. We believe this test shows the linearity of the system over the range 20 – 2 ppm. The residual plot confirms this linearity nicely.

- A water correction coefficient derived at approximately 81 ppm was applied to all sampling data. This seems like a strange approach. It is indeed a useful test to confirm that the coefficient remains constant over a large methane mole fraction range, but this is not useful to derive a single universal coefficient, for application to all methane mole fraction levels. It would be better to derive such a coefficient at an ambient methane mole fraction level, where most UAV sampling takes place. Unless the authors can confirm that the coefficient is independent of methane mole fraction, it should be re-evaluated at roughly 2 ppm, with the 2 ppm coefficient instead used.

- Dear reviewer, you are right we asked ourselves the wrong question. Instead of asking ourselves "how clearly can we see the effect at 20 ppm?" we should have asked "can we see any effect at 2 ppm?". We therefore carried out and report an additional new experiment performed at a concentration of 2 ppm. This test did show no visible effect (within the measurement noise) at this concentration level, suggesting additional correction is not needed.
- This stands to reason; the magnitude of the effect is expected (based either on first principles or the 20-ppm experiment) to be 20 to 40 ppb of CH_4 over a H_2O range of 0 to 100%, on a sensor noise of some 65 ppb. The original experiment, performed at higher concentration does clearly indicate that no dilution correction is applied by the instrument. We therefore still opt to implement the correction found during the high-concentration experiment.

- A calibration conducted between 0 ppm and 2 ppm (according to Figure B1) was applied to UAV sampling conducted up to 3 ppm. It would make more sense to sample higher methane mole fraction levels during this calibration. Furthermore, the mole fraction range presented in Figure B1 is inconsistent with the description of the experiment provided in the text, which instead suggests that sampling was conducted up to 81 ppm. This is clearly not the case from Figure B1.
  - Dear reviewer, thank you for the comment. The curve depicted in Figure B1 shows a range of 2 to 20 ppm, and therefore also incorporates higher methane mole fractions. We changed the axis labels from ppm to ppb to state this more cleary. The description is still correct but due to the mixing of the high concentration air with ambien air, the actual recorded starting concentrations is closer to 20 ppm instead of the 81 ppm inside of the tank. Additionally, please note that this procedure does not constitute a calibration, but rather a characterization of the sensor's linearity.

- What is missing from this manuscript is a comparison of emission flux estimates derived using raw measurements versus those with a water correction and calibration applied to them. The authors claim that the laboratory testing helped to improve their measurements, but it would be interesting to evaluate the importance of this effect on final emission fluxes. Perhaps, the Axetris data does not require any supplementary correction and raw measurements would still be competitive with the AirCore (provided that the heating coil is used).
  - Dear reviewer, we performed these experiments hoping to ascertain the absence of negative effects caused by water sensitivity and non-linearity. We found that these effects were indeed absent, indicating that the results are trustworhty, much more so than they would have been without testing. From your comment we learned that the point we tried to make (the sensor performs adequatly "out of the box" provided one controls temperature carefully). We have now attempted to make this point more clearly throughout the text.

- This manuscript presents a comparison in methane emission flux estimates using AirCore data and Axetris data. The most logical way in which to test the effectiveness of a sensor would be to compare mole fraction measurements directly, as identical mole fractions would result in identical fluxes. This alterative sensor comparison approach must be clearly justified throughout the manuscript. I think I understand the reason. The AirCore data is smoother and a background must be derived in a different way. But please make this clearer.

  - Dear reviewer, the reasons are actually more mundane. A true test would require comparison to a high cost and high precision in situ sensor and drift free, but we do not have access to one, if one exists at all. The adverse effects of baseline drift affects background determination to a large degree, demonstrably being the largest sources of error. The AirCore does not suffer from that, making it an ideal reference instrument (despite its inherent smoothing). We have added additional sentences to reflect this.

- This paper then compares flux uncertainties from this work to flux uncertainties from other studies, as a means to evaluate the sensor performance. This is incorrect as there are many factors determining flux uncertainties, as the authors themselves show in this work. This comparison to other studies instead allows the overall UAV flight strategy to be evaluated, including the sensor as a component of the overall sampling framework. This is useful in the context of evaluating the overall the UAV flux quantification strategy, but cannot be used to draw conclusions on the quality of the sensor.
  - This is correct. We used the examples of other studies to highlight where other studies find that the largest part of their flux uncertainty comes from. We are aware that direct, numerical comparison between the uncertainties found in this study with uncertainties found in other studies does not allow for quantitative comparison due to varying methodologies. However, it does allow for qualitative comparison. We tried to clarify this by incorporating the following sentences: "During field deployment, the variability of the observed plume was 39% (1σ, n=4). The observed simulation uncertainties and the field example agree well. Furthermore, the main contributors to the uncertainty identified with the simulation are consistent with those reported in other studies (Andersen et al., 2021; Karion et al., 2013; Morales et al., 2022; Nathan et al., 2015). To the point of our study, drift and sensor noise do not emerge from the uncertainty analysis as dominant."

- A key advantage of the Axetris is that it offers a higher temporal resolution than the AirCore. It may actually be better than the AirCore in general, if the appropriate laboratory-derived corrections are applied and a heating coil is used. This may be tested with a controlled release in future work. It would be interesting to see. Maybe the authors can include this thought in the discussion section, if they agree.
  - Yes you are right, we would love to do this and we included this to the discussion sentence as future work.

Specific comments

Line 14: "accurate quantification of its emissions is critical for mitigating climate change"

- Strictly speaking, quantifying methane emissions does not mitigate climate change. It allows for the improvement of climate change mitigation strategies.
  - We revised the sentence to: "accurate quantification of emissions is critical to the improvement of climate mitigation policies"

Line 16: “in sity CH4 sensor (Axetris)”.

- Please provide details on which specific sensor model is used. This is and important detail in the abstract
    - The sensor model is added to the Abstract and now the sentence reads:
    - “… a commercially available in situ CH_4 sensor (LGD-compact A CH_4; Axetris AG, Kaegiswil, Switzerland)…”

Line 19: “provided a water vapour correction term”.

- How can the tests “provide” a water correction “term”? Does this mean that the authors devised a water correction by conducting laboratory testing? I don’t think it is right to say that the tests provided a “term”. Perhaps a water correction was applied to raw measurements using a laboratory-derived water correction coefficient. But that is not clear from what is said here.
    - Noted, we changed the sentence to: “…and a water vapour correction was applied to ensure accurate measurements.”

Line 21: “This mean flux was comparable to the value of 4.2 ± 1.1 gCH4/s obtained from the established AirCore technique.”

- Does this mean that the AirCore sampled simultaneously to the Axetris on a UAV?
    - Correct. This will become more evident in the rest of the paper, but for clarity we added this sentence to the abstract

- It seems a bit strange to compare fluxes when evaluating the performance of the sensor. It would be better to compare direct mole fraction measurements. The flux technique is irrelevant. If the two sensors provide identical mole fraction measurements then they will inevitably provide identical fluxes. Some justification is required.
    - Thanks for bringing this up, this is a valuable point to make more explicitly. We believe it to be a misrepresentation of a complex interplay between noise, drift, flight speed, enhancement height etc to state that “identical concentrations ==> identical fluxes”. Surely the Axetris performance is not identical to a high-cost in in situ sensor. But under certain conditions (e.g. gentle, long flights with fair plume coverage, enhancements of several 100's of ppbs, etc.) it does yield comparable results. For flux calculations, the enhancements instead of the raw concentrations are used. Restricting ourselves to only comparison sensor output would not illustrate this point well enough. By contrast, deriving flux esimtates from our noisy results does demonstrate the sensor’s capability.

Line 22: “Ornstein-Uhlenbeck method”

- Maybe a few short words on this method could be good here, but I understand if the authors prefer to avoid this here.
    - We’ve certainly attempted this, but believe that merely a few short words would not provide sufficient explanation of the method, and we believe it is best to leave it as it is and provide more in-depth information in the text.

Line 33: “1942 ppb”

- This is quite precise, considering it represents mole fraction for the entire planet for a full year. Maybe it is better to say 1.9 ppm. But it's fine if the authors insist.
    - This is the mole fraction presented by the WMO and we believe this is a good representation of the average mole fraction. However, we added the uncertainty as presented by the WMO to address this comment.

Line 35: "Saunois et al., 2020"

- There is a new methane budget paper: https://doi.org/10.5194/essd-17-1873-2025. It is better to cite and use this instead of or in addition to the 2020 paper.
    - Revised and we updated the reference to "Saunois et al., 2025" instead.

Line 37: "climate mitigation"

- This doesn't make much sense (mitigating the climate). I thing the authors mean to say that reducing emissions is essential to mitigate climate change. "Climate change" and "climate" do not mean the same thing.
    - This is correct, thank you for pointing this out. We changed "climate mitigation" to "mitigating climate change"

Line 43: "underreporting"

- There can also be over-reporting (or double-counting) taking place.
    - This is correct, we incorporated these suggestions to the sentence. It currently reads: "These uncertainties arise from limitations in available data as well as under- and over-reporting in national inventories."

Line 44: "scarce"

- This is a strong word. While it is true that data can be scarce in places, some regions on Earth have dense sampling networks.
    - We revised the sentence to: "Current CH_4 emission inventories suffer from inconsistencies due to spatial variability in observational data and methodological disparities." The wording is less strong, but the sentence still highlights the differences in the density of sampling networks.

Line 48: "exhibit complex emission patterns with substantial diurnal and seasonal variability"

- Some citations are required here to support this statement.
    - We added two references to this statement

Line 65: "max"

- Please replace this with "maximum".
    - Revised and changed to maximum.

Line 66: "relatively low implementation cost"

- UAVs may be cheaper than aircraft, but not necessarily cheaper than ground-based vehicular sampling. This statement should be qualified.
    - Dear reviewer, we have added the following table, showing our interpretation of the order of magnitude to the supplementary material. The qualatiy class column identifies the classification we give to a sensor based on the other categories presented. We hope we addressed the critique given with the implementation of this table and you understand our reasoning.

	Quality class	*Purchase cost*	*Per-day use cost†*	*Conc. precision / accuracy**	*Emission uncertainty*	*Comment*
Electrochemical sensors + (small) UAV	*Low*	*0.5 + 2 = 2.5 k€*	*0.5 k€*	*±1000 ppb (with limit of detection often >>5 ppm)*	*(very high)*	*Not useful except for (indoor) leak detection*
Axetris + UAV	*Medium*	*5 + 10 = 15 k€*	*0.5 k€*	*<80 ppb / <50 ppb*	*20%*	*This study*
AirCore + Licor + UAV	*High*	*25 + 45 + 10 = 80 k€*	*1 k€*	*<1 ppb / <2 ppb*	*20%*	*This study*
High-end sensor + UAV	*High*	*50 + 10 = 60 k€*	*1 k€*	*<1 ppb / <2 ppb*	*20%*	*May require sensor specialists during deployment*
Licor + Vehicle	*High*	*45 + rental = 45 k€*	*1 k€*	*<1 ppb / <2 ppb*	*50%*	*Mass balance not possible (emissions obtained through OTM 33A). Access limitations.*
Licor + small aircraft	*High*	*45 k€ + rental = 45 k€*	*3 k€*	*<1 ppb / <2 ppb*	*20%*	*Hard to do facility scale*

† All high quality class instruments are assumed to require two operators, while low and medium class sensors are assumed to require one.
** note: accuracy is operationally approximated here as the instrumental drift that may roughly be expected over the duration of a emissions quantification deployment (e.g., between calibrations / during a flight). This is <50 ppb for the Axetris (see Figure 2 of the manuscript) and typically is much smaller for high-end sensors. Note that these figures are meant to indicate performance ranges, not to represent precise technical specifications.*

Line 72: "these techniques"

- Which techniques is this referring to? The previous sentence refers to three "deployment strategies". This next sentence is describing how sampling acquired from any of the three deployment strategies can be used to derive a methane emission flux, which seems like a logical order to present this information.
    - We refered to the deployment strategies. We changed the wording to make it clearer: "Although these deployment strategies differ operationally…"

Line 73: "they all rely on the mass balance approach (MBA; Nathan et al., 2015; Vinković et al., 2022) or the inverse Gaussian approach (IGA; Andersen et al., 2021) for flux quantification"

- This seems like a major oversimplification. All methods are effectively a type of mass balancing as they compare downwind enhancements (due to source emissions) to an upwind background. Both approaches subtract a downwind plane from an upwind plane. The difference is how measurements on the downwind plane are integrated. Gaussian plume modelling typically applies Gaussian statistics to downwind sampling to obtain a smooth sampling plane (hence yielding an inverted flux). By contrast most other so-called simple "mass-balancing" approaches do not apply Gaussian statistics and simply spatially interpolate available measurements or average them into grid squares (yielding a flux through spatial integration). I would therefore argue that Gaussian plume inversions are a more advanced approach to (or subset of) mass-balancing.

