Ms. Ref. No.: EGUSPHERE-2024-3161

Title: Evaluating the feasibility of using downwind methods to quantify point source oil and gas emissions using continuous monitoring fence-line sensors

The Powerhouse Energy Campus
Colorado State University
430 North College Avenue
Fort Collins, CO 80524

E-mail: Mercy.Mbua@colostate.edu

June 10th, 2025

Dear Professor Presto,

We appreciate the time and effort you, the reviewers and community dedicated to providing feedback on our manuscript. We have closely followed the suggestions and have revised the manuscript accordingly. We believe the revisions have improved our manuscript and that it is ready for publication. We have listed the original comments and our response in blue below for our reply to comments. All figure numbers, tables, and lines refer to the updated manuscript. Data for reviewers' is available at Reviewer URL: http://datadryad.org/stash/share/7s42bajRb2czaY9hr6NbTKDAolCWZ_pNrty0o54qXBE.

We look forward to hearing from you.
Please find our detailed responses below.

Yours sincerely,

Mercy Mbua (corresponding author)

and co-authors: Stuart N. Riddick, Elijah Kiplimo, Kira B. Shonkwiler, Anna Hodshire and Daniel Zimmerle

Reviewer comments

Eben Thoma (RC1)

General comments

The authors added additional detail on eddy covariance (EC) and modified the analysis approach for EC, aerodynamic flux (AF), and the Gaussian plume inverse method (GPIM). A backwards Lagrangian stochastic (bLs) analysis based on WindTrax was added. The authors also provided links to supporting information on the specifics of the controlled release trials and other data. The revised manuscript carries new discussion and conclusions. There are two major concerns with the revised manuscript:

(1) Although an improved EC data analysis approach is presented, there remains concerns with the configuration and some aspects of method quality assurance (detailed below).

(2) The GPIM analysis changed from the original manuscript now showing large overestimates, especially for the multi-release single-point scenarios (MSRP). A similar result is found for bLs. The GPIM equation (SI Equation 7) differs from the original form, and it is unclear if ground reflection is considered (see identical terms in denominator). Please check for possible calculation error that could drive bias.

Author's response:

Thank you for your comment. The GPIM equation had a typo, and we have corrected it. We have redone our analysis using the original submission equation that accounts for the height of the boundary, and using the standard Gaussian equation (Supplementary Information Equations 7 & 8). Based on your concern about using the bootstrapped mean as a metric for evaluating the models in the comment below, we have now presented our data using a linear regression and this has ensured correct models evaluation in alignment with our original submission.

For the Gaussian Plume Inverse model, we tested the model in six different scenarios to evaluate both the equations and generation of dispersion coefficients using single-release single-point emissions to test parameters where the model works best, and this was used for multi-release single-point emissions quantification.

For the backward Lagrangian stochastic model, data where the measurement point is on the edge of touchdown or outside of the touchdown is flagged by the model as (-9999) and this was filtered out.

Further, models comparison using a data subset, 15-minutes 10 degrees wind sector range has been added to the manuscript.

Changes to Manuscript

Section 2.4.2.2

Lines 290-295

The GPIM was evaluated under six scenarios (two equations and three different dispersion coefficients generations) using single-release single-point emissions to test when the model works best (Supplementary Information Section 2a: Equation 7 and 8). Dispersion coefficients were generated based on (1) high frequency sonic anemometer data at ~ 10 Hz, (2) EPA point-source dispersion coefficients (US EPA, 2013), and (3) 1 Hz sonic anemometer data. The

scenario with the slope closest to 1, and highest R^2 across averaging durations, and wind sector ranges was selected and used for multi-release single-point emissions quantification.

From the controlled release data file, there are 713 individual releases. This reviewer counts N=69 clearly defined single point single source trials with a mean (min, max) emission rate of 1.3 (0.2, 6.6) kg/hr and a release duration of 2 hrs and 41 minutes (0:19, 7:10). This represents a very reasonable dataset that can be understood. The remainder of the dataset consists of multiple release points grouped as trials, some very complex in design with many source locations and mixtures of long and short release durations. It is not clear how the MSRP set is formed and screened since almost all multipoint release trials involve multiple release locations. Even with use of mean wind direction as an upwind source screening tool, the potential influence of off-axis sources under meandering winds over the duration of a trial is difficult to eliminate. The authors acknowledge this possibility in Section 4.3(4.4) for both the GPIM and bLs result. However, this has as a direct bearing on the conclusions drawn as the "interfering sources" may dominate the MSRP result for GPIM and bLs. Additionally, the bLs point source analysis by WindTrax can strongly diverge when source is located near the edge of the touchdown cone (similar to Figure 2B). This is nonphysical effect that is the cause of the extreme outlier results (e.g., MRF = 11958) and should be evaluated.

Author's response:

Thank you for your comment. We agree that using the average wind direction as an upwind screening tool is effective for clearly isolated, single-source emissions, but more problematic in multi-source trials where wind meandering and overlapping plumes introduce significant uncertainty. We acknowledge that in such cases, plume interference can compromise the accuracy of quantification, particularly for point-source dispersion models such as GPIM and bLs.

