Reviewer 2

General comments:

This study deploys two UAV platforms equipped with open-path and closed-path methane analyzers to quantify emissions from a known geological CH4 seep at the Mackenzie River Delta, Canada. Vertical curtain flights at two downwind distances are used to compare three flux quantification methods: direct mass balance, cluster Kriging mass balance, and Gaussian plume inversion, while evaluating the influence of sensor types on flux estimates. The paper reports a mean seep emission rate of 11.4 $\pm$ 6.8 kg CH4 / hr across methods and concludes that direct mass balance yields the most consistent estimates across platforms.

The paper addresses an important problem: quantifying geological CH4 seeps in remote Arctic environments where conventional monitoring is impractical. The study site is interesting, and the simultaneous comparison of three quantification methods to the same real-world dataset is a valuable contribution, as is the assessment of how open-path and closed-path sensor characteristics influence flux estimates. The manuscript is well-written and transparent about the assumptions underlying each approach.

My main concern relates to the scope of the conclusions relative to the dataset. The study comprises four flights at one site under one set of atmospheric conditions, with no independent ground-truth emission rate available. This supports an informative method comparison and sensor comparison for this specific case, but I find that the evidence is insufficient to conclusively support broader claims about method robustness or generalizability beyond this specific test case. Moreover, the three methods operate within different uncertainty frameworks, making direct comparison of uncertainty magnitudes difficult in my opinion. Furthermore, the two platforms differed not only in sensor type but also in flight strategy, making it also difficult to attribute differences in flux estimates solely to sensor characteristics.

I recommend reframing the manuscript as a detailed case study. The authors apply established methods under conditions that are likely more challenging than most previous applications, which typically involved temporally stable (industrial) sources, homogeneous terrain, and sources above ground level. Explicitly situating the results within this context would enable the authors to discuss method performance, sensor behavior, and experimental design insights in this test case relative to prior studies. This would make the manuscript a valuable resource for the community, especially for the design of future UAV-based methane flux studies of natural seeps. Correspondingly, I suggest revising the title as well.

Thank you for the reviewer's feedback. We think that the reviewer comments helped significantly to improve our manuscript. We agree with the reviewer that there are only four flights with no independent ground-truth to support the findings of the manuscript. As such, in the revised manuscript we have avoided making broader claims as the reviewer suggested. We also thank the reviewer to point out the differences between uncertainty estimations of GPI and MB methods. In revised manuscript, we avoided directly comparing them and only highlight the difference between the methods. The reviewer is correct in stating that the flight strategies are also different, and the flux estimation differences cannot only be attributed to sensor characteristics; we have also revised the manuscript by acknowledging the impact of different flight strategies as well. Finally, we applied the reviewers' suggestions by reframing the manuscript and avoided directly comparing the methods. Please see our detailed responses to the reviewer's comments below and the new title for our study. Our responses are shown in blue color, and statements added or revised in the manuscript are indicated by italics.

New manuscript title:

Application of UAV-based methods for quantifying methane point source emissions over an Arctic geological seep

Specific comments:

1. P1, L16-17: "finding that mass-balance approaches yielded the most robust quantification with smaller uncertainties": From my point of view, this conclusion is not sufficiently supported given the limited dataset. The three methods rely on different uncertainty frameworks, and the mass balance approaches omit an important source of uncertainty: "the turbulent nature of atmospheric transport" (P8, L173-175).

Thanks for the reviewer's comment. The reviewer is correct, we now avoid making direct comparisons of the methods in revised manuscripts. Per the reviewer's comment, we revised the manuscript lines 15 - 17 as follows.

We then applied two widely used quantification techniques (mass-balance and Gaussian plume inversion) finding that mass-balance approaches were better suited for flux estimation over the seep area as the measurements were not close to a Gaussian profile with an average model-data mismatch of 81%.

2. P1, L18: "average estimated rate": I would clarify here that this average is taken across different methods applied to four drone flights, as the current wording could be interpreted (as I did at first) as an average over repeated experiments.

   Thanks for the reviewer's comment. Per reviewer suggestion, we revised the respective manuscript lines 18 - 19 as follows

   We estimate that the seep emission rate falls in the range of 7.8 to 16.0 $\mathrm{kgCH_4\ h^{-1}}$ *and the average across all methods and flights were estimated as 12.3* $\pm$ *8.0* $\mathrm{kgCH_4\ h^{-1}}$.

3. P2, L37-39: "Accurately estimating the emission rates from point or localized sources is only possible if the locations are already known": This statement may overstate the limitation as there are UAV-based measurement approaches that have shown the ability to locate and quantify unknown point sources simultaneously.

   Thanks for the reviewer's comment. We agree with the reviewer that there are existing UAV-based approaches to locate and quantify unknown point sources. As the reviewer suggested, we revised the manuscript lines 38 - 40 as follows.

   Accurately estimating the emission rates from point or localized sources if the locations are already known is less complex than unknown point sources, since this allows the monitoring systems to be positioned such that sources fall within their spatial coverage.

4. P4, L91:From my understanding, Dallimore et al. identify two seeps close to each other at the Channel Seep 2 site. It would be helpful to specify which seep is investigated here. This would make it clear whether the second seep was downwind or upwind, and address any potential concerns about whether the other source could have influenced the measurements.

   Thanks for the reviewer comment. Channel Seep 2 was designated with a white arrow in Figure 1 in Dallimore et al. We agree with the reviewer that there is one more seep that is very close, but Channel Seep 2, which is designated with the arrow, is located downwind of the other seep that the reviewer is mentioning. However, there were rare ebullition events from a pond close by which may have an impact on the flux calculations in here. In comparison with the main seep, the impact of these ebullition is expected to be rather small. Per reviewer comment, we have added a sentence to the manuscript lines 100 - 104 as follows

   Although there are several other seeps in this region, to the best of the author's knowledge, there is no close strong source in the downwind direction that can significantly influence the measurements of this study. We rarely observed sporadic ebullition events from a pond close to Channel Seep 2; however, these ebullition events compared to main seep were minuscule and expected to have a negligible impact on the flux estimations.

5. P4, L92:"high ebullition rate": This description would benefit from quantification or a reference to an observed rate. Please state explicitly whether the source is assumed to be time-invariant over the measurement period, and discuss how plausible this assumption is.

   Thanks for the reviewer comment. Per the reviewer's request, we have added the seepage rate from Dallimore et al. in the manuscript lines 94 - 96. We have assumed that the source is time-invariant during our measuring period, as this allowed us to use a simplified mass balance technique as well as the Gaussian plume inversion method without considering temporal variation (i.e., the Gaussian Puff model). Per the reviewer's comment, we revised the manuscript lines 135 - 138 as follows.

