Response to Anonymous Referees

We thank the Referees for their helpful and constructive comments and suggestions, which have helped improve the manuscript. Below we provide our answers and responses to specific comments and questions, and detail the changes we have made in the revised manuscript. In the document below, the referee's comments are shown in blue, while the authors' responses are provided in black.

Responses to Anonymous Referee #1

This paper evaluates the methane quantification of UAV-based mass-balance method using single-blind controlled-release experiment at two UK sites. The manuscript demonstrates strong practical value and provides useful experimental validation data for better characterizing the performance of UAV-based mass-balance methane emission quantification method and providing insights on methodological and environmental requirements to deliver robust measurements. This paper is generally well written, and the experimental design appears carefully executed. However, several methodological clarification and results discussions should be strengthened before publication.

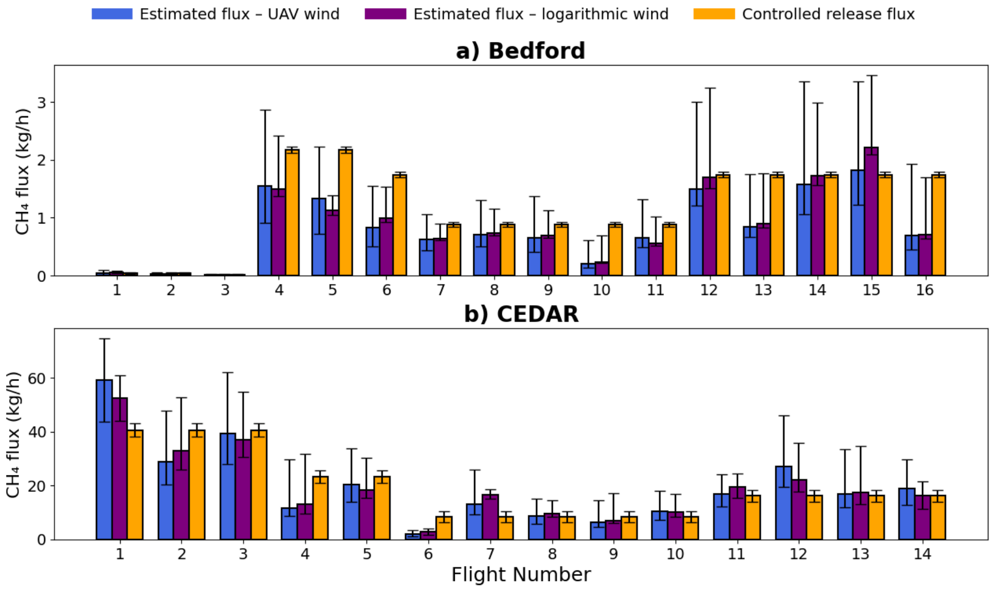
The major comments are summarized below:

(1) While the manuscript investigates the influence of varying wind speeds on methane quantification, there is limited discussion regarding the impact of wind measurements obtained from different measurement platforms. Since the mass-balance flux calculation is directly proportional to wind speed, systematic differences between mast-based and UAV-mounted wind measurements could introduce additional uncertainty or bias into the methane emission estimates. The authors are encouraged to provide a more quantitative comparison of emission estimates derived using different wind datasets, or at minimum discuss how discrepancies between these wind measurement approaches may propagate into methane quantification uncertainty.

We realise that this question is important, as it can inform on the important problem of measuring winds for UAV surveys and whether there is a need to measure winds onboard a UAV (which can be challenging), or whether a nearby mast might be sufficient for example. To address this comment, we estimated methane emissions using the logarithmic wind profile to compare it with our estimations using wind data from the UAV-mounted anemometer. Near-surface wind conditions were characterised using a four-level mast equipped with ultrasonic anemometers installed at 3, 6, 9, and 12 m above ground level at Bedford, and a one-level mast equipped with ultrasonic anemometers installed at 3m in CEDAR. The anemometers were recording 1-min vector averages at each site. For an approximate estimation of wind speed, we applied the logarithmic wind profile:

$$u(z) = u_{ref} * \frac{ln(\frac{z}{z_0})}{ln(\frac{z_{ref}}{z_0})}$$

where u(z) is the horizontal wind speed at height z, u_{ref} is the wind speed at the reference height z_{ref} and z_0 is the surface roughness length. For the local terrain, we used z_0=0.008 m at the Bedford Aerodrome site (lawn grass) and z_0=0.010 m at the Reading site (rough pasture) (Anjum, 2014).