As noted, these complexities are not limitations of the modeling approaches per se, but rather reflect the constraints of our experimental setup. Our single-sensor configuration and the decision to include multi-source trials were intentional, aiming to evaluate model performance under realistic but challenging conditions—such as those encountered in continuous emissions monitoring. In response to Reviewer 2's suggestion, we have revised the manuscript to more explicitly clarify the influence of experimental design on source interference and model performance. We now emphasize that, in practical applications, improved source localization and screening can often be achieved using multiple sensors, which were beyond the scope of this study. Regarding the extreme outliers in the bLs point source analysis (e.g., MRF = 11958, in the revised plot, extreme estimates of up to 2*10^5 kg h^{-1}), we agree this divergence can occur when the estimated source lies near the edge of the touchdown cone—a known artifact of the model's geometry and assumptions. This can also occur due to unsteady emissions due to plume meandering, or inaccurate atmospheric stability (L) over such a short time. This could be the case as the estimated emissions begin to stabilize at 15 and 30 minutes duration shown in the revised plot.

Changes to Manuscript

Section 4.2

Lines 518-527

The MSRP emission profiles tested in this study were complex challenging the GPIM application as the method is a point-source specific quantification approach and works best in open areas, free of obstacles, and when the background

concentration is well defined. For multiple emissions, even when the sensor is nominally downwind of a single source based on the average wind direction, quantification can be complicated by interference from neighboring sources. However, it is important to emphasize that such complexity is not a fundamental limitation of quantification itself, but rather a function of the experimental design and study objectives. For example, plume interference can often be minimized through strategic localization and optimization using multiple sensors—an approach that differs from the single-instrument setup used in this study. This study's design involves defining plumes based on wind sector ranges, as opposed to using multiple sensors to localize sources, and therefore does not replicate how various continuous monitoring solutions typically operate.

Section 4.4

Lines 560-566

Oil and gas point sources could either be single emissions or multiple emissions occurring concurrently. In this study's design, cases involving multiple emissions with more than one release point located upwind posed challenges for the specific Gaussian and backward Lagrangian stochastic (bLs) model implementations, which were applied assuming a single active source at a time. While these models can be extended to handle multi-source scenarios, the assumptions used here limited their ability to distinguish individual contributions when plumes overlapped. As a result, interference from neighboring emissions introduced ambiguity in model-observation alignment, particularly under complex wind conditions.

Comments on Main Paper:

Comment 1

Specific comments regarding EC.

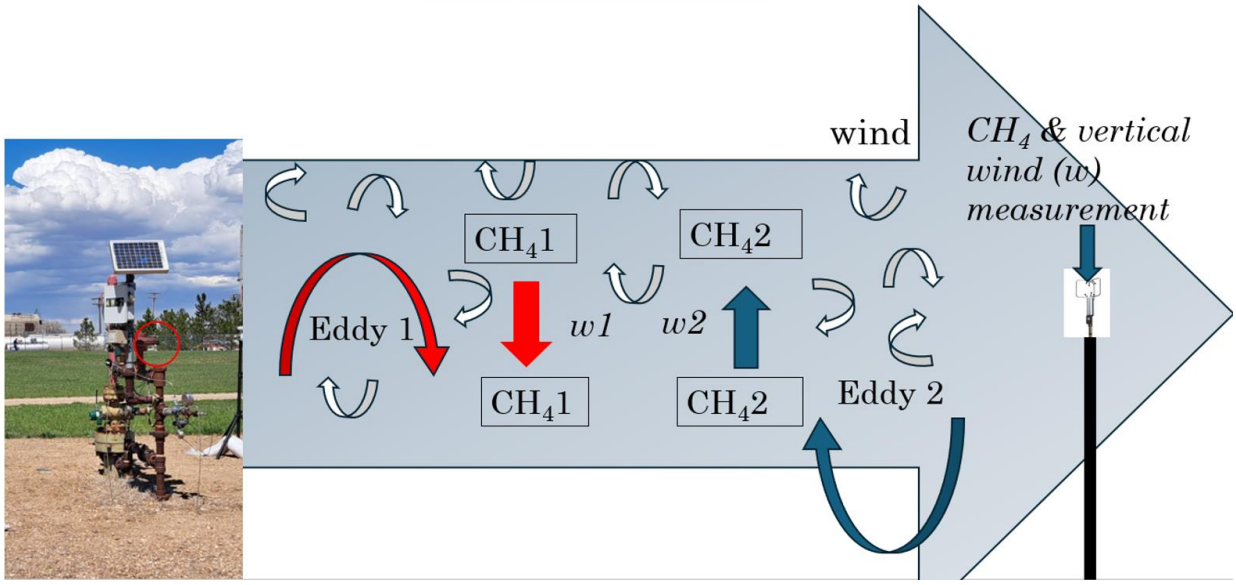
Despite being improved from the original submission, there seems to be a fundamental lack of understanding of the eddy covariance method that undermines the experimental design, QAQC, and analysis in this study.

1. Figure 1A has an incorrect depiction of the eddy covariance method. It seems to imply downward eddies are negative fluxes and upward eddies are positive. In fact there are always both downward and upward eddies in contexts with either negative and positive fluxes.

Author's response:

Thank you for your comment. We have accordingly modified the depiction for eddy covariance and aerodynamic flux gradient

Eddy Covariance



2. From the responses from the authors, it is clear the data were not collected in a manner suitable for eddy covariance. For example, both the CH4 and the wind data were not sampled at 10 Hz. Instead they were sampled at variable frequencies from 4 - 9 Hz. At best the authors could average down to the lowest common frequency and caveat their results to acknowledge they are missing all of the high frequency eddies. It is inappropriate to interpolate up to a higher frequency as the authors have done here.

Author's response:

Thank you for your comment. Yes, our system did not sample data at the recommended >10 Hz frequency (varied between 4 and 12 Hz) suitable for eddy covariance. However, we did not interpolate the data to the highest frequency, instead, we filtered out our dataset when the frequency was 8 Hz and above. We down sampled all 8Hz and greater data to an 8 Hz common frequency. To make this clear to our readers, we have made the following changes to the manuscript.