   We focus on a seep previously identified and named Channel Seep 2 (Wesley et al. 2023, Dallimore et al. 2024), which originates from the riverbed. The maximum measured gas seepage rate was reported as $0.4\ \mathrm{m^3\ min^{-1}}$, *which is equiva-*

lent to $16.1\ \mathrm{kg\ hr^{-1}}$*, assuming all the seepage consists of* $\mathrm{CH_4}$ *with a density of* $0.67\ \mathrm{kg\ m^{-3}}$ *(see Fig. 1 (b)) (Dallimore et al. 2024).*

The mean wind incidence angle was < 20°in all cases (see Fig. 2 and Table 1) and the non-zero wind incidence angle is expected to have a negligible impact on the emission rate calculation (Mohammadloo et al. 2025). For all our quantification methods, we assumed that the seep source as well as the channel flow rate is time-invariant during the sampling period.

6. P4, L92:Because the seep is located within a water channel, I wonder whether channel flow could have introduced advective transport of CH4, influencing the plume. Could the authors discuss this?

Thanks for the reviewer for pointing this out, as we have not considered this. There may be some impact of the channel flow on the plume morphology, it is not possible to distinguish this effect from the impact of the atmospheric turbulence without very detailed information. However, we don't think that this will significantly impact our emission rate estimation, as the flow rate of the channel does not change significantly over the measurement period of 20 mins. Nevertheless, the plume morphology is an important part of our uncertainty budget for the GPI method, as it is based on model-data mismatch. This further supports our conclusion that the that Gaussian plume inversion is less well-suited to this application, as the plume morphology may be impacted by the channel flow. Please see our response to comment #5 for the revised part of the manuscript, as well as manuscript lines 388 - 391 and 413 - 416 as below.

In this study, deviation from the Gaussian plume shape is attributed to limited sampling time which cannot adequately represent time-averaged well-defined plumes, micro-topographical features that introduce additional disturbances, and channel flow rates that may distort the plume morphology (Shah et al. 2019, Allen 2019).

While our analysis assumes the seep releases $\mathrm{CH_4}$ *directly into the atmosphere without significant interaction or advection within the water column, we acknowledge that water flow dynamics could potentially play a role in dispersing* $\mathrm{CH_4}$ *prior to its atmospheric emission. Future investigations could benefit from measurements of water flow velocities and dissolved methane concentrations to quantify potential advection of* $\mathrm{CH_4}$ *within the channel.*

7. P4, L109-111 and Table 1:"Two additional ground-based wind sensors were deployed to verify the UAV-based measurements": Please provide more details on the sensor type, measurement height, and location. Furthermore, the cruising speeds of the UAVs were relatively high (2-3 m/s, 5 m/s, and 7 m/s), potentially affecting the reliability of wind speed and direction measurements. The statement "UAV-based wind speed and direction measurements showed good agreement with the ground-based measurements (data not shown)" is important and interesting. I suggest adding the ground-based wind data and a discussion of the observed agreement in the manuscript. Please also report the distance between the wind sensor and the rotor plane or body of the drone.

Thanks for the reviewer comment. Two ground sensors, Gill WindUltra, Gill Instruments and Windsonic4, Campbell Scientific were placed near the seep location at about 1.5 and 2.7 m above the ground level, respectively. The primary goal of these sensors was to support flight planning, as the curtains should be oriented orthogonal to the wind direction. The second goal was to verify our on-board wind measurements. Both of the ground sensors measure 2D wind speed and wind direction. As the reviewer suggested, we compared the wind measurements recorded by all of the sensors (two ground-based and two on-board), during the flights and added these as an Appendix in the revised manuscript. As can be seen from these figures, the wind direction is consistent among all sensors. When comparing the wind speeds, it should be noted that the anemometers are each measuring at different altitudes. Also note the spatial difference between CP-1 and OP-2 and vice versa in wind speed measurements. Overall, while the two UAVs utilized different anemometer models, our analysis in Appendix D demonstrates that the measurements are consistent within the combined uncertainty ranges of the sensors. The observed variations in wind speed are attributed to the spatial gradients between the two curtain locations (150 m vs 80 m downwind) and the increase in wind speed with altitude, rather than sensor bias. Per reviewer request, we added this comparison in Appendix D in the revised manuscript as follows. The distance between the rotors and the anemometer was 67 cm for UAV-MPI and 75 cm for UAV-NRCan. Regarding the cruising speed of the UAVs, we don't think these speeds are high enough to significantly impact the measurement of wind speed and direction, especially when it is below 5 m/s. To correct the wind speed measurements from UAV-NRCan that had flight speeds of 5 and 7 m/s, the pitch and roll angles were also accounted. Per reviewer request, we have added an Appendix D as follows.

The anemometer deployed on UAV-MPI, 0.67 m above the rotor plane, recorded wind measurements at 2 Hz with a reported accuracy of $\pm 0.2\ \mathrm{m\,s^{-1}}$ *for speed and* ± 1.0 *°for direction. UAV-NRCan is a DJI Matrice 300 RTK quadcopter equipped with a* $\mathrm{CH_4}$ *gas analyzer custom-built by the National Research Council of Canada (NRC), and a 2D anemometer (WindUltra, Gill Instruments) placed 0.75 m above the rotor plane to measure wind speed and direction.*

The wind speed measurements are relatively consistent between the UAV platforms, and the wind conditions are similar for the far- and near-curtain flights. More details can be found in Appendix D, where the measurements of ground-based sensors were also included.

Measured wind speeds from all four curtain flights are illustrated as a function of altitude in Fig. 3 (see also Appendix D).

Appendix D: Wind speed and direction comparison

Figure D1 shows the measured wind direction and speed during all curtain flights and from ground-based sensors. Note that UAV-based wind measurements depart from the ground-based measurements as UAVs ascend during the curtain flight. The differences between the on-board wind measurements (CP-1 and OP-2, CP-2 and OP-1) are primarily due to spatial differences of the curtains (i.e. 150 vs 80 m downwind distance from seep location) and different flight altitudes. This further demonstrates that all wind speed measurements are consistent throughout the measurement duration. We placed the ground-based sensors close to the seep location at two different heights, one at about 1.5 m (Windsonic4) above ground level and other (Gill Windultra) at about 2.7 m, while UAV-based anemometers were placed 0.67 and 0.75 m above the rotor plane in UAV-MPI and UAV-NRCan, respectively. Wind direction measurements agree well with the ground-based sensor after correcting the measurements from UAV-MPI and ground sensor (Windsonic4) north alignment.