- I would recommend replacing this statement with the something like "Although these deployment strategies differ operationally, mole fraction measurements from any approach rely on mass balancing to derive a flux, where downwind sampling is either spatially integrated (following spatial interpolation or grid-square averaging), or modelled using a Gaussian plume inversion.".
    - Thank you for this clear feedback. We understand where you are coming from and we agree that your suggestion improves the argument. We therefore replaced the statement to: "Although these deployment strategies differ operationally, the mole fractions (observed during flights) require subsequent processing to derive flux estimates from them, where downwind sampling can be spatially integrated (e.g. following spatial interpolation or grid-square averaging), or modelled using the inverse Gaussian method"

Line 75: "it can directly convert atmospheric concentration measurements into fluxes"

- How is the so-called MBA method more "direct" than IGA approach? Both methods require some computation to convert mole fraction into flux. So I disagree with the idea that one approach is more "direct" than the other.
    - We concur, "directly" may not have been the correct wording for this. What we are trying to convey with this statement is that the MBA is a more straightforward framework for flux estimation, since the IGA usually requires additional assumptions on atmospheric state / plume geometry (that may not hold up in real life scenarios). We have amended the sentence to read: "This study focuses on the mass balance approach (MBA) because it converts atmospheric mole fractions to fluxes following fewer intermediate modelling steps compared to the IGA."

Line 79: "20% to 75%"

- Multiple citations are required here to support this.
    - Revised, the citations were placed at the end of the paragraph, but we copied them to a few sentences earlier and also retained them at the end of the paragraph.

Line 80: "non-uniform sampling"

- What does this mean?
    - In this setting, non-uniform refers to uneven sampling along the flight track, e.g. when one area is overrepresented compared to another. Ofcourse getting perfectly uniform sampling is difficult (if not impossible), but the larger the gap in sampling, the larger the error can become. We implemented this comment by altering the manusctipt to: ""non-uniform sampling (sampling taken at irregular intervals)."

Line 86: "Tuszon et al., 2020"

- I can’t find this in the reference list.
    - Thank you for pointing this out, we revised this and added the reference to the reference list.

Line 86: “Allen et al., 2019”

- Allen et al. did not conduct any sensor development. They tested pre-existing commercially available sensors.
    - You are correct, we removed the citation from this summation.

Line 87: “high cost and maintenance demands”

- High precision sensors do not necessarily have higher maintenance demands. Quite often, they are easier to maintain.
    - This may well be true for certain types of high-cost sensors. However, for various published open-path QCL systems, we are aware of the significant effort required in maintaining optical performance. However, this information is not directly citable, and we therefor haved toned down the the statement to “conceivably high maintenance demands (particularly for open-path optical designs)”.
- It may be worth noting here that better sensors are often heavier, potentially resulting in reduced flight endurance.
    - Indeed, and these require the use of costlier drones, etc. We further clarified this in the text by adding the following sentence: “In addition, high-precision sensors tend to be heavier, which can reduce flight endurance or require larger and more expensive UAS platforms to accommodate the increase in payload.”

Line 88: “low-cost”

- According to my personal understanding a low-cost sensor is less than 100 €. So I would not call an Axetris low-cost, in my opinion.
    - You are right. We already realized this, but we did not change all the mentions of low-cost ot medium cost. Besides, we prefer the term cost-effective over the classification of medium/ low cost.
- A rough cost range should be given here.
    - See table above/ and attached as supplementary material.

Line 88: “medium-precision”

- What does medium-precision mean? I think the authors should provide a rough parts-per-million range here.
    - See table, where we clarified our reasoning. However, as mentioned it is difficult to make this classification a part of the official manuscript since most is based on assumptions and experience with the field.

Line 92: “low-cost”

- See the comment above. It is fine to call this sensor “low-cost” if the authors insist (although I advise against this), but then a rough cost range definition of “low cost” should also be provided in the manuscript.

- Thank you for your comment. We do agree that this sensor falls more within the medium-cost category and we changed instances of low-cost to medium-cost.

Line 94: “minimise drift”

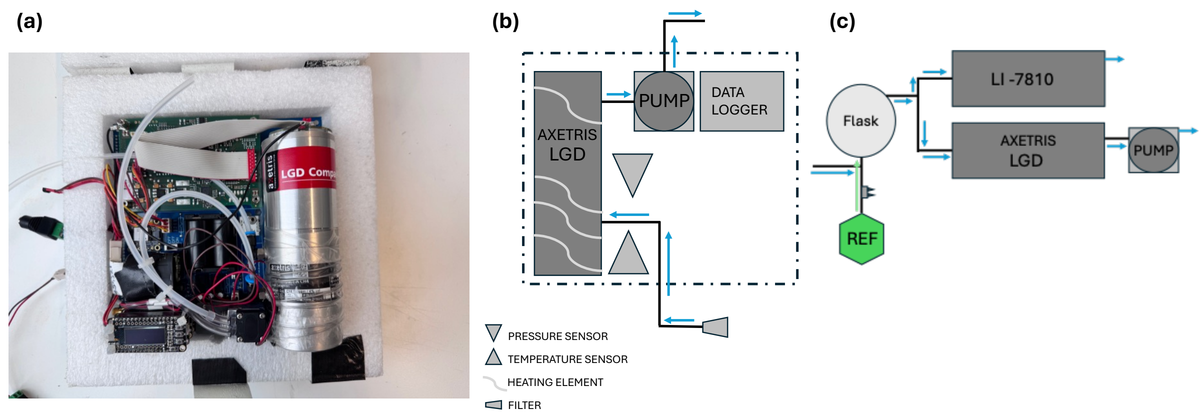
- Laboratory testing cannot be conducted to minimise drift. It is only possible to measure drift or to conduct testing to “account” for drift. But the sensor will naturally drift regardless of whether it is tested in the laboratory or not.
    - Revised and we changed the wording to “identify drift”.
- The authors add a heating coil to the sensor, which improves stability by reducing temperature sensitivity. However, this is not a part of laboratory characterisation testing. This is instead hardware development and sensor optimisation.
    We only partly concur with this statement. Specifically, we do *not* reduce temperature sensitivity, but rather we reduce the temperature variability (that the sensor is sensitive to). The identification of temperature sensitivity and the quantification of the results of alleviating measures (e.g. the heating coil) would count as characterisation. The addition of the heater is, indeed, development. We make that distinction more carefully in the manuscript.

Line 102: “in situ CH4 sensor”

- It might be nice to introduce the Axetris here and discuss the sensor technology here (i.e. the second sentence of subsection 2.1). This would make this subsection seem a little less abstract. But it is fine if the authors want to leave it as it is.
    - The purpose of this section is as a leading paragraph, showing an overview of the complete paragraph. We decided to leave it more abstract for that specific reason and have the more detailed description in subsection 2.1.

Line 113: “We developed a UAV-mounted sensor package”

- Consider including an annotated photograph of the sensor package (highlighting the inlet and outlet). This may be included as a subfigure of Figure 1.
    - Revised and added to Figure 1. We decided to leave out the highlighting of the inlet and outlet since this did not make the Figure clearer. Figure one is updated to:



Line 120: “the interface PCB”

- I don't follow this. How can the Axetris report the interface PCB on which it is mounted?
    - Thank you for pointing this out. There are actually two PCBs; one PCB manufactured by our group and an additional one that is part of the Axetris instrument, which was mounted on the self-built PCB. We have clarified this in the main text as follows: "We integrated the Axetris onto a custom printed circuit board (PCB) designed around an ESP32 microprocessor, enabling logging of data from the Axetris and ancillary sensors. The Axetris reports CH_4 mole fractions, the temperatures of the measurement cell and the interface PCB (which is mounted onto the custom-built PCB)."

Line 131: "ground time"

- What is the ground time?
    - Thanks, this was a spelling mistake. Revised to "ground team"

Line 137: "PID"

- This acronym has not been defined.
    - Revised and defined in the main text as: "A proportional-integral-derivative (PID) controller controlled…"

Line 144: "Figure 1"

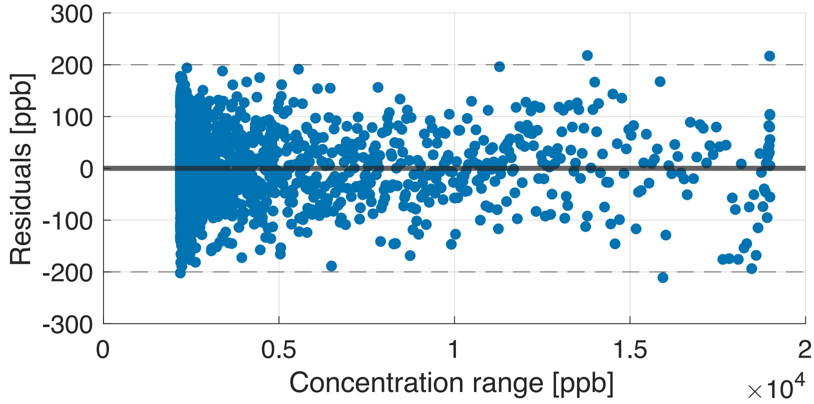
- What does "REF" stand for in this figure? Is it defined in the main text?
    - We added the meaning of REF to the figure caption. The complete caption of Figure 1 now reads: "Figure 1: a) Photograph of the Axetris package. The different colours visualise the flow of air through the sensor. Blue represents the path of the air entering the Axetris sensor. Orange represents the connection between the Axetris and the pump and green represents the outflow of the air. b) schematic overview of the Axetris package for UAS applications. The blue arrows show the airflow through the system during measurements. Components in the dotted and dashed lines enclosed area are integrated onto the PCB and enclosed in the PE foam. The datalogger communicates with the Axetris sensor and provides active control of the heating element, including setting and maintaining the desired temperature. c) Experimental overview for linearity test setup. First, the flask is filled with the reference gas (REF), which has a high CH4 concentration (indicated by the green arrow). Then, both systems start to sample the air from the flask, while ambient air was drawn in to mix with the high concentration (blue arrows), slowly decreasing the concentration in the flask.

Line 162: "short-term drift"

- Was pressure and temperature controlled when making this assessment, with sufficient warm-up time?
    - This was indeed the case, we clarified this in the main text as: "From our laboratory tests under controlled conditions and with sufficient warm-up time, we found 1-second, …"

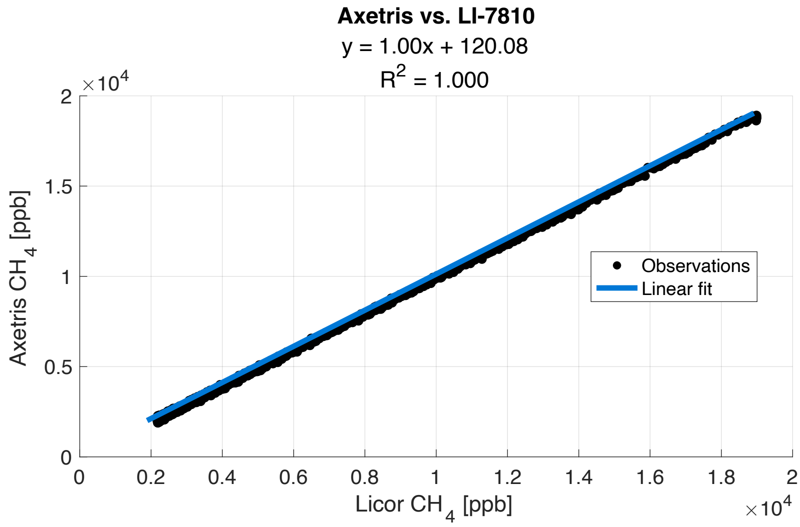
Line 166: "Linearity of the Axetris"

- This subsection (in its current format) does not provide an evaluation of the linearity of the measurement, but rather provides a calibration of the instrument. A linearity test shows whether the increase in one measurement is consistent with increase in a second measurement, following a linear regression. The magnitude of any derived linear coefficients is independent of the standard of linearity.
    - We added a plot of the residuals between observations and the linear fit to show the sensors linearity (i.e., neutral, non-wavey noise around the linear fit) over the observed concentration range. We believe this sufficiently demonstrates the linearity of the Axetris. The following figure is added to the Appendix as Figure B2:



Line 167: "high CH4 concentration (~81ppm)"

- If this is the case, why is Figure B1 plotted between 0 ppm and 2 ppm? Why not plot it up to 81 ppm?
    - We used 81 ppm to increase [CH_4] in a mixing vessel to ~21 ppm. Gradual reduction back to background concentrations yields a range from 20 ppm to 2 ppm. For clarity we have removed from the figure the initial (at the start of the experiment) rise from 2 to 21 ppm that, confusingly, was originally visible as a S-shaped trace (due to difference response times of the two sensors). Only the gradual, monotonous decay of concentrations back to ambient is now plotted.
    - The updated Figure (Figure B1) is as follows:



- If a high concentration was diluted with ambient air, how was 0 ppm (lower than the ambient background) achieved in Figure B1?
    - See comment above, a minimum concentration of 2 ppm was recorded.
- Where did this high concentration gas come from? Was a specially prepared gas blend purchased? If so, did the authors consider the presence of interfering trace gases in the overall gas matrix?
    - The used air mixture is simple dried, compressed ambient air spiked (pre-compression) with pure CH_4 and prepared in our laboratory. Our lab is excellently equiped for this procedure (including ICOS certification). We are confident of the absence of contaminant gases. 81 ppm was derrived using the LI-7810, assuming linearity.