Changes to Manuscript:

Section 2.4.1.1 Lines 195-201

Evaluating the MGGA CH_4 data showed that actual sampling was between 4 and 12 Hz (majority of the data collected at approximately 6 Hz), even though the analyzer had been configured to sample at 10 Hz (Supplementary Information Section 2b). To account for this sampling variability, data were filtered to when sampling was equal to or greater than 8 Hz. Data sampled at frequencies above 8 Hz were down sampled to 8 Hz. The 8 Hz frequency threshold was selected to ensure uniform sampling, enough data for model evaluation as most sampling was at lower frequencies, and to preserve as much temporal resolution as possible given the system limitations.

Section 4.1 Lines 475-485

Our results were derived from data filtered to include only periods with sampling frequencies ≥8 Hz, which significantly reduced the number of usable emission measurements. Although the instrument was configured to sample at 10 Hz, it did not consistently achieve this rate. This discrepancy may be attributed to instrument-related factors such

as the 0.4-second gas flow response time, which could delay analysis of the drawn air sample in the cavity, or the use of a 3 lpm pump with 3 meters of tubing, which reduced the effective turnover rate. The dataset used for eddy covariance evaluation was predominantly flagged as low quality (flag 2) according to the Mauder and Foken (2004) quality control test, which classifies flux data based on steady-state conditions and the presence of well-developed turbulence (flags 0 = high, 1 = intermediate, 2 = low quality). Many of the low-quality flags were likely driven by wide deviation in w/CH_4 stationarity reflecting intermittent plume capture, where the EC system alternated between sampling emitting and non-emitting regions. The EC model produced negative emission rates associated with negative fluxes during periods of high non-stationarity (Supplementary Information, Section 2c. iv).

3. The authors did not sync the dataloggers between the sonic anemometer and CH4 gas analyzer, so they have no reliable way to match the CH4 series averaged down to 8 Hz with the sonic data. In an ideal circumstance, out of sync time series can be compensated for if there is a clear point of maximum covariance when determining the time lag. However, the authors used an insufficient pump speed in their closed path system with a long tube length, which would induce both attenuation and massive time lag.

Author's response:

We acknowledge that asynchronous CH_4 and sonic data collection with no accurate way of matching time series due to insufficient pump speed and a long tube is a significant source of error. We acknowledge this in our manuscript with the following changes:

Changes to Manuscript:

Section 2.4.1.1 Lines 203-209

As the MGGA gas analyzer and sonic anemometer were not designed to clock synchronously, using the MGGA CH4 clock time as a reference, meteorological data from the sonic anemometer were matched to the MGGA CH4 data using linear interpolation to generate concentration-meteorological 8 Hz data. While in an ideal circumstance of a fast pump and short tube length a correct timeseries matching can be achieved through establishing a clear point of maximum covariance when determining the time lag, this is difficult for our system due to a 3 lpm pump flowrate and a 3 m tubing that caused both attenuation and time lag.

Section 4.1 Lines 493-495

We acknowledge that the system used in this study was not designed or configured for standard eddy covariance analysis, and that this limitation impacts the interpretation of our results in the context of EC-based flux quantification.

4. The authors report that their closed path intake was 10 cm below the location of their sonic anemometer. This level of vertical separation would not matter for an instrument height of 40 meters for example, but for short towers such as theirs, vertical separation should be avoided, and is an additional source of error.

Author's response:

Thank you for your comment. The intake for the closed-path system was positioned approximately 10 cm below the sonic anemometer to protect the inlet tubing from debris and precipitation by mounting it on an aluminum shield facing downward. We recognize that even this small vertical separation can introduce additional errors in flux

measurements when using short towers. This design choice was a compromise to ensure instrument protection while maintaining data collection in field conditions. We have now clarified this limitation in the manuscript and acknowledge its potential impact on our results.

Changes to Manuscript:

Section 4.1 Lines 489-493

The intake for the closed-path system was positioned approximately 10 cm below the sonic anemometer to protect the inlet tubing from debris and precipitation by mounting it on an aluminum shield facing downward. We recognize that even this small vertical separation can introduce additional errors in flux measurements when using short towers. This design choice was a compromise to ensure instrument protection while maintaining data collection in field conditions. We acknowledge that the system used in this study was not designed or configured for standard eddy covariance analysis, and that this limitation impacts the interpretation of our results in the context of EC-based flux quantification.

5. Given all of these issues with the data collection, the transfer function corrections are likely substantial and probably outside of acceptable bounds. The original review recommended an analysis of the cospectra, which the authors did not do.

Author's response:

We have added cospectra analysis to the supplementary information.

Changes to Manuscript:

Section 2.4.1.3 Lines 239-242

Cospectral analysis revealed that the EC system in this study smoothed out low-frequency eddies, as the cospectra lack the ideal shape characterized by a low-frequency rise, a peak region, and a high-frequency decay (Supplementary Information Section 2c.i). While the slope in the high-frequency region varies around the theoretical −4/3 slope, the cospectral data followed the 1:1 line, indicating consistent spectral shape across sampling periods.

6. The authors need to provide evidence for their u* threshold choice based on the independence of fluxes from u*. This is a routine analysis and figure included in supplements for any eddy covariance study.

Author's response:

We evaluated the relationship between CH_4 flux and u* to identify the point where the fluxes are independent of u_*. We have added the CH_4 flux vs u_* plots under supplementary information.