The figure shows the estimated methane fluxes using the UAV-based winds (blue), using the logarithmic wind profile (purple) and the released rates (orange). The difference between the estimated fluxes using different wind inputs remains within error, both between wind inputs and with the true controlled release fluxes. But we note that the onboard winds give the closest agreement with controlled releases when averaged over all flights. Thus, we would continue to recommend that onboard winds are preferable for accuracy, but that mast winds can be useful, so long as errors are well accounted for and included in uncertainty budgeting. The flights we performed for the controlled release experiment reached relatively low altitude, especially for Bedford where the UAV reached 15m approximately. This indicates that the vertical resolution of the mast-mounted anemometers (at 3, 6, 9 and 12m) can represent the realistic wind that is measured on board the UAV. In real cases though with more complex emissions, the emission distribution might require the flights to reach higher altitudes up to 120m. In this case, the discrepancies between the real wind and the logarithmic wind might be larger as the wind is increasing with height.

To address this comment, we added a summary of this answer in the revised manuscript (lines 367-375, page 17):

To assess the sensitivity of the mass balance flux estimate to the wind data source, we additionally estimated methane emissions using wind speeds derived from the ground-based meteorological mast, equipped with anemometers at 3, 6, 9, and 12 m at Bedford, and at 3 m at CEDAR, extrapolated to UAV flight height using a logarithmic wind profile ($u(z) = u_{ref} \cdot \ln(z/z_0)/\ln(z_{ref}/z_0)$), where u_{ref}: wind speed measured at the reference height z_{ref}, and z_0: surface roughness length (0.008 m at Bedford and 0.010 m at CEDAR; Anjum, 2014). The difference between the estimated fluxes using different wind inputs remains within error, both between wind inputs and with the true controlled release fluxes. We note, however, that the onboard winds give the closest agreement with controlled releases when averaged over all flights. Thus, we recommend that onboard winds are preferable for accuracy, but that mast winds can be useful, so long as errors are well accounted for and included in the uncertainty budgeting.

(2) The manuscript should include an additional caveat regarding the broader applicability of the controlled release experiments. In particular, caution should be exercised when extrapolating the reported uncertainties and performance metrics to other real-world applications, since the experimental sites, especially the Bedford site, were conducted on a former runway with relatively open and simplified terrain conditions, such environments may not fully represent the complexity of actual methane emission sites, which often involve buildings, equipment, surface roughness, and

more complicated fields. A discussion of these limitations would help clarify the scope and transferability of the presented results.

We agree. This is an important insight and guidance for others, which we have also made in our prior works (and when reviewing others' work), and we should do so here as well - thanks for the reminder. The primary objective of this study was not to derive a universal quantification bias for the UAV mass balance method or extrapolate the reported performance metrics and uncertainty estimates to other operational environments, but rather to evaluate the method under controlled conditions where the true emission rate is known, as a test of the methodological systematic performance. In most real-world applications, the true emission rate is unknown and therefore the accuracy of a quantification method cannot be assessed directly. For this reason, controlled release experiments provide a unique opportunity to evaluate method performance, investigate the drivers of quantification bias, and improve uncertainty characterisation by comparing estimated emissions against a known release rate and inform future operational use of the method (Allen et al., 2018).

The controlled release experiments were intentionally conducted in environments where emissions originated from within a well-defined source area and where the released flux could be quantified precisely. This minimises interference from neighbouring methane sources and allows a direct assessment of the measurement methodology. We agree that the Bedford site, in particular, represents a relatively simplified environment compared to many operational methane-emitting facilities, which may contain buildings, processing equipment, flares, wells, complex terrain, and multiple simultaneous emission sources. Such features can make methane quantification more challenging (e.g. Sherwin et al., 2024).

Nevertheless, the experiments were conducted under a range of realistic meteorological conditions, including neutral and convective boundary-layer conditions and wind speeds ranging from approximately 2–7 m s^{-1}. In addition, the CEDAR site provided a more complex terrain setting than Bedford. While these conditions help assess method performance under varying atmospheric conditions, the resulting uncertainty estimates should not be interpreted as universally applicable to all sites. Site-specific environmental and operational characteristics might affect the quantification performance and uncertainty.

