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

March 18th, 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. The authors would also like to emphasis that the paper aims to inform real-world oil and gas fence-line sensors deployment on feasible quantification approaches as oil and gas emissions can be complex (multiple sources could be emitting at the same time). As a result, we have accordingly modified our manuscript and title to "Evaluating the feasibility of using downwind methods to quantify point source oil and gas emissions using continuous monitoring fence-line sensors".

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 compare three approaches to quantify point source emissions (Gaussian, eddy covariance, and aerodynamic flux gradient methods). This manuscript is not recommended for publication in current form. Primary concerns center on the method application, transparency of the experiment conditions and data corrections, and quality assurance (QA). In addition to the comments provided here, the authors are encouraged to consider CC1.

Author's response:

Thank you for reviewing our paper. We have accordingly addressed the concerns in the specific comments below.

Comments on Main Paper:

Comment 1

It is assumed that this work was opportunistically executed as part of other controlled release activities performed by METEC to test, for example, fence line sensor network leak detection capability. If this is true, this point should be made clear to the reader as this may represent a study limitation. The controlled release profiles are complex and a primary issue with this work resides in the QA of the individual release trails in the context of this comparison. Even when the mean wind direction during a 20-minute or 30-minute observation is acceptable, a short duration release event (as low as 10 seconds, line 148) producing a spatially underdeveloped plume may yield zero methane signal enhancement on the instruments, as the release could have occurred while winds were off axis. If there is insufficient sampling of the plume for a given trial, the emission calculation performance is irrelevant. A similar argument holds for the 2m and 4 m vertical positions. The work would be improved by adding additional detail on the controlled release trials along with methane concentration statistics (including background corrections) by trial for each instrument. Ideally this could occur on a uniform 20-minute time base so direct comparisons across the techniques are possible.

a) Complex controlled release profiles; QA of individual release trails

Author's response:

Thank you for your comment. Measurements were conducted as part of the METEC Spring 2024 Advancing Development of Emissions Detection (ADED) Campaign, as this study aimed to investigate the accuracy of downwind quantification models in complex oil and gas settings for their application in continuous monitoring solutions using point sensors. We have reanalyzed our data to uniform 5 minutes, 10 minutes and 15-minute durations for all models where there were continuous durations. Even though there were short releases such as 10 s that represent intermittent emissions in oil and gas such as a venting of a gas pneumatic, our duration tables represent minute durations when there were continuous releases for that period. For example, a 5-minute event table is for a continuous release in a 5-minute duration. To clarify this, the following changes have been made to the manuscript.

Changes to manuscript:

Section 1

Lines 131-149:

Continuous monitoring of CH_4 emissions using fence line sensors requires proper quantification of intermittent and persistent releases from oil and gas during all release (complex emission profiles) and atmospheric conditions (unstable, neutral and stable). Oil and gas emissions are characterized by intermittent, non-uniform, single or multiple point source emissions, varying in leak size, location, height and distance between the source and sensor, and are typically in complex aerodynamic environments (i.e. not flat). An ideal quantification model should always quantify emissions and should capture short and long-lasting emission events. Most models have been validated to work best during neutral conditions for single point sources. However, it is important to test and apply these models during non-neutral conditions as well as these are part of real-world conditions where continuous monitoring is applied. In this study, we evaluate if using a readily available CH_4 cavity ring down analyzer for models' quantification such as the closed-path EC is a feasible solution to quantify point source emissions.

This study aims to inform the feasibility of downwind quantification models in oil and gas settings by investigating which models are likely to work most of the time with instrumentation that is typically available for fence-line deployment. Fence-line sensor deployments involve multiple sensors, continuously running in all conditions and providing emissions data. Using robust releases and environmental conditions, this study aims to investigate the performance of these methods in quantifying emissions for known gas release rates and evaluating uncertainties that could result in incorrect CH_4 reporting. Specifically, the study will (1) evaluate the overall quantification accuracy of closed-path EC, AFG, bLs model, and the GPIM method in quantifying single-release single-point and multi-release single-point emissions that simulate oil and gas emissions, (2determine the mean relative factor (estimated emissions over actual emission) for these models.

Section 2.2

Lines 188-189: Controlled releases were part of the METEC Spring 2024 Advancing Development of Emissions Detection (ADED) Campaign conducted between February 6 and April 29, 2024 (Colorado State University, 2024).

Section 2.4.1

Lines 221-223: The concentration-meteorological-release event tables were split into 5, 10 and 15-minute release event tables (i.e. there was a continuous release in the duration).

Comment 2

As it is currently written, there are issues with the eddy covariance data collection, pre-processing, flux calculations, post-processing, and footprint modeling that will cause erroneous results. These are detailed below: The authors use a non-standard system and instrument configuration for eddy covariance. These choices have significant consequences for flux quantification that need to be addressed and justified.

Author's response:

Thank you for your comment. We have accordingly reanalyzed our data using EddyPro software and clearly stated our system configuration and pre-processing of the data using our non-standard eddy covariance system.

Changes to manuscript:

Section 2.4.1

Lines 205-246:

2.4.1 Eddy Covariance

2.4.1.1 Data pre-processing

Evaluating the MGGA CH_4 data showed that actual sampling was between 4 and 12 Hz (highest sampling at 6 Hz), even though it had been set to sample at 10 Hz (Supplementary Information Section 2b). To account for this sampling variability, data were filtered to when sampling was equal/greater than 8 Hz. Data where the frequency was greater than 8 Hz were down sampled to 8 Hz. The sonic anemometer meteorological data (horizontal wind vectors (u, v), vertical wind vector (w), temperature (T), and pressure (P)) actual sampling varied between 7 and 9 Hz with the most frequent frequency at 8 Hz (Supplementary Information Section 2b). As the MGGA gas analyzer and sonic anemometer were not designed to clock synchronously, using the MGGA CH_4 clock time as a reference, meteorological data from the sonic anemometer were matched to the MGGA CH_4 data using linear interpolation to generate concentration-meteorological 8 Hz data.

The aggregated concentration-meteorological data were then merged with METEC's release data and metadata, and release event tables created. Release event tables were aggregated tables of concentration, meteorology and release (emission source location, duration and rate) information for all defined release events at METEC. The concentration-meteorological -release event data were then separated into single-release and multi-release events. Single-release events were when there was a single emission point at the site level, while multi-release events were when there was more than one emission point at the site level. The concentration-meteorological-release event tables were split into 5, 10 and 15-minute release event tables (i.e. there was a continuous release in the duration). 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, ±20º, and ±45º.

2.4.1.2 Flux calculation

Turbulent fluxes were calculated using the open software EddyPro® version 7 (LI-COR, Nebraska, USA, n.d.). Acquisition frequency was set at 8 Hz, while file duration and the flux interval were set at 5, 10, and 15 minutes, respectively, depending on the file being processed. Table 1 shows the instruments input to the software.

Table 1. Anemometer and Gas Analyzer Input into EddyPro

Anemometer Information		**Gas Analyzer Information**	
Manufacturer	Young	Manufacturer	Other
Model	81000	Model	Generic closed path
Height	3 m	Tube length	300 cm
Wind data format	u, v, w	Tube inner diameter	3.275 mm
North alignment		Nominal tube flow rate	3.2 l/m
North off-set	0.0	Northward separation	0.00 cm
Northward separation	Reference	Eastward separation	0.00 cm
Eastward separation	Reference	Vertical separation	-10.00 cm
Vertical separation	Reference	Longitudinal path length	10.00 cm
Longitudinal path length		Transversal path length	2.54 cm
Transversal path length		Time response	0.4 s

In raw data processing, axis rotations for tilt correction under wind speed measurement offsets was checked. Under turbulent fluctuations, double rotation and block average detrend methods were used. Covariance maximization with default was used for time lag detection; time lags detection was checked. Compensation for density fluctuations (Webb-Pearman-Leuning terms) was unchecked as the MGGA analyzer synchronously reported dry CH_4 and water mole fractions, cell temperature and pressure. Mauder and Foken (2004) (0-1-2 system) were used for quality check. All statistical tests for raw data screening, Vickers and Mahrt (1997)– spike count/removal, amplitude resolution, drop-outs, absolute limits, skewness and kurtosis, discontinuities, time lags, angle of attack and steadiness of horizontal wind were checked. The default values for all these tests were used. Similarly, default settings for spectral analysis and corrections were used. Analytic correction of high-pass filtering effects (Moncrieff et al., 2005) for low frequency range; and correction of low-pass filtering effects (Fratini et al., 2012 - In situ analytic) and instruments separation (Horst and Lenschow, 2009 - only crosswind and vertical) in the high frequency range were used.

2.4.1.3 Post-processing

During post-processing, flux data were filtered based on (1) quality flags, Mauder and Foken (2004) (0-1-2 system), and (2) surface friction velocity ($u^* > 0.13$ m/s). Data that were flagged "2" were first filtered out as they were considered poor quality fluxes (LICOR, 2025), and the remaining dataset were filtered for high turbulence data. All data was filtered out as low quality and no further post-processing was done.

<u>Comment 3</u>

The gas analyzer (MGGA) used here and 3D sonic anemometer are not designed to be logged together with clocks synced. Properly synced clocks at 10 Hz is an essential step and possible large source of error in data acquisition, and the authors need to address how they achieved this with their non-standard system. For example, LICOR and Campbell Scientific instrument systems are designed to log all fast data (10 Hz) simultaneously on the same data logger or microcomputer with PIP LAN networking to ensure clocks remain synced.

<u>Author's response:</u>

Thank you for your comment. We have added this to our manuscript.

<u>Changes to manuscript:</u>

Section 2.4.1.1

Lines 212-215:

As the MGGA gas analyzer and sonic anemometer were not designed to clock synchronously, using the MGGA CH_4 clock time as a reference, meteorological data from the sonic anemometer were matched to the MGGA CH_4 data using linear interpolation to generate concentration-meteorological 8 Hz data.

Comment 4

The authors use a closed-path sampling system with a long (3 meter) tube. This choice creates significant errors that need to be addressed and corrected, if possible. Eddy covariance measurements require at least 10 Hz system response time in order to properly characterize turbulent eddies. Sampling at 10 Hz is not the same as a system response of 10 Hz. Closed path systems introduce a time-lag, laminar flow issues, and lack the response time of open-path systems. Slower system response is one of the main reasons the vast majority of methane eddy covariance is done using open-path sensors, and when closed-path sensors are used the system performance needs to be optimized and carefully justified. A well designed closed-path system with sufficient pump speed might effectively have a system response of 5 Hz while sampling at 10 Hz. This leads to an under estimation of fluxes, and in the best case scenarios can be corrected using transfer functions. At a minimum, the authors need to investigate the cospectra of their results to demonstrate they have a sufficient overall system response and properly apply transfer functions to correct their data. Related to this issue, when using a closed-path system the pump speed needs to be reported. With a longer tube such as this, a very fast pump speed, possibly also with a vacuum induced, is required in order to have a fast enough system response. Without a fast enough pump speed, the data cannot be used for eddy covariance methods. Pump speeds for closed path systems range from 8 l/m at normal pressure for short tubes, to over 100 l/m with partial vacuum for longer tubes.