Changes to Manuscript:

Section 2.4.1.3 Lines 243-248

We also examined the relationship between CH_4 flux and friction velocity (u*) to identify a u threshold below which flux estimates may be unreliable (Supplementary Information Section 2c. ii). However, no consistent relationship was observed across atmospheric stability classes (unstable, stable, and neutral). CH_4 fluxes varied widely—including both positive and negative values—across the full range of u* (~0 to 1 m s^{-1}), with no discernible threshold beyond which

fluxes stabilized. This indicated that CH_4 fluxes were effectively independent of u*, and thus, data from all u* values were retained.

7. The authors use time intervals shorter than the typical 30 minutes. While 15 minutes can sometimes be sufficient, smaller intervals than that almost always are neglecting low frequencies of eddies and result in incorrect fluxes. Anytime an eddy covariance study deviates from the 30 minute standard, it has to be justified with an ogive analysis, as was requested in the initial review.

Author's response:

We have reanalyzed our data at 5-, 15- and 30-minute averaging periods. For all these averaging durations, ogive plots have been added under supplementary information.

Changes to Manuscript:

Section 2.4.1.3 Lines 248-257

Ogive analysis was conducted to assess whether averaging durations of 5, 15, and 30 minutes were sufficient for capturing the full turbulent flux. The resulting ogive curves deviated from the ideal asymptotic shape, particularly at the highest and lowest frequencies. Notably, the curves did not exhibit a clear plateau near the low-frequency end, where cumulative flux should approach unity. This indicates incomplete flux capture. Furthermore, the similarity in ogive shapes across different frequencies—mirroring patterns seen in the cospectra—suggests a lack of significant turbulent contributions and the influence of non-turbulent, possibly advective, processes. These results imply that the EC system may not have fully resolved the flux due to either insufficient averaging time, non-stationarity, or instrument-related limitations (Supplementary Information Section 2c.iii). As positive fluxes are generally considered emissions, and negative fluxes depositions, data were further filtered for positive fluxes which were then quantified to emission rates.

8. The authors interpret the large amount of data flagged as poor quality as a consequence of low turbulence or bad system design. It is most likely that non-stationarity is the main problem, as they are including extremely brief emissions that are shorter than the averaging interval.

Author's response:

We used release-events where release durations were equal or longer than the averaging interval. Shorter releases were excluded from the analysis. We also acknowledge that even though the release might have been steady, changes in wind direction, flow distortion caused by presence of nearby equipment, and differences between the source and measurement height could contribute to non-stationarity as evidenced by presence of negative fluxes in some cases.

Changes to Manuscript:

Section 4.1 Lines 479-485

The dataset used for eddy covariance evaluation was predominantly flagged as low quality (flag 2) according to the Mauder and Foken (2004) quality control test, which classifies flux data based on steady-state conditions and the presence of well-developed turbulence (flags 0 = high, 1 = intermediate, 2 = low quality). Many of the low-quality flags were likely driven by wide deviation in w/CH4 stationarity reflecting intermittent plume capture, where the EC system alternated between sampling emitting and non-emitting regions. The EC model produced negative emission

rates associated with negative fluxes during periods of high non-stationarity (Supplementary Information, Section 2c. iv).

9. The authors used a footprint model that previous studies have found to perform poorly in controlled release experiments, against the recommendation of the original review. There are several other widely used footprint models, one or more of which should at least be used in conjunction with the model the authors have chosen to use. The authors claim they cannot implement the Kormann and Meixner footprint model without multiple sonic anemometers, but they are incorrect.

Author's response:

We have accordingly recalculated the footprint using both the Klujn, and Kormann and Meixner footprint models.

Changes to Manuscript:

Section 2.4.1.4 Lines 259-267

Eddy covariance footprints were calculated using the Kljun et al. (2015) and the Kormann and Meixner (2001) footprint models. For the Kljun et al. (2015), the freely online MATLAB code of the model was used, while the Kormann and Meixner (2001) was coded in MATLAB. To determine the point source footprint contribution, the study first calculated the area that contributed 90% of the vertical flux; and based on the location (x and y coordinates based on wind direction and distance from source) of the point source, the source was determined if it was within the 90% footprint area. Point source emissions of sources within this region were then calculated based on the approach by Dumortier et al. (2019). This approach assumes all measured flux is equal to flux resulting from a single point source. In case of the mast being downwind of more than one source, more sonic anemometers are needed to solve the two unknown point source fluxes.

10. The authors use footprints generated from 3 m height instruments to interpret aerodynamic flux gradients from 2 and 4 m heights. This is incorrect. As was pointed out in the first review, an entire meter of height difference especially at these short heights has a massive effect on the size, extent, and peak location of footprints, and the 3 m footprint cannot be used to interpret at 2 m and 4 m instrument heights. By this same logic, the aerodynamic flux gradient method cannot be used for point sources, since the footprint influence at the source location is necessarily different from the two measurement heights.

Author's response:

We appreciate the point and agree that using the 3 m footprint to interpret flux gradients between 2 m and 4 m is not valid, particularly for a point source. We have removed the aerodynamic flux gradient analysis from the manuscript and revised the relevant sections accordingly.

11. The authors only report a single metric of performance in their results: bootstrapped relative mean factor. This is not the most meaningful metric of method performance, and bootstrapping it obfuscates the actual performance. What the authors should do is report the slope, adjusted r-squared, and significance of a linear regression through the actual and estimated emissions. The authors demonstrate this fit in the top row of their results figures, but the choice to log-transform one axis obfuscates the results. Using a log-plot artificially reduces the variability in results (visually). It is clear despite this, that there is very little agreement between the estimate and actual emissions for any of their results.