We have added new text to the manuscript to clarify the intended scope of the controlled release experiments and the limitations associated with transferring the reported performance metrics to more complex operational environments (lines 234-239, page 9) as follows:

The experimental design, particularly at the Bedford site, represents a less complex setting than many methane-emitting facilities (e.g. landfills and oil and gas sites), which often contain buildings, processing equipment, flares, wells, and other infrastructure that can make methane quantification more challenging (Chen et al., 2024; Sherwin et al., 2024). Nevertheless, controlled-release experiments provide a unique opportunity to validate quantification methods against a known emission rate under realistic meteorological conditions while minimizing influences from external methane sources (Allen et al., 2018). Future studies in more complex industrial settings would help evaluate any associated differences in performance.

(3) The manuscript currently provides relatively limited detail regarding the configuration of the controlled releases. Additional clarification would improve reproducibility and interpretation of the results. For example, it is unclear whether each site involved only a single emission source operating at a given time, or whether multiple emission sources were released simultaneously. The duration and timing of individual release events should also be described more explicitly. In addition, the physical arrangement, geometry, and locations of the release sources would be valuable to show directly in Figure 3, as these factors may influence plume structure and methane quantification performance.

We agree that more detail on config would be useful. To address this comment, we have added two tables (Tables 1 and 2, pages 9 and 11) providing additional details regarding the configuration of the controlled releases, including the release period and rate, source type, and source location (latitude and longitude coordinates). We also clarified explicitly that the release events consisted of a single emission source operating at any given time, with UAV flights conducted while each emission rate and source configuration remained constant.

The tables are also now introduced in line 208, page 8, as follows:

Further details on the experimental sites and the prevailing meteorological conditions during the campaigns are provided in the following subsections. Detailed characteristics of the controlled release experiments at each site, including release timing, emission rates, and source configurations, are provided in Tables 1 and 2, while an overview of the experiments conducted at each site is summarised in Table 3.

In the revised manuscript, we added the following text (lines 202-206, page 8):

At each site, only a single controlled emission source was operated at any given time. Individual release events were conducted as discrete time periods with constant emission rates maintained throughout each event. Details can be found in Tables 1 and 2. UAV flights were performed while the release was ongoing, allowing sampling of fully developed plumes. Between release events, the emission system was adjusted to the next target flow rate or configuration, and no simultaneous multi-source emissions were performed.

We also added the following sentence to the Bedford Aerodrome subsection (lines 227-228, page 9):

Each release event consisted of a single emission source operated at a constant methane release rate over a defined time period (Table 1), during which UAV flights were conducted. Simultaneous operation of multiple emission sources was not performed.

In the CEDAR subsection, we added the following clarification (lines 256-257, page 11):

A total of 14 flights were conducted, ~~for~~ covering four different release rates from ~~and one source type (Table 2~~ a single ground-based linear emission source of 60 m length (Tables 2, 3). Similar to Bedford, the methane release rate was kept constant over a defined time period, and no simultaneous emission sources were operated.

The locations and configurations of the release sources have also been added to the revised version of Figure 3 (page 10). We adjusted the caption of the figure accordingly:

Figure 3: (a) Bedford Aerodrome showing methane release points (RP2–RP4), with one active at a time. (b) Aerial view of the CEDAR site showing the two locations of the ground-based linear release source (blue: 08 September 2025; red: 09–10 September 2025). The corresponding source boundary coordinates are provided in Table 2. The source was rotated by approximately 45º between measurement days to maintain a configuration approximately perpendicular to the prevailing wind direction. (c) Example UAV flight path downwind of a release point, illustrating the vertical "ladder" transect pattern used for CH_4 measurements. (d) Wind rose corresponding to the same flight. Imagery:© Landsat / Copernicus, Map data: © 2026 Google.

(4) The description around line 201 ("In total, 24 flights were carried out, for seven different release rates and three different types of emission sources") requires additional clarification. It is currently unclear whether the seven different release rates refer to temporal variations of the site-level aggregated emission plume, or to different emission sources operating within the site. Similarly, the "three different types of emission sources" appear to refer to different controlled-release geometries, but this should be explicitly stated. The authors should also provide justification for introducing these

three different release configurations and explain what types of real-world methane emission scenarios they are intended to represent. Such discussion would help readers better understand the practical relevance of the experimental design. Similar concerns apply to the description around line 218.

The seven release rates in Bedford (and the four release rates at the CEDAR site) refer to distinct controlled-release emission rates tested during the campaign. To clarify, only one release rate was active during each experiment, and the reported rate corresponds to the total methane emission released during that test. We have clarified this in the revised manuscript (see also response to Comment 3 above).