Line 168: "ambient air"

- Where did the ambient air come from? Was a compressed gas cylinder used? If it was natural laboratory air, can the authors be sure that is was not contaminated?
    - You are in principle right to worry about the appropriateness of the methods employed to prepare our standards. However, our laboratory is excellently equipped for high-accuracy work with compressed gases (incl. ICOS certification). We employ well-documented, community-standard procedures for collecting, filtering, drying, compressing, storing and assigning (values to) our standards. Also, TDLS like the Axetris are typically insensitive to cross sensitivity. We therefore assume effects of unexpected contamination to be of insignificant concern.

Line 169: "LI-7810 system linearity"

- Please make it clear that this linearity was tested for a different LI-7810 to the one being used in this work. Linearity results from a previous study for a different unit are being assumed to be valid here.
    - Linearity of high-quality instruments like the LI-7810 is assumed to be rather homogeneous between individual instruments. Insofar as this is not the case, the deviations from this assumptions are considered by us to be far from relevant to our application. To stress that point, we included a figure showing the residuals of the fit between the LI-7810 and the Axetris (Figure B2). These

residuals are neutrally distributed along the range CH_4 concentrations observed. Either both instruments are linear, or they happen to both be non-linear to the exact same degree, which is the less likely option.
    - However, to address the point of the reviewer, we included the following clarification in the text: "The precision, accuracy, and linearity of the LI-7810 analyser under standard operating conditions were evaluated in an extensive report executed by ICOS (ICOS, 2020; we generalise these findings to our specific instrument)."

Line 171: "wide range of concentrations"

- Please provide the concentration range here (assuming this is referring to the ICOS testing).
    - Revised and added the observed concentration range
    - "These tests enabled direct comparison of the systems' response across a wide range of concentrations (2 to 20 ppm)."

Line 178: "The slope confirms linearity between the two systems"

- Calculating a gain factor value for the slope of a linear regression does not provide an indication of the linearity of the fit. A gain factor of exactly 1 or a gain factor of 100 can both be derived from perfectly linear data, assuming all of the data points follow a straight line. Linearity can instead be evaluated by the root-mean-squared error of the fit, indicating that the gain factor (regardless of its magnitude) remains valid over a large mole fraction range.
    - To address the majority of the comments we added a residual plot to the results. We hope these results will strenghten our statement that the sensor behaves linearily.

Line 184: "significantly"

- This word is unnecessary. Sometimes the effect is large, sometimes it is small.
- A dilution effect is, by definition, only as large as the mole fraction of water. As water mole fraction is typically around 1% (and no more than around 3%), it is hardly "significant".
    - Revised and we removed the word "significatly"

Line 191: "high concentration of CH4 (~81 ppm)"

- Why use a high concentration source? It would make more sense to characterise the water effect at an ambient methane mole fraction. The UAV will rarely sample at 81 ppm, so it seems strange to obtain a universal water correction coefficient for extreme sampling conditions.
    - As mentioned with the general comments we chose for this high concentration since we believe the water vapour effect is not noticable at lower concentrations. We do understand the critique here, and we therefore decided to perform an additional water vapour test at ambient levels.

Line 195: "1.5% – 61 % at 25 ºC"

- At which pressure were these values derived? I don't think relative humidity is important here, but if the authors insist in including it, they should include the pressure that this corresponds to.

    - Revised and removed the mention of the relative humidity.

Line 199: "Chen et al. (2010)"

- Is this citation necessary? In this work, the wet and dry data is simply compared and a linear fit is applied to it. It's not really a new "method" devised by Chen et al.. Furthermore, Chen et al. used a polynomial fit, while a linear fit is used in this work.
    - Revised and removed the reference

Line 207: "Chen et al., 2010"

- Rella et al. (https://doi.org/10.5194/amt-6-837-2013) should also be cited here.
    - Revised and added Rella et al. to the reference list and updated the citation in the text.

Line 212: "With this, we assume that the non-dilution part of the water vapour correction remains the same at ambient levels as at high concentrations."

- But the spectral interference and overlap between the water and methane peaks can change depending on the methane mole fraction level. Methane affects the water peak and vice-versa. So, the most standard condition in which to evaluate the influence of water mole fraction variability is at ambient methane mole fraction levels of 2 ppm (see Figure 2b in Chen et al. (2010), for example).
    - This is correct, but we assume it does not change much and since we cannot see the water vapour effect at lower levels we need to make some assumptions. However, we did perform an additional water vapour test and added the results to the supplementary material.

Line 218: "System precision"

- Was the data used for this test calibrated and water-corrected? This is especially important for the UAV precision test, where varying water mole fraction may have influenced instrumental measurements.
    - In all cases we used a dry gas (tank gas in the laboratory) and a calibration bag filled with tank gas during the UAV precision test, so no water correction was applied. The manuscript was amended to read: "The system's precision can be determined by operating it under stable conditions while sampling a dry reference gas with a known concentration over a longer period."

Line 222: "e.g. the Axetris"

- Is ths necesarry?
    - Revised and removed this mention. The sentence now reads: "It provides insight into the noise, stability, and drift of the Axetris,…"

Line 229: "observed"

- Where were these observations made? Were these measurements from the Axetris cell or from somewhere in the ambient laboratory?
    - Revised and added the following: "… (as recorded by the sensors on the PCB), …"

Line 230: "UAV's precision"

- Please rephrase this. The UAV precision is not being measured but rather the sensor precision during UAV deployment.
    - Revised to: "The sensors precision during UAS deployment was determined for each of three consecutive flights…"

Line 231: "three consecutive flights"

- Please clarify that an Allan variance test was applied to each flight individually. This is an important detail, as it would otherwise be incorrect to merge different datasets into a single Allan variance analysis.
    - Revised to: "… each of the three consecutive flights"

Line 241: "To maintain measurement precision below 10 ppb, calibration every hour is required."

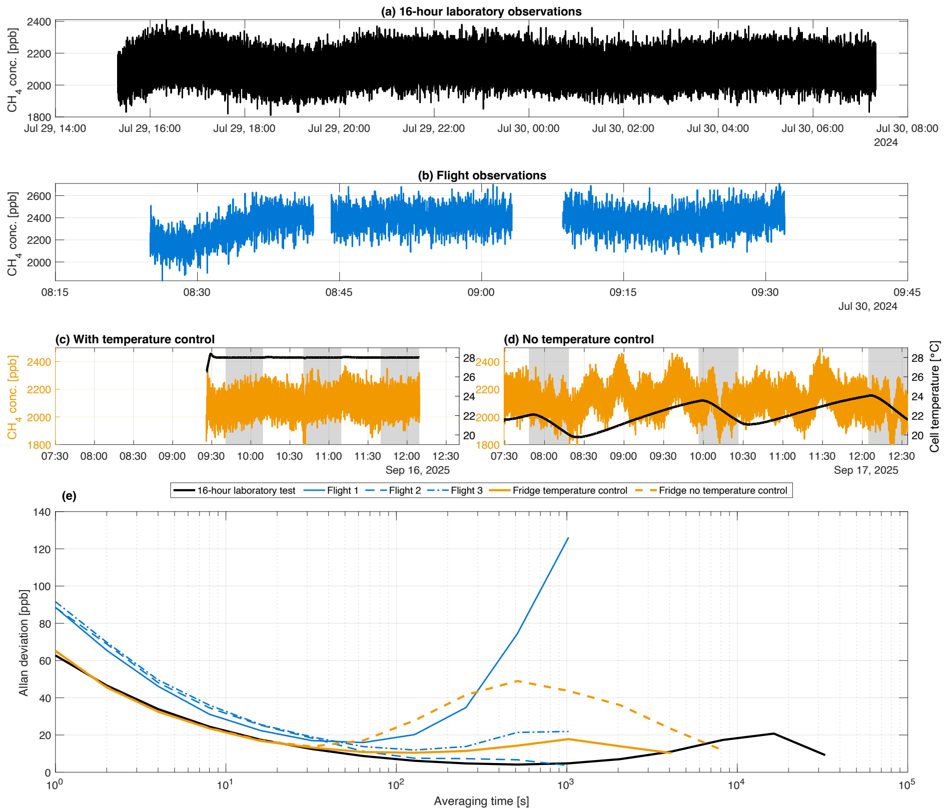
- I do not understand the basis of this statement.
- The minimum Allan deviation is 60 ppb, at 1 Hz. So how can a precision of less than 60 ppb (i.e. 10 ppb) possibly be achieved?
    - Thank you for pointing this out. This confusing likely stems from the meaning of precision in this example. The Allan variance is a way to measure the system precision over a longer period of time. However, in our case, this precision is directly translated to the observed drift.
- How is this time of 3600 s between calibrations derived from this Allan variance test?
    - This is based on the averaging time over a longer period of time. The 60 ppb precision is minimum when you would average over 1 second. When increasing the averaging time, the sensor precision can improve. However, when the allan deviation is going up again, this usually means that sensor or instrumental drift starts to play a role as well. We believe that within the hour, these effects are not pronounced yet and the precision is still below 10 ppb. We added an enlarged version of the Allan deviation plot (Figure 2e) to the supplementary material and showed a visualization of the allan deviation after an hour (Figure D1).

Line 246: "without additional temperature control"

- Please provide some temperature measurements (and standard deviation) values here please for the three UAV flights.
    - Revised and added the standard deviation of the three flights as: "…pressure and temperature variations (1.01 ± 0.004 bar; 28.4 ± 0.7 °C),…".

Line 249: "Figure 2"

- The title for the horizontal axis title should either be the correct Greek symbol (τ) or be replaced with the text "averaging time".
    - Revised and changed the xlabel to averaging time.
- It might make more sense to change the colour of subplot b and then to use consistent colours in subplot d. Otherwise it is confusing because different colours mean different things in the four subplots.
    - Revised and changed the colours to match. The revised figure



- ○

Line 287: "~ 100 minutes; Figure 2c"

- I don't see this. ?
  - ○ We tried to visualize this by plotting each time series over the same length (so each from 7:30 until 12:40. However, since this cannot be read clearly, we decided to remove the specific "~ 100 minutes" statement.

Line 293: "Both instruments were deployed"

- Please clarify here whether they were both deployed simultaneously or if they were deployed individually for separate flights.
  - ○ Revised and added "simultaneously" to stress that they were deployed at the same time.

Line 298: "dairy cow farm"

- It would be good if site coordinates can be included here (although I understand if this is not possible).
  - ○ We believe the description of the site location (including the farm outline in our flight maps and the aerial photography in Vinkovic et al., 2022) is sufficient and adding the exact coordinates will not improve the message of this manuscript. We therefore decided to not include this.

Line 300: "The farm was"

- Please replace this with "Emissions from the farm were".
  - ○ Revised and replaced with "Emissions from the farm were…"

Line 306: "senor package was mounted"

- But where was the air inlet? Was the air inlet also below the plane of the propellers? If so, has the influence of downwash been considered?
    - Revised and added information about the location of the sampling inlet as: "…UAS using two straps, and the sampling inlet was connected to a sampling tube pulled through a 1 m long carbon fibre rod. The inlet is located 50 cm upwind of the propeller downwash column."
    - The positioning of the sampling inlet ensures it is not influenced by the drone's propeller downwash.

Line 314: "drone"
- Replace this word with "UAV".
    - Revised and changed to "UAS" to match the entire manuscript

Line 316: "altitude"
- Please replace this with "height above ground level". Altitude generally refers to a measurement with reference to mean sea level.
    - Revised and changed all instances of altitude to "height above ground level"

Line 324: "as the average the"
- Please replace this with "as the average of"
    - Revised and replaced "as the average the" with "as the average of"

Line 324: "during the duration"
- Consider replacing this with "for the duration".
    - Revised and replaced "during the duration" with "for the duration"

Line 325: "height"
- Please replace this with "height above ground level".
    - Revised and replaced "height" with "height above ground level"

Line 354: "shifting the assigned timestamps of the project AirCore data"
- Was the lag time through the air inlet of the Axetris accounted for?
    - It was indeed. Prior to field testing, we determined the time lag from the sampling inlet to both the LI-7180 and the Axetris. For both instruments, this was 3 seconds, and we accounted for this time lag before spatial interpolation and linking to the correct GPS coordinates.
- Were Axetris measurements shifted to the AirCore time-stamp or vice-versa?
    - The AirCore measurements were linked to the Axetris since the AirCore samples are analysed on the ground. However, we updated the linking method to be more robust and less dependent on visual confirmation by including the AirCore pump status for linking the Licor observations to the geospatial data from the AirCore. The following sentences were added to the manuscript: "Information about the AirCore pump status helped to correctly link the AirCore observations to the UAS's geospatial position. Incorrectly selecting the LI-7810 timestamps corresponding to the ambient sample introduced a larger uncertainty."

- The best way to adjust concentration measurements is to ensure that they correspond to the time stamp of geospatial position (with lag time and time lag accounted for). Shifting mole fractions so they agree with each other, rather than adjusting them independently to geospatial position measurements, seems suspicious.
    - We updated the explanation to include the pump status of the AirCore and highlight the importance of correclty selecting the timestamps of the LI7810 corresponding to the ambient sample.