Author's response:

We have accordingly reported our results based on metrics of a linear regression between the estimated and actual emissions, (the slope, adjusted r-squared, and p-value).

Changes to Manuscript

Section 3

12. The authors justify their choice of gas analyzer by pointing out the only open-path CH4 analyzer on the market has a more limited range (0-40 ppm). However, that LICOR instrument can be extended to 100 ppm which would put it on the same range as the analyzer used in this study. Both the LICOR and Picarro instruments the authors mention in their discussion are specifically designed for eddy covariance and to operate in adverse weather. For example, the closed path Picarro system they mention is designed with a faster pump speed, shorter tube length, and synchronized datalogger, such that the system has a known system response time that while short of the 10-Hz that can be achieved with open-path systems, is still sufficient in many circumstances. There are comparative studies of closed path and open path methane eddy covariance that can serve as examples of closed-path performance when set up well. The majority of studies use open path sensors for methane eddy covariance because of the extra effort involved in avoiding the time-lag, attenuation, clock syncing, etc issues that the authors encounter here. Closed path eddy covariance can be done well, but it was not implemented correctly in this case.

Author's response:

We acknowledge that the instrument used in this study was not suitable for EC and that limits the EC conclusions drawn in the study.

Changes to Manuscript:

Section 4.1 Lines 502-509

This study acknowledges the limitations of the eddy covariance (EC) setup used, particularly that the ABB MGGA GLA131 Series analyzer is not designed specifically for EC applications. As a result, the conclusions drawn from the EC data are constrained. The study recommends further EC testing with instruments specifically designed for EC, ideally featuring a wide measurement range (0 to ~500 ppm), faster pump speeds, shorter tubing, synchronized data logging, sampling frequencies above 10 Hz, and rugged designs suitable for field deployment. Additionally, the study recognizes that environmental factors—such as obstructions, intermittent emissions, and variable wind directions causing plume meandering—can degrade EC data quality and complicate its application in oil and gas field studies.

Ali Lashgari

Comment 6:

For transparency, it would be helpful to the reader to have a section on validation where at least a few cherry-picked cases from the dataset are discussed after being processed through the different modeling frameworks in addition to having a few figures on how the measurements made by the different sensors compare in an author-defined common context.

Authors response

We removed the aerodynamic flux gradient from our study which used the two separate analyzers as Reviewer 1 stated the model cannot be used for point source quantification. As a result, the models we are evaluating are eddy covariance, Gaussian Plume Inverse Method and the backward Lagrangian stochastic model all using one instrument to avoid instrument bias.

For the eddy covariance data, we filtered out cases where negative fluxes were estimated by EddyPro and used two footprint models, Klujn et al (2015) and Kormann and Meixner (2001) to validate the footprint models. However, emissions were largely underestimated, and this has been acknowledged as our study limitation that could have been caused by non-stationarity and instrument limitations with our non-standard eddy covariance system.

For the Gaussian Plume Inverse model, we tested the model in six different scenarios to evaluate both the equations and generation of dispersion coefficients using single-release single-point emissions to test parameters where the model works best, and this was used for multi-release single-point emissions quantification.

For the backward Lagrangian stochastic model, data where the measurement point is on the edge of touchdown or outside of the touchdown is flagged by the model as (-9999) and this was filtered out.

Further, models comparison using a data subset, 15-minutes 10 degrees wind sector range has been added to the manuscript.

Changes to Manuscript

Section 2.4.4.2

Lines 290-295

The GPIM was evaluated under six scenarios (two equations and three different dispersion coefficients generations) using single-release single-point emissions to test when the model works best (Supplementary Information Section 2a: Equation 7 and 8). Dispersion coefficients were generated based on (1) high frequency sonic anemometer data at ~ 10 Hz, (2) EPA point-source dispersion coefficients (US EPA, 2013), and (3) 1 Hz sonic anemometer data. The scenario with the slope closest to 1, and highest R^2 across averaging durations, and wind sector ranges was selected and used for multi-release single-point emissions quantification.

Section 3.4

3.4 Models Comparison – Subset Data

Using a subset of the data (SRSP), filtered by 15-minute intervals within a 10-degree wind sector range where each model provided an emission estimate, the bLs model exhibited the best performance, with its linear regression closely aligned with the 1:1 line (Figure 10). The slope of the regression line for the GPIM was 1.6,

indicating an overestimation, while the bLs had a slope of 0.95, suggesting high accuracy. In contrast, the EC model produced slopes of 0.08 and 0.10 when using the Kormann and Meixner (2001) and Kljun et al. (2015) footprint parameterizations, respectively, indicating significant underestimation. When emission estimates were categorized by emission point, the GPIM notably overestimated emissions at locations 4W-22 and 4W-51 (identified in Figure 3), both situated approximately 10 m from the measurement location. The EC model consistently underestimated emissions across all sites, while the bLs model provided estimates closest to the expected values. The EC model produced negative emission rates associated with negative fluxes during periods of high non-stationarity (Supplementary Information, Section 2c. iv). These deviations from stationarity reflect intermittent plume capture, where the EC system alternated between sampling emitting and non-emitting regions. Overall, these findings indicate that for source-receptor distances ranging from approximately 10 to 90 meters, the bLs model demonstrated the highest accuracy in quantifying emissions.

Comment 7:

Regarding the following statement: “In this study, the precision to which downwind methods (closed-path EC, AFG, GPIM and bLs) could quantify the emission rate … introduced real-world scenarios that have not previously been tested”; the claim that the effect of obstacles is included in these experiments may need a reconsideration. To my knowledge, the tank battery is the only noticeable on-site obstacle and even its effect on concentration measurements by the sampling devices is expected to be non-existent: the tanks with a length-scale of ~4-5m are easily around 50m+ from the location of the sampling devices (figure 2). In the current revision, no evidence is included to substantiate the claim that the influence of obstacles is indeed present.