Indeed, the “three different types of emission sources” refer to three distinct controlled-release source geometries used during the validation experiments. We have also now clarified this in the revised manuscript and provided additional details of the source configurations in Tables 1 and 2 (see also response to Comment 3).

Regarding the last part of the comment, in the original manuscript, we described the experimental configurations at the Bedford site, where point and elevated sources were tested over a range of release rates (0.021–2.16 kg h^{-1}) and emission heights (ground level, 1 m, and 3 m), representing realistic emission scenarios such as equipment leaks and manure sources. We also noted that these emission rates are consistent with those reported for oil and gas infrastructure and lower-emission production facilities. Similarly, for the CEDAR site, we described the use of a 60 m linear ground-level source designed to mimic distributed landfill-type emissions, with higher release rates (8.3–40.6 kg h^{-1}) extending the tested emission range.

In the revised manuscript, we have clarified the experimental design by explicitly stating the rationale for the source configurations and their intended representation of real-world methane emission scenarios.

For the Bedford section, we added the following statement (lines 229-233, page 9):

The three source configurations were selected to represent different classes of methane emissions encountered in practice. The point source was intended to mimic localized equipment leaks, the elevated ring source to represent spatially distributed diffuse emissions (e.g. from waste or manure storage surfaces), and the vertical linear source to simulate emissions distributed over a vertical structure or volume (e.g. manure piles). Together, these configurations allowed evaluation of the UAV-based mass balance method across a range of source geometries and resulting plume structures.

For the CEDAR section, we added (lines 257-259, page 11):

This configuration was selected to represent spatially distributed surface emissions typical of landfill environments, where methane is released over extended areas rather than from isolated point sources.

(5) In the “Results and Discussion” section, the authors should consider adding additional subheadings to improve the organization and readability of the manuscript, such as subsections “Flight-level methane quantification” and “Analysis of factors affecting methane quantification”

Good idea. In the revised manuscript, we added the following subsections to improve the organisation and readability of the “Results and Discussion” section:

3.1 Per-flight methane quantification and validation

3.2 Factors affecting quantification performance

(6) The statement around line 262 ("we see that there is no strong fundamental dependence on the release rate or the source type") should be revised for consistency with the rest of the manuscript. Based on Figure 5 and other discussions in the paper, the bias at lower release rates appears noticeably larger than that at higher release rates, and the authors also acknowledge this trend elsewhere in the manuscript. The authors are encouraged to revise this statement to more accurately reflect the observed trends and maintain consistency throughout the paper.

We agree that the original wording was not fully consistent with the trends shown in Fig. 5 across the whole emissions range. The manuscript has been revised to clarify that no systematic dependence on source type is observed, noting that the bins contain a mixture of different source types. We now explicitly state that the largest negative biases occur at the lowest release rates and generally reduced bias at higher release rates (Table 4). We modified the revised manuscript (lines 315-323, page 15) as follows:

Across all flights, the average bias is approximately -25% compared to the controlled released rates. Focusing on the different bins in Fig. 5, which include a mixture of source types, there is no evidence of systematic differences in bias attributable to source type within the observed variability. ~~we see that there is no strong fundamental dependence on the release rate or the source type. We can conclude from this~~ This suggests that the quantification method ~~is capable of capturing both~~ performs consistently across the range of source geometries considered, including small localised emissions close to the ground (ring or point emissions located at 1 m), larger extended ground-level emissions (60 m linear sources) and vertically distributed emissions (linear vertical sources from ground to 3 m). However, the largest negative biases occur at the lowest release rates, while higher release rates generally show reduced bias (Table 4). ~~Indeed, the lowest bias is observed for higher release rates (Table 4), as~~This is consistent with larger concentration enhancements relative to background variability at larger emission rates, resulting in a higher signal-to-noise ratio and reduced sensitivity to turbulence and wind variability. This leads to a more coherent plume structure and more robust flux estimation.

(7) The statement regarding the "largest bias of -60.55" is somewhat unclear. It is not evident whether this value refers to single flight-level observation or an average across flights 20-24. The authors should clarify the statistical basis of this reported bias value to avoid ambiguity.