<u>Author's response:</u>

Thank you for your comment. Our system has a relatively slow pump flow rate of ~ 3 lpm and we understand this could create laminar flow issues with a 3 m tubing. However, our study aimed to investigate if we can use data from instruments readily available to us to quantify using eddy covariance. Methane concentration readings from oil and gas emissions could be up to 300 ppm. The current readily available sensors have a range of up to 25 ppm. For example, the LI-COR LI-7700 open path CH_4 analyzer has a measurement range of 0 to 25 ppm at -25 ºC and 0 to 40 ppm at 25ºC; PICARRO G2311-f, a closed-path analyzer has an operating range of 0 to 20 ppm. The instrument used in this study is a field instrument (ABB MGGA GLA131 Series has a measurement range of 0 to 100 ppm for CH_4 but can be extended to 0 to 1%). We have attempted to account for the gas flow response time (0.4 seconds), pump flow rate (3 lpm) and the long tubing (3 m) using Eddy Pro corrections of time lag detection. Based on our reanalyzed results, our eddy covariance system quantified emissions with a mean relative factor (estimated emission over actual emission) of 0.7 to 1 for single-release single-point emissions, and within a mean relative of 1 and 2.4 for multi-release single-point emissions. Our system limitations have been clearly stated in the updated manuscript.

<u>Changes to manuscript:</u>

Section 2.4.1.2

Lines 226-229

Turbulent fluxes were calculated using the open software EddyPro® version 7 (LI-COR, Nebraska, USA, n.d.). Acquisition frequency was set at 8 Hz, while file duration and the flux interval were set at 5, 10, and 15 minutes, respectively, depending on the file being processed. Table 1 shows the instruments input to the software.

Table 1. Anemometer and Gas Analyzer Input into EddyPro

Anemometer Information		**Gas Analyzer Information**	
Manufacturer	Young	Manufacturer	Other
Model	81000	Model	Generic closed path
Height	3 m	Tube length	300 cm
Wind data format	u, v, w	Tube inner diameter	3.2 mm
North alignment		Nominal tube flow rate	3.2 l/m
North off-set	0.0	Northward separation	0.00 cm
Northward separation	Reference	Eastward separation	0.00 cm
Eastward separation	Reference	Vertical separation	-10.00 cm
Vertical separation	Reference	Longitudinal path length	10.00 cm
Longitudinal path length		Transversal path length	2.54 cm
Transversal path length		Time response	0.4 s

Lines 244-245: Covariance maximization with default was used for time lag detection; time lags detection was checked.

Section 4.1

Lines 464-486

Eddy covariance was tested using a closed-path analyzer, cavity ring-down spectroscopy, with a 3.2 lpm pump flowrate and a 0.4 s gas flow response time. The closed-path EC estimated emissions between a factor of 0.67 and 0.97 for SRSP emissions, and between 1.02 and 2.43 for MRSP emissions at ±45° wind sector range (Section 3.1). This was a wider uncertainty in estimated emissions than one reported by Dumortier et al. (2019), who estimated emissions at between 90 and 113% of true emission (~1.5 kg day^{-1}) with concentrations between 2 and 3 ppm. Our study tested closed-path EC at emission rates between 0.005 and 8.5 kg h^{-1}.

Our study's results were when the data was filtered for frequencies greater than 8 Hz, hence, largely reducing the sampled emissions. The 10 Hz sampling frequency set for this instrument was not a true 10 Hz and this could have been due to the 0.4 gas flow response time that delayed analysis of the drawn air sample in the cavity, or the 3 lpm pump flow rate for a 3 m tubing that might have varied the effective sample turnover rate. The rest of the data were flagged as low quality by Mauder and Foken (2004) (0-1-2 system), which flags based on steady state and well-developed turbulence. This could have been due to low turbulence as experiments were carried out in winter, and instrumentation limitations (low pump flow rate and asynchronous configuration of the gas analyzer and meteorological instrument).

Continuous monitoring requires deployment of multiple sensors which create limitations of cost and requires instrumentation with a wide measurement range as concentrations for oil and gas emissions can range between 0 to 250 ppm, as in this study (Supplementary Information Section 1). The currently available EC instruments have a narrow measurement range (LI-COR LI-7700 open path CH_4 analyzer has a measurement range of 0 to 25 ppm at -25 ºC and 0 to 40 ppm at 25ºC; PICARRO G2311-f, a closed-path analyzer has an operating range of 0 to 20 ppm). Also, the instrumentation should be environmentally robust and not lab-grade (be able to run smoothly in adverse weather conditions). Given these parameters, market available EC instruments that can currently be deployed in oil and gas are limited. The instrument used in this study is a field instrument (ABB MGGA GLA131 Series has a measurement range of 0 to 100 ppm for CH_4 but can be extended to 0 to 1%).

Comment 5

The authors have not implemented numerous steps that are standard in the pre-processing of eddy covariance data. There are community standards for the QA/QC of the 10-20 Hz fast data before it is used for flux calculations. These include but are not limited to: Spike removal, absolute limits thresholding, corrections for skewness and kurtosis, dropouts removal, amplitude resolution, time lag corrections (absolutely critical for closed path systems such as this), steady-state tests and the mean removal through block averaging or detrending, and coordinate rotation or planar fit.

Author's response:

Thank you for your comment. We have reprocessed our data through these statistical tests using EddyPro and has been included in the manuscript.

Changes to manuscript:

Lines 230-241: In raw data processing, axis rotations for tilt correction under wind speed measurement offsets was checked. Under turbulent fluctuations, double rotation and block average detrend methods were used. Covariance maximization with default was used for time lag detection; time lags detection was checked. Compensation for density fluctuations (Webb-Pearman-Leuning terms) was unchecked as the MGGA analyzer synchronously reported dry CH_4 and water mole fractions, cell temperature and pressure. Mauder and Foken (2004) (0-1-2 system) were used for quality check. All statistical tests for raw data screening, Vickers and Mahrt (1997)– spike count/removal, amplitude resolution, drop-outs, absolute limits, skewness and kurtosis, discontinuities, time lags, angle of attack and steadiness of horizontal wind were checked. The default values for all these tests were used. Similarly, default settings for spectral analysis and corrections were used. Analytic correction of high-pass filtering effects (Moncrieff et al., 2005) for low frequency range; and correction of low-pass filtering effects (Fratini et al., 2012 - In situ analytic) and instruments separation (Horst and Lenschow, 2009 - only crosswind and vertical) in the high frequency range were used.

Comment 6

There are additional steps during the flux processing the authors do not implement. These include: high-pass and low-pass corrections, transfer functions (absolutely required given their system configuration), and WPL corrections.

WPL corrections might not be necessary for a closed path system if the instrument is measuring methane, water, cell temperature, and cell pressure at synced 10 Hz rate. If this is the case, the reported dry mixing ratio of methane can be used without WPL corrections. Since it can be the case that methane and water are measured at 10 Hz but cell temperature and pressure are not, the authors need to report how their instrumentation sampling was configured in this regard. Without synced 10 Hz cell temperature and pressure, the dry mixing ratio of methane is not truly at 10 Hz, and instead the authors should use the non-dry gas measurement and apply WPL corrections.

Author's response:

Thank you for your comment. We have reprocessed our data EddyPro and these corrections have been included. Our instrument was measuring methane, water, cell temperature and pressure synchronously hence, we used the reported dry methane measurement.

Changes to manuscript:

Lines 232-234: Compensation for density fluctuations (Webb-Pearman-Leuning terms) was unchecked as the MGGA analyzer synchronously reported dry CH_4 and water mole fractions, cell temperature and pressure.

Lines 237-241: Similarly, default settings for spectral analysis and corrections were used. Analytic correction of high-pass filtering effects (Moncrieff et al., 2005) for low frequency range; and correction of low-pass filtering effects (Fratini et al., 2012 - In situ analytic) and instruments separation (Horst and Lenschow, 2009 - only crosswind and vertical) in the high frequency range were used.

Comment 7

There are post-processed QA/QC steps the authors have not implemented. These include:

- Quality tests for developed turbulence and stationarity. These are standard across the eddy covariance scientific community.

- Friction velocity thresholding. This eliminates low turbulence times and is assessed for each site/system.

- Energy balance closure. A good energy balance closure does not necessarily mean good fluxes, but a bad energy balance closure can immediately identify issues with the sonic anemometer or instrument configuration.

Author's response:

Thank you for your comment. We have reprocessed out data through Mauder and Foken (2004) (0-1-2 system) quality check and filtering for surface friction velocity > 0.13 m/s. All data were flagged by quality check and friction velocity thresholding hence no further post-processing was done.

Changes to manuscript:

Lines 255 - 258

2.4.1.3 Post-processing

Flux data were filtered based on (1) quality flags, Mauder and Foken (2004) (0-1-2 system), and (2) surface friction velocity (u* > 0.13 m/s). Data that were flagged 2 were first filtered out as were considered poor quality fluxes (LICOR, 2025) and the remaining dataset were filtered for high turbulence data. Almost all data were filtered out and no further calculations were done.

Comment 8

Given the experimental design, any controlled releases with durations shorter than the averaging period used in eddy covariance flux calculations should not be used because they will by definition be non-stationary. The authors should consider using a shorter averaging period. 30 minutes is standard for ecosystem fluxes which are spatially homogeneous and change relatively slowly in time. Eddy covariance controlled release experiments commonly use shorter averaging periods in order to minimize non-stationarity issues. The authors should process their fluxes using 5 minutes, 10 minutes, and 15 minutes for example. They should then investigate the ogives to determine how much if any of the fluxes are underestimated by using a shorter averaging window. Previous controlled release experiments with eddy covariance have shown that the under sampling of large eddies by using a shorter averaging window affected flux estimates by <10% and considered this a worthwhile trade off to ensure stationarity in the observations.

Author's response

Thank you for your comment. We have reanalyzed our data to uniform 5 minutes, 10 minutes and 15-minute durations for all models where there were continuous durations. All data were flagged by quality check and friction velocity thresholding hence no further calculations were done.

Changes to Manuscript

Section 2.4.1

Lines 221-223: The concentration-meteorological-release event tables were split into 5, 10 and 15-minute release event tables (i.e. there was a continuous release in the duration).

Lines 243-246

2.4.1.3 Post-processing

Flux data were filtered based on (1) quality flags, Mauder and Foken (2004) (0-1-2 system), and (2) surface friction velocity (u* > 0.13 m/s). Data that were flagged 2 were first filtered out as were considered poor quality fluxes (LICOR, 2025) and the remaining dataset were filtered for high turbulence data. Almost all data were filtered out and no further calculations were done.

Comment 9

The cospectra should be investigated and included in the SI. This is a standard QA/QC step that can illuminate issues such as slow system response times, aliasing, and interference near the sonic anemometer. A good cospectra will not necessarily mean good fluxes, but a bad cospectra will certainly mean there are issues that need to be corrected.

Author's response

Thank you for your comment. All data were flagged by quality check and friction velocity thresholding hence no further post-processing such as cospectra investigation was done.

Comment 10

Footprint modeling: This is an essential step in interpreting eddy covariance data in a system with point sources. However, the authors incorrectly parameterized the model and used the footprint area to normalize the fluxes to emissions incorrectly. Firstly, the choice of a roughness length of 1 is physically improbable for their site, and given it is a key parameter in the model, the footprints are likely incorrect. The roughness length should not be estimated or chosen randomly. It can be derived from the authors' 3D sonic anemometer data under neutral conditions. In a scenario with perfectly homogeneous fluxes surrounding the eddy covariance tower, then the footprint area can be multiplied by the flux as the authors have done here. If there is any heterogeneity present (such as a point source) then this is incorrect. The footprint area is weighted with a small area of peak influence that asymptotically declines to zero influence in all directions away from it. The correct approach, given that the location of the point source is known, is to multiply the weight of the pixel representing the point source by the measured flux, and by the area of the pixel containing the point source. See Rey-Sanchez et al. 2022 for a description of how footprints should be used to calculate point source eddy covariance emissions and compared to controlled releases. Given that the authors can likely assume a zero flux for all non-point source pixels within the footprint, this is a straightforward calculation. Lastly, the authors chose to use a footprint model which previously studies have shown to perform the worst in controlled release experiments. Rey-Sanchez et al. 2022 compared three commonly used models and at a minimum, the authors here should do so as well. The exact footprint peak influence and location matters much more for correctly calculating emissions from point sources, so the variation in this result between footprint models is a source of uncertainty that should be quantified for a methods comparison paper such as this.