Line 364: "altitude"

- Please replace this with "height above ground level". Altitude generally refers to a measurement with reference to mean sea level.
    - Revised and replaced "altitude" with "height above ground level"

Line 369: "complete"

- Remove this word.
    - Revised and removed the word "complete"

Line 376: "Spatial interpolation"

- This work presents a very sensible approach to spatial integration. Perhaps it can be alluded to in the abstract or introduction. This approach is much better than geospatial kriging or simple grid-square averaging, with no weighting.
    - Thank you for the acknowledgment. We agree with this and decided to introduce the method a bit more broadly in the abstract. We revised the abstract to include: "Additionally, the sensor was used to quantify whole-farm CH_4 emissions by spatially interpolating measured mole fractions using a Gaussian weighting scheme. This yielded a mean flux of 4.1 ± 1.6 gCH_4/s averaged over four flights, which was comparable to the value of 4.2 ± 1.1 gCH_4/s obtained from simultaneous flights with the established AirCore technique."

Line 378: "observed CH4 enhancements were interpolated onto a grid"

- A fundamental requirement of this approach is the assumption that the emission plume does not move during sampling. This method assumes that turbulence does not change the location of the plume and that the wind direction does not change. Please discuss this here as a caveat to this approach.
    - This is not the case. Turbulence and location variability would yield a broader, more sparse or even non-contiguous and less intense contour plot of identical integral. We acknowledge that this result may not be intuitive. We attempt and clarify and illustrate this in Section 4.2.1: we added a six-panel figure in the Appendix (Figure K2) showing the integral approaching a clean Gaussian after six repeated samplings of the same variable plume.

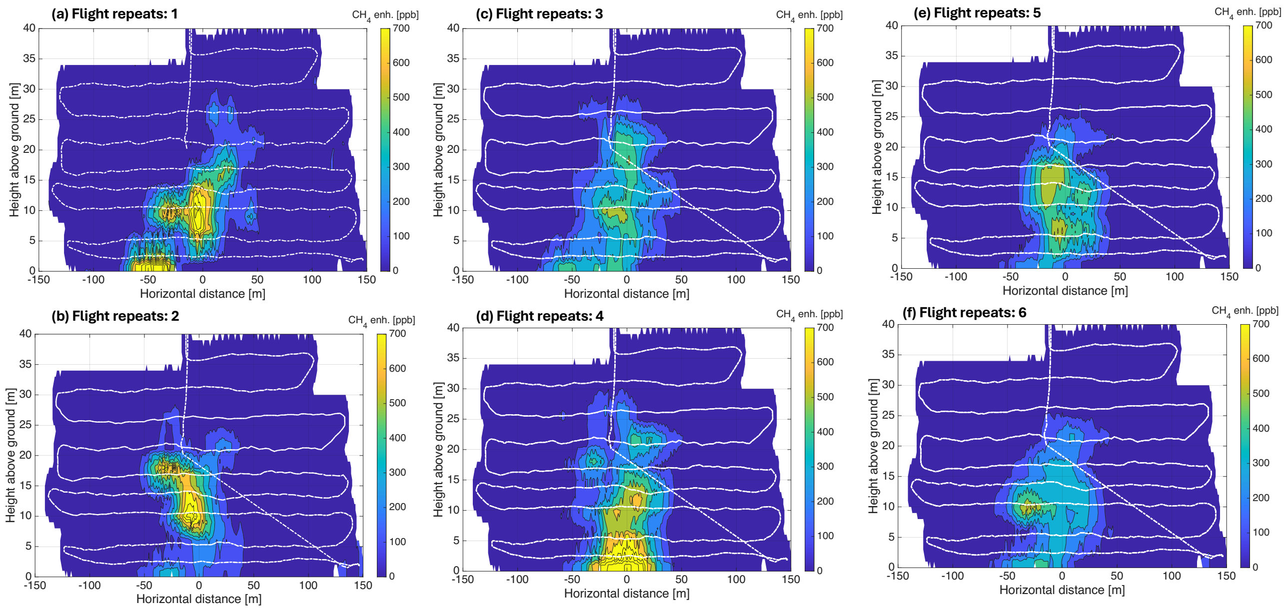


Figure K2: Illustration of sampling of a stochastic plume with an increasing number of flight repeats. Panels (a)-(f) correspond to 1-6 flight repeats, respectively. Increasing flight repeats progressively converge toward a Gaussian plume shape. The determined fluxes from panel (a) to (f) are: 13.9 kgCH_4/hr, 12.3 kgCH_4/hr, 10.4 kgCH_4/hr, 9.5 kgCH_4/hr, 11.1 kgCH_4/hr, 10.0 kgCH_4/hr, respectively.

Line 397: "Grid cells with no contributing observations (i.e., beyond the cutoff radius of the nearest measurement) are assigned missing values (NaN) and excluded from subsequent analysis."

- An inevitable consequence of this is that if a grid-square is under-sampled, but it potentially contains parts of the emission plume, this will inevitably result in flux underestimation. Please comment on this in the main text.
    - This is correct, thank you for your sharp observation.
- Unless the authors can be sure that each grid square containing flux emissions has been sampled, the entire plume will not be captured, resulting in flux underestimation following this approach.
    - Of course, this is a possibility, and is known as a source of uncertainty in UAS-based sampling. Our flight strategy was designed to ensure that sampling was performed well beyond the expected plume boundary. With this, we ensure we captured the entire plume and do not underestimate flux estimates.

Line 401: "spatially interpolated CH4 concentration"

- Please describe exactly which corrections were applied to the data. Was a calibration and water correction applied, as alluded to in section 2? Perhaps this detail can go somewhere else in this section. But it needs to be stated clearly somewhere.
- I am not totally sure if the sensor data is calibrated and water corrected or not. It should be totally clear if the corrections from Section 2 are used and how.
    - Revised, we added the following sentence to Line 394: "After applying the corrections as stated in Section 2 (i.e. compensation for water vapour dilution and thermal stability), we interpolated the corrected and observed CH_4 enhancements onto a grid using a Gaussian weighting scheme with finite spatial influence."

Line 408: "Δx and Δz"

- Please provide the exact size of these values chosen for this work.
  - Revised and added the following clarification: "… (2 by 2 meters in this study), …"

Line 411: "Mvol gives the molar volume"

- So this equation actually provides molar volume in dry conditions (excluding all water molecules), as the partial pressure of air in dry conditions is used here. This is very nice to see and often overlooked in most research. I am glad that the authors have included this detail. However, this only works if a water correction is applied to raw mole fraction measurements to obtain dry mole fractions. This really must be clarified throughout the manuscript.
  - As a reply to the specific comment on Line 401 we decided to include clarification of the applied correction. We do indeed correct for water vapour based on the correction term presented in Section 2.

Line 437: "The averaging background method is not ideal for the Active AirCore due to its intrinsic signal smoothing (Morales et al., 2022)."

- Please can a bit more explanation be provided here? I don't really understand why the Aircore data cannot be used in the same way as the Axetris data. This is an important point.
  - We added a figure to the appendix (F) showing that the "tail" of the AirCore data is much longer than that of the Axetris data. To use the same strategy for the AirCore means increasing the length of each transect, which is not something we took into account.

Line 450: "the integral of raw and smoothed Axetris data is identical"

- What exactly does this mean? Does this mean that adding up all of the numbers in the smoothed Axetris and Aircore time series is identical?
  - I am sorry for the confusion; this statement actually refers to the Axetris data only (smoothed vs. non-smoothed) not the AirCore. With the statement we try to clarify that smoothing the data does not appreciably change the integral (e.g. the flux) calculated. We acknowledge that the discussion of the integral creates confusion, and removed this statement.
- Please provide the result here, showing how identical this result is. I think "identical" is a strong word, unless it can be shown that it is exactly the same.
  - (Does not apply)

Line 460: "in subsequent analysis"

- Please replace this with either "in subsequent analyses" or "in the subsequent analysis".
  - Revised and replaced "in subsequent analysis" to "in subsequent analyses"

Line 481: "Figure 3"

- What size are the grid cells? I think this detail needs to be included here and in the main text. It is difficult to see any grids in this image.

    - The grid sizes in the figure are not the same as the grid sizes used for the spatial interpolation (which are 2x2 meters). This information was added to subsection 3.7
- What do the black contour lines represent? If this data is averaged into grid squares, there shouldn't be any continuous contour lines.
    - The black contour lines are a feature of the contourf function in MATLAB. The choice between displaying gridded data as "squares" or "contours" based on an assumed smooth field is, for our purposes, arbitrary, and we prefer the contour method. We note that with an infinitely detailed grid (say, of length 1x1cm), the "squares" approach would approximate the "contour" approach.

- The green and purple colours in panel a are impossible to distinguish. They both look grey to me. I suggest using two alternative bolder colours.
    - Revised and changed to deep blue and orange and updated all the graphs in the process.

- How was sampling conducted so close to ground level with a UAV? What was the minimum sampling height (excluding take-off)? Perhaps provide this detail in Table 1.
    - The cirlces in the plot show the actual sampling points where the lowest is at 5 meters above take-off. We note that ground level in The Netherlands typically is exceedingly horizontal; our field site does not show elevation changes more than 70 cm.

Line 535: "θ is the mean-reverting term"

- I would recommend using a different symbol here, as θ is already used in section 3, to describe the flux quantification method.
    - Revised and changed θ to Θ

Line 541: "drone"

- Replace this word with "UAV".
    - Revised and replaced "drone" with "UAS"

Line 457: "4D"

- Please define this acronym.
    - Revised and replaced "4D" with "4-Dimensional"

Line 555: "using a typical flight trajectory from the previous field campaign"

- This means that this analysis is only valid for the type of sampling conducted in this specific UAV sampling campaign. The sampling extent and sampling spatial density are specific to flight 1 in this work. Different uncertainty analysis results may occur if using a different UAV piloted in a different way. I think this point needs to be explicitly made in this section.
    - These results are representative for any sampling attempt that has a sufficiently dense sampling grid as we elaborate on in Section 5. Of course we concur that flying a too sparse sampling grid will result in strongly elevated errors.We do acknowledge the outcome can differ, but this will not result in significant variations. We clarified this in the updated manuscript by adding the following paragraph: "The statistical analysis provides insight into

potential sources of uncertainty in UAS-based sampling. The OU-simulation is appropriate for the specific flight trajectory presented in this study. The uncertainties derived using the simulation remain reasonable when real-world flights are conducted comparable to those simulated. Deviations may arise when observed fluxes are highly variable over a short time (i.e. one hour) or when the flight path (specifically: sampling density) is sparser than the modelled configuration. In those cases, the simulation results may no longer be representative. Flight paths of all flights in this study are comparable to flight 1, therefore OU-derived uncertainties can be applied to all flights."

Line 573: "Flight duration"

- This analysis requires the flux, wind conditions, atmospheric turbulence and position of the plume to remain constant throughout the full duration of each of the six flights. In reality the plume will be influenced by all of these things. Therefore, I am not sure how useful the conclusions of this subsection are in the real world.
  - Even though this is correct, it still shows how multiple flights can lower the uncertainty observed. Of course, in real life scenarios a lot of additional factors play a role, but this point is addressed in subsection 4.2.5 where we combine the uncertainties.

Line 596: "Figure 4"

- Why not combine Figure 4 and Figure 5 into a single figure with four subplots?
  - We decided to combine both Figures, resulting in the following Figure and caption:

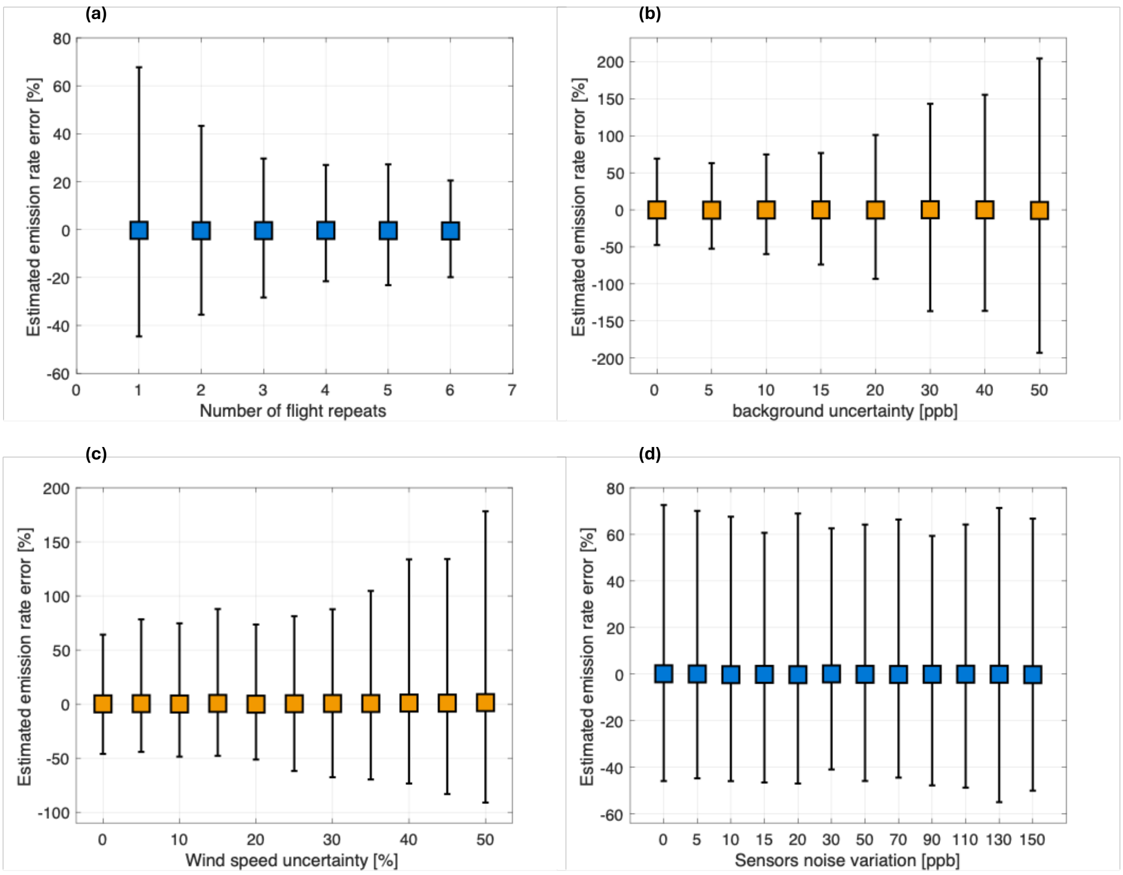


Figure 4: The marker shows the mean error, and the error bars present the 95% confidence interval. (a) An overview of the emission flux estimates and the variability from 500 repeat simulations with an increase in flight repeats. (b) An overview of the uncertainties linked to the background determination. (c) An overview of the emission flux estimates and the effect of the wind speed uncertainty. (d) An overview of the uncertainties linked to the instrumental noise.