The reported value refers to the mean bias across flights 20–24 (all flights conducted at the 0.45 kg h^{-1} release rate, Table 4 of the revised manuscript), rather than a single flight-level observation. The manuscript has been revised to clarify this point in lines 359-362, page 16:

The largest mean bias of -60.5% is observed for flights 20-24 in Bedford (Fig. 6 a), corresponding to the average bias across ~~all 5~~ the five flights conducted at the ~~in this rate~~ 0.45 kg h^{-1} release rate on ~~are conducted in~~ 16/05/2024. During these flights, ~~when~~ the average wind speed ~~during the flights~~ ranges from 2.3-3.2 m s^{-1}, representing the lowest wind speed compared to the rest of the flights.

(8) The manuscript concludes that extrapolating the lowest transect measurements to ground level improves agreement between estimated and known methane fluxes, based on the increased percentage of flights whose known release rates fall within the estimated uncertainty bounds. However, this analysis alone may not fully demonstrate improved quantification accuracy. In particular, it is unclear whether the apparent improvement arise from a genuine reduction in bias or simply from enlarged uncertainty intervals after extrapolation. The authors should provide additional analysis (changes in mean bias) to better support this conclusion.

As stated in the original manuscript, the extrapolated value is not included (added to) in the central estimate of the flux, but it is taken into account by extending the (positive only) uncertainty range of the estimated flux. Therefore, the central flux estimate, the bias and the root mean square error (RMSE) relative to the controlled release rate remain unchanged. We do not believe that adding any surface-extrapolated flux to the main estimate would be sensible (as we just don't know the concentration gradient there for sure), and that it should only ever be reflected in the uncertainty budget in any mass balance method. Indeed, the percentage of flights for which the known release rate falls within the estimated uncertainty bound increases when the extrapolated value is taken into account. This increase shows that there is better uncertainty coverage in this case, rather than lower bias or higher accuracy for the method (which remains unchanged). To better clarify this part, we revised the manuscript (lines 347-355, page 16), as follows:

In order to deal with this underestimation issue, we extrapolated the measurements of the lowest transect to the ground level to fill the gap above the surface (Sect. 2.2). As discussed in Sect. 2.2, this extrapolated contribution is considered as an additional uncertainty term, rather than being added to the central emissions estimate. Table 5 shows the percentage of flights where the known release rate lies within the uncertainty bounds of the estimated emission rates for each site. Indeed, ~~the agreement between the known and the estimated fluxes~~ this percentage has increased after ~~this~~ the modification of the method~~, as the percentage increased~~ from 41.6 to 70.8% and from 50 to 71.4% for Bedford and CEDAR, respectively. This indicates improved coverage of uncertainty intervals and a more complete representation of the uncertainty associated with the potential incomplete sampling of the plume near the surface. However, the central flux estimates (thus the bias and RMSE) remain unchanged, and discrepancies still remain between the known and the estimated flux.

(9) One key assumption of mass-balance method is that the entire plume cross-section is adequately sampled. The manuscript should provide more evidence that the UAV flight paths fully captured that methane plume under all experimental conditions. It would be helpful to include additional information, such as a simple schematic showing the flight paths relative to the plume, explanation of how the transect spacing was selected, discussion of possible underestimation if part of the plume was missed during measurements.

Ideally, continuous and simultaneous three-dimensional measurements would be required throughout the plume volume; however, this is not achievable with a single in situ platform (e.g. Yong et al., 2024). Consequently, all UAV-based mass-balance approaches rely on sampling strategies and associated assumptions regarding plume coverage. Complete verification that the entire plume cross-section has been captured is therefore inherently challenging. In addition, it can be difficult to determine whether part of the plume has been missed, particularly in cases where methane remains close to the surface and passes beneath the measurement plane due to operational restrictions that prevent UAV flights only a few metres above ground level (e.g. McManemin et al., 2025). Nevertheless, individual flight surveys can be assessed for indications of incomplete plume interception by examining the spatial distribution of methane enhancements within the measurement plane and the resulting flux estimates, which was the approach taken here.

One such case is already discussed in the manuscript, where a negative bias (-60.5%) is observed for flights 20–24 conducted on 16/05/2024. We attribute this underestimation to low wind conditions (2.3–3.2 m s^{-1}), which may have reduced plume advection and led to incomplete interception of the plume by the UAV transects located approximately 100 m downwind. Under such conditions, the plume is more likely to remain closer to the source and near the ground, increasing the probability of partial sampling and underestimation

Regarding the flight paths and sampling geometry, Figures 2 and 3 already provide complementary information on the sampling strategy (and flight path etc). Figure 2 illustrates the measurement plane, UAV transects, and the interpolated methane concentration field used for the mass-balance calculation, while Figure 3 shows an example flight path in relation to the source location. Together, these figures demonstrate the spatial sampling approach adopted throughout the study and the relationship between the UAV measurements and the emission source.