<u>Author's response</u>

Thank you for your comment. We have recalculated the surface roughness and footprint.

Comment 10 – Question 1 Calculation of surface roughness

Surface roughness was calculated using the high frequency sonic anemometer data during neutral conditions and was determined to be 0.1 m.

<u>Changes to Manuscript:</u>

Section 2.3

2.3 Calculation of Roughness Length

Surface roughness length (z0) was calculated from friction velocity (Supplementary Information Section 2a: Equations 1 and 2) by splitting the high frequency sonic anemometer data into 15-minute tables and filtering for those in neutral conditions, |L| > 500 (Supplementary Information Section 2a: Equation 3). The overall roughness length selected as the median of all the calculated z0 was 0.1 m (Rey-Sanchez et al., 2022).

Comment 10-Question 2: Emission calculation from the footprint for point sources

Point source emission was calculated based on the approach by Dumortier et al (2019) where emission per source = measured flux / value of the footprint at the cell. The source location was first determined if it was within the 90% footprint region, and the footprint value extracted.

<u>Changes to Manuscript:</u>

<u>Section 2.4.3</u>

<u>Lines 271-275</u>

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.

Comment 10 – Question 3 Footprint Choice

Our study was limited to a single sonic anemometer which limited site-based calculations of exponential wind velocity power law and the exponential eddy diffusivity power law that are inputs to the Kormann and Meixner (2001) footprint model. Our set up provided enough inputs for the Kljun footprint model. This limitation has been stated in our manuscript.

Changes to Manuscript

Lines 254-259

Eddy covariance and AFG footprints were calculated using the Kljun et al. (2015) footprint model. Even though Rey-Sanchez et al. (2022) reported the Kljun et al. (2015) footprint model to be less accurate compared to the Kormann and Meixner (2001), Kormann and Meixner (2001) was too complex for our study because it required multiple sonic anemometers or tracer release experiments to calculate the exponential wind velocity power law, and the exponential eddy diffusivity power law for site specific data. Our study was limited to a single sonic anemometer, and this provided enough inputs for the Kljun et al. (2015) footprint model.

Comment 11

It is unclear with the way the methods are written how the footprint normalization for the aerodynamic flux gradient method was done. It seems as though the area of the eddy covariance footprint was used to convert the AFG emissions into kg/hr. This is incorrect. The sensors are in a different location at different heights and will not have the same footprint as the eddy covariance measurements and will additionally have the same issues of normalizing to the whole area for a point source that was described above. In general, the AFG method should not be used to measure point source emissions. There will be a different footprint for each height along the profile. This means each sensor will see (or maybe not see at all) the point source differently. This will obviously lead to issues when calculating a flux from the profile gradient. The AFG method only works in homogeneous environments where the differing footprints at different heights don't matter.

Author's response

Thank you for your comment. We used the sonic anemometer collocated with the eddy covariance tower to calculate the aerodynamic footprint. Even though the anemometer was 9.4 m away at 3 m height (average of the two aerodynamic flux gradient sensors, 2 and 4 m), the inputs to the Kljun et al (2015) model are atmospheric stability, friction velocity, standard deviation of v wind speed and surface roughness. We do not anticipate these variables to change significantly between the two distances. The distance to the point source in footprint calculation was from the aerodynamic flux gradient locations. Further, emissions calculation using the corrected method for point source calculations provided promising results for aerodynamic flux gradient quantification in complex emissions profiles where there were multiple emitting sources. This limitation of non-collocated sonic anemometer and the flux gradient tower has been acknowledged and recommended for consideration in future studies.

Changes to Manuscript

Lines 161-164

For AFG, CH_4 concentration data was collected at 2 and 4 m using two Aeris (Hayward, CA, USA) MIRA Ultra Series analyzers connected to tubing with a 3.275 mm inner diameter (Figure 3-2). As we had only one sonic anemometer, data from the sonic anemometer collocated with the MGGA were used for the AFG quantification.

Comment 12

Line 147 "The gas release rates ranged between 0.005 kg h-1 and 8.5 kg h-1, and the release durations ranged from 10 seconds to 8 hours, simulating both fugitive and large emission events. The releases were run both during the day and night". A table in SI that describes the release experiments and the meteorological conditions to make this transparent to the reader would be very helpful. Development of a QA metric that summarized the methane concentration fields observed during each trial would provide important supporting information. As currently written, the reader cannot understand if the under performance of the techniques is related to the method or to non-representative concentration fields or instrument factors.

Author's response

Thank you for comment. After reanalyzing our eddy covariance data, we realized the underestimation is because the data is not suitable for eddy covariance calculations as it failed the quality test. For aerodynamic flux gradient, the underestimation was due to using the area source for footprint calculation. Correctly calculating the footprint for a point source has improved our results. We have uploaded our data and the METEC releases as well. The Reviewer link is here: http://datadryad.org/stash/share/7s42bajRb2czaY9hr6NbTKDAolCWZ_pNrty0o54qXBE.

Comment 13

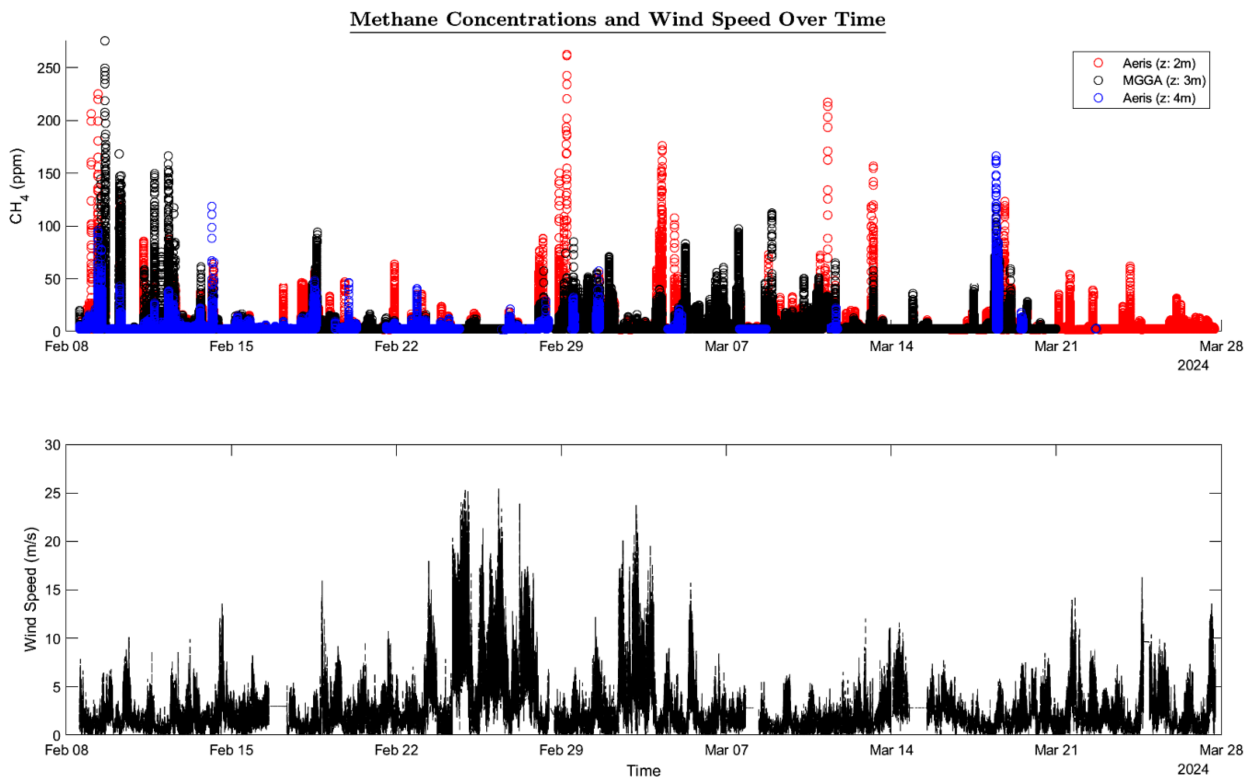
Extending comment 12, because the measurements were simultaneous and almost colocated, a direct comparison of the concentrations measured would be very useful. For example, at 2 m and 4 m on the same mast, what this the ratio of concentrations before and after background correction? For proximate releases at ground level, there may be little signal at 4 m potentially invalidating method assumptions for those cases. At the eddy covariance unit position at 3 m mast height there should be some agreement in measured concentrations (for like time periods) with the other mast (perhaps the average of 2 and 4 m values). Concentration correlation plots by trial would inform both the degree of colocation of the slightly separated masts and the factors of instrument performance, apart from inverse modeling complexity.

Author's response

Thank you for your comment. The data from the eddy covariance mast was used for eddy covariance, backward Lagrangian stochastic and the Gaussian plume inverse model. This ensured direct comparison of these three models. The aerodynamic flux gradient was the only one calculated based on measurements from the 2 and 4 m Aeris analyzers. This study acknowledges the limitation of the sonic anemometer being 9.4 m for aerodynamic measurements. The inputs to the Kljun et al (2015) model are atmospheric stability, friction velocity, standard deviation of v wind speed and surface roughness, and these variables should not change significantly at this distance. However, with the point source footprint calculation, the aerodynamic flux gradient performed better that the Gaussian and backward Lagrangian models for point source quantification when multiple sources were happening. We have also provided the concentration plots in the Supplementary Information Section 1.

<u>Changes to Manuscript</u>

Supplementary Information Section 1



<u>Comment 14</u>

Line 125: Out of curiosity, why were the two masts not located together to facilitate direct comparison? By the photos, this seems that it would have been an easy thing to do.

<u>Author's response</u>

Thank you for your comment. The 2 and 4 m mast was set up before the 3 m one (which could only be set up at the placed location due to internet connectivity).

<u>Comment 15</u>

Basic information on the release trials should be included in SI. For example, how many releases of each type were conducted. At what levels, duration, time of day/night, etc.

<u>Author's response</u>

Thank you for your comment. The release data has been uploaded.

<u>Comment 16</u>

Line 159-165: It is unclear why uniform 20-minute event tables were not used for al measurements to facilitate direct comparisons between the techniques for discrete experiments. The authors make the argument that for two of the approaches, 15 min is not long enough, and 30 min may capture excess atmospheric variability but then just choose the default 30 min homogenous source eddy covariance default period without explanation. For this custom use of EC, 20-minute averages temporally aligned with the other approaches would be the preferred starting position.

Author's response

Thank you for your comment. We have reanalyzed our data to uniform 5 minutes, 10 minutes and 15-minute durations for all models where there were continuous durations.

Changes to Manuscript

Lines 221-223: The concentration-meteorological-release event tables were split into 5, 10 and 15-minute release event tables (i.e. there was a continuous release in the duration).

Comment 17

Line 172: Why was this form of atmospheric stability estimate utilized for the inverse Gaussian approach instead of more robust and comparative estimates derived from high frequency sonic anemometry? This stability estimate was developed to support a specific mobile source assessment approach designed to provide approximate emissions estimates with the derived stability index values limited to daytime measurement. How were nighttime data treated?

Author's response

Thank you for your comment. We have reanalyzed data by generating our dispersion coefficients using the sonic anemometer data. This means that dispersion parameters can be generated at all times.