Additionally, while it is true that the field experiments were agnostic to the local conditions during the selected time period, the individual experiments were not stratified based on these conditions to offer a nuanced and in-depth discussion. Thus to claim that “In this study, the precision to which downwind methods could quantify the emission rate of point source(s) were tested in different atmospheric conditions (rain, sunny, snow, windy, calm etc.), and aerodynamic scenarios (emissions sources in open areas, behind obstacles, changing atmospheric stability, and day/night” appears to generalize beyond the specific experimental design.

Authors response

Thank you for this comment. We agree that our earlier statement overgeneralized the influence of obstacles and atmospheric conditions. In the current study, although field measurements were collected during a range of weather conditions, these were not explicitly stratified or controlled in the analysis. Furthermore, while on-site structures such as tank batteries were present, their distance from the sampling systems (>50 m) makes any direct aerodynamic influence on measurements unlikely. We have revised the manuscript to remove these claims and have rephrased the discussion to more accurately reflect the scope and limitations of the study.

Changes to Manuscript

Section 4

Lines 457-464

In this study, we evaluated the ability of downwind methods—including a non-standard closed-path EC system, the GPIM, and the bLs model—to quantify emissions from single-release and multi-release point sources. While the field measurements took place under naturally varying meteorological conditions, these were not explicitly stratified or analyzed as experimental factors. Additionally, although on-site infrastructure such as storage tanks were present, their distance from the sampling instruments (~50 m) likely rendered their aerodynamic influence negligible. As such, the analysis focuses on quantification of performance under realistic but uncontrolled field scenarios, without attributing model behavior to specific atmospheric or obstacle-related conditions.

Comment 9:

Although appreciated, I believe the explanation does not fully address the original comment. The authors may reduce the discussion around hardware specifications and instead provide a context as to how the authors recover the parity plots (figures 4-11). Since this research is an investigation on assessing the pros/cons of the different modeling frameworks, this context can be very valuable for a reader. As noted earlier, a validation exercise for a few cherry-picked cases when a direct connection is drawn going from raw measurements to rate estimates for the different approaches would go a long way in making this work more transparent.

Authors response

I have added supplementary data 'MATLAB Code & Software Configuration' folder that contains MATLAB scripts used detailing how primary measurement was converted to quantified estimates, WindTrax and EddyPro configuration, and a subset data used to generate the models comparison' using a data subset, 15-minutes 10 degrees wind sector range has been added to the manuscript.

I have a added a section of Traceability Example

I have also added more details on the quantification for the GPIM and bLs.

Changes to Manuscript

Section 3.5

3.5 Traceability Example

To illustrate how raw data were converted into model-based emission estimates, we present one representative 15-minute interval used in Figure 10. During a controlled release at point 4W-22 (wellhead), located approximately 10.5 m from the mast, the ground-truth release rate was 3 kg CH_4 h^{-1}. Over this interval, the average CH_4 concentration enhancement was 8.3 ppm above the background (determined using the 5th percentile method, see Section 2.4.2.1). The cross wiwind direction was 153° (0.3 m crosswind distance), with an average wind speed of 5.8 m s^{-1}. The same interval was processed through the three modeling frameworks:

- The bLs model (WindTrax), using measured concentration, geometry, and meteorological data, estimated 3.5 kg h^{-1}.
- The GPIM model, using Equation 7 and 8 (Supplementary Information) and dispersion coefficients, estimated 10.3 kg h^{-1}.

- The EC method (using the (Kormann and Meixner, 2001) footprint) estimated –0.004 kg h^{-1} due to a negative flux under high non-stationarity conditions.

This example illustrates how the bLs model reproduced the true emission most closely, GPIM overestimated, and EC underestimated the emission. More examples of data presented in Figure 10 are available under supplementary data "MATLAB Code & Software Configuration – Validation.xlsx".

Section 2.4.2.1

Lines 270-279

Methane concentration data from the MGGA analyzer and meteorology data from the sonic anemometer were averaged to 1 Hz and aggregated. The aggregated concentration-meteorological data were merged with METEC's release data and metadata, and release event tables created. The concentration-meteorological-release event data were then separated into single-release and multi-release events. For single-release events, the concentration-meteorological-release event tables were split into 5, 15 and 30-minute release event tables. Based on the bearing of the emission point to the measurement point and the average wind direction in the duration, the data was further filtered to downwind data, ±5º, ±10º and ±20º wind sector ranges. Multi-release events were further classified into multi-release single-point emissions (i.e., there were multiple emissions at the site level, but the mast was downwind of a single source) and multi-release multi-point emissions (i.e. there were multiple emissions at the site level and the mast was downwind of more than one source).

Section 2.4.3

The GPIM was evaluated under six scenarios (two equations and three different dispersion coefficients generations) using single-release single-point emissions to test when the model works best (Supplementary Information Section 2a: Equation 7 and 8). Dispersion coefficients were generated based on (1) high frequency sonic anemometer data at ~ 10 Hz, (2) EPA point-source dispersion coefficients (US EPA, 2013), and (3) 1 Hz sonic anemometer data. The scenario with the slope closest to 1, and highest R^2 across averaging durations, and wind sector ranges was selected and used for multi-release single-point emissions quantification.