Regarding the transect spacing more generally, the selected configuration represents a compromise between spatial resolution, plume coverage, and flight endurance constraints. The spacing and dimensions of the measurement plane were chosen to maximise the likelihood of intercepting the full plume while ensuring sufficient vertical coverage and allowing the survey to be completed within a single battery cycle. Under most experimental conditions, the selected transect spacing appeared adequate (on inspection of the mapped data) to characterise the plume cross-section, as indicated by the overall agreement between the estimated and controlled release rates; however, the potential chance for under-sampling cannot be completely excluded, particularly under weak and variable wind conditions - conditions we would advise are challenging for UAV surveys using mass balancing.

(10) The background CH4 determination method requires additional justification and discussion. Since both plume and background measurements are included within the same dataset, the resulting histogram may be influenced by plume-enhanced concentrations, potentially biasing the estimated background concentration. This issue may become particularly significant for low-emission cases where plume enhancements are comparable to background variability. In addition, the manuscript states that the background range was "manually selected" based on visual identification of the distribution tail transition. This approach appears somewhat subjective and may introduce operator-dependent uncertainty. The authors should clarify the selection criteria in a more quantitative and reproducible manner and discuss the sensitivity of the results to the chosen background range.

The determination of background CH_4 concentrations remains an open topic in UAV-based emission quantification, and no universally accepted approach currently exists. As discussed by Parrish et al. (2012), the definition of background concentrations inherently involves a degree of subjectivity. For UAV-based quantifications, one ongoing discussion concerns whether background concentrations should be derived from dedicated upwind measurements or estimated directly from the downwind survey data when it can be assumed sampling was outside of the plume of an emitter (e.g. Shaw et al., 2021, Fosco et al., 2025)

In this study, background concentrations were derived from measurements collected during the downwind survey (when out of plume) rather than from separate upwind flights. This choice was motivated by both practical and scientific considerations. First, field observations indicated that ambient CH_4 concentrations can vary on timescales of approximately one hour, meaning that a flight-specific background derived from measurements collected with plume sampling likely better represent the atmospheric conditions during the survey. Second, deriving the background from the downwind survey increases the number of available measurements and eliminates the need for additional flight time dedicated to upwind characterization.

To estimate the background concentration, histograms were constructed from all CH_4 measurements collected during each flight, and the modal concentration was selected as the representative background value following Yong et al. (2024). Previous studies have shown that, when a plume is present, background air typically constitutes the largest fraction of observations and forms the dominant peak of the concentration distribution, whereas plume-affected measurements generate a right-skewed tail toward elevated concentrations (Allen et al., 2019; Golston et al., 2018; Yang et al., 2018; Fosco et al., 2025; Yong et al., 2024). Under these conditions, the mode is less influenced by plume-enhanced observations than statistics such as the mean and therefore provides a robust estimate of the ambient background concentration. The resulting background concentrations were generally close to 2 ppm, consistent with the global atmospheric CH4 background concentration (NOAA Global Monitoring Laboratory, 2026) and supporting the validity of the adopted approach.

We acknowledge that the determination of the background uncertainty range contains a degree of expert judgement. And we believe it always will (and should), so long as surveyors are open and clear about their approach and understand the issues above. Put simply, there is no one-size-fits-all approach to this problem and some well-informed flexibility is preferable for accuracy over rigidity (even if at the expense of standardization). Because plume strength and histogram shape varied among surveys, a fully automated threshold was not found to be suitable for all cases. Instead, we selected a conservative range around the background peak for each flight to encompass both ambient variability and uncertainty associated with the background definition itself.

We have clarified this procedure in the revised manuscript and added the following parts in the manuscript (line 491-500, page 24):

To determine the background CH_4 concentration, we analysed the distribution of CH_4 measurements from each flight by constructing a histogram of all recorded data points (Fig. A1). The background concentration was identified as the mode of the distribution—the value with the highest frequency, following Yong et al., 2024. Background measurements were obtained from the edges of the gas plume, as the flight path was extended toward the plume's outer limits.