Line 642: "improved spatial resolution the nominal plume"

- Consider rewriting this, as I don't follow the text.
  - Revised and rewritten to: "Low-noise sensors enable improved detection of smaller plumes and enhance the spatial resolution of the nominal plume"

Line 663: "standard deviation of ± 40%"

- How does this compare to the standard deviation between the four UAV fluxes from this work? It is interesting to compare here, in this subsection?
  - We addressed this comment by adding the following sentences: "During field deployment, the variability of the observed plume was 39% (1σ, n=4). The observed simulation uncertainties and the field example agree well. Furthermore, the main contributors to the uncertainty identified with the simulation are consistent with those reported in other studies (Andersen et al., 2021; Karion et al., 2013; Morales et al., 2022; Nathan et al., 2015). To the point of our study, drift and sensor noise do not emerge from the uncertainty analysis as dominant."

Line 690: "a persistent uncertainty of approximately 20%"

- But this 20% minimum threshold is specific to the UAV sampling strategy in flight 1. If sampling over a greater area or with a higher spatial density, this uncertainty threshold may be lower. It is worth making this point.
  - This is indeed correct, we added the following statement in the discussion to address this comment: "The statistical analysis provides insight into potential sources of uncertainty in UAS-based sampling. The OU-simulation is appropriate for the specific flight trajectory presented in this study. The uncertainties derived using the simulation remain reasonable when real-world flights are conducted comparable to those simulated. Deviations may arise when observed mass emission rates are highly variable over a short time (i.e. one hour) or when the flight path (specifically: sampling density) is sparser than the modelled configuration. In those cases, the simulation results may no longer be representative. Flight paths of all flights in this study are comparable to flight 1, therefore OU-derived uncertainties can be applied to all flights."

Line 693: "cost-effective CH_4 sensor"

- How much does the Axetris cost? What is the total cost of the finalised UAV enclosure, in terms of raw materials? The authors cannot make such a statement unless they provide at least a rough idea of how much their system costs.
- How much does the AirCore cost for comparison? This also requires a LI-7810 so the net cost is inevitably higher, for this work.
  - This comment is addressed by the table presented above (which is also added in the Appendix (Table A1).

Line 697: “exceeds 39%”

- Why say that it exceeds 39%? It is better to just provide the actual value.
    - Revised. “Exceeds” was not the correct choice of words, since it is equal to 39%. Thank you for pointing this out.

Line 700: “The results”

- Please specify which results are being referred to here.
    - Changed “the results” to “The differences in the standard deviation of all flights indicate that…”

Line 702: “mid-cost sensor”

- In the introduction it is called a low-cost sensor. The cost of the sensor is not consistent, even in this manuscript.
    - Thank you for pointing this out. We realized the Axetris falls more within the medium cost range and we missed some of the metions in our final revision. We made sure the manuscript is consistent
- It is better to provide the cost (even roughly) and stick to either mid-cost or low-cost, with a defined price range for the chosen term.
    - As mentioned before, we decided against this since the total cost of a sensor depends on a wide variety of costs and not a lot is published on this matter.

Line 702: “the Axetris sensors uncertainties are similar to those found during prior UAV-based studies
(Andersen et al., 2021; Karion et al., 2013; Morales et al., 2022; Nathan et al., 2015) using higher precision and accuracy sensors.”

- I disagree with the way the sensor uncertainty is evaluated here. This statement does not represent the sensor uncertainty. It represents the uncertainty in fluxes derived using a specific sensor with a specific UAV sampling strategy. As the authors show in Figure 6, the flux uncertainty in a function of the emission flux rate. Unless all of the previous papers cited in the sentence sampled the same flux rate with exactly the same UAV sampling strategy (and identical atmospheric conditions), this comparison cannot be made. This is not a way to evaluate sensor performance as it conflates other factors specific to each sampling campaign.
    - You are correct that we cannot compare the uncertainties found in this study with other UAS based studies. We agree that a direct comparison is not possible and this was also not our intention. With our uncertainty analysis we try to show that even though the conditions vary, the Axetris is capable of determining the flux within the same uncertainty ranges as found in other studies. We changed the sentence to: “With active temperature control, the Axetris sensors is capable of determining the fluxes within 10% of the AirCore findings and the total uncertainty during the campaign fall within the uncertainty ranges reported by other UAS-based studies (Andersen et al., 2021; Karion et al., 2013; Morales et al., 2022; Nathan et al., 2015) using higher precision and accuracy sensors.”

Line 732: “creating favourable conditions for both IGA and MBA”

- As mentioned in the introduction, I don't think mass balancing and Gaussian plume modelling can be split into two totally different approaches, as in this paragraph. Both methods are a form of mass balancing as they compare upwind measurements (assumed background) to downwind plume enhancements. The key difference is that Gaussian plume modelling attempts to model a continuous plane from available sampling using Gaussian statistics whereas simple "mass-balancing" integration methods do not. These simple mass-balancing methods either spatially interpolate available data or average the data into a grid, as in the method presented here.
    - We concur that both methods are "mass balance" approaches that produce, each in its own way, an interpolated cross-section of a plume. For both methods, this cross-section must be multiplied by the (often ill-constrained) wind profile, making both methods equally susceptible to wind speed variability. Indeed, these methods do not at all differ in the latter half of their application, but merely in the initial inference of the cross-section. Perhaps these methods should be distinguished between only in that respect, and be referred to as, say, "local interpolation MBA" vs. "gaussian fitting MBA" (we'll note that we don't make things any easier by performing our local interpolation by gaussian smoothing, which has nothing to do with IGA). However, while we agree with your rationale, we consider that proposing new terminology would also create inconsistency with existing literature and therefore prefer to clarify it instead.
- All flux methods (including Gaussian plume modelling) are affected by wind speed. This is not exclusively a problem for the grid-square averaging integration method presented in this work. This paragraph may therefore seem to be misleading.
    - Having said all of the above, we do not believe that this makes the "IGA" and "MBA" methods equally suited to all circumstances. Specifically, we believe it important to consider both the degree to which a plume is represented by the sampling along the flight path and the conformance of its shape to "gaussian". With non-complete sampling of the plume, IGA may "fill in the blanks" in ways that local interpolation (such as used in this study) does not. When the plume *is* completely sampled, *and* is of gaussian shape, local interpolation and IGA will produce similar results. In the situation of a completely sampled but *non-gaussian* plume (due to, say, excessive turbulence or wind *direction* variability), we believe that IGA may produce overly simplistic results. We write "Wind direction during the flights was variable, leading to horizontally stochastic plume observations, making appropriate the use of a method that does not rely on Gaussian plume assumptions."

Line 739: "making appropriate the use if a method that does not rely on Gaussian plume assumptions."

- I disagree. If enough sampling is conducted, Gaussian plume methods can also deduce the average emission flux, as a turbulently moving plume and the UAV randomly intersect. These random intersections should average out over time. It certainly isn't better for a non-Gaussian approach, such as for the grid-square averaging approach in this work. The assumption in this work is that each grid square will include a statistical sample of the moving plume as well as the background.
    - Certainly, given enough sampling, both methods converge. Indeed, we show in an added figure FigureK2 that our 'local interpolation' method does converge

towards a Gaussian shape with more sampling (in this case of a modelled plume). We remain of the opinion that for a well-sampled, non-gaussian-shaped plume, regular local interpolation produces a more appropriate result that gaussian fitting. We cannot follow reviewer's assertion about each grid cell including a sample of both the plume and the background.

Line 753: “in situ CH4 sensor”

- I think the sensor should be named here. Otherwise it is a bit of an empty statement.
    - Revised and added: “…in situ CH_4 sensor (referred to as the Axetris)…”

Line 761: “lab”

- Replace this with “laboratory”.
    - Revised and replaced “lab” with “laboratory”

Line 768: “the close agreement”

- Please add a sentence here stating that different background methods were used for the difference sensors. This is an important detail, to emphasise that the data was not processed in the exact same way.
    - Revised to: “The close agreement between the techniques, despite using different background determination methods tailored to each technique, validates the use of the Axetris sensor for reliable flux quantification.”

Line 768: “robust”

- Consider the implications of this strong word. A 20% uncertainty may not be that robust for some.
    - Thank you for pointing this out, we decided to replace “robust” with “reliable”

Line 778: “Multiple flights”

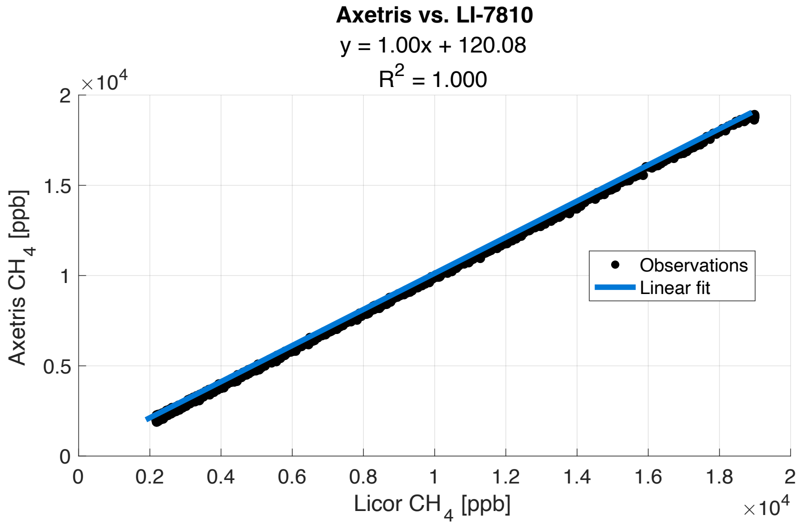
- It is important to include the caveat here that this requires the emission flux to remain the same and atmospheric conditions to remain exactly the same.
    - We do not fully agree with the caveat pointed out. With this type of research we try to approach facility scale emissions of the chosen anthropogenic source. In reality, there are a lot of factors influencing the observed flux. However, changing wind does not directly affect the obtained flux, since enhancements scale with wind speed and multiple flights will definitely improve the effect of changing wind directions.
    - Due to inherent sensor noise of the sensor and time varying fluxes, your flux estimate always has a certain uncertainty from the start. Doing multiple flights will approach the “true” average of that day. The average is therefore better approached by conducting more flights.
    - If instead we wanted to identify and quantify a short lived event, it will become important that the emission flux remains constant, since doing multiple flights would indeed be problematic.

Line 848: “https://icos-atc.lsce.ipsl.fr/docs”

- This link does not seem to work.
    - Reference has been updated with a working link

Line 933: "Figure B1"

- It is really difficult to distinguish the line from the points. Perhaps make the line bright red with small black dots for data.
    - Revised and changed the line to blue to match the colour scheme of the manuscript



Line 954: "drone"

- Replace this word with "UAV".
    - Revised and replaced "drone" with "UAS"

Line 961: "drone"

- Replace this word with "UAV".
    - Revised and replaced "drone" with "UAS"

Line 968: "drone"

- Replace this word with "UAV".
    - Revised and replaced "drone" with "UAS"

Line 974: "drone"

- Replace this word with "UAV".
    - Revised and replaced "drone" with "UAS"

Line 980: "drone"

- Replace this word with "UAV".
    - Revised and replaced "drone" with "UAS"

Reply to reviewer number 2, Anonymous review

Dear Anonymous Referee,

Thank you for your thorough review of our manuscript. Hereby, I present the replies to all your comments (in green).

General comments

Dear Editor,

In this article, the authors present a declination of an already existing methodology of methane emissions quantification with the use of an alternative instrumentation. They present laboratory and real fields tests of this instrumentation, to better characterize its specificities. They also performed real field quantifications of methane sources, with both the reference instrumentation and the newly tested instrumentation in parallel, which allows a direct comparison of the performances of each method. They also performed a detailed statistical analysis of some of the main sources of uncertainties associated with this emissions quantification, based on one of the tests performed on the field.