Changes to Manuscript

Supplementary Information Section 2a

σ_y (m) = horizontal dispersion coefficient calculated as $\sigma_v(x/U)$: σ_v is the standard deviation of v wind speed from the sonic anemometer, x is the downwind distance from the point source and U is the mean wind speed

σ_z (m) = vertical dispersion coefficient calculated as $\sigma_w(x/U)$: σ_w is the standard deviation of w wind speed from the sonic anemometer, x is the downwind distance from the point source and U is the mean wind speed

Comment 18 and 19

Line 192: The background subtraction procedure may represent a significant issue, or at minimum requires further clarification and supporting detail. The authors reference a background correction procedure utilized for a specific mobile source assessment approach designed to provide approximate emissions estimates. That background correction procedure is based on the lowest 5th percentile of a single 15-minute observation. The authors state "CH4 background was calculated as the average of the lowest 5th percentile of all continuous concentration readings (US EPA, 2013)." If the authors combined all data for the series of experiments for this calculation, there are many issues with that approach related to drift and other factors. If the background for each 20/30 minute trial was determined, this needs to be more clearly stated. These data (the corrections) should ideally be provided as SI so the reader can better understand the fidelity of the concentration measurement itself. Currently, the concentration measurements are assumed to be accurate and precise with operation in challenging field conditions.

Related to comment (18), even if the background correction technique utilized was applied to each trial, it may be inappropriate for the inverse forms utilized. This approach can be useful for single point measures of short duration that employ high precision instrumentation (e.g. CRDS). However, is the ambient precision envelope is somewhat broad and if the calibration offset exhibits trend (drift), the background compensation for the 2m vs 4m units for the Aerodynamic Flux Gradient, for example, can have a large impact on the calculation. Currently the reader has no sense for the levels of these corrections or the basic stability and comparability of these instruments over time.

Author's response

Thank you for your comment. We recalculated our background concentration as lowest 5th percentile, 5 minutes before each release started to capture concentrations from the residual methane in air during the preceding release especially during stable conditions. However, in cases where this background was higher than the mean concentration in the quantifying duration, the minimum concentration for that duration was used as background. This was done to ensure we capture the background noise due to the complexity of multiple emitting points, using a constant background from a long emission event, and to capture sensor drift. The following changes were made to the manuscript.

Changes to Manuscript

Section 2.4.3.1

Lines 269-277

For continuous monitoring sensors, background concentration can be determined from CH_4 concentrations measured by a sensor upwind of the emission source, or by sampling when the wind is blowing away from the source. However, for continuous monitoring sensors, using an upwind sensor has the limitation of missing downwind background noise resulting from emissions in the preceding emission event where there is residual CH_4 in air especially during stable conditions, and capturing sensors drift in the downwind sensor. In this study, background CH_4 was calculated as the average of the lowest 5th percentile, 5 minutes before each release started. In cases where this background was greater than the mean CH_4 concentration in the quantifying duration, the minimum CH_4 concentration for that duration was used as the background. Methane enhancement was then calculated as CH_4 concentration minus the background.

Comment 20

The introduction and discussion lack context and comparison to previous research. There are numerous controlled release experiments and other methods comparisons for these three approaches that the authors' results should be put in context with.

Author's response

Thank you for your comment. We have added previous controlled studies to our introduction. The discussion section has also been rewritten.

Changes to Manuscript

4.1 Eddy Covariance

Eddy covariance was tested using a closed-path analyzer, cavity ring-down spectroscopy, with a 3.2 lpm pump flowrate and a 0.4 s gas flow response time. The closed-path EC estimated emissions between a factor of 0.67 and 0.97 for SRSP emissions, and between 1.02 and 2.43 for MRSP emissions at ±45° wind sector range (Section 3.1). This was a wider uncertainty in estimated emissions than one reported by Dumortier et al. (2019), who estimated emissions at between 90 and 113% of true emission (~1.5 kg day^{-1}) with concentrations between 2 and 3 ppm. Our study tested closed-path EC at emission rates between 0.005 and 8.5 kg h^{-1}.

Our study's results were when the data was filtered for frequencies greater than 8 Hz, hence, largely reducing the sampled emissions. The 10 Hz sampling frequency set for this instrument was not a true 10 Hz and this could have been due to the 0.4 gas flow response time that delayed analysis of the drawn air sample in the cavity, or the 3 lpm pump flow rate for a 3 m tubing that might have varied the effective sample turnover rate. The rest of the data were flagged as low quality by Mauder and Foken (2004) (0-1-2 system), which flags based on steady state and well-developed turbulence. This could have been due to low turbulence as experiments were carried out in winter, and instrumentation limitations (low pump flow rate and asynchronous configuration of the gas analyzer and meteorological instrument).

Continuous monitoring requires deployment of multiple sensors which create limitations of cost and requires instrumentation with a wide measurement range as concentrations for oil and gas emissions can range between 0 to 250 ppm, as in this study (Supplementary Information Section 1). The currently available EC instruments have a narrow measurement range (LI-COR LI-7700 open path CH_4 analyzer has a measurement range of 0 to 25 ppm at -25 ºC and 0 to 40 ppm at 25ºC; PICARRO G2311-f, a closed-path analyzer has an operating range of 0 to 20 ppm). Also, the instrumentation should be environmentally robust and not lab-grade (be able to run smoothly in adverse weather conditions). Given these parameters, market available EC instruments that can currently be deployed in oil and gas are limited. The instrument used in this study is a field instrument (ABB MGGA GLA131 Series has a measurement range of 0 to 100 ppm for CH_4 but can be extended to 0 to 1%).

4.2 Aerodynamic Flux Gradient

Overall, the AFG method quantified emissions within an MRF of 1.3 to 1.7 for SRSP emissions and between an MRF of 2.4 and 3.3 for MRSP emissions (Section 3.2). The uncertainties in AFG were higher than EC especially for MRSP emissions but lower than the GPIM and bLs methods. The differences between AFG and EC could have been due to different instrumentation and analytical approaches that limited the exact comparison of the methods i.e. EC data was filtered for frequencies less than 8 Hz, and AFG instrumentations required both analyzers to be running, periods when one analyzer was down, were not tested. To our knowledge, this is the first study to test AFG for point source quantification and results are promising.

Compared to the EC method that requires a very fast analyzer which may be difficult to deploy in oil and gas, the AFG requires at least 2 analyzers sampling at 1 Hz frequency, which is currently possible with the range of sensors available in the market. The main advantage of flux gradient methods (EC and AFG) tested in this study is that they do not require background control as background CH_4 concentration is a highly variable parameter that cannot be controlled in open air especially when multiple emissions are happening. The AFG method relies on differences in CH_4 concentrations between two heights and this study shows that in complex sites where there are multiple sources, the method quantifies better than the point-source GPIM and bLs methods. The main limitation of the AFG and EC methods for point-source quantification is that when the measurement point is downwind of more than one source in a wind sector range, more than one sonic anemometers are required to estimate the flux of each source for footprint calculation based on Dumortier et al. (2019)'s calculations.

4.3 Gaussian Plume Inverse Method

The GPIM method quantified emissions within an MRF of 2.4 and 2.6 for SRSP and between 15.7 and 25 for MRSP emissions (Section 3.3). The GPIM 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, in aerodynamically complex environments, even though the sensor is downwind of a single source based on average wind direction, quantification is complexed by interference from other neighboring sources. The GPIM has previously been reported to quantify emissions within 40.7 and 60% error for a single point-source, (Riddick et al., 2022b). However, GPIM correct quantification has been suggested to be better for longer distances where the plume is well mixed. This is typically a challenge for fence-line sensors that have to be deployed within the facility boundaries where large downwind distances may not be practical.

4.3 Backward Lagrangian Stochastic Model

The bLs method quantified emissions within an MRF of 0.8 to 1.05 for SRSP emissions and between 3.85 and 11958.8 MRF for MRSP emissions (Section 3.4). Similarly to the GPIM method, the bLs method used in this study is a point-source specific quantification method that simulates transport of molecules in open area and where the background concentration is defined. In this case, as with the SRSP test scenario, the bLs approach quantified within 20% uncertainty. However, for MRSP emissions, the bLs largely overestimated emissions and this could have been due to the interference of neighboring sources, that even though the measurement point is downwind of a single source, actual plumes are not distinct and model-simulated plumes may not be representative. The point-source bLs approach in WindTrax is also not designed for more than one downwind source.

Ali Lashgari (RC2)

General comments

This manuscript compares three methods of point source emissions quantification (eddy covariance-EC, aerodynamic flux gradient method-AFGM, and the inverse Gaussian plume method-IGPM), investigating the performance of these methods in quantifying emissions for known gas release rates and evaluating uncertainties involved in the estimates. Although this is a timely topic, I recommend a revision to the current form of the manuscript before its publication. Here are my comments:

Authors response

Thank you for reviewing our paper. We have accordingly addressed the concerns in the specific comments below. We have also added the backward Lagrangian point-source quantification model.

Comment 1

Please provide a sufficient explanation of the three employed approaches. For instance, it is unclear how the inferred rates are estimated in Figures 4-6. Providing sufficient details about the different modeling choices can make significant improvements to the paper. For instance, the equation for the inverse Gaussian plume model (equation 7) only applies to a single source. It will be helpful to explain how this approach is used to model multiple release scenarios. Another example is the definition of the downwind distance (x) under such conditions.

Authors response

Thank you for your comment. We have revised our manuscript accordingly. In this study, we evaluated the models for single-release single-point emissions (single emission at the site level) and multi-release single-point emissions (multiple emissions at the site level, but the mast was downwind of one source. For the eddy covariance and aerodynamic flux gradients models, to estimate footprint of each source when there is more than one source in wind sector range, more than one sonic anemometer is needed to solve the footprint equation as calculated by Dumortier et al (2019). In this study, we had a single sonic anemometer. The backward Lagrangian stochastic model in WindTrax can only simulate one upwind source at a time. As a result, we limited our comparison to these two test scenarios.

Changes to Manuscript

Lines 136-139

Specifically, the study will (1) evaluate the overall quantification accuracy of closed-path EC, AFG, bLs model, and the GPIM method in quantifying single-release single-point and multi-release single-point emissions that simulate oil and gas emissions, (2) determine the mean relative factor (estimated emissions over actual emission) for these models.

(1) Eddy covariance and aerodynamic flux gradient footprint calculation

Lines 254-265

Eddy covariance and AFG footprints were calculated using the Kljun et al. (2015) footprint model. Even though Rey-Sanchez et al. (2022) reported the Kljun et al. (2015) footprint model to be less accurate compared to the Kormann and Meixner (2001), Kormann and Meixner (2001) was too complex for our study because it required multiple sonic anemometers or tracer release experiments to calculate the exponential wind velocity power law, and the exponential eddy diffusivity power law for site specific data. Our study was limited to a single sonic anemometer, and this provided enough inputs for the Kljun et al. (2015) footprint model. The default pixel size 2 * 2 m was used in this study. This study first calculated the area that contributed 90% of the vertical flux; and based on the location 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.

(2) Backward Lagrangian stochastic model

Lines 288-293

Pre-processed data from the GPIM method was used for bLs quantification. Quantification was done using the open-source software WindTrax 2.0 (Crenna, 2006; WindTrax 2.0, n.d.). For every 5-, 10- and 15-minute duration in the ±5º, ±10º, and ±20º, respectively, inputs included roughness length (z0), Monin-Obukhov length (*L*), mean (wind speed, wind direction, concentration, pressure, temperature), background concentration, source height, and distance from the emission point to sensor. WindTrax is also designed to quantify a single point source at a time, and hence, was only used to quantify single-point single emissions and multi-point single emissions.