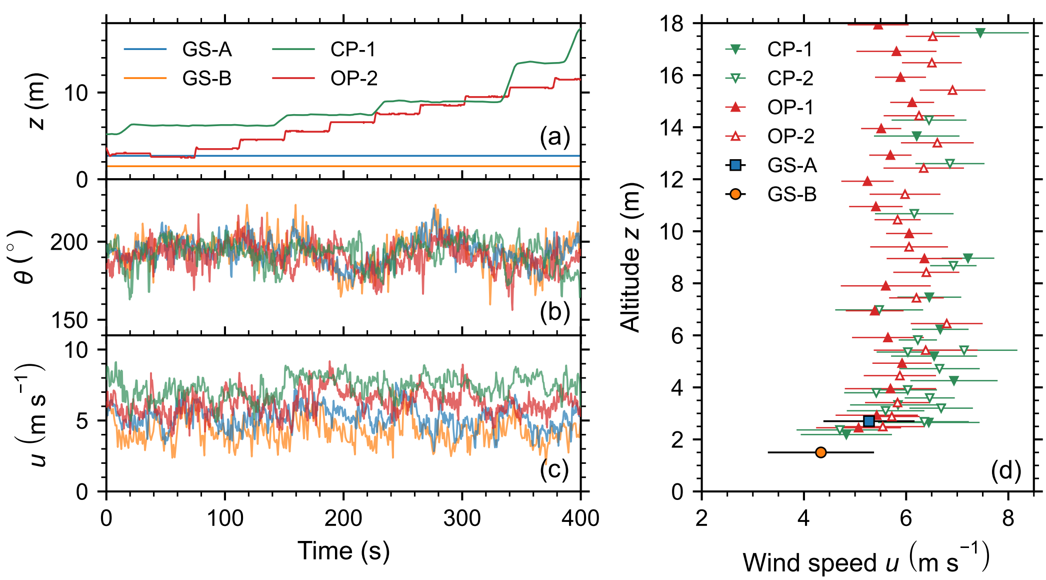

Figure 1. Comparison of wind measurements during CP-1 and OP-2 flights: (a) the altitudes of ground sensors and both UAVs, (b) wind direction measurements, and (c) wind speed measurements from both UAV platforms and ground sensors (GS-1: WindSonic4; GS-2: Gill WindUltra). (d) Wind speed measurements from all platforms during all four flights at different altitudes, where each symbol represents the mean wind speed and horizontal bars represent the standard deviations.

8. P7, Eq. 2 & P13, L270:Please clarify how the background mixing ratio is determined and what is meant by "sampling the background concentration once". It would be helpful to add whether and how its uncertainty is included in the overall uncertainty quantification.

Thank you for the reviewer comment. The background mixing ratio was derived from the average of sampling points outside of the plume-affected regions, determined independently for each of the gas analyzers. The phrase referred to by the authors was a suggestion to save flight time by repeating background mixing ratio measurements. As background mixing ratios were mostly determined by large-scale weather patterns, the changes over short time scales will be minor. Therefore, sampling background mixing ratio once and using it for the emission rates estimations for several curtain flights may be an alternative way of saving flight time, which is crucial in such remote locations. Per reviewer request we clarify the manuscript lines 298 - 308 as follows

We have not included the uncertainty of the background mixing ratio in our calculations, but thanks to reviewer comments, we have now added this into our uncertainty budget as well. The background concentration can vary with altitude, which can cause biases in our emission rate estimation, nevertheless we do not anticipate significant background variation from the altitudes covered in this study. To account for this, we now added the background uncertainty in our uncertainty budget estimation. To be conservative, we have defined this error based on the $\pm 1\sigma$ of the points we used to estimate the background concentration and propagated this to our flux calculations as was practiced in Yong et al. (2024). This additional uncertainty accounts for 3 % of CP flux calculations and between 10-13 % of OP flux calculations as $\pm 1\sigma$ was 3 ppb for CP and 9 ppb for OP analyzer. The difference between the background concentration after applying the water and temperature corrections to OP measurements which was omitted in the original submission, became very small. The revised background methane dry mole fractions are 2031.8 and 2029.1 ppb for CP and OP, respectively. We have revised the manuscript lines 155 - 158, Fig. B1 in revised manuscript, and 194 - 199 as follows.

In CP-2 flight, after sampling the curtain up to an altitude of about 15 m, the remaining part of the battery was used to re-sample the section of the curtain that is close to the ground (i.e., between about 2 and 6 m). Treating CP-2 as two individual curtains indicates the benefits of repetitions in reducing the uncertainties due to atmospheric turbulence, since the estimations deviate by $\pm 5\%$ from the one using all the data. Excluding the repeated transects, the emission rate is estimated to be $13.1 \pm 6.5\ \mathrm{kgCH_4\ h^{-1}}$. As the additional transect measurements in CP-2 do not capture the full vertical extent of the plume, it therefore cannot be treated as an independent curtain. We combine the repeated transects close to the ground with the segment of the original curtain above 6 m to construct a new curtain (i.e., a second curtain patched with the repeated transects and the portion above 6 m). The emission rate calculated from this new curtain is $14.1 \pm 7.6\ \mathrm{kgCH_4\ h^{-1}}$. Under challenging conditions where the repetitions are limited as in this study, after sampling the curtain once, re-sampling sections close to the ground may be beneficial as long as the battery lasts.

A constant background $\mathrm{CH_4}$ of 2031.8 ppb and 2029.1 ppb is removed from the measured dry mole fractions for UAV-MPI and UAV-NRCan, respectively, where these background values were estimated by averaging sampling points outside of the plume-affected regions (see Fig. B1).

Using constant background concentration may cause bias in our flux estimations as background concentration may vary. To account for this in our uncertainty budget, we used $\pm 1\sigma$ of the background concentration and propagated this as a systematic error source in our flux estimations (Yong et al. 2024). This yielded about 3 % additional uncertainty for CP flux estimates and between 10-13% for OP flux estimates as $\pm 1\sigma$ corresponds to about 3 ppb for CP and 9 ppb for OP analyzer. We note that this difference between the sensors originates from different noise characteristics of the sensors rather than being a physical difference in background.

9. P7, L156-157:"At the lowest level of the sampling plane, we used a logarithmic function to complete the vertical profile, assuming zero flux at ground level (Bonne et al., 2024).": This is not intuitive to me, considering the source is at ground level. The flux profiles in Fig. 6, as well as in Bonne et al., are concave (bowing towards the upper left for z(q(z)). In neutral conditions, the wind profile is typically convex (bowing towards the lower right for z(u(z)). Since q(z)=u(z)c(z), a concave flux profile would require concentrations to increase with height. Bonne et al. studied elevated sources; however, in the present study, the source is at ground level, and high concentrations near the surface are expected (e.g., by the Gaussian plume model), implying a convex rather than concave profile is fitting.