They highlight the strengths of this methodology and the strong interest of such low-cost instrumentation for a generalization of this type of measurements of methane emissions.

This paper has a great interest in the current context where many groups are developing airborne methods based on Uncrewed Aircraft Systems (UAS) to monitor methane emissions at the facility scale, filling a significant gap in the greenhouse gases emissions reporting. Furthermore, the emphasis on the uncertainty analysis is very beneficial as this is crucial for the validation of such method.

The topic of this article corresponds to the scope of Atmospheric Measurement Techniques. This paper is generally well written and very clear. However, there are still minor aspects which could be improved before final publication.

General comments

The authors presented intercomparaison of emissions quantifications based on Axetris and active AirCore measurements. For the data treatment of the Axetris measurements, the authors applied an H2O correction based on AirCore H2O measurements.

- What is the reliability of H2O measurements of an active AirCore? In the case where the Axetris would be used as a stand-alone instrument without AirCore, the authors suggest that the H2O values can be derived from model or nearby meteorological station values. But they don't present any estimates of the performance of the quantification method using such values, compared to AirCore H2O measurements. It would be interesting to evaluate the performances of the method without having access to AirCore H2O measurement, as the Axetris should at the end be employed as a standalone instrument. An alternative could also be to embark an extra H2O sensor on-board the UAS on top of the Axetris. It would also be interesting to compare the

AirCore H2O measurements with the humidity values measured by the embarked Trisonica mini, if they have been logged.

- Thank you for your comment, and that is a fair point of critique. However, it should be noted that the water vapour concentration observed by the LI-7810 were used, and not from the AirCore directly. During field deployment, the AirCore samples ambient air through a chemical dryer, removing most of the water before analysis. The AirCore measurements are therefore corrected differenty compared to the Axeteris, which still needs a water vapour dilution correction.
- The water vapour concentration during this field experiment is determined by the LI-7810, which was able to sample ambient air in between flights. The average value obtained from this was used for the general water vapour dilution correction. We added the additional suggestion to obtain the approximate [H2O] from RH measurements reported by local meteorological sites, in case one does not have a LI-7810 at hand. This convenient approximation contributes only the tiniest error to the method (e.g., if using, at 10°C, 65% RH instead of a true 70% RH, the commensurate [H2O] is 0.79% vs 0.85% – resulting in an error in the mass emission rate estimate of 0.06%).
- The addition of an extra H_2O sensor is a really good suggestion and is also something that we are considering for the next version of the Axetris package. We do log the humidity values from the TriSonica mini but believe the meteorological data from the nearby stations is sufficient.

- Despite the good quality of the monitoring methodology and the uncertainty analysis, there are still some gaps in the protocol and the analysis, which are still limiting the scope of this study. One would expect more emphasis on these limits in the discussion.
  - We added some ideas for future studies in the Discussion to address how to ensure some of these limitations can be brought to light: "Future studies should concentrate on a systematic comparison between the Axetris and high-precision, high-frequency in situ instrument. Similar to this study, both sensors should be deployed simultaneously, to allow direct comparison to mass emission rate estimates using comparable data analytics. Furthermore, large-scale controlled release experiments could provide a quantitative assessment of the sensors' performance."
  - Additionally, we added some additional points to the discussion of out uncertainty analysis to highlight what was not considered during this analysis: "The statistical analysis provides insight into potential sources of uncertainty in UAS-based sampling. The OU-simulation is appropriate for the specific flight trajectory presented in this study. The uncertainties derived using the simulation remain reasonable when real-world flights are conducted comparable to those simulated. Deviations may arise when observed mass emission rates are highly variable over a short time (i.e. one hour) or when the flight path (specifically: sampling density) is sparser than the modelled configuration. In those cases, the simulation results may no longer be representative. Flight paths of all flights in this study are comparable to flight 1, therefore OU-derived uncertainties can be applied to all flights.
    However, not all sources of uncertainty have been considered. First, this study did not include the effects of spatial sampling density, which can also affect

the accuracy of the mass emission rate measurements. Significant data gaps negatively affect the MBA, leading to systematic underestimation of the mass emission rate. This limitation has been addressed in detail by Mohammadloo et al. (2025). Higher-frequency observations in the horizontal and vertical directions decrease the estimated emission rate error. Based on their findings and our own flight track featuring relatively dense horizontal and vertical coverage, we expect to have only an additional 5-10% error. Furthermore, wind direction uncertainties are also not included in the statistical analysis. Changes in wind direction influence the dispersion and detection of atmospheric plumes and is often referred to as the meandering effect (Wietzel et al., 2025). As mentioned before, wind variability (e.g. changed in wind speed and wind direction) propagate non-linearly through the MB calculations, therefore affecting the mass emission rate estimates in varying ways, depending on the error. Future work should combine optimised flight-path design with real-time plume detection to further minimise spatial sampling bias."

- Concerning the characterization of the instrument: the authors performed an experiment to evaluate the stability of the analyser to variations of the temperature of the air surrounding the instrument, but the temperature of the incoming air is not mentioned (probably laboratory temperature) and also not varying. This would be useful to test the influence of the temperature of the incoming air, as it might probably influence the cell temperature stability: different temperature gaps between the instrument regulated temperature and the incoming air temperature could be tested, to simulate the behaviour of the instrument in different types of environments or seasonality. One could also extent these tests to measure the impact of temporally varying temperatures on one side (simulating the variations which might be encountered if measurements are performed in an industrial environment with warm plumes (under the wind of natural gas flares for example). Regarding the humidity sensitivity experiment, the authors proposed a humidity generation method based on a heated wet paper towel. This allows a strong variation of the humidity of the incoming air, but the variation of the temperature of the incoming air is not discussed. Could it have an impact of the spectroscopic response of the instrument? Ideally, humidity sensitivity tests rather be performed at constant temperature. This also justifies the need of the already mentioned sensitivity test to incoming temperature.
    - Thank you for your comment. This is indeed something we did not consider directly. However, we believe that having the active temperature control to stabalise the cell temperature and therefore also limit the temperature variability. Of course it can be possible that the sensor is sensitive to the temperature of the incomming air, but seeing the great improvement with this temperature control we believe this effect to be negligible. Even when varying air temperatures are sampled, the sensor readings are more stable with the active control compared to without.
    - Numerical example: at a pumping rate of 200 sccm, air of 10°C flowing into the cavity of 25°C would have a cooling capacity of 0.05 W. If the incoming air temperature would vary by 5°C (with time, altitude, flight direction or similar), that would vary the cooling capacity between 0.03 and 0.07 W. That variability certainly is of absolutely no consequence. Any very gradually accumulating effect is effectively compensated by the controlled heater (max. power: ~2 W).

    - Besides the smapling pathway during the water vapour experiment was more than 2 meters, which already equilibrates the air temperature before it reaches the cell temperature.

- Concerning the uncertainty analysis, it is based on the analysis of one case based on a single flight. Although it provides an overall good interpretation of the relative contributions of the different sources of uncertainty for this case, the conclusions are therefore difficult to generalize to any monitoring case, where many parameters might be different such as types of sources (punctual or diffuse, low or high elevations, with/without ejection speed, single or multiple sources), different topographies, presence of buildings/obstacles, varying distances between source and monitoring plane, different wind conditions (mean speed and turbulence). This uncertainty analysis could also be further developed to study the influence of other sources of uncertainties in this case, such as the quality of the wind direction measurements, or of the quality of the H2O measurements.
    - Thank you for your comment, this is of course correct. We tried to implement as many possible sources of uncertainty that are known sources of uncertainty with higher precision sensors to determine how this might influence the sensors uncertainty during UAS deployment. You don't optimise flight for the terrain roughness or similar, but for plume shape and variability thereof. If that's wide or complex (e.g. due to topography or ejection velocity), that plume is in principle still well-quantifiable with our method, provided sufficient repeats or flight duration.
    - It is true that this uncertainty analysis is optimised based on one specific case and each case is slightly different. However, we believe that an uncertainty analysis like this can still give a good approximation of the biggest sources of uncertainty and does give a good uncertainty range for the UAS applications with a single source and curtain flight pattern.

- To put the uncertainty analysis of this single case in a broader context, it would be interesting to compare these uncertainties with the variability of the quantifications obtained between repeated flights on the same source. Are the estimated levels of uncertainties coherent with the real observed variability?
    - This is correct, and indeed a good addition to the manuscript. We did not explicitly state this in the first version of the manuscript, but we do believe this to be a nice bridge between the statistical analysis and real-life scenarios. We specified this in the updated version and implemented as:
    - "This simulation was used to estimate the overall combined uncertainty during our field campaign for a single flight. With the above-mentioned setup, the obtained flux was 10.0 kgCH_4/hr ([3.3–19.1 kgCH_4/hr; 95% CI), with a standard deviation of ± 40% (1σ). During field deployment, the variability of the observed plume was 39% (1σ, n=4). The observed simulation uncertainties and the field example agree well."

- The proposed uncertainty analysis is limited to one case and does not allow a quantification of uncertainty for any single flight. I would also like to see in the discussions if the authors think that the proposed uncertainty analysis method could be applied to any single flight to evaluate the uncertainty of the flux quantification for each individual flight.
    - Thank you for this comment. We added an additional paragraph to the discussion to address this topic. We believe that our uncertainty analysis does give a good representation of sensor uncertainty during UAS deployment. However, applying this uncertainty to any flight is not possible. (There are limitations). The following paragraph was added to the discussion to address this point:
    - "The statistical analysis provides insight into potential sources of uncertainty in UAS-based sampling. The OU-simulation is appropriate for the specific flight trajectory presented in this study. The uncertainties derived using the simulation remain reasonable when real-world flights are conducted, comparable to those simulated. Deviations may arise when observed fluxes are highly variable over a short time (i.e. one hour) or when the flight path (specifically: sampling density) is sparser than the modelled configuration. In those cases, the simulation results may no longer be representative. Flight paths of all flights in this study are comparable to flight 1, therefore OU-derived uncertainties can be applied to all flights."

Specific comments

Wording: I think the gender-neutral policies of EGU suggests using the term "Uncrewed Aircraft Systems (UAS)" instead of "UAV". Also replace "drone" with the appropriate term (either UAV or UAS) throughout the text.

- Revised and replaced all instances of "UAV" with "UAS" where applicable. The reference list still mentions the term "UAV" due to references used

Line 131: "the ground time » > the ground team?

- Revised and replaced "the ground time" with "the ground team". This was indeed a spelling mistake.

Line 200: It seems that the LI-7810 Is used to monitor the dry-air CH_4 concentrations, but the Licor is technically also measuring wet air. What correction did you use to calculate the dry-air CH_4 concentrations from Licor measurements? Is it a manufacturer correction? Did you validate this function?

- This is indeed something to consider. The analyser has a build-in water vapour dilution correctoin. We validated this correctoin works appropriately (i.e. the instrument reports "bone-dry" mole fractions, irrespective of the actual water vapour content).

Lines 212-215: I agree with the following remarks that the errors introduced by water corrections would be negligible compared to other uncertainties. However, wouldn't it be possible to use the same protocol at different CH_4 concentrations to check the stability of this correction?

- This would indeed be a good possibility. We performed the same protocol at a different CH_4 concentration, close to ambient, to determine the effect at lower concentrations. This additional experiment is added to the Appendix (Appendix; Figure C3 and C4) and shows that at lower concentrations the water vapour effect cannot be distinguished from the sensors noise.

Line 250 (Figure 2): It is regrettable that the time series of the in-flight tests are not shown here, as for the laboratory tests.

- Thank you for your comment. We added the time series of the in-flight test to Figure 2, in a similar manner as the laboratory tests. We also updated the color scheme so the time series match the Allan deviation graphs.

Line 255-265: It is not clear whether the incoming air sample temperature was stable or varying as well during the tests. On the field, one does not necessarily expect strong air temperature changes during a flight (at least as long as no warm source such as fires or industrial sources are monitored), but there might also be a strong temperature difference between the cell temperature and the outside air temperature. Would it also affect the measurements or is the cell temperature regulation sufficient to compensate important outside-air to cell-temperature gradients?

- Thank you for this comment. During the temperature dependency experiments, we sampled the air from a cylinder. The temperature of the incomming air is not significantly different compared to the cell temperature and we therefore do not believe this influences the results much. The active temperature control is sufficient to compensate for this difference, since we did not experience any visible effects.

Line 324: "was derived as the average the WindMaster Pro observations" > "the average of the"

- Revised and replaced "was derived as the average the..." to "was derived as the average of the…"

Line 358-370: Information is missing here about the distance between source and observational plane.

- Thank you for pointing this out. We added a sentence stating the average distance between the observational plane and the source.
- "The flights were conducted under light to moderate wind conditions (3.7 ± 1.4 m/s, NW to WNW; Appendix E) and consisted of 12 transects between 4 and 60 m above ground level at approximately 120 m downwind of the farm (Appendix H)."

Line 430: This method might work but is subject to interpretation and is time consuming. Alternative baseline fitting methods exist which could be employed to estimate the background concentrations, even with noisy datasets once the appropriate parameters have been found. They could be applied either to individual transects or the complete time series of observations at once.