Comment 2

As part of this study, critical choices are made related to the physics embedded in the covariance w'c' (equation 5) under different scenarios that an eddy covariance (EC) tower provides. More in-depth explanations of those choices will be helpful.

Authors Response

We have accordingly reanalyzed our data using EddyPro software and clearly stated our system configuration and pre-processing of the data using our non-standard eddy covariance system.

Changes to manuscript:

Section 2.4.1

2.4.1 Eddy Covariance

2.4.1.1 Data pre-processing

Evaluating the MGGA CH_4 data showed that actual sampling was between 4 and 12 Hz (highest sampling at 6 Hz), even though it had been set to sample at 10 Hz (Supplementary Information Section 2b). To account for this sampling variability, data were filtered to when sampling was equal/greater than 8 Hz. Data where the frequency was greater than 8 Hz were down sampled to 8 Hz. The sonic anemometer meteorological data (horizontal wind vectors (u, v), vertical wind vector (w), temperature (T), and pressure (P)) actual sampling varied between 7 and 9 Hz with the most frequent frequency at 8 Hz (Supplementary Information Section 2b). As the MGGA gas analyzer and sonic anemometer were not designed to clock synchronously, using the MGGA CH_4 clock time as a reference, meteorological data from the sonic anemometer were matched to the MGGA CH_4 data using linear interpolation to generate concentration-meteorological 8 Hz data.

The aggregated concentration-meteorological data were then merged with METEC's release data and metadata, and release event tables created. Release event tables were aggregated tables of concentration, meteorology and release (emission source location, duration and rate) information for all defined release events at METEC. The concentration-meteorological -release event data were then separated into single-release and multi-release events. Single-release events were when there was a single emission point at the site level, while multi-release events were when there was more than one emission point at the site level. The concentration-meteorological-release event tables were split into 5, 10 and 15-minute release event tables (i.e. there was a continuous release in the duration). 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, ±20º, and ±45º.

2.4.1.2 Flux calculation

Turbulent fluxes were calculated using the open software EddyPro® version 7 (LI-COR, Nebraska, USA, n.d.). Acquisition frequency was set at 8 Hz, while file duration and the flux interval were set at 5, 10, and 15 minutes, respectively, depending on the file being processed. Table 1 shows the instruments input to the software.

Table 1. Anemometer and Gas Analyzer Input into EddyPro

Anemometer Information		**Gas Analyzer Information**	
Manufacturer	Young	Manufacturer	Other
Model	81000	Model	Generic closed path
Height	3 m	Tube length	300 cm
Wind data format	u, v, w	Tube inner diameter	3.275 mm
North alignment		Nominal tube flow rate	3.2 l/m
North off-set	0.0	Northward separation	0.00 cm
Northward separation	Reference	Eastward separation	0.00 cm
Eastward separation	Reference	Vertical separation	-10.00 cm
Vertical separation	Reference	Longitudinal path length	10.00 cm
Longitudinal path length		Transversal path length	2.54 cm
Transversal path length		Time response	0.4 s

In raw data processing, axis rotations for tilt correction under wind speed measurement offsets was checked. Under turbulent fluctuations, double rotation and block average detrend methods were used. Covariance maximization with default was used for time lag detection; time lags detection was checked. Compensation for density fluctuations (Webb-Pearman-Leuning terms) was unchecked as the MGGA analyzer synchronously reported dry CH_4 and water mole fractions, cell temperature and pressure. Mauder and Foken (2004) (0-1-2 system) were used for quality check. All statistical tests for raw data screening, Vickers and Mahrt (1997)– spike count/removal, amplitude resolution, drop-outs, absolute limits, skewness and kurtosis, discontinuities, time lags, angle of attack and steadiness of horizontal wind were checked. The default values for all these tests were used. Similarly, default settings for spectral analysis and corrections were used. Analytic correction of high-pass filtering effects (Moncrieff et al., 2005) for low frequency range; and correction of low-pass filtering effects (Fratini et al., 2012 - In situ analytic) and instruments separation (Horst and Lenschow, 2009 - only crosswind and vertical) in the high frequency range were used.

2.4.1.3 Post-processing

During post-processing, flux data were filtered based on (1) quality flags, Mauder and Foken (2004) (0-1-2 system), and (2) surface friction velocity ($u^* > 0.13$ m/s). Data that were flagged "2" were first filtered out as they were considered poor quality fluxes (LICOR, 2025), and the remaining dataset were filtered for high turbulence data. All data was filtered out as low quality and no further post-processing was done.

Comment 3

Equations 1-4 and their associated text can be moved to Supplemental Information, as they provide more introductory content. I also suggest moving the Gaussian Plume Model (equation 7) along with detailing the dispersion coefficients $\boldsymbol{\sigma}$ and other model parameters in the text to the SI.

Authors response

Thanks for your comment. All equations have been moved to Supplementary Information Section 2a.

Comment 4

While a detailed criticism of the assumptions embedded in the EC and other approaches is added in the discussion section by the authors, very little quantitative effort is made to confirm how several of these assumptions are being violated. For instance, when using the EC approach, no plots are provided, attempting to explain whether the local surface layer at the point of measurement is well mixed or not.

Authors response

Thank you for your comment. We have reanalyzed our data by accounting for these assumptions. For eddy covariance, EddyPro was used to account for factors such as time lags, stationarity and steady tests, and statistical tests that determined whether the data was usable for eddy covariance. Changes to manuscript are in comment 2 above. For the aerodynamic flux gradient, backward Lagrangian and the Gaussian plume inverse models, raw data are used by the assumption that by accounting for the distance, atmospheric stability, wind speed and height which are inputs to the model, we can estimate what the model quantifies in continuous measurements. The variability such as whether the plume is well mixed at the point of measurement is captured in the uncertainty. The point of the study is to check what uncertainties are associated with using continuous monitoring data as inputs for fence-line quantification as inputs to a model.

Comment 5

The decision to use an averaging time window of 30 minutes remains suspect as there are many METEC experiments with durations shorter than half an hour. Reynolds averaging is a critical piece behind the use of eddy covariance concepts and a choice of 30 mins based on discussion from a prior article on agricultural emissions isn't well justified.

Authors response

Thank you for your comment. We have reanalyzed our data to uniform 5 minutes, 10 minutes and 15-minute durations for all models where there were continuous durations. Even though there were short releases such as 10 s that represent intermittent emissions in oil and gas such as a venting of a gas pneumatic, our duration tables represent minute durations when there were continuous releases for that period. For example, a 5-minute event table is for a continuous release in a 5-minute duration. To clarify this, the following changes have been made to the manuscript.

Changes to Manuscript

Section 2.4.1.1

Lines 211-213: The concentration-meteorological-release event tables were split into 5, 10 and 15-minute release event tables (i.e. there was a continuous release in the duration).

Comment 6

I recommend providing a reference dataset of the raw measurements along with a comprehensive validation exercise. This helps readers in the evaluation of the accuracy of the data reduction procedures and to appreciate the inverse modeling choices.

Authors response

Thank you for comment. A comprehensive dataset used to evaluate the models has been uploaded.

URL: http://datadryad.org/stash/share/7s42bajRb2czaY9hr6NbTKDAolCWZ_pNrty0o54qXBE

Comment 7

I recommend a better flow of text and linkage between sections 4.1 and 4.2 (where authors refer to results) and section 3 to draw their conclusions.

Authors response

Thank you for your comment. The discussion section has been rewritten to reflect our results.

Changes to Manuscript

4 Discussion

Methane emissions quantification from oil and gas is a complex system comprising of gas emissions from different heights, different locations, encountering aerodynamic obstacles of different sizes, and of varying emissions duration, amongst others. The ability to precisely quantify emissions using data collected by a point sensor, downwind of a source is directly influenced by plume dynamics. The CH_4 plume downwind of a source will change in size and shape in different atmospheric conditions, in open areas versus areas with obstacles, diurnally, and in different seasons (Casal, 2008). In this study, the precision to which downwind methods (closed-path EC, AFG, GPIM and bLs) 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). As a result, testing the predicted emission rates to controlled release rates in different conditions introduced real-world scenarios that have not previously been tested, hence better understanding model uncertainty in the application of quantifying emissions from oil and gas production infrastructure.

4.1 Eddy Covariance

Eddy covariance was tested using a closed-path analyzer, cavity ring-down spectroscopy, with a 3.2 lpm pump flowrate and a 0.4 s gas flow response time. The closed-path EC estimated emissions between a factor of 0.67 and 0.97 for SRSP emissions, and between 1.02 and 2.43 for MRSP emissions at ±45° wind sector range (Section 3.1). This was a wider uncertainty in estimated emissions than one reported by Dumortier et al. (2019), who estimated emissions at between 90 and 113% of true emission (~1.5 kg day^{-1}) with concentrations between 2 and 3 ppm. Our study tested closed-path EC at emission rates between 0.005 and 8.5 kg h^{-1}.

Our study's results were when the data was filtered for frequencies greater than 8 Hz, hence, largely reducing the sampled emissions. The 10 Hz sampling frequency set for this instrument was not a true 10 Hz and this could have been due to the 0.4 gas flow response time that delayed analysis of the drawn air sample in the cavity, or the 3 lpm pump flow rate for a 3 m tubing that might have varied the effective sample turnover rate. The rest of the data were flagged as low quality by Mauder and Foken (2004) (0-1-2 system), which flags based on steady state and well-developed turbulence. This could have been due to low turbulence as experiments were carried out in winter, and instrumentation limitations (low pump flow rate and asynchronous configuration of the gas analyzer and meteorological instrument).

Continuous monitoring requires deployment of multiple sensors which create limitations of cost and requires instrumentation with a wide measurement range as concentrations for oil and gas emissions can range between 0 to 250 ppm, as in this study (Supplementary Information Section 1). The currently available EC instruments have a narrow measurement range (LI-COR LI-7700 open path CH_4 analyzer has a measurement range of 0 to 25 ppm at -25 ºC and 0 to 40 ppm at 25ºC; PICARRO G2311-f, a closed-path analyzer has an operating range of 0 to 20 ppm). Also, the instrumentation should be environmentally robust and not lab-grade (be able to run smoothly in adverse weather conditions). Given these parameters, market available EC instruments that can currently be deployed in oil and gas are limited. The instrument used in this study is a field instrument (ABB MGGA GLA131 Series has a measurement range of 0 to 100 ppm for CH_4 but can be extended to 0 to 1%).

4.2 Aerodynamic Flux Gradient

Overall, the AFG method quantified emissions within an MRF of 1.3 to 1.7 for SRSP emissions and between an MRF of 2.4 and 3.3 for MRSP emissions (Section 3.2). The uncertainties in AFG were higher than EC especially for MRSP emissions but lower than the GPIM and bLs methods. The differences between AFG and EC could have been due to different instrumentation and analytical approaches that limited the exact comparison of the methods i.e. EC data was filtered for frequencies less than 8 Hz, and AFG instrumentations required both analyzers to be running, periods when one analyzer was down, were not tested. To our knowledge, this is the first study to test AFG for point source quantification and results are promising.

Compared to the EC method that requires a very fast analyzer which may be difficult to deploy in oil and gas, the AFG requires at least 2 analyzers sampling at 1 Hz frequency, which is currently possible with the range of sensors available in the market. The main advantage of flux gradient methods (EC and AFG) tested in this study is that they do not require background control as background CH_4 concentration is a highly variable parameter that cannot be controlled in open air especially when multiple emissions are happening. The AFG method relies on differences in CH_4 concentrations between two heights and this study shows that in complex sites where there are multiple sources, the method quantifies better than the point-source GPIM and bLs methods. The main limitation of the AFG and EC methods for point-source quantification is that when the measurement point is downwind of more than one source in a wind sector range, more than one sonic anemometers are required to estimate the flux of each source for footprint calculation based on Dumortier et al. (2019)'s calculations.