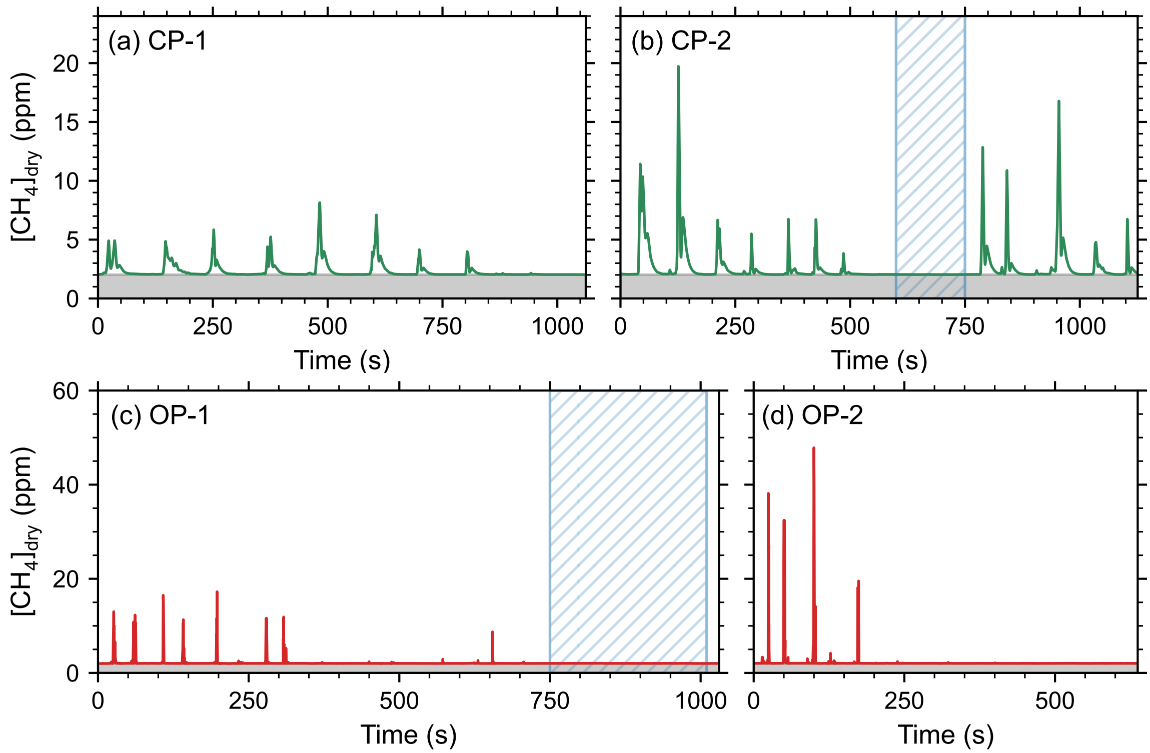

Figure 2. Measured methane concentration timeseries for the four curtain flights, labeled CP-1, CP-2, OP-1, and OP-2. The shaded regions indicate the background methane concentration, recorded as 2.0318 ppm for the closed-path sensor (a,b) and 2.0291 ppm for the open-path sensor (c,d). The background concentrations were estimated by averaging the measurements within the dashed region.

Thanks for the reviewer comments. The reviewer is correct, the concentrations close to ground level are expected to be high and the profile should be convex rather than concave, and we have revised the analysis accordingly. To fit the lowest measurement level to the ground, we used the estimated friction velocity (u_*) and roughness length (z_0) from the logarithmic wind profile and again assumed the transect-integrated flux density at the ground, where z = (z_0) will be zero. This resulted a convex shape fitting as reviewer stated. The changes of the fitting impacts the flux estimation which were also revised accordingly. Per reviewer comment, we now revised the fitting assuming high concentrations close to ground and revised the emission rate calculations accordingly.

At the lowest level of the sampling plane, we used a logarithmic function to complete the vertical profile, assuming zero flux at the ground level as in (Bonne et al. 2024, Dooley et al. 2024). Given that the methane source is effectively at ground level, high CH_4 mixing ratios are expected close to the ground, and so the shape of the extrapolation curve is convex.

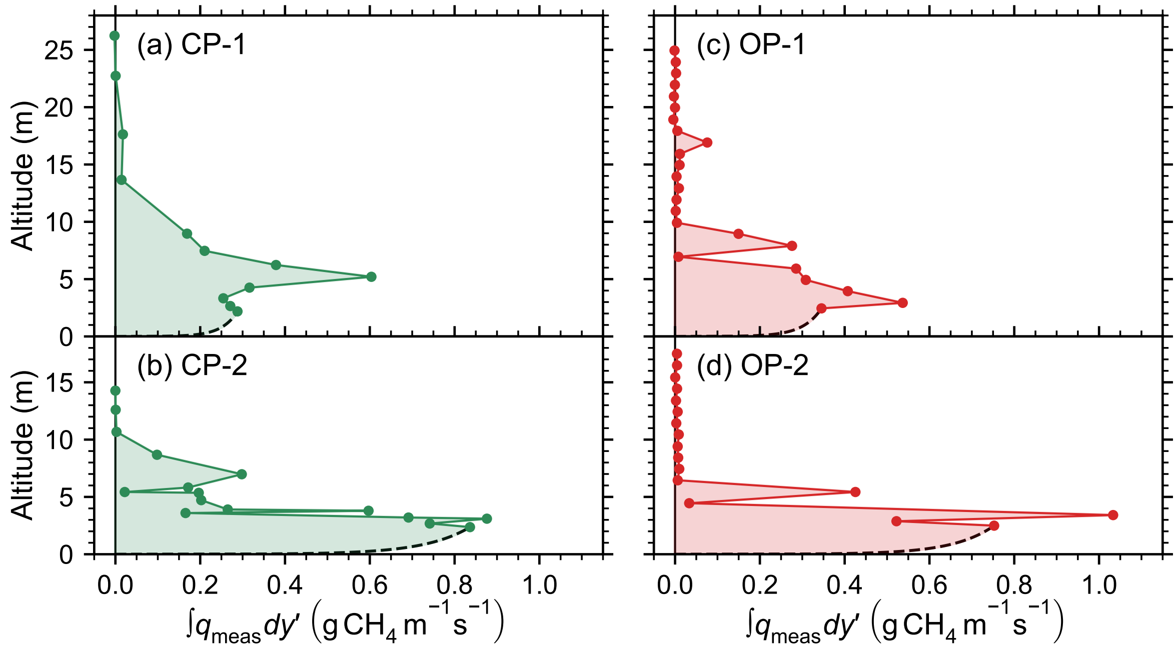

Figure 3. Calculated transect-integrated flux densities ($\int q_{\mathrm{meas}} dy'$) for each transect used in the DMB approach. The dashed lines indicate the logarithmic fitting that was employed to extrapolate the profile from the ground to the height of the first measurement transect.