- Thank you for this suggestion. We concur that the baseline fitting method possibly is more robust than the method used in this manuscript and less prone to errors. An additional benefit is the applicability of this method for the Axetris and the AirCore. However, due to our limited testing of the baseline method, we opt to retain in this manuscript the originally presented method but will certainly strive to incorporate the baseline method into future work.

Line 460: « completed eliminated » > « completely eliminated »

- Revised and replaced "completed eliminated" to "completely eliminated"

Line 467-468: In this complete section, the authors present the example of the second flight only, but here there is a mix of results from all flights and individual flights. It should be more coherent. Furthermore, the mean difference of all flights is discussed, but the average value does not appear either in the text at this point or in the Table. I suggest moving these remarks to the following paragraphs treating all flights (lines 486-499).

- I am not entirely sure what this comment refers to. The lines mentioned contain information about the average wind conditions of the day and naturally contain information about all the flights. The paragraph above (Lines 461- 466 in the original manuscript contains information about Figure 3 and refers to flight 2. However, the last sentence might be the one you refer to, stating: "The emissions estimates derived by the two methods (AirCore, Axetris) differed by less than 10%, both for individual flights as for the mean of flights." We therefore removed this sentence here and merged it to form the following sentence: "The correlation between the individual flight observations is strong and the emissions estimates derived by the two methods (AirCore, Axetris) differed by less than 10%, both for individual flights as for the mean of all flights (Table 1)."

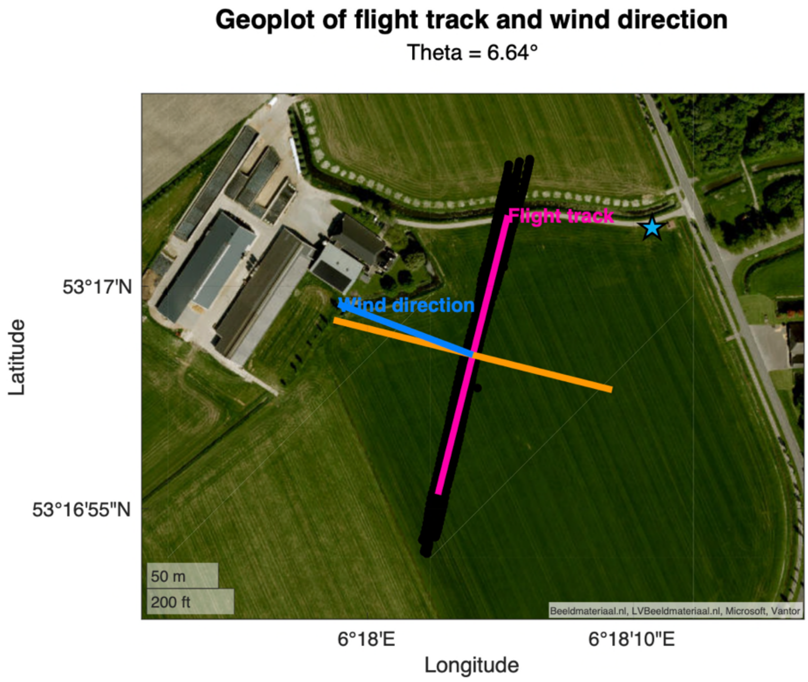
Line 470: The authors are referring to Figure D1 (appendix), where the legend could be more precise in terms of period of observations: what time period is exactly represented (it is the complete day or only the period when the flights occurred)? From observations at which frequency?

- We changed the caption of Figure F1 (Figure D1 in the original version) to include the notion of the Windmaster Pro data. The caption now reads: "*Figure F1: Wind rose of Windmaster pro observations during the campaign day at 29-07-2025*". As stated, the observations are from the complete day, and the Windmaster Pro collects observations at 10 Hz. The additional figures (Figures E2 to E6) are flight-specific.

  They could also refer to Table 1 where the mean wind speed and directions of each flight is presented.
- Table 1 gives flight-specific data, while Appendix E1 shows the observations during the entire day and shows that the wind speed and direction did not vary significantly during the day. Since we are discussing the daily average, we will reference Figure E1 rather than Table 1.

Line 471-472: The information about flight plan should be presented earlier in the "Fight strategy" section, in particular the distance between source and observations which is lacking in this section. You are also referring to Appendix G, where the figure are lacking a colorbar for CH4 concentrations (or you could only present UAS positions instead of CH4 values, which are hardly readable).

- Revised and we moved the information about the flight plan to section 3.5 (flight strategy). This section now mentions the distance between the source and the observations and also the number of transects and height above ground: "The flights were conducted under light to moderate wind conditions (3.7 ± 1.4 m/s, NW to WNW; Appendix E) and consisted of 12 transects between 4 and 60 m above ground level at approximately 120 m downwind of the farm (Appendix H)."
- Thank you for pointing it out. We changed all figures in Appendix H to only show the UAS position (black) instead of the CH_4 concentrations, and we changed the caption to complement this. We agree that the CH_4 concentrations were hard to read and adding them to this top down view does not add any valuable information. All figures now look something like this:



-

Line 615-625: Wind speed uncertainty here only represents the uncertainty of the mean wind speed. However, wind uncertainty can also affect the wind direction, which is not considered here. This would be interesting to add this additional uncertainty into the analysis.

- You are indeed correct and this is certainly something that should be discussed more clearly. We spend some sentence to address this by adding onto the paragraph describing the influence of spatial sampling density to the discussion. We decided to leave it out of the analysis due to difficulty modeling the correct interpretation onto the statistical analysis to represent real life movements of the plume.
- We did add some discussion about this point into the revised manuscript: "Furthermore, wind direction uncertainties are also not included in the statistical analysis. Changes in wind direction influence the dispersion and detection of atmospheric plumes and is often referred to as the meandering effect (Wietzel et al., 2025). As mentioned before, wind variability (e.g. changed in wind speed and wind direction) propagate non-linearly through the MB calculations, therefore affecting the mass emission rate estimates in

varying ways, depending on the error. Future work should combine optimised flight-path design with real-time plume detection to further minimise spatial sampling bias."

Line 631: correct "may stem from to changing"

- Revised and replaced "may stem from to changing" to "may stem from changing"

Line 639: "to accurate resolve" > "to accurately resolve"

- Revised and replaced "to accurate resolve" to "to accurately resolve"

Line 642: "improved spatial resolution the nominal plume" > "improved spatial resolution of the nominal plume"?

- Revised and changed the complete sentence to: "Low-noise sensors enable improved detection of smaller plumes and enhance the spatial resolution of the nominal plume"

Line 676-691: I agree with the interpretation that the uncertainties associated with noise level and background concentrations estimate will probably be the factors that play a role in the dependency of the uncertainty to the emission rate. However, this would have been preferable to perform the same analysis for each individual source of uncertainty, to really justify this assumption.

- Thank you for pointing this out and we understand the critique. However, we tried to address this concern by pointing out that by increasing the flight repeats (e.g. fly more during the same day) the uncertainty at low emission rates benefits more compared to higher emission rates by stating "random errors are effectively averaged out through repeated sampling." Additional meteorological errors play a bigger role when emission rates increase, and noise and background uncertainty have less of an influence (since percentually, it would be difficult to be off more). The uncertainty of each individual source was conducted in section 4.2, but at lower emission rates.
- We did add an additional Figure to the Appendix L showing the effect of each individual source of uncertainty for a source of an emission mass rate of 1 kg/hr and refered to it at the end of the statement:

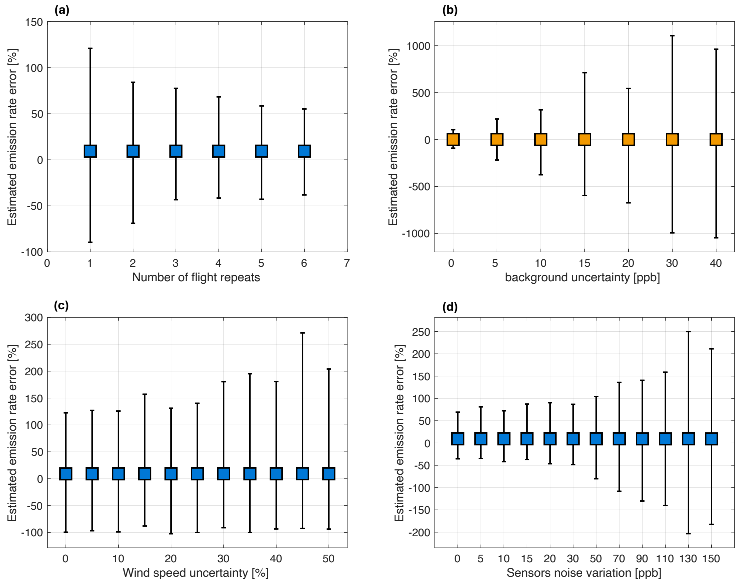


Figure L1: Overview of the individual sources of uncertainty and their respective effect on the estimated error based on an emission mass emission rate of 1kg/hr.

Line 732: define “IGA »

- IGA is already defined earlier in the paper (line 72)

Line 740 : « the use if a method » > “the use of a method”

- Revised and replaced “the use if a method” with “the use of a method”

Reply to Reviewer number 3, Anynymous review

Dear Anonymous Referee,

Thank you for your thorough review of our manuscript. Hereby I present the replies to all your comments (in green).

General comments

The paper discusses initial validation tests with the cost-effective medium-precision Axetris sensor compared to a high-precision established sensor. Lab tests by Van Ettinger et al highlighted the importance of temperature control to reduce sensor noise. The Axetris sensor was then deployed on a drone to measure the methane emission rate from a dairy farm; the testing was conducted in parallel with the established AirCore method and the results show potential of the Axetris sensor for use in the mass balance approach. The authors conduct a detailed sensitivity analysis, which highlights the main factors of error in the method.

I think the work is relevant for many researchers and environmental scientists in the field of methane emission measurement. The novelty of the paper lies in the insight that a cost-effective medium-precision sensor is capable of providing accurate results on methane emission quantification. The uncertainty analysis is more extensive than what is usually found in the literature and that's very helpful. The paper is very well written in general, although I think that some more deeper discussion can be provided on a few specific topics (listed below). I also find that a few graphs are unclear at the moment; a better description may be needed there. Overall, I would like to recommend the following minor revisions before publication in AMT.

Specific comments

The title is very generic at the moment, and does not capture some of the main points of the paper. From a practical perspective, I think the key point is that the new Axetris sensor is cost-effective but still can provide promising results; I would therefore add something along these lines of "a cost-effective in situ methane sensor". Furthermore, on the title, I think "point source emissions" is too limiting: first of all, the Mass Balance Approach also applies to diffuse sources within a certain domain. And second, one could argue that a dairy farm used in the experiment produces more diffuse emissions (less of a point source, indeed) than, say, a vent stack in an oil & gas facility. I would therefore recommend to replace "quantify point source emissions" to something like "quantify emissions at facility level" or just "quantify emissions".

- We changed the title to be more specific. It currently reads: Evaluating the performance of a cost-effective in situ methane sensor for UAS-based systems and its ability to quantify facility-scale emissions.

Drone-based methane sensing is a rapidly developing field, and I see that many references to existing measurement campaigns are from several years ago. For some latest references, I would recommend to read and refer to, e.g., the following recent papers and references therein:

- AMT – Controlled release testing of commercially available methane emission measurement technologies at the TADI facility (Audrey McManemin, Catherine Juéry, Vincent Blandin, James L. France, Philippine Burdeau, and Adam R. Brandt, 2026)

- Validation and demonstration of a drone-based method for quantifying fugitive methane emissions - ScienceDirect
(C. Scheutz, J.E. Knudsen, N.T. Vechi, J. Knudsen, 2025)

I will come back to these references below, because some of the conclusions in these papers are in line with what Van Ettinger et al reported so it will be good to make that connection.

- We added some newer references to the paper. Based on the references suggested above. We decided to only include references talking about specific sensor development and no papers talking about a wider application, but we did read through the suggestions above. The paper from Scheutz et al. 2025 also uses the Axetris sensor, but leaves out the sensor characterization and focuses more on a "new" flux quantification technique.

If the key novelty in van Ettinger et al's paper, over other established methods using expensive high-precision sensors used in the references above, is in the Axetris sensor that is low-cost but requires temperature control, then a few considerations need to be discussed in further depth. First, it is unclear now in how far Axetris themselves have already done lab testing before bringing the sensor on the market. What has already been tested, and how does the lab testing by the Ettinger et al build on that? Second, for anyone wanting to deploy the sensor on a drone themselves, it will be important to know how the temperature control is implemented. What is required? Is this available on the market or do people have to develop it themselves? What is the complexity of this solution, i.e. how much would it cost approximately?

- Thank you for your comment. This is indeed something to consider. We have had some contact with Axetris during the writing process, and they helped us better understand the sensor. However, as fas as we know, Axetris themselves have not published a sensor characterisation details pertaining to temperature stability. Most of the information about the sensor is available in their brochure, where they state the detection range etc. They are aware of the need for temperature stability and mention (in the brochure) that with the help of an included temperature sensor, the cell temperature is measured and the concentration observations are consequently compensated. We establish that this "compensation" is not a functional part of the final design, which is why we advice the additional active temperature control.