4.3 Gaussian Plume Inverse Method

The GPIM method quantified emissions within an MRF of 2.4 and 2.6 for SRSP and between 15.7 and 25 for MRSP emissions (Section 3.3). The GPIM 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, in aerodynamically complex environments, even though the sensor is downwind of a single source based on average wind direction, quantification is complexed by interference from other neighboring sources. The GPIM has previously been reported to quantify emissions within 40.7 and 60% error for a single point-source, (Riddick et al., 2022b). However, GPIM correct quantification has been suggested to be better for longer distances where the plume is well mixed. This is typically a challenge for fence-line sensors that have to be deployed within the facility boundaries where large downwind distances may not be practical.

4.3 Backward Lagrangian Stochastic Model

The bLs method quantified emissions within an MRF of 0.8 to 1.05 for SRSP emissions and between 3.85 and 11958.8 MRF for MRSP emissions (Section 3.4). Similarly to the GPIM method, the bLs method used in this study is a point-source specific quantification method that simulates transport of molecules in open area and where the background concentration is defined. In this case, as with the SRSP test scenario, the bLs approach quantified within 20% uncertainty. However, for MRSP emissions, the bLs largely overestimated emissions and this could have been due to the interference of neighboring sources, that even though the measurement point is downwind of a single source, actual plumes are not distinct and model-simulated plumes may not be representative. The point-source bLs approach in WindTrax is also not designed for more than one downwind source.

Comment 8

In dispersion modeling efforts, the modeler connects raw measurements with their understanding of the complex boundary layer physics and then uses this to judiciously fix the free parameters of the chosen model while acknowledging the underlying limitations of the constitutive equations all along. Additional explanations are appreciated to provide details on how authors tackled this challenge.

Authors response

Thank you for your comment. This study investigates if the four established downwind models can be applied in oil and gas continuous monitoring i.e., can the model quantify an emission every time there is a detection for a 5, 10 or 15-minute duration. As such, we have fixed the parameters that are inputs to the model based on measurement conditions and our site in our revised manuscript. For example, for the Gaussian plume inverse model, we have reanalyzed our data such that the dispersion parameters are calculated based on sonic anemometer data, as opposed to literature values. This way, we have used site and atmospheric specific data for our calculations ensuring we have correctly used the model.

Comment 9

Line 390: “Oil and gas point sources violate assumptions (1), (2) and (4)”. Assumption 2 requires the CH4 fluxes to be turbulent. If they are not turbulent at the point of measurement, are they laminar? I’d recommend that the authors confirm this by plotting the power spectrum of the CH4 concentration at the point of measurement and investigating the absence of the characteristic turbulence cascade in response. Our experience with continuous monitoring methane concentration data from METEC experiments referenced in the study unequivocally suggests that ambient concentration levels at measuring stations could be turbulent in nature.

Authors response

Thank you for your comment. After reprocessing our data using the EddyPro software, all data were flagged by quality check and friction velocity thresholding hence no further post-processing such as cospectra investigation was done. The Eddy covariance discussion was rewritten as below.

Changes to Manuscript

4.1 Eddy Covariance

Eddy covariance was tested using a closed-path analyzer, cavity ring-down spectroscopy, with a 3.2 lpm pump flowrate and a 0.4 s gas flow response time. The closed-path EC estimated emissions between a factor of 0.67 and 0.97 for SRSP emissions, and between 1.02 and 2.43 for MRSP emissions at ±45° wind sector range (Section 3.1). This was a wider uncertainty in estimated emissions than one reported by Dumortier et al. (2019), who estimated emissions at between 90 and 113% of true emission (~1.5 kg day^{-1}) with concentrations between 2 and 3 ppm. Our study tested closed-path EC at emission rates between 0.005 and 8.5 kg h^{-1}.

Our study's results were when the data was filtered for frequencies greater than 8 Hz, hence, largely reducing the sampled emissions. The 10 Hz sampling frequency set for this instrument was not a true 10 Hz and this could have been due to the 0.4 gas flow response time that delayed analysis of the drawn air sample in the cavity, or the 3 lpm pump flow rate for a 3 m tubing that might have varied the effective sample turnover rate. The rest of the data were flagged as low quality by Mauder and Foken (2004) (0-1-2 system), which flags based on steady state and well-developed turbulence. This could have been due to low turbulence as experiments were carried out in winter, and instrumentation limitations (low pump flow rate and asynchronous configuration of the gas analyzer and meteorological instrument).

Continuous monitoring requires deployment of multiple sensors which create limitations of cost and requires instrumentation with a wide measurement range as concentrations for oil and gas emissions can range between 0 to 250 ppm, as in this study (Supplementary Information Section 1). The currently available EC instruments have a narrow measurement range (LI-COR LI-7700 open path CH_4 analyzer has a measurement range of 0 to 25 ppm at -25 ºC and 0 to 40 ppm at 25ºC; PICARRO G2311-f, a closed-path analyzer has an operating range of 0 to 20 ppm). Also, the instrumentation should be environmentally robust and not lab-grade (be able to run smoothly in adverse weather conditions). Given these parameters, market available EC instruments that can currently be deployed in oil and gas are limited. The instrument used in this study is a field instrument (ABB MGGA GLA131 Series has a measurement range of 0 to 100 ppm for CH_4 but can be extended to 0 to 1%).

Comment 10

Assumption 4 states that measurements should be inside the boundary layer and in the constant stress layer. If the choice of RL from section 2.4.4 is to be taken at face value, then the constant stress layer extends up to 1m upward thus implying that measurements satisfy assumption 4.

Authors response

This statement has been withdrawn from our discussion as in the response to Comment 7 above. Also, we recalculated the roughness length from our sonic anemometer measurements in neutral conditions.

Changes to Manuscript

2.3 Calculation of Roughness Length

Surface roughness length (z0) was calculated from friction velocity (Supplementary Information Section 2a: Equations 1 and 2) by splitting the high frequency sonic anemometer data into 15-minute tables and filtering for those in neutral conditions, |L| > 500 (Supplementary Information Section 2a: Equation 3). The overall roughness length selected as the median of all the calculated z0 was 0.1 m (Rey-Sanchez et al., 2022).

Comment 11

Line 395-398: "Even though current eddy covariance application assumes the vertical flux at a point is independent of atmospheric stability (Denmead, 2008), atmospheric stability has an impact on point source gas …". While Denmaead (2008) notes that EC "is independent of atmospheric stability", it merely implies that EC by virtue of being a more fundamental approach relies on direct measurement of turbulent fluxes like <u'w'> and <u'θ'>, i.e. the information on atmospheric stability is implicitly built into the EC approaches. Apriori estimates of the Obhukov length scale that divide the surface layer into multiple regimes as required by the AFGM and Gaussian models are therefore not needed. The EC does not imply that the vertical flux is independent of atmospheric stability.

Authors response

Thank you for your comment. After reprocessing our data using the EddyPro software, all data were flagged by quality check and friction velocity thresholding hence no further post-processing. This statement has been withdrawn from our discussion as in the response to Comment 7 above.

Comment 12

Line 384-385 & 411: "Positive fluxes represent emissions and downward fluxes represent absorptions". This is not an assumption, but rather a statement of fact and an unnecessary one at that.

Authors response

This statement has been withdrawn from our discussion as in the response to Comment 7 above.

Comment 13

Line 391: "Emissions are collimated plumes instead of turbulent fluxes". It requires clarification as to what the authors mean when mentioning the phrase 'turbulent fluxes'. Collimated plumes are simply the smoothing effect of time averaging on an underlying turbulent structure. Please see the following references

- https://www.cambridge.org/highereducation/books/turbulent-flows/C58EFF59AF9B81AE6CFAC9ED16486B3A#overview
- https://epubs.siam.org/doi/pdf/10.1137/10080991X

Authors response

This statement has been withdrawn from our discussion as in the response to Comment 7 above.

Comment 14

Line 72: I wonder how are major shortcomings identified using fence-line approaches by Ilonze et al. (2024)?

Authors response

Ilonze et al. (2024) study compared continuous monitoring solutions and their reported quantification estimates based on what solutions reported. Details on specific factors that caused some solutions to overestimate are not provided. The study provides the overall performance of fence-line continuous monitoring sensors.

Comment 15

Line 219: "The roughness sublayer is set to 1" is not well justified. I suppose it means 1m? A roughness sublayer (RL) is a region immediately above the surface and below the inertial layer where horizontal in-homogeneity in flow variables persists. In other words, within the RL variables like <w'c'> (where c is the local methane concentration, w is the vertical wind speed, and ' represents the fluctuating field) can be assumed to be horizontally homogeneous. While there are well-established results in the literature on how to estimate the RL height, having access to direct point measurements at multiple heights offers a direct route to estimate the RL height. Please see the following reference.

- https://journals.aps.org/prfluids/abstract/10.1103/PhysRevFluids.3.114603

Authors response

We recalculated the roughness length from our sonic anemometer measurements in neutral conditions.

Changes to Manuscript

2.3 Calculation of Roughness Length

Surface roughness length (z0) was calculated from friction velocity (Supplementary Information Section 2a: Equations 1 and 2) by splitting the high frequency sonic anemometer data into 15-minute tables and filtering for those in neutral conditions, $|L| > 500$ (Supplementary Information Section 2a: Equation 3). The overall roughness length selected as the median of all the calculated z0 was 0.1 m (Rey-Sanchez et al., 2022).

Comment 16 and 17

Lines 430-431: "The main assumption of the Gaussian plume model is that CH4 emitted from a point source enters the airflow, disperses vertically and laterally, forming a conical plume (Riddick et al., 2022b; US EPA, 2013)". This is not an assumption but instead the standard response of the tracer being advected horizontally and vertically by the air.

Line 431-433: "The formation of a conical plume is hindered at oil and gas facilities by obstacles (equipment) and is affected by atmospheric stability." While strictly speaking sizable obstacles may hinder the plume, whether this is indeed the case for the METEC site is unclear. The manuscript does not offer proper evidence to support this argument. Given the sparsity of the few obstacles on the METEC site and the focus of the current study on distant, fenceline monitoring, it seems to reason that obstacles in fact do not hinder plume features.

Authors response

This statement has been withdrawn from our Gaussian plume discussion has been corrected as below.

Changes to manuscript

4.3 Gaussian Plume Inverse Method

The GPIM method quantified emissions within an MRF of 2.4 and 2.6 for SRSP and between 15.7 and 25 for MRSP emissions (Section 3.3). The GPIM 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, in aerodynamically complex environments, even though the sensor is downwind of a single source based on average wind direction, quantification is complexed by interference from other neighboring sources. The GPIM has previously been reported to quantify emissions within 40.7 and 60% error for a single point-source, (Riddick et al., 2022b). However, GPIM correct quantification has been suggested to be better for longer distances where the plume is well mixed. This is typically a challenge for fence-line sensors that have to be deployed within the facility boundaries where large downwind distances may not be practical.

Comment 18

Line 463-465: "Even though these modeling approaches have been reported to work elsewhere (e.g., agricultural and landfill emissions), it does not necessarily mean it could work in the intended area of application." This blanket statement is unwarranted. Compared to many complex scenarios observed in agricultural and landfill emission cases, the current controlled-release tests present a much-simplified dataset both in terms of emission characteristics (discrete constant emissions rates from only 5 equipment groups all starting and stopping simultaneously) and site features (flat terrain with few sparsely spaced obstacles).

Authors response

This statement has been withdrawn and the implications section has been corrected as below.