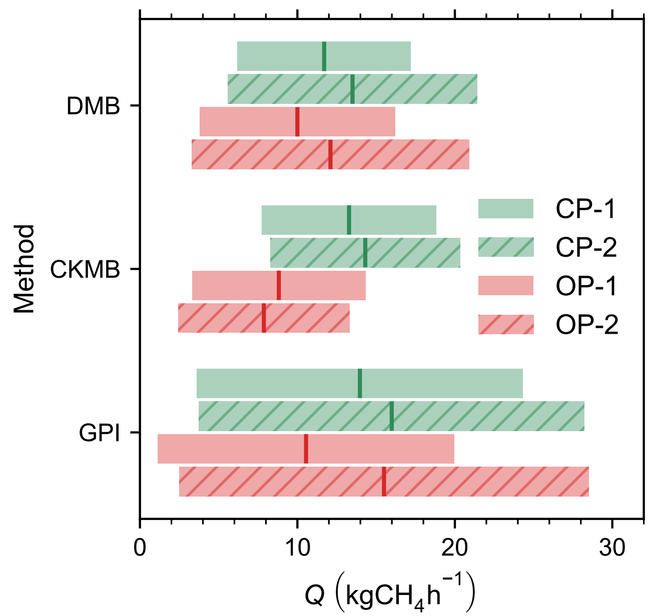

Figure 4. Emission rate estimations (Q) from different models and UAV platforms. Here, DMB, CKMB, and GPI denote Direct Mass Balance, Cluster Kriging Mass Balance, and Gaussian Plume Inversion approaches. Far curtains OP-1 and CP-1 are represented as solid red and green bars and near curtains OP-2 and CP-2 are represented with hashed red and green bars, respectively.

10. P7, L166:"We quantified the instrument errors using error propagation based on field measurement data as well as laboratory tests". Please specify which components were included (e.g., instrument precision, wind speed uncertainty, background concentration uncertainty), and what values were assigned to each.

Thank you for the reviewer comment. We provided the instrument uncertainties between lines 108 - 118 with accuracy of each wind sensor and gas analyzers. More details regarding the instrument uncertainties for open-path can be found in Beattie et al. (2026) and for CP in Appendix A, as well as in Bolek et al. (2024). We have now added the background uncertainties in the budget with corresponding values and the information were provided in revised manuscript, see our response to comment #8. Per the reviewer's request, we revised the manuscript lines 108 - 118 and 180 - 182 as follows.

The $\mathrm{CH_4}$ analyzer is a closed path (CP) analyzer (Strato, Aeris Technologies) that measures the dry mole fraction of $\mathrm{CH_4}$ at 2 Hz sampling frequency with $< 1\ \mathrm{ppb\,s^{-1}}$ of sensitivity. The CP instrument was customized by adding a thermally controlled enclosure to control the temperature of the measuring cell, which reduces the instrument drift (see Appendix A). The anemometer deployed on UAV-MPI, 0.67 m above the rotor plane, recorded wind measurements at 2 Hz with a reported accuracy of $\pm 0.2\ \mathrm{m\,s^{-1}}$ for speed and ± 1.0 °for direction. UAV-NRCan is a DJI Matrice 300 RTK quadcopter equipped with a $\mathrm{CH_4}$ gas analyzer custom-built by the National Research Council of Canada (NRC), and a 2D anemometer (WindUltra, Gill Instruments) placed 0.75 m above the rotor plane to measure wind speed and direction. The NRC $\mathrm{CH_4}$ analyzer is a mid-infrared tunable diode laser absorption spectroscopic system with an open path (OP) gas cell, a sampling rate of 100 Hz, and a resolution of 26 ppb at 10 Hz, calibrated over the concentration range of 2 ppm to 50 ppm (Beattie et al. 2026), and the anemometer has an accuracy of $< 2\%$ RMSE for wind speed and < 1.0 °RMSE for wind direction. Two additional ground-based wind sensors were deployed to verify the UAV-based measurements.

The uncertainties are attributed to instrument errors, interpolation errors, plume capture uncertainty, background concentration estimation, and non-stationary plume dynamics. The instrument errors of OP and CP analyzers were detailed in Beattie et al. (2026) and in Bolek et al. (2024), as well as in Appendix A, respectively.

11. P8, L166-170:"In the CKMB method, the uncertainties are quantified using the covariance matrices provided by the Kriging algorithms, which were below 10% of the calculated emission rates for all cases. For the DMB method, the uncertainties associated with linear interpolation along the vertical axis are estimated to be around 10% based on the Kriging algorithm uncertainties.": I wonder about the reasoning behind applying Kriging-based uncertainty estimate to the DMB approach, which does not rely on Kriging. Can the authors please clarify this.

Thanks for the reviewer's comment. The reasoning behind this was that with Kriging, we can directly quantify the interpolation error using the covariance matrix, however with linear interpolation, as was practiced in the DMB approach, this is not as straightforward since the function is unknown. The choice of interpolation technique was reported to have a negligible impact on the emission rate estimation in Mohammadloo et al. (2025), which is expected with high-resolution sampling of UAV platforms, especially in horizontal axis, therefore we assumed that the interpolation uncertainty may be similar with DMB and CKMB methods and used the range from Kriging, as it can be easily obtained. However, this assumption can be significantly violated with low-resolution sampling. Per the reviewer's request, we revised the manuscript lines 184 - 190 as follows

For the DMB method, the uncertainties associated with linear interpolation along the vertical and horizontal axes are assumed to be around 10% as was found in the Kriging algorithm uncertainties. Due to the high-resolution sampling and because the choice of interpolation techniques has been shown to have a negligible impact on the emission rate estimation (Mohammadloo et al. 2025), we assigned similar-magnitude interpolation uncertainties for both mass balance methods. Since the interpolation uncertainties were readily available from the CKMB method, we applied the same range for the DMB method. This assumption could be significantly violated in the case of low-resolution UAV sampling. In this case, a more rigorous uncertainty quantification approach would be required.

12. P8, L182:"incorporating variable wind direction into the model": To my knowledge, temporal variability is not included in this model. Rather, it allows for fixed misalignment between the mean wind vector and the coordinate framework and computes dispersion differently from the Gaussian plume formulation.

Thanks for the reviewer's comment. We have revised the manuscript lines 206 - 208 as follows.

Several formulations have been proposed to overcome this issue, such as replacing the diffusivity parameter with a near-field mixing factor (Shah et al. 2019) or applying a different dispersion calculation than the Gaussian plume formulation (Vergassola et al. 2007, van Hove et al. 2025).

13. P8, Eq. 4:Please clarify if the surface is treated as a flat, homogeneous surface. Dallimore et al. mention that "prolonged activity has caused an erosional niche to form on the river bank and the formation of a 5 m deep pockmark" at the Channel Seep 2 site. I wonder whether and how this topography could have influenced plume dispersion, and therefore affected the flux estimates. It would be helpful if the authors could discuss this.