- To address the second part of this comment, we are planning on making the design of the hardware involved in the Axetris package open source, allowing other laboratories and companies to implement it into their system design. Currently, we are making some improvements to the design to clean it up and add some additional sensors to broaden its use. As soon as this process is complete we will make the final design available. As stated in the description of the Axetris package, the required hardware is absolutely minimal. Component costs are less than €100,= (MOSFET + heating wire + tape). The only requirement is that a microprocessor is available to execute the control loop. In practice, this criterion is commonly met.

Another issue is the wind measurement implementation. I understand from Van Ettinger et al that they did measure the wind at the location of the UAV, but in the end the data was not used for the mass balance calculations. Some discussion is warranted about the implementation that Van Ettinger et al tested. Why did it not give useful results? Is this something that can be improved in future? Let me say that it is not easy to measure wind at the location of the UAV accurately, but e.g. Scheutz et al (2025) do manage it and show the value of the data in the methane emission quantification analyses. Some discussion is needed about how important this may be and what can be done in the future to improve wind measurements at the drone.

- Thank you for your comment. We decided to use the 'trisonica' wind measurements at the UAS location only for illustrative purposes, since we are not confident that our UAV-movement correction method yields sufficiently accurate data. Using the coordinate-specific wind data can be a great improvement if implemented correctly (as can be read in Scheutz et al., 2025). And we are also planning on implementing this for the next publication. However, we still used the uncorrected TriSonica wind measurements to verify our logarithmic profile (as can be seen in Appendix E).

In the field trial, the Axetris sensor is compared to the AirCore method. I understand that this may have been a practical choice due to availability of the equipment, but it does raise the problem that there are two distinct differences have to be evaluated: the difference in performance between the sensors, and the potential difference induced by the use of the data analytics required on the AirCore data (sampling, mapping of concentration data onto the original UAV trajectory, etcetera). In an ideal scenario, I think it would have been preferable to compare the Axetris sensor to another (expensive but established) in situ sensor that gives high-precision and high-frequency data, mounted on the same UAV: then the readings for the concentration measurements can be directly compared to each other, and this will probably lead to further insights on theperformance of the Axetris sensor. Of course I am not suggesting that such an experiment should be included in the current paper, but I think it does deserve to be mentioned as a suggestion for future work.

- Yes you are correct. Unfortunatley we were limited to the equipment that we had access to. In other work done by Morales et al. the Active Aircore was compared to an open path in situ QCLS. We added the following sentence to the discussion to addres this point: “Future studies should concentrate on a systematic comparison between the Axetris and high-precision, high-frequency in situ instrument. Similar to this study, both sensors should be deployed simultaneously, to allow direct comparison to flux estimates using comparable data analytics. Furthermore, large-scale controlled release experiments could provide a quantitative assessment of the sensors’ performance.”

Based on the observations above, I feel that some conclusions are a bit overstated, in particular “The close agreement between the techniques validates the use of the Axetris sensor for robust flux quantification” (line 769). I think it is premature to use words like “validates” and “robust” for a method that has been tested on only five flights downwind (of which four flights were successful) of a dairy farm, from which the mass emission rate was not known at source level. I agree with the authors that the Axetris implementation has potential and that the results are worth reporting on in a scientific paper, but further testing will be required in different wind conditions and at different facilities before a robust validation can be claimed. In this light, I think further test campaigns, especially with controlled releases of methane, can be suggested as next steps in the Discussion or the Conclusions, to further evaluate the Axetris implementation – in addition to the side-by-side testing with a high-precision sensor as suggested above.

- We changed the tone of this statement so we do not overstate the conclusion. Right now it reads: “The close agreement between the techniques, despite using different background determination methods tailored to each technique, validates the use of the Axetris sensor for reliable flux quantification.”

- The following sentences were added to the discussion and mentions future study objectives: “Future studies should concentrate on a systematic comparison between the Axetris and high-precision, high-frequency in situ instrument. Similar to this study, both sensors should be deployed simultaneously, to allow direct comparison to flux estimates using comparable data analytics. Furthermore, large-scale controlled release experiments could provide a quantitative assessment of the sensors’ performance.”

Technical question/ issues/ suggestions

- Line 66: examples are from 2016 and 2021. Please consider adding more recent papers on UAV deployments. E.g. the two papers mentioned above and references therein.
  - Revised and added some additional references.
- Line 74: I know that some papers in the literature do use the word “flux” as a synonym for “mass emission rate” but from a fluid dynamics perspective the concepts are distinct. A flux is a mass emission rate per unit of area (e.g. in g/(m2s), for a diffuse emission from a large area like a landfill or biological processes). Here and henceforth, I would therefore recommend to use “mass emission rate” for the quantity that has actually been measured from the farm – not “flux”.
  - Revised and changed all the instances of “flux” to “mass emission rate”

- Line 131: what does “ground time” mean?
  - This was a spelling mistake, and we revised it to “ground team”

- Line 175 (equation): the equation does not go through (0,0). Is this a problem, for instance, when extrapolating these results to higher or lower concentrations than observed in the field trial?
  - We understand your concern, but this should not be a problem. The test was conducted over a large range of concentrations (2 – 20 ppm). 2 ppm is a background level one would usually measure during ambient air experiments, since this is close to the mean global concentration. You would therefore not really encounter any concentrations that are significanlty lower as the those considered during this test. Of course, at very high concentrations, the system may not behave linearly, but since we measure emission spikes, these events are short-lived and should not be strongly affected by nonlinearity.

- A related question on the Figure B1 (page 30): why do the outlier points in Figure B1 seem to lie on a hyperbolic tangent-like curve rather than on random distribution around the straight fitting line? Perhaps it’s my limited understanding, but I don’t think there’s an explanation in lines 167-181 for this hyperbolic tangent-like data.
  - This is an artifact of the difference in the response time. The data included the entire time series of the experiment, which starts with sampling ambient air and then rises to the high emission inside the flask. The Axetris simply reaches that plateau sooner. To remove the hyperbolic-tangent-like data (which are not meaningful for the experiment), we decided to remove the observations associated with the rise. This resulted in an updated Figure B1, which can be found in the updated manuscript. Removing the tangent-like behaviour also changed the linear equation, which is updated as a result of the change.

- Line 187: as I mentioned under the "Specific comments" , I am curious to know if Axetris hasn't performed any such measurement themselves on their sensor before bringing it to market. If they did do these tests, what were the results?
    - To our knowledge, they did not do such tests themselves. We have been in frequent contact with them and brought the extent of the temperature dependency to their attention. Axetris AG mentions being aware of some effect, but were unaware of its severity. They state that the oscillations we describe are negligible for the (i.e., their) intended use case of the sensor (at much higher concentrations than we see)

- Lines 301: "for direct comparison of our results with theirs". It should be added that there is no evidence that the methane emission rate from the site was the same though. In fact, this point is made later in the paper, but it would be appropriate to mention it here already as well. This is not a controlled release test where the emissions are controlled and known to be the same as what Vinkovic measured.
    - Revised to: "Emissions from the farm were previously quantified by Vinković et al. (2022), allowing us to identify similarities or differences in the emission pattern of the farm." As you mentioned, it is not possible to do a direct comparison, but comparing the data can still give us insight into the differences between the observation days.

- Line 326 and below: "This logarithmic wind profile is used for calculation of fluxes" . As mentioned under "Specific comments" , some discussion may be needed about the causes of the noise in the TriSonica wind measurements that made that these in situ measurements could not be used in the mass balance approach (compare for instance with the discussion in Scheutz et al (2025)).
    - Thank you for pointing this out. As stated before, we flew with the TriSonica attached, but we did not use the direct observations from the TriSonica because they still required additional corrections that were not ready before we sent this manuscript for publication. We will add an additional sentence, also referring to Scheutz et al. (2025), stating that including point-specific wind data can help improve emission quantification.

- Line 385: Equation uses x and y for coordinates in vertical plane; z is used for the observed value in equation 3. On the next page (Equation 4), the parameter z is used for the vertical coordinate. This needs to be made consistent.
    - Revised and ensured to keep referring to y as a coordinate in the vertical plane.

- Line 405, equation 4: The original mass balance method has cos(theta) inside the summation, because the method allows for changing wind directions to be included (see e.g. Mohammadloo (2025) for a derivation, which shows the dot product with the wind velocity and the vertical curtain's orientation). Taking cos(theta) outside of the summation is an additional assumption that needs to be noted down.
    - Thank you for this comment. You are indeed correct. In the original method (as explained in Mohammadloo et al., 2025), variations in wind direction along the flight track are allowed, which is why the term appears inside the summation. In our case, we assume a constant wind direction and use a single value for the entire flight. This results in the cosine term being taken outside of the summation. We clarified this assumption in Section 3: "In this work, a

constant wind direction was assumed during flight, allowing the corresponding term to be placed outside of the summation."

- Line 430 "on visual inspection": this approach looks subjective, and a possible cause for further uncertainty in the results. I didn't see a clear discussion in Chapter 4 on this: can the visual inspection approach lead to errors or uncertainties not already accounted for in the sensitivity analysis?
    - Of course, this method can introduce additional error, especially when plume boundaries are not reached. However, each background determination method has its flaws, since the exact background is never known and can also change on both sides of the plume. Even using the 10th-percentile method can introduce an offset relative to the "true" background. In a way, this uncertainty is included in section 4.2.2, where we try to quantify the added error for a false representation of the background. As mentioned there, we assume we can determine the background with a 10 ppb uncertainty.

- Line 475: "Performing multiple passes across different altitudes, these plume-displacements can be captured and accounted for in the final flux estimate." This observation is in line with Scheutz et al (2025), see e.g. their Fig 7; a reference can be made to that paper.
    - Thank you for pointing this out, we added Scheutz et al . (2025) to this statement to improve the strenght of this argument.

- Figure 3 top plot: How was the LOWESS smoothing done? In Figure 3's top plot, it seems that the original Axetris data has much higher peaks than the smoothed data, but the troughs are similar. Was the LOWESS smoothing really mass-conserving? Perhaps an equation can be provided to show this.
    - The LOWESS smoothing was done within MATLAB. You are correct, and the function is not inherently mass-conserving. The difference in the determined integral of the smoothed Axetris data and the raw Axetris data is 103 ppb, which constitutes a percentual difference of 0.001% over the complete time series.

- Figure 3 caption: "The middle of the barn is at 0m horizontal distance". Many people will intuitively understand what is meant here, but the wording a bit sloppy and can be improved. The barn is not actually located in the plane flown by the drone, of which we are seeing the integrated contour plots. It may also be nice to mention the drone's downwind distance to the farm here.
    - Thank you for pointing this out, we updated the caption to make it a bit more clear and also mention the downwind distance. It now reads: "*The horizontal distance is referenced to the projected location of the barn centre (defined as 0 m). The sampled plane is located 120 m downwind of the farm.*"

- Line 577: "Increasing the flight duration lowers the CI bounds and the standard deviation". This finding seems to be again in line with Scheutz et al (2025).
    - Thank you for pointing this out, we added the reference to this statement.

- Lines 627-632: The current explanation for the difference with the observations by Mohammadloo et al is not clear. In particular, does "changing atmospheric conditions" relate to a spatial difference between the ground-based anemometer and the wind speed at the location of the drone? Otherwise, could it perhaps be related to how Ettinger et al have used a mean wind direction in the MBA rather than the actual time-varying wind direction (see my earlier comment on line 405)?

  - This contradiction pertains solely to wind speed uncertainty (and not wind direction). Mohammadloo et al separated the effects of wind speed and wind direction (before combining them). Our results show a clear increase in wind speed variability, whereas Mohammadloo et al. stated that this should not be a problem since concentrations scale directly with increasing wind speed. This statement is correct, but it fails to account for the use of an incorrect wind speed in ones flux calculations.
  - Say you think the wind speed is 3 m/sbased, since on your on-site wind measurements, while in reality it is 4 m/s. This would mean you would multiply the measured mole fractions by a wind speed that is too low, resulting in an underestimation of the true flux. We believe this was not properly considered in Mohammadloo et al., resulting in the difference.
  - The final statement was poorly worded and updated to: "However, this overlooks the impact of an erroneous wind speed on emission flux determinations. If the wind speed is underestimated (e.g. 3m/s instead of the true wind speed 4 m/s), the resulting flux will also be underestimated, since the flux scales proportionally with the wind speed."

- Line 747-751: I completely agree with this paragraph, but the authors could perhaps add that the effectiveness of their current solution can also be improved in future by simply removing the heavy Aircore equipment from the drone. This will enhance the flight endurance of the system and would allow for more repeated measurements.
  - This is a fair point and is certainly something to consider. We therefore added the following sentence: "Besides, the sensor's effectiveness can be further improved when used as a stand-alone, since flight duration and flight repeats increase due to a lower payload and less post-flight processing."

- Line 756: "We improved the sensor's performance by properly insulating the sensor and applying active thermal regulation to maintain a constant cell temperature" . As noted under the Specific comments, some discussion is needed on how difficult/costly it is to implement such an enhancement for temperature control.
  - Thank you for this comment. As mentioned, we plan to make the design open source to address this. However, the implementation cost is very low (< 100 €).

- Line 1003: this is presumably flight 2, not 1.
  - You are correct. Revised and changed it from flight 1 to flight 2.