Changes to manuscript

Lines 522-535

Oil and gas point sources could either be single emissions or multiple emissions occurring concurrently. In cases of multiple emissions with more than one release point being upwind, the 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. As a result, flux quantification models used in other applications such as eddy covariance and aerodynamic flux gradient have been proposed as the solution. This study's results show that generally reasonable quantification estimates are achieved with flux approaches (eddy covariance and aerodynamic flux gradient), but these methods require more instrumentation effort (fast sampling analyzer for eddy covariance, and multiple collocated sensors for aerodynamic flux gradient). Even though the widely applied Gaussian plume inverse method and the backward Lagrangian stochastic models are widely used for single-point emissions, this study shows aerodynamic complexities, the difficulty in defining the background, and interference from neighboring sources challenge the application of these models for fence-line continuous monitoring. This study recommends more testing of flux quantification models for oil and gas quantification as they could improve emissions quantification for leak repair prioritization and methane reporting.

Anonymous Referee #3 (RC3)

This paper makes a valuable contribution towards quantifying emissions from point sources. The problem is important, and the work is thoughtful. Mbua et al. use controlled release experiments to test three widely used methodologies and find all three methods are inaccurate and two methods seriously underestimate emissions. In Mbua et al's controlled release experiments, using fence line measurements, both eddy covariance and flux gradient methods were found to underestimate emissions. Gaussian plume methods performed better but with very wide scatter. I note the points already raised by RC1 and RC2, particularly RC1's point 4 on closed path measurement leading to underestimation. More generally, this paper is tackling a major problem. Accurate quantification of emissions (especially methane) from point sources such as natural gas production facilities is a tough nut to crack, more so if the "point: source is somewhat disseminated (e.g. a gas production facility with pipes, pumps and valves scattered over areas up to a hectare, or similarly, a farm complex over a similar area with manure lagoons, biodigesters and animals. Very large sources (eg >100 kg/hr methane) can be studied from space, and small sources (e.g. single cows) can be isolated, but quantifying complex aggregated sources in the 1-50 kg/hr range (like gas production facilities) is not easy. Use of methods similar to Mbua et al's systems is widespread, and thus these experiments are potentially important in devising better protocols and better regulation and mitigation of methane emissions from fossil fuel installations (gas production facilities, venting oil wells, coal mine vents, etc). The work also has wide applicability for tracking biogenic emissions from landfill and waste, around biodigesters, and on farms with large animal populations. The scientific methodology and experimental procedures are clearly explained and the work is well presented. Overall the paper is well written. Controlled release experiments like this are likely to be influential in designing better quantification methods and in devising regulation protocols. Thus, the paper should be accepted. That said, I note the comments made by other referees and have some additional minor comments.

Authors response

Thank you for reviewing our paper. We have accordingly addressed the concerns in the specific comments below.

Comment 1

Abstract. Lines 1-11 are really introduction and might be omitted here, and covered in Section 1 instead. Line 20-21 'input to……..method' – minor English problem? Maybe methods?

Authors response

Thank you for your comment. The first sentences aim at providing context for our study, and why we have to evaluate quantification accuracy for methane reporting.

Changes to Manuscript

Abstract. The dependable reporting of methane (CH_4) emissions from point sources, such as fugitive leaks from oil and gas infrastructure, is important for profit maximization (retaining more hydrocarbons), evaluating climate impacts, assessing CH_4 fees for regulatory programs, and validating CH_4 intensity in differentiated gas programs. Currently, there are disagreements between emissions reported by different quantification techniques for the same sources. It has been suggested that downwind CH_4 quantification methods using CH_4 measurements on the fence-line of production facilities could be used to generate emission estimates from oil and gas operations at the site level, but it is currently unclear how accurate the quantified emissions are. To investigate downwind methods' accuracy, this study uses fence-line simulated data collected during controlled release experiments as input for closed-path eddy covariance, aerodynamic flux gradient, the Gaussian plume inverse method, and the backward Lagrangian stochastic model in a range of atmospheric conditions. Generally, results show that flux quantification methods provide more reasonable estimates compared to point-source specific models especially when multiple releases are happening at the facility level. The closed-path eddy covariance quantified emissions with a mean relative factor (estimated emission over actual emission) of 0.7 to 1 for single-release single-point emissions, and within a mean relative of 1 and 2.4 for multi-release single-point emissions. The aerodynamic flux gradient method quantified emissions within a mean relative factor of 1.3 to 1.7 for single-release single-point emissions, and between 2.4 and 3.3 for multi-release single-point emissions. The Gaussian plume inverse model quantified emissions within a mean relative factor of between 2.4 and 2.6 for single-release single-point emissions, but largely overestimated emissions when multiple releases were happening; mean relative factor between 16 and 25. Similarly to the Gaussian plume inverse method, the backward Lagrangian stochastic model for point sources using WindTrax quantified within a mean relative factor of between 0.8 to 1 for single-release single-point emissions, but largely overestimated emissions for multi-release single-point emissions; mean relative factor of 3.9 and 11958. As continuous monitoring of oil and gas sites can involve complex emissions where plumes are not defined due to multiple sources, this study shows that common downwind point source dispersion models could largely overestimate emissions. This study recommends more testing of flux quantification models for oil and gas continuous monitoring quantification.

Comment 2

Lines 1-32 could be rewritten somewhat, maybe to mention the Global Methane Pledge and to introduce the tension between bottom-up and top-down methodologies.

<u>Authors response</u>

Thank you for your comment. We have rewritten the abstract section based on our results as in Comment 2 above. The discrepancy between bottom up and top down technologies is included in the introduction.

Comment 3

Line 50 – maybe give some details about very rough emission ranges (e.g. satellite detection of point sources >100 kg/h, low flying aircraft (say 10-100 kg, SUV-mounted instruments (say 1-20 kg/hr). Also maybe mention the scale of 'disseminated' 'point' sources – e.g. a complex gas facility on a 100mx100m pad, or a farm barn/lagoon complex of the same size.

<u>Authors response</u>

Thank you for your comment. Detection limits have been added to the manuscript.

<u>Changes to Manuscript</u>

<u>Lines 52-60</u>

Top-down methods, including using aircraft such as Bridger Photonics LiDAR (Light Detection and Ranging; 90% detection limit of ~ 2 kg h^{-1}) (Johnson et al., 2021) and satellites such as Carbon Mapper (predicted 90% detection limit of about 100 kg h^{-1}) ("Carbon Mapper - Science & Technology," n.d.) can also be used to infer emissions. However, these survey methods only quantify emissions over a very short period of time (< 10 s) and observations are typically made during the day which can often coincide with maintenance activities that can bias emissions and result in overestimation (Riddick et al., 2024a; Zimmerle et al., 2024). Additionally, different top-down technologies measuring the same source have disagreed in their reported emissions which has called into question the credibility of these methods (Brown et al., 2023; Conrad et al., 2023). As a result, ensuring accuracy in models and technologies used in CH_4 emissions quantification has been a complex issue.

Comment 4

Line 56 – 'NG acronym may not be understood in a global journal. Better spell it out.

Authors response

Thanks for your comment. NG has been spelled out.

Changes to Manuscript

Comment 5

Line 96 – Fig 1 is good. Maybe add a Gaussian plume panel also?

Thanks for your comment. The figures for the Gaussian plume inverse model and backward Lagrangian models have been added.

Changes to Manuscript

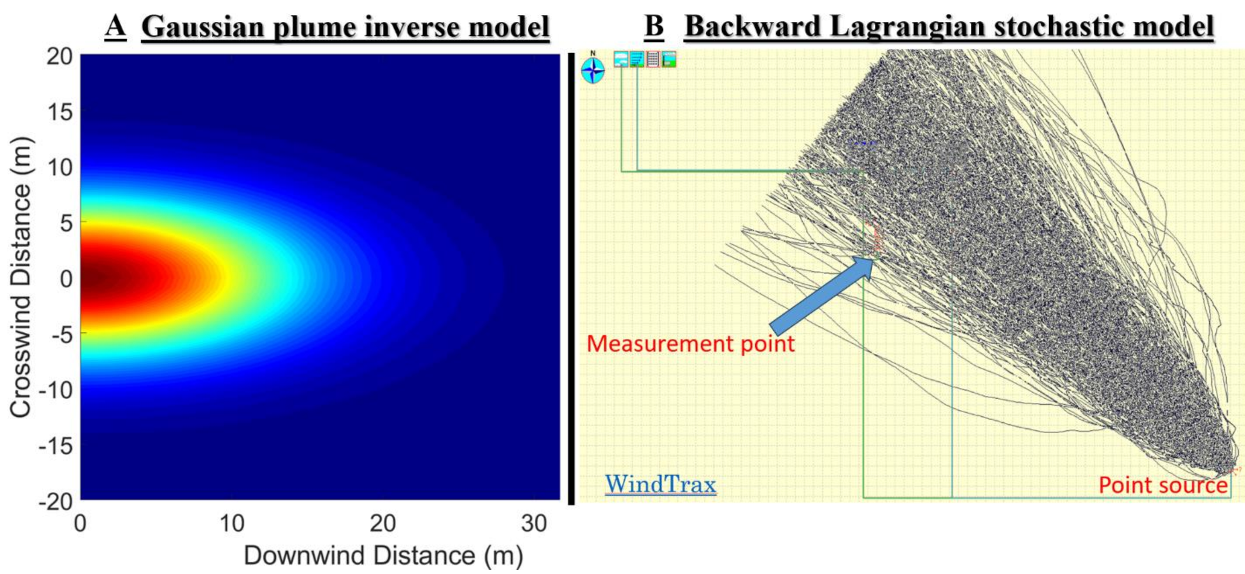


Figure 2. A: An illustration of a plume that follows a Gaussian plume inverse model where emission rate can be inferred from concentrations at different downwind distances and crosswind distances. B: An illustration of how the backward Lagrangian stochastic model traces particles to the source.

Comment 6

Line 115 – maybe say how big the METEC facility is and give more details for non-US readers.

Authors response

Thank you for your comment. A description of METEC has been added.

Changes to Manuscript

Section 2.1

Lines 142-146

Controlled release experiments were conducted at the Colorado State University's Methane Emissions Technology Evaluation Center (METEC) in Fort Collins, CO (USA, 65 miles north of Denver) between February 8, and March 20, 2024. The METEC center is a simulated oil and gas facility that does controlled testing for emissions leak detection and quantification technology development, field demonstration, leak detection protocol and best practices development (METEC, 2025).

Comment 7

Line 181 – masts' locations? Plural masts on many locations? Text as it is now has many mast on one location.

Authors response

Thank you for your comment. We had two masts 9.4 m apart, one for eddy covariance, backward Lagrangian stochastic model and the Gaussian plume measurements, and the other for the aerodynamic flux gradient quantification.

Changes to Manuscript

Lines 163-173

Methane concentration data for closed-path EC, GPIM and bLs methods were collected through an inlet tubing (3.275 mm inner diameter) at 3 m height, connected to the ABB (Zurich, Switzerland) GLA131 Series Microportable Greenhouse Gas Analyzer (MGGA) set to sample at 10 Hz. The MGGA is a closed-path greenhouse gas analyzer with a ~3.2 lpm pump flowrate, 10 cm cell length, 1 inch cell diameter (~0.23 standard cubic centimeters per minute (sccm) effective volume), and 0.4 s gas flow response time. The inlet tubing was collocated with an R. M. Young (Traverse City, MI, USA) 81000 sonic anemometer (R.M. Young Company, 2023) which measured micrometeorology at 10 Hz (Figure 3-1). The northward, eastward and vertical separation of the inlet tubing from the sonic anemometer was 0, 0, -10 cm, respectively. For AFG, CH_4 concentration data was collected at 2 and 4 m using two Aeris (Hayward, CA, USA) MIRA Ultra Series analyzers connected to tubing with a 3.275 mm inner diameter (Figure 3-2). As we had only one sonic anemometer, data from the sonic anemometer collocated with the MGGA were used for the AFG quantification. The two sampling points are 9.4 m apart.