Thanks for the reviewer's comment. We assumed a flat, homogeneous surface for simplicity. However, topographical disturbances—such as riverbank erosion and micro-topographical features—would likely alter plume dynamics, causing deviations from the idealized Gaussian distribution. Furthermore, these disturbances, combined with atmospheric turbulence, may necessitate additional measurement repetitions to obtain well-defined, time-averaged plume statistics for robust mass balance and Gaussian plume inversion analyses. Given the remote nature of the emission source, number of repetitions is very limited. In the case of a pockmark below the water level, it is hard to provide more discussion on this as we do not have any information about the dissolved part of this CH_4 within the water. As CH_4 bubbles rise, they dissolve into water, and as this process increases, the amount of dissolved CH_4 is expected to increase, however we would not want to speculate about this process without additional information. Per the reviewer's comment, we revised the manuscript lines 179 and 203 - 205 as follows also please see our response to comment #20.

In both mass balance methods, the surface is treated as homogeneous and flat.

In the Gaussian plume model, the concentration field is assumed to be steady-state, meaning that the wind field is stable over time, such that the concentration field is time invariant and the surface is treated as homogeneous and flat.

14. P9, L225:"The measured peak CH4 enhancements for both OP curtains are 2-3 times larger than those measured in the CP curtains": I think that this is an interesting observation. The authors note in the introduction that OP analyzers allow "near-instantaneous response" but do not characterize the effective response time, smoothing, or time lag of CP analyzers. Is it possible to include anything about this based on the experiment? For example, Morales et al. characterize the smoothing and lagging of the AirCore system (but both instruments were deployed on the same drone in their study). I wonder if prior studies also observed 2-3 times larger enhancements?

Thanks for the reviewer comment. As the reviewer already stated, it is not possible to directly compare this using this experiment since the analyzers were carried by different UAV platforms. From our lab analysis, we can say that the time lag of the CP analyzer is about 2 s and the air exchange rate is 0.7 Hz at 0.6 l/min of flow rate. This may deviate in field conditions with changing UAV flight speed. In Morales et al. (2022) Fig. 4, the comparison between QCLAS and Aircore appears similar to our observation in this study between CP and OP analyzers. Per the reviewer's comment, we revised the manuscript lines 250 - 255 as follows.

The measured peak CH_4 enhancements for both OP curtains are 2-3 times larger than those measured in the CP curtains. Morales et al. (2022) observed similar differences between an active AirCore system and a custom-built, open-path quantum cascade laser absorption spectrometer (QCLAS). These differences in the measured mixing ratios can mainly be attributed to instrument characteristics of the CP analyzer, including cell volume and flow rate. Although the measured peak mixing ratios are affected by the instrument characteristics, the integrated area should not be affected (Tettenborn et al. 2025). In our analyses, we verify this by comparing the emission rate estimations between these sensors.

15. P11: I find the observation of asymmetric tails interesting. The authors attribute the asymmetry to two potential factors. By comparing data from transects flown in opposite directions, as shown in Fig. 5, the wind incidence effect can be isolated from the instrumental effect. The data appear to suggest that the instrumental effect is dominant. For future studies, I wonder whether the asymmetry can introduce a systematic bias in the flux estimates or whether flying alternating transect directions (partly) mitigates this effect. I think that a comparison with observations in the literature (if any, Morales et al.?) would strengthen the discussion.

Thanks for the reviewer comment. We have been looking into several possible causes of these asymmetric tails; however, apart from the instrumental artifacts and wind incidence angle, we do not see any additional possible reasons. As we do not have any direct reference measurements, we cannot conclusively explain this behavior. Asymmetrical tailing occurs as the reviewer points out due to the closed-path analyzer's working principle. We do not expect that this tailing would cause a significant bias in our emission rate estimations because, although the CP analyzers smooths the data, the integrated area should not be affected by this (see our response to comment #14). The cause of this gradual tail to a large extent is due to the slow air exchange rate of the analyzer. In literature, as the reviewer pointed out, Morales et al. (2022) had a similar comparison, although with the Aircore instead of a closed path analyzer. Although the working principle of Aircore differs from CP analyzers, there are minor asymmetry in the measurements in Aircore measurements. Additionally, with the Aircore, molecular diffusion may also play an important role, which does not exist to that extent in closed-path analyzers. Therefore, we leave that part out. Per the reviewer's request, we revised the manuscript lines 274 - 280 as follows.

This asymmetrical tailing towards the flight direction is attributed to air mixing in the analyzer cell and, to a large extent, due to the slow air exchange rate of the CP analyzer (i.e. 0.7 Hz at 0.6 L/min of flow rate). We do not expect that this tailing would cause significant bias in our emission rate estimations, as integrating along a transect should yield similar results, irrespective of the instrument response time (Tettenborn et al. 2025). Apart from the smoothing and broadening of the measured data, the extended tail of the CP measurement has additional smaller secondary peaks along the flight direction which may be attributed to (i) a slightly larger wind-incidence angle—supported by the shorter tail in the opposite flight direction, (ii) limited pump speed and/or friction within the tubing that prevents complete flush of the sampled air effectively.

16. P13, L260-272 & P17, L314-320 & P17, L322-323:"The calculated emission rates [...] agree within their estimated uncertainties.": I would avoid directly comparing uncertainty ranges across methods, as their underlying frameworks are fundamentally different. Furthermore, with uncertainty ranges of 45-54% for the mass balance methods and up to 89% for GPI, I think that these ranges are too wide to discriminate between methods based on a single test case and should not be used as evidence for relative method performance. Instead, I think that it would be more interesting to compare the observed method and sensor differences with those reported in prior literature.

Thanks for the reviewer comment, our main aim in here is not to compare method performance but to emphasize the consistent performance of UAV-based methods across different quantification techniques. The reviewer is correct in stating that there are not enough data to make a methodological comparison, especially when the error bounds are very wide. Per the reviewer's request, we have added more discussion from the literature about the methods and sensor performance and have avoided any direct comparison between the two methods. We revised the manuscript lines 81 - 85 and 383 - 401 as follows.

We applied three emission rate quantification methods to the measured data: mass balance with Kriging interpolation, direct mass balance, and Gaussian plume inversion. Finally, we discussed the results obtained from the two gas analyzers and the different quantification methods, evaluating their advantages and disadvantages for UAV-based applications over the seep site, which are relevant to applications in other sectors where $\mathrm{CH_4}$ *emissions are a concern.*

Similar to the CKMB method, the GPI method tends to smooth the observed methane enhancements and widen the extent of the modelled plume both vertically and horizontally. When the morphology of the measured plume deviates strongly from a Gaussian shape, the GPI approach may be less suitable than other quantification methods, as the model cannot sufficiently explain the observation. This can prevent the optimization from converging or from finding the optimum parameters, as was reported in (Andersen et al. 2021), even under relatively homogeneous conditions without any impact of topography and channel flow. In this study, deviation from the Gaussian plume shape is attributed to variations in the wind field close to the emission source, micro-topography, water channel flow rate, and insufficient averaging time, which is particularly important for the OP sensor. This can be to some extent mitigated by using both curtains simultaneously. Given the 3-dimensional nature of the Gaussian plume model, it is also possible to fit the data from both the far and the near curtains simultaneously. For the CP sensor, fitting both curtains together leads to an estimated emission rate of $12.8 \pm 9.9\ \mathrm{kgCH_4h^{-1}}$*, which is slightly less than the estimated flux rates for CP-1 and CP-2 individually. For the OP* *sensor, the combined estimate is $12.4 \pm 10.9\ \mathrm{kgCH_4h^{-1}}$, falling between the estimates for OP-1 and OP-2. Additionally, the two-dimensional curtain flight patterns employed in this study may not be optimum choice to retrieve the plume shape parameters required for the GPI method. Sampling strategies that use adaptive path planning to maximize the information gain in a limited time may be more suitable, as described in (van Hove et al. 2026)*

17. Section 3, especially Section 3.3:I suggest reframing the findings as insights into method behavior rather than as evidence of method superiority. Please consider adding a specific 'lessons learned' and/or 'outlook' section, as this could provide a lot of value to the community when designing future studies.