Section 3.4

3.4 Models Comparison – Subset Data

Using a subset of the data (SRSP), filtered by 15-minute intervals within a 10-degree wind sector range where each model provided an emission estimate, the bLs model exhibited the best performance, with its linear regression closely aligned with the 1:1 line (Figure 10). The slope of the regression line for the GPIM was 1.6, indicating an overestimation, while the bLs had a slope of 0.95, suggesting high accuracy. In contrast, the EC model produced slopes of 0.08 and 0.10 when using the Kormann and Meixner (2001) and Kljun et al. (2015) footprint parameterizations, respectively, indicating significant underestimation. When emission estimates were categorized by emission point, the GPIM notably overestimated emissions at locations 4W-22 and 4W-51 (identified in Figure 3), both situated approximately 10 m from the measurement location. The EC model consistently underestimated emissions across all sites, while the bLs model provided estimates closest to the expected values. The EC model produced negative emission rates associated with negative fluxes during periods of high non-stationarity (Supplementary Information, Section 2c. iv). These deviations from stationarity reflect intermittent plume capture,

where the EC system alternated between sampling emitting and non-emitting regions. Overall, these findings indicate that for source-receptor distances ranging from approximately 10 to 90 meters, the bLs model demonstrated the highest accuracy in quantifying emissions.

Additionally, it is stated that: ”continuous monitoring requires deployment of multiple sensors which create limitations of cost …”. The term “continuous monitoring (CM)” is quite broad. Although the statement is true when multiple sensors are deployed at different points on a site to perform quantification and localization, the present framework of using effectively a single device (with vertically-separated inlets) in a fence-line role is clearly a different use of the CM term. It would be helpful for the reader if this aspect is clarified to remind the reader this investigation deals with a CM fence-line type system that is only one of the possibilities in which a CM system can be constructed and deployed.

Authors response

We agree that “continuous monitoring” encompasses a wide range of deployment strategies. In this study, the CM system consisted of a single sensor with an inlet in a fence-line configuration. To clarify this distinction, we have revised the manuscript to explicitly state that the results and limitations discussed apply specifically to this single-instrument fence-line CM approach, which differs from multi-sensor CM networks that may offer broader spatial coverage and source localization capabilities but with increased cost and complexity.

Changes to Manuscript

Lines 496-502

In this study, continuous monitoring was conducted using a single sensor with an inlet deployed at a fence-line distance. This system requires instrumentation capable of measuring a wide concentration range, as emissions from oil and gas sites can vary between 0 and 250 ppm (Supplementary Information Section 1). While continuous monitoring systems, comprising multiple sensors can offer enhanced spatial coverage and source localization, they also introduce higher costs. The limitations and findings reported here therefore apply specifically to this single-sensor fence-line continuous monitoring approach and may not be representative of all continuous monitoring frameworks.

Comment 14:

Please note that Ilonze et al. (2024) compared the results of the solution provided, most of which employed multiple sensors for quantification. It may lack the equivalency with the fence-line framework used in this study with the analysis of the solution providers’ performance by Ilonze et al. (2024). My advice would be to withdraw this comparison or allaborate on the differences and their consequenses.

Authors response

Thank you for the suggestion. We have retained the reference to Ilonze et al. (2024) but have clarified in the manuscript that the comparison is qualitative rather than direct. The differences in monitoring frameworks—multi-sensor vs. single-sensor fence-line setup—are now explicitly acknowledged, and we have discussed how these differences may influence quantification performance and interpretation.

Changes to Manuscript

Lines 61-71

Detection and localization of simulated fugitive emissions using this approach have been demonstrated successfully in controlled release studies. For example, Ilonze et al. (2024) reported a 90% probability of detection for emissions between 3.9 and 18.2 kg CH_4 h^{-1} using multi-sensor and scanning/imaging systems. However, significant uncertainty in quantification remains; their study reported emissions being misestimated by a factor of 0.2 to 42 for releases between 0.1 and 1 kg CH_4 h^{-1}, and by a factor of 0.08 to 18 for emissions above 1 kg CH_4 h^{-1}. While informative, the methods in Ilonze et al. (2024) differ in keyways from those employed here—specifically, their use of multiple sensors and a distributed monitoring configuration as opposed to the single-instrument, fence-line-based framework used in our study—limiting direct comparison of quantification accuracy. This study will evaluate the quantification accuracy of the closed-path EC, Gaussian plume inverse model (GPIM), and the backward Lagrangian stochastic model (bLs) for oil and gas point source quantification using a single-instrument deployed at a fence line distance.

Comment 16-17:

I recommend clarification that the error range (40-60%) from the cited reference applies to operational emissions (not controlled releases) from a driving survey (i.e. not fence line motoring in the same sense of the word as continuous monitoring) of an entire basin that includes agriculture emissions among other things. It may be over-generalization to use this reference to support the argument that GPIM is fundamentally limited, as appears to be the case right now.

Authors response

Thank you for your comment. We acknowledge that Riddick et al. (2022b) primarily reports on mobile survey-based quantification across a basin. However, the error range cited in our manuscript (40.7–60%) was derived from a controlled release experiment within that study, not from the basin-scale results. Specifically, the controlled release involved 10 replicate measurements of compressed natural gas released at 1.5 m above ground level and quantified using the same Gaussian plume inverse framework applied to mobile survey data. This provides a relevant and controlled reference point for understanding GPIM performance, even if the deployment mode differs from continuous fence-line monitoring. We have revised the manuscript to clarify this distinction and avoid overgeneralizing the results.