One advantage of this approach is that it eliminates the need for upwind flights to estimate background concentrations, allowing for increased data collection within the plume itself (Yong et al., 2024). In the presence of a plume, background air typically accounts for most measurements, whereas plume-affected observations form a right-skewed tail at higher concentrations (Allen et al., 2019; Golston et al., 2018; Yang et al., 2018; Fosco et al., 2025). Consequently, the mode is generally representative of background conditions and is less sensitive to plume-enhanced observations than the mean. ~~However, because both plume and background points are included within the same dataset, the resulting distribution is skewed, making it difficult to clearly separate background from plume measurements. To address this, we manually selected the background range by identifying the point where the left side of the distribution transitions into the tail. This range was selected to assess background variability, which allows for estimating the uncertainty in the background concentration and, consequently, the flux calculation (see Equation 1).~~

Background uncertainty was estimated from the spread of measurements around the modal peak. Because plume strength and histogram shape varied between surveys, a single automated threshold was not suitable for all flights. Instead, the background range was selected from the dominant background peak while excluding the plume-enhanced tail. This range was chosen to capture both ambient CH_4 variability and uncertainty associated with the background determination and was subsequently propagated through the flux calculations (see Equation 1).

Minor comments:

(1) Section "2.2 UAV-based mass balance flux method" currently contains only a single subsection. The authors may consider removing the subsection heading and integrating the content directly into the main section for improved formatting consistency and readability.

To improve readability and formatting consistency, we followed the reviewer's suggestion and removed the separate subsection heading. The content of "Adaptation of the UAV-Based Mass Balance Flux Method" has been integrated into Section 2.2, resulting in a single section entitled "UAV-based mass balance flux method and adaptation."

(2) Equation (3) may be difficult to read in its current format. The authors may consider separating the upper and lower uncertainty bounds into two lines to improve clarity and readability.

To improve readability, Equation (3) has been reformatted by separating the upper and lower uncertainty bounds into two lines.

(3) In the right y-axis label of Figure 2, please correct the methane notation from “ch4” to the proper chemical format “CH4”

Done

(4) At line 311, the symbol Δθ appears for the time without definition or explanation. The authors should clearly define this parameter upon its first occurrence to improve clarity and readability.

To address this comment, we have clarified the definition of Δθ at its first occurrence in the revised manuscript (line 380-383, page 18), as follows:

To identify potential factors that most strongly influence the quality of the agreement between the estimated and controlled release rates, Fig. 7 presents linear regression analyses between the absolute flux bias (%) and six explanatory parameters related to wind conditions, source geometry, and release characteristics: mean wind speed, standard deviation of wind direction, distance from the source, controlled release rate, Δθ (defined as the angular range defined at the source by lines connecting the source to the endpoints of the UAV measurement transect), and the percentage of measurements outside Δθ. ~~We define a~~ This source–transect angular ~~window, Δθ, as the angular range defined at the source by lines connecting the source to the endpoints of the UAV measurement transect. This~~ window represents the range of wind directions for which emissions released at the source are advected towards and intersect the UAV sampling plane.

(5) At line 291, the authors should provide a clear definition or explanation of the “dead sampling” when the term is first introduced, as its meaning may not be immediately clear to all readers.

We have added the definition of the term ‘dead sampling’ (sampling of background air (i.e. regions outside the plume))’ in the revised manuscript (line 346, page 16):

In case this is not feasible due to safety constraints or because it is rather complicated and (cost and logistically) expensive, it is important to develop the measurement plane further away from the source, thus allowing sufficient mixing of the gas and reducing any near-surface impact due to dead sampling, defined as sampling of background air (i.e. regions outside the plume).

(6) All abbreviations and acronyms should be defined in full upon their first appearance in the manuscript to improve clarity and readability for a broader audience.

Our apologies for this. We have carefully reviewed the manuscript and ensured that all abbreviations and acronyms are defined at their first occurrence. The following revisions we made in the manuscript:

Line 10, page 1:

Overall, fluxes calculated using the method show good correlation but substantial scatter at the individual flight level (CEDAR: slope = 1.01, Pearson correlation coefficient (r) = 0.84, root mean squared error (RMSE) = 38.8%; Bedford: slope =0.71, r = 0.85, RMSE = 46.2%) and substantially improved agreement when averaged by release rate (CEDAR: slope = 1.01, r = 0.94, RMSE = 19.9%; Bedford: slope = 0.68, r = 0.99, RMSE = 38.4%).

Line 270, page 12:

At Bedford, the ordinary least squares (OLS) fit yields a slope of 0.71 with Pearson correlation coefficient (r) = 0.85, indicating a systematic underestimation and moderate dispersion (root mean square error (RMSE) = 46.2%, bias = −38.5%).