Comment 8

Line 194 – background. 5th percentile - A steady time-invariant small leak could thus be included in 'background' Are there any measurements of upwind background?

Authors response

Thanks for your comment. We did not have an upwind mast. Background concentrations were recalculated to account for time variation and residual methane from previous releases.

Changes to Manuscript

Lines 269-277

For continuous monitoring sensors, background concentration can be determined from CH_4 concentrations measured by a sensor upwind of the emission source, or by sampling when the wind is blowing away from the source. However, for continuous monitoring sensors, using an upwind sensor has the limitation of missing downwind background noise resulting from emissions in the preceding emission event where there is residual CH_4 in air especially during stable conditions, and capturing sensors drift in the downwind sensor. In this study, background CH_4 was calculated as the average of the lowest 5th percentile, 5 minutes before each release started. In cases where this background was greater than the mean CH_4 concentration in the quantifying duration, the minimum CH_4 concentration for that duration was used as the background. Methane enhancement was then calculated as CH_4 concentration minus the background.

Comment 9

Line 215 – footprint 80%-90% contour contrast. This is interesting: space for a slightly longer comment?

Authors response

After evaluating the approach for footprint calculation for point sources, we used the 90% footprint area. We identified if the point source was located in the 90% footprint, and using the footprint value, we calculated the emission.

Changes to Manuscript

Lines 254-265

Eddy covariance and AFG footprints were calculated using the Kljun et al. (2015) footprint model. Even though Rey-Sanchez et al. (2022) reported the Kljun et al. (2015) footprint model to be less accurate compared to the Kormann and Meixner (2001), Kormann and Meixner (2001) was too complex for our study because it required multiple sonic anemometers or tracer release experiments to calculate the exponential wind velocity power law, and the exponential eddy diffusivity power law for site specific data. Our study was limited to a single sonic anemometer, and this provided enough inputs for the Kljun et al. (2015) footprint model. The default pixel size 2 * 2 m was used in this study. This study first calculated the area that contributed 90% of the vertical flux; and based on the location 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.

Comment 10

Line 235 – Line 346 'largely underestimated' – see also Fig 4. I note the comment posted earlier by Brian Lamb. Maybe that question could be picked up further? Mbua et al's finding is surprising and immediately makes the reader worry about eddy covariance quantification in wetlands also!

Authors response

Thanks for your comment. Point source emission was recalculated based on the approach by Dumortier et al (2019) where emission per source = measured flux / value of the footprint at the cell. The source location was first determined if it was within the 90% footprint region, and the footprint value extracted.

Changes to Manuscript:

Section 2.4.3

Lines 261-265

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.

Comment 11

Lines 388- 401. This is important and perhaps needs also later to be put in the wetland context where atmospheric stability can also be a problem.

Authors response

Thanks for your comment. This statement was withdrawn and eddy covariance discussed as follows.

Changes to Manuscript

Section 4.1

Eddy covariance was tested using a closed-path analyzer, cavity ring-down spectroscopy, with a 3.2 lpm pump flowrate and a 0.4 s gas flow response time. The closed-path EC estimated emissions between a factor of 0.67 and 0.97 for SRSP emissions, and between 1.02 and 2.43 for MRSP emissions at ±45° wind sector range (Section 3.1). This was a wider uncertainty in estimated emissions than one reported by Dumortier et al. (2019), who estimated emissions at between 90 and 113% of true emission (~1.5 kg day^{-1}) with concentrations between 2 and 3 ppm. Our study tested closed-path EC at emission rates between 0.005 and 8.5 kg h^{-1}.

Our study's results were when the data was filtered for frequencies greater than 8 Hz, hence, largely reducing the sampled emissions. The 10 Hz sampling frequency set for this instrument was not a true 10 Hz and this could have been due to the 0.4 gas flow response time that delayed analysis of the drawn air sample in the cavity, or the 3 lpm pump flow rate for a 3 m tubing that might have varied the effective sample turnover rate. The rest of the data were flagged as low quality by Mauder and Foken (2004) (0-1-2 system), which flags based on steady state and well-developed turbulence. This could have been due to low turbulence as experiments were carried out in winter, and instrumentation limitations (low pump flow rate and asynchronous configuration of the gas analyzer and meteorological instrument).

Continuous monitoring requires deployment of multiple sensors which create limitations of cost and requires instrumentation with a wide measurement range as concentrations for oil and gas emissions can range between 0 to 250 ppm, as in this study (Supplementary Information Section 1). The currently available EC instruments have a narrow measurement range (LI-COR LI-7700 open path CH_4 analyzer has a measurement range of 0 to 25 ppm at -25 ºC and 0 to 40 ppm at 25ºC; PICARRO G2311-f, a closed-path analyzer has an operating range of 0 to 20 ppm). Also, the instrumentation should be environmentally robust and not lab-grade (be able to run smoothly in adverse weather conditions). Given these parameters, market available EC instruments that can currently be deployed in oil and gas are limited. The instrument used in this study is a field instrument (ABB MGGA GLA131 Series has a measurement range of 0 to 100 ppm for CH_4 but can be extended to 0 to 1%).

Comment 12

Line 434 – daytime insolation - in facilities working all night as well as all day, nocturnal factors like low inversion height and fogs may also be a factor, especially if the facility is installing automated systems.

Authors response

Thanks for your comment. We recalculated the Gaussian plume inverse model based on sonic measurements as opposed to Pasquil Gifford classification system hence, the daytime insolation is no longer considered. This means that dispersion parameters can be generated at all times.

Changes to Manuscript

Supplementary Information Sectio 2a

σy (m) = horizontal dispersion coefficient calculated as $\sigma v(x/U)$: σv is the standard deviation of v wind speed from the sonic anemometer, x is the downwind distance from the point source and U is the mean wind speed

σz (m) = vertical dispersion coefficient calculated as $\sigma w(x/U)$: σw is the standard deviation of w wind speed from the sonic anemometer, x is the downwind distance from the point source and U is the mean wind speed

Comment 13

Line 465 – maybe a paragraph here in 'future fixes' – how to make EC work better perhaps, and especially how to make Gaussian methods better. For example, by using drones it may be possible to map plumes much more accurately. Also perhaps a digression somewhere to mention wetlands (e.g. L386) in more detail.

Authors response

Based on our results, the discussion and conclusion was changed as below.

Changes to Manuscript

4.1 Eddy Covariance

Eddy covariance was tested using a closed-path analyzer, cavity ring-down spectroscopy, with a 3.2 lpm pump flowrate and a 0.4 s gas flow response time. The closed-path EC estimated emissions between a factor of 0.67 and 0.97 for SRSP emissions, and between 1.02 and 2.43 for MRSP emissions at ±45° wind sector range (Section 3.1). This was a wider uncertainty in estimated emissions than one reported by Dumortier et al. (2019), who estimated emissions at between 90 and 113% of true emission (~1.5 kg day^{-1}) with concentrations between 2 and 3 ppm. Our study tested closed-path EC at emission rates between 0.005 and 8.5 kg h^{-1}.

Our study's results were when the data was filtered for frequencies greater than 8 Hz, hence, largely reducing the sampled emissions. The 10 Hz sampling frequency set for this instrument was not a true 10 Hz and this could have been due to the 0.4 gas flow response time that delayed analysis of the drawn air sample in the cavity, or the 3 lpm pump flow rate for a 3 m tubing that might have varied the effective sample turnover rate. The rest of the data were flagged as low quality by Mauder and Foken (2004) (0-1-2 system), which flags based on steady state and well-developed turbulence. This could have been due to low turbulence as experiments were carried out in winter, and instrumentation limitations (low pump flow rate and asynchronous configuration of the gas analyzer and meteorological instrument).

Continuous monitoring requires deployment of multiple sensors which create limitations of cost and requires instrumentation with a wide measurement range as concentrations for oil and gas emissions can range between 0 to 250 ppm, as in this study (Supplementary Information Section 1). The currently available EC instruments have a narrow measurement range (LI-COR LI-7700 open path CH_4 analyzer has a measurement range of 0 to 25 ppm at -25 ºC and 0 to 40 ppm at 25ºC; PICARRO G2311-f, a closed-path analyzer has an operating range of 0 to 20 ppm). Also, the instrumentation should be environmentally robust and not lab-grade (be able to run smoothly in adverse weather conditions). Given these parameters, market available EC instruments that can currently be deployed in oil and gas are limited. The instrument used in this study is a field instrument (ABB MGGA GLA131 Series has a measurement range of 0 to 100 ppm for CH_4 but can be extended to 0 to 1%).

4.2 Aerodynamic Flux Gradient

Overall, the AFG method quantified emissions within an MRF of 1.3 to 1.7 for SRSP emissions and between an MRF of 2.4 and 3.3 for MRSP emissions (Section 3.2). The uncertainties in AFG were higher than EC especially for MRSP emissions but lower than the GPIM and bLs methods. The differences between AFG and EC could have been due to different instrumentation and analytical approaches that limited the exact comparison of the methods i.e. EC data was filtered for frequencies less than 8 Hz, and AFG instrumentations required both analyzers to be running, periods when one analyzer was down, were not tested. To our knowledge, this is the first study to test AFG for point source quantification and results are promising.

Compared to the EC method that requires a very fast analyzer which may be difficult to deploy in oil and gas, the AFG requires at least 2 analyzers sampling at 1 Hz frequency, which is currently possible with the range of sensors available in the market. The main advantage of flux gradient methods (EC and AFG) tested in this study is that they do not require background control as background CH_4 concentration is a highly variable parameter that cannot be controlled in open air especially when multiple emissions are happening. The AFG method relies on differences in CH_4 concentrations between two heights and this study shows that in complex sites where there are multiple sources, the method quantifies better than the point-source GPIM and bLs methods. The main limitation of the AFG and EC methods for point-source quantification is that when the measurement point is downwind of more than one source in a wind sector range, more than one sonic anemometers are required to estimate the flux of each source for footprint calculation based on Dumortier et al. (2019)'s calculations.

4.3 Gaussian Plume Inverse Method

The GPIM method quantified emissions within an MRF of 2.4 and 2.6 for SRSP and between 15.7 and 25 for MRSP emissions (Section 3.3). The GPIM 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, in aerodynamically complex environments, even though the sensor is downwind of a single source based on average wind direction, quantification is complexed by interference from other neighboring sources. The GPIM has previously been reported to quantify emissions within 40.7 and 60% error for a single point-source, (Riddick et al., 2022b).

However, GPIM correct quantification has been suggested to be better for longer distances where the plume is well mixed. This is typically a challenge for fence-line sensors that have to be deployed within the facility boundaries where large downwind distances may not be practical.

4.3 Backward Lagrangian Stochastic Model

The bLs method quantified emissions within an MRF of 0.8 to 1.05 for SRSP emissions and between 3.85 and 11958.8 MRF for MRSP emissions (Section 3.4). Similarly to the GPIM method, the bLs method used in this study is a point-source specific quantification method that simulates transport of molecules in open area and where the background concentration is defined. In this case, as with the SRSP test scenario, the bLs approach quantified within 20% uncertainty. However, for MRSP emissions, the bLs largely overestimated emissions and this could have been due to the interference of neighboring sources, that even though the measurement point is downwind of a single source, actual plumes are not distinct and model-simulated plumes may not be representative. The point-source bLs approach in WindTrax is also not designed for more than one downwind source.