Thanks for the reviewer comment. We agree that, instead of method comparison, this section can be better suited to provide more insights into method behavior. Per reviewer comment, we have revised this section as follows.

3.3 Insights and lessons learned Comparison between the methane concentration measurements for both UAV platforms shows that the OP analyzer often records sharper and larger $\mathrm{CH_4}$ peaks, while the peaks are much smoother and damped for the CP analyzer. The OP analyzer may be preferable for plume tracking due to its higher sampling rate and lower latency compared to the CP analyzer. However, plume morphologies are discrete due to sharp peaks in OP, and rather smooth in the CP. Nevertheless, the quantification methods presented in this study regardless of analyzer type were able to produce consistent emission rate estimates for the seep, overlapping within their respective method uncertainty ranges. The estimated emission rates from all of these methods are provided in Fig. 10.

The application of DMB is straightforward and produces the most consistent emission rate estimates across both curtains and UAV platforms. However, adequate sampling density is essential, particularly in the vertical direction. Sparse sampling in the vertical direction can lead to large uncertainties and underestimation of the emission flux, especially when the plume center is missed (the uncertainties are estimated to be about 25%). The uncertainties associated with the linear interpolations between transects are largely unknown and may be underestimated in this study. Further investigation is needed to rigorously quantify these uncertainties. The choice of the extrapolation method between ground and the first measuring height for DMB can significantly impact the resulting emission rate estimates. Collecting additional data while UAVs are on the ground may help to improve the extrapolation. The CKMB method provides excellent agreement in the emission rate estimates when comparing the far and near curtains for each UAV platform individually, and the interpolation uncertainty can be directly and conveniently quantified using covariance matrices. However, the variation between the platforms is greatest for CKMB compared to the other quantification methods. This is likely due to sensor response times and differences in flight execution: the OP platform utilized autonomous flight paths (resulting in regular sampling interval), whereas the CP platform was manually piloted (resulting in irregular sampling interval). Compared to DMB, the CKMB method is more complicated to apply and computationally more demanding as the curtain area increases. Additionally, fitting a variogram to the measured data usually requires optimizing the variance and length scales, which may not always converge (Andersen et al. 2021).

The near-field Gaussian Plume Inversion (GPI) generates emission rate estimates that are similar to the DMB and CKMB method estimates. Across all of the methods, there was less variation in the flux estimations for UAV-MPI (CP-1 and CP-2) compared to those for UAV-NRCan (OP-1 and OP-2). This variation in the OP sensor estimates was most notable in the GPI method, for which there was 38% difference between the estimates for OP-1 and OP-2. Similar to the CKMB method, the GPI method tends to smooth the observed methane enhancements and widen the extent of the modelled plume both vertically and horizontally. When the morphology of the measured plume deviates strongly from a Gaussian shape, the GPI approach may be less suitable than other quantification methods, as the model cannot sufficiently explain the observation. This can prevent the optimization from converging or from finding the optimum parameters, as was reported in (Andersen et al. 2021), even under relatively homogeneous conditions without any impact of topography and channel flow. In this study, deviation from the Gaussian plume shape is attributed to limited sampling time which cannot adequately represent time-averaged well-defined plumes, micro-topographical features that introduce additional disturbances, and channel flow rates that may distort the plume morphology (Shah et al. 2019, Allen et al. 2019). Therefore, over the seep area studied here with high-density sampling, mass balance approaches (CKMB and DMB) should be preferred when on-board wind measurements are available to avoid the large model - data mismatch as observed in GPI method. This can be to some extent mitigated by using both curtains simultaneously. Given the 3-dimensional *nature of the Gaussian plume model, it is also possible to fit the data from both the far and the near curtains simultaneously. For the CP sensor, fitting both curtains together leads to an estimated emission rate of* $12.8 \pm 9.9\ \mathrm{kgCH_4 h^{-1}}$*, which is slightly less than the estimated flux rates for CP-1 and CP-2 individually. For the OP sensor, the combined estimate is* $12.4 \pm 10.9\ \mathrm{kgCH_4 h^{-1}}$*, falling between the estimates for OP-1 and OP-2. GPI method can become practical whenever an on-board wind measurement is not available although this may lead to higher uncertainties. Additionally, the two-dimensional curtain flight patterns employed in this study may not be the optimum choice to retrieve the plume shape parameters required for the GPI method. Sampling strategies that use adaptive path planning to maximize the information gain in a limited time may be more suitable, as described in (van Hove et al. 2026).*

Overall discrepancies between applied quantification methods for each individual flights was on average 12%, except for OP-2 where the difference was on the average 35%. CMKB method estimations for OP sensor, compared to other two methods were lower especially for OP-2. This may be due to plume dynamics not being statistically stationary as indicated by the differences in $\mathrm{CH_4}$ *enhancements (see Fig. 7). OP-2 measurements do not show any enhancements above 6 m above ground level, while in CP-2, enhancements were observed even about 9 m above ground level. These differences affect the vertical interpolation of Kriging which resulted as smaller emission rate estimations compared to other methods. Furthermore, GPI estimation of OP-2 was almost double the estimation of CKMB. Since the CKMB method clusters the observational data as either enhanced or background, and interpolates these two clusters separately, the vertical extent of the predicted plume in CKMB compared to GPI method is smaller, hence the estimated emission rate. The observed differences (about 16%) between mass balance methods can be attributed to the different interpolation schemes applied as well as UAVs sampling instantaneous plume dynamics rather than a static, theoretical representation of the plume. Moreover, these differences can also be partly attributed to the different flight strategies. While our analysis assumes the seep releases* $\mathrm{CH_4}$ *directly into the atmosphere without significant interaction or advection within the water column, we acknowledge that water flow dynamics could potentially play a role in dispersing* $\mathrm{CH_4}$ *prior to its atmospheric emission. Future investigations could benefit from measurements of water flow velocities and dissolved methane concentrations to quantify potential advection of* $\mathrm{CH_4}$ *within the channel.*

18. P15: Whereas curtain flights make sense for the mass balance methods, the flight path may not be the most informative for constraining the plume shape parameters in the GPI method. Other flight patterns may provide more information about the 3D structure of the plume. Consider adding this to the discussion.