Line 32, page 2:

Responsible for nearly 30% of observed global warming (Shan et al., 2025), atmospheric CH_4 concentrations have increased 2.6-fold since 1750, reaching 1942 ±2 parts per billion (ppb) in 2024 (WMO, 2025).

Line 100, page 4:

This ensured accurate geolocation of gas concentration and wind measurements. Automated flight paths were pre-planned using the Universal Ground Control Software (UgCS; SPH Engineering, 2025), providing repeatable surveys with constant speed, altitude, and coverage, and reducing operator bias (which can translate into emissions bias as described in d in Shah et al. (2020a, b)).

Caption in Fig. A1, page 23:

Figure A1. Histogram of CH4 measurements from Flight 16 on 15 May 2024. The modal concentration (red dashed line; 2.070 parts per million (ppm)) was selected as the representative background CH4 concentration. The uncertainty range (black dashed line; 2.125 ppm) corresponds to ±1 standard deviation, visually estimated around the mode, and was used to propagate uncertainty in the flux calculations (see Sect. 2.2).

References

Allen, G., Williams, P., Ricketts, H., Shah, A., Hollingsworth, P., Kabbabe, K., Helmore, J., Finlayson, A., Robinson, R., Rees-White, T., Beaven, R., Scheutz, C., and Fredenslund, A. M.: Validation of landfill methane measurements from an unmanned aerial system: Project SC 160006, Tech. rep., Environment Agency, UK, https://www.gov.uk/government/publications/validation-of-landfill-methane-measurements-from-an-unmanned-aerial-system, 2018.

Chen, Z., El Abbadi, S. H., Sherwin, E. D., Burdeau, P. M., Rutherford, J. S., Chen, Y., ... & Brandt, A. R. (2024). Comparing continuous methane monitoring technologies for high-volume emissions: A single-blind controlled release study. *ACS ES&T Air*, *1*(8), 871-884.

Fosco, D., et al. "Innovative drone-based methodology for quantifying methane emissions from landfills." *Waste Management* 195 (2025): 79-91.

Golston, L.M., Aubut, N.F., Frish, M.B., Yang, S., Talbot, R.W., Gretencord, C., Mc Spiritt, J., Zondlo, M.A., 2018. Natural Gas Fugitive Leak Detection Using an Unmanned Aerial Vehicle: Localization and Quantification of Emission. Rate. Atmosphere 9, 333. https://doi.org/10.3390/atmos9090333.

McManemin, A., Juéry, C., Blandin, V., France, J. L., Burdeau, P., and Brandt, A. R.: Controlled release testing of commercially available methane emission measurement technologies at the TADI facility, EGUsphere, 2025, 1–21, 2025

NOAA Global Monitoring Laboratory – The NOAA Annual Greenhouse Gas Index (AGGI): https://gml.noaa.gov/aggi/aggi.html, last access: 05 June 2026.

Parrish, D. D., Law, K. S., Staehelin, J., Derwent, R., Cooper, O. R., Tanimoto, H., Volz-Thomas, A., Gilge, S., Scheel, H.-E., Steinbacher, M., and Chan, E.: Long-term changes in lower tropospheric baseline ozone concentrations at northern mid-latitudes, Atmos. Chem. Phys., 12, 11485–11504, https://doi.org/10.5194/acp-12-11485-2012, 2012.

Shaw, J. T., Shah, A., Yong, H., and Allen, G.: Methods for quantifying methane emissions using unmanned aerial vehicles: a review, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379, 20200 450, https://doi.org/10.1098/rsta.2020.0450, 2021.

Sherwin, E. D., El Abbadi, S. H., Burdeau, P. M., Zhang, Z., Chen, Z., Rutherford, J. S., Chen, Y., and Brandt, A. R.: Single-blind test of nine methane-sensing satellite systems from three continents, Atmos. Meas. Tech., 17, 765–782, https://doi.org/10.5194/amt-17-765-2024, 2024.

Yang, S., Talbot, R.W., Frish, M.B., Golston, L.M., Aubut, N.F., Zondlo, M.A., Gretencord, C., McSpiritt, J., 2018. Natural Gas Fugitive Leak Detection Using an Unmanned Aerial Vehicle: Measurement System Description and Mass Balance Approach. Atmosphere 9, 383. https://doi.org/10.3390/atmos9100383.

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, 47–54, https://doi.org/10.1016/j.wasman.2024.03.025, 2024.