Changes to Manuscript

Lines 101-105

Riddick et al. (2022b) reported absolute uncertainties of between 40.7 and 60% in a controlled release experiment involving 10 replicate measurements of compressed natural gas (1.5 m release height), with concentrations measured using a mobile vehicle survey. While this differs from continuous fence-line deployment, it offers insight into the inherent uncertainty of the GPIM method in field conditions.

Second, here and elsewhere, the authors suggest that “quantification is complexed by interference from other neighboring sources”. These statements may give the impression that quantification is impossible in such scenarios whereas it is clearly a question of the choices made in experimental design and the study objectives, both of which

are defined at the outset. Fence-line monitoring used in the way used here to model the relatively short controlled-release experiment is limited by the constraints imposed during the design phase rather than some larger theoretical limitation. I recommend that the explanation here and elsewhere in the manuscript should highlight this basic fact to ensure that the discussion is in good faith.

Authors response

The authors have modified the manuscript to address this concern regarding the framing of quantification challenges in the presence of neighboring sources. The revised text now appropriately distinguishes between theoretical limitations and practical constraints arising from experimental design.

Changes to Manuscript

Lines 520-527

For multiple emissions, even when the sensor is nominally downwind of a single source based on the average wind direction, quantification can be complicated by interference from neighboring sources. However, it is important to emphasize that such complexity is not a fundamental limitation of quantification itself, but rather a function of the experimental design and study objectives. For example, plume interference can often be minimized through strategic localization and optimization using multiple sensors—an approach that differs from the single-instrument setup used in this study. This study's design involves defining plumes based on wind sector ranges, as opposed to using multiple sensors to localize sources, and therefore does not replicate how various continuous monitoring solutions typically operate.

Lines 541-551

This discrepancy may be due to design-related challenges—specifically, interference from neighboring sources and the lack of distinct plume separation in complex flow conditions. Although the measurement point was nominally downwind of a single source, the real-world plume structure may not align with model assumptions. Additionally, the bLs implementation in WindTrax is designed for single-source scenarios and applying it in multi-source environments without adaptation can lead to inaccuracies. The GPIM and bLs methods are sensitive to background correction, which in this study was complicated by temporal overlap between release events and residual CH_4 accumulation, particularly under stable atmospheric conditions. Although this is a controlled-release study, residual methane from prior emissions and the presence of multiple plumes can affect the CH_4 concentration during a candidate event, challenging the assumptions used to define background and isolate a single-source plume using wind-sector-based criteria. These findings highlight the importance of aligning modeling assumptions with the experimental context rather than pointing to a fundamental limitation of the method itself.

Comment 18:

The authors state that this study shows the difficulty in defining the background. To my knowledge, one major point of pursuing controlled-release testing like the ones discussed here was to have test cases without interference from operational background sources. As such, this statement may be confusing to a reader.

Authors response

Indeed, one of the motivations for using controlled-release studies is to limit interference from external background sources. To clarify, our reference to the difficulty in defining the background pertains not to interference from external operational sources, but to challenges internal to the experimental design—specifically, the presence of residual CH_4 from prior release events and the potential for multiple overlapping plumes during some test periods. While controlled-release removes background uncertainty from unrelated field operations, it does not fully eliminate the complexity of defining background in cases where emissions are frequent, closely spaced in time, or where meteorological conditions (e.g., stable stratification) limit dispersion. We have updated the manuscript (Lines 712–716) to reflect this clarification, emphasizing that the challenge stems from the temporal and spatial overlap of test releases, rather than from uncontrolled ambient sources.

Changes to Manuscript

Lines 545-551

The GPIM and bLs methods are sensitive to background correction, which in this study was complicated by temporal overlap between release events and residual CH_4 accumulation, particularly under stable atmospheric conditions. Although this is a controlled-release study, residual methane from prior emissions and the presence of multiple plumes can affect the CH_4 concentration during a candidate event, challenging the assumptions used to define background and isolate a single-source plume using wind-sector-based criteria.

Secondly, it is mentioned that “Gaussian model and the backward Lagrangian stochastic models are limited, as they can only quantify one source at a time; and interference from neighboring emissions affects the underlying principles of dispersion on which these models were developed”. While it is true that the specific choice made in this paper restricts the Gaussian plume approach to a single source, there is numerous evidence of employing the Gaussian plume approach for multi-source problems and there are no limits imposed by the so-called “principles of dispersion” for such synchronous, constant-rate releases rates. As an example, a highly referenced article titled “The Mathematics of Atmospheric Dispersion Modeling” (2011) by JM Stockie provides examples of using the Gaussian plume model for multi-source problems.

Authors response

Thank you for the comment. We agree that the Gaussian plume and bLs models are not inherently limited to single-source applications and have been successfully applied to multi-source scenarios in the literature. Our original wording unintentionally implied a theoretical constraint, when the actual limitation stems from the specific model application choices in our study—namely, that each quantification event was treated under the assumption of a single dominant source upwind. This was an intentional simplification aligned with our controlled-release test design and quantification framework. To address this, we have revised the manuscript to clarify that the constraints discussed are due to implementation choices and experimental design, not to the underlying principles of dispersion modeling.

Changes to Manuscript

Lines 560-566

Oil and gas point sources could either be single emissions or multiple emissions occurring concurrently. In this study's design, cases involving multiple emissions with more than one release point located upwind posed challenges for the specific Gaussian and backward Lagrangian stochastic (bLs) model implementations, which were applied assuming a single active source at a time. While these models can be extended to handle multi-source scenarios, the assumptions used here limited their ability to distinguish individual contributions when plumes overlapped. As a result, interference from neighboring emissions introduced ambiguity in model-observation alignment, particularly under complex wind conditions.