    Thanks for the reviewer comment. We have revised the discussion regarding the GPI method. Please see our response to comment # 17.

19. P18: I think that Dallimore et al. report an observed maximum gas seepage rate at the Channel Seep 2 site. Consider including their findings in the figure or text.

    Thanks for the reviewer comment, please see our response to comment # 5.

20. P19, L367-368: "If UAV-based wind measurements are available, mass balance approaches (CKMB and DMB) to quantification should be preferred over the GPI because they reduce the number of assumptions involved in the calculations.": To my knowledge, both methods share many of the same assumptions, importantly: spatial and temporal stationarity of the plume during sampling. MB methods require interpolation and extrapolation of the concentration and wind field, while GPI assumes a Gaussian profile. To me, it is not clear that MB methods are based on fewer assumptions than GPI. Moreover, I consider the number of assumptions to be less important than the total uncertainty and bias they introduce.

    Thanks for the reviewer comment. The reviewer is correct, the number of assumptions should not matter, but the total uncertainty and bias they introduce. Here, we intend to state that using on-board wind measurements, while also considering denser sampling associated with UAV-based surveys, would make Mass Balance a better alternative compared to the Gaussian plume model. The idea behind this was that the uncertainties in these flux rate estimations are primarily governed by random error, such as atmospheric turbulence, which may be mitigated by repeated curtains. In the mass balance method, we think that the uncertainties caused by interpolation and extrapolation can be minimized by denser sampling. In the case of GPI method, however, we introduce additional systematic error due to model-data mismatch as the measurements deviate from a Gaussian profile. Our measurements showed that the distribution is not close to

Gaussian, as confirmed by the large model - data mismatch, ranging between 74 - 90%. The deviation from the Gaussian profile may be due to limited sampling time that cannot represent a time-averaged Gaussian plume, micro-topographical features that introduce additional disturbances, and the channel flow rate that can perturb the plume morphology and deviate it from Gaussian shape (Shah et al. 2019, Allan et al. 2019). Per reviewer comment, we have revised the manuscript lines 385 - 393 as follows.

When the morphology of the measured plume deviates strongly from a Gaussian shape, the GPI approach may be less suitable than other quantification methods, as the model cannot sufficiently explain the observation. This can prevent the optimization from converging or from finding the optimum parameters, as was reported in (Andersen et al. 2021), even under relatively homogeneous conditions without any impact of topography and channel flow. In this study, deviation from the Gaussian plume shape is attributed to limited sampling time which cannot adequately represent time-averaged well-defined plumes, micro-topographical features that introduce additional disturbances, and channel flow rates that may distort the plume morphology (Shah et al. 2019, Allen et al, 2019). Therefore, over the seep area studied here with high-density sampling, mass balance approaches (CKMB and DMB) should be preferred when on-board wind measurements are available to avoid the large model - data mismatch as observed in GPI method.

21. Appendix A: I think that this is an interesting finding, and the authors could consider moving it to the main body of the paper.

Thanks for the reviewer comment. We agree that this is an interesting finding; however, we do not want to shift the focus of the manuscript to instrument performance improvements, but rather to discuss the methodologies used to quantify $\mathrm{CH_4}$ point sources. Therefore, we would like to keep this in the Appendix.

Technical corrections:

1. P7, L147: Typo "whichis".

Thanks for the reviewer comment, we have now revised the manuscript lines 159 as follows.

$u_\perp$ is the component of the recorded wind speed perpendicular to the flux curtain, and $\rho_{\mathrm{CH_4}}(z)$ is the density of $\mathrm{CH_4}$ gas, used to convert the $\mathrm{CH_4}$ mixing ratio [ppm] to mass concentration [$\mathrm{gCH_4m^{-3}}$], which is calculated according to

2. P20, L392: Missing units.

Thanks for the reviewer comment, we have now revised the manuscript lines 439 - 440 as follows.

Prior to testing the impact of the temperature controller, we observed large temperature fluctuations ($\pm 10\ ^\circ\mathrm{C}$) within the analyzer cell under relatively stable conditions (see Fig. A1 (b)).

1 References

Beattie, M. N., Sun, C., MacLeod, R., Sabourin, N., Morse, P. D., Smallwood, G. J., Corbin, J. C., and Norooz Oliaee, J.: Ultra-Lightweight Mid-IR Methane Sensor for UAV-based Measurements, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2026-137, 2026.

Yong, H., Allen, G., Mcquilkin, J., Ricketts, H. and Shaw, J.T.: Lessons learned from a UAV survey and methane emissions calculation at a UK landfill. Waste Management, 180, pp.47-54, https://doi.org/10.1016/j.wasman.2024.03.025, 2024.

Allen, G., Hollingsworth, P., Kabbabe, K., Pitt, J.R., Mead, M.I., Illingworth, S., Roberts, G., Bourn, M., Shallcross, D.E. and Percival, C.J. The development and trial of an unmanned aerial system for the measurement of methane flux from landfill and greenhouse gas emission hotspots. Waste Management, 87, pp.883-892, https://doi.org/10.1016/j.wasman.2017.12.024, 2019.

Andersen, T., Vinkovic, K., de Vries, M., Kers, B., Necki, J., Swolkien, J., Roiger, A., Peters, W. and Chen, H.: Quantifying methane emissions from coal mining ventilation shafts using an unmanned aerial vehicle (UAV)-based active AirCore system. Atmospheric Environment: X, 12, p.100135, https://doi.org/10.1016/j.aeaoa.2021.100135, 2021.

van Hove A, Aalstad K, Pirk N.: Actively inferring methane sources with drones. Environmental Data Science, 5:e2. doi: 10.1017/eds.2026.10029, 2026.

Tettenborn, J., Zavala-Araiza, D., Stroeken, D., Maazallahi, H., van der Veen, C., Hensen, A., Velzeboer, I., van den Bulk, P., Vogel, F., Gillespie, L., Ars, S., France, J., Lowry, D., Fisher, R., and Röckmann, T.: Improving consistency in methane emission quantification from the natural gas distribution systems across measurement devices, Atmos. Meas. Tech., 18, 3569–3584, https://doi.org/10.5194/amt-18-3569-2025, 2025.

Dooley, J. F., Minschwaner, K., Dubey, M. K., El Abbadi, S. H., Sherwin, E. D., Meyer, A. G., Follansbee, E., and Lee, J. E.: A new aerial approach for quantifying and attributing methane emissions: implementation and validation, Atmos. Meas. Tech., 17, 5091–5111, https://doi.org/10.5194/amt-17-5091-2024 2024.