Reviewer 2

Title: Quantitative comparison of methods used to estimate methane emissions from small point sources

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

E-mail: Stuart.Riddick@colostate.edu

26th April 2022

Dear Dr Chen,

We thank reviewer 2 for their comments. As suggested, we have amended the manuscript to address the reviewers' comments and have indicated changes to the manuscript in red text.

Please find our detailed responses below.

Yours sincerely,

Stuart Riddick

Reviewer 2 General comment 1:

The state of the art in the introduction doesn't acknowledge other available techniques. The selection of methods reproduced in the study is not explicit, and the reasons for ignoring/discarding other techniques is not clarified.

Response to reviewer:

The genesis of this publication was the repeated question asked of METEC research scientists: "how good are these methods at measuring emissions typically seen from abandoned oil and gas wells and what are the benefits of repeat experiments?". Here, we present the methods most commonly used to quantify emissions and conduct blinded experiments to how representative a single quantification measurement can be and then investigate the utility of repeat measurements. Other methods that the reviewer has listed below are simply not suitable for measuring emissions typical of the majority of abandoned wells. FLIR is not a quantification approach, mass balance could be used but I have never come across it being used to measure individual point sources, tracer release is a long a technically tricky approach and ill-suited to this sort of hit-and-run measurement approach, and remote sensing has typical detection limits of 10+ kg CH_4 h^{-1} for drones and 100+ kg CH_4 h^{-1} for aircraft and completely unsuitable.

Changes to the manuscript:

To clarify our objective we have added text at L 27:

"Methane (CH_4) gas is a powerful greenhouse gas with a greenhouse warning potential 84 times larger than carbon dioxide over 100 years. Quantification of CH_4 emissions from abandoned wells has recently become an area of interest as studies suggest over 200 Gg CH_4 yr^{-1} is emitted from 2.2 million abandoned wells in the US alone (US EPA, 2021). Quantifying and then plugging these wells makes them an attractive target for achieving goals set out in the Paris Agreement (Nisbet et al., 2020). Additionally, private companies are beginning initiatives to generate revenue through carbon credits gained by plugging wells and accurate quantification is essential for realizing the capital.

As there are millions of abandoned wells globally, there is a growing need to measure as many wells as quickly as possible to identify the most emissive wells. Typically, an emission from an abandoned well can be considered as an above-ground point source that is relatively small in emission size, up to 180 g CH_4 $hour^{-1}$ (Riddick et al., 2019a; Pekney et al., 2018; Townsend-Small et al., 2016; Boothroyd et al., 2016). Other emission sources, such as emissions pipeline leakage, are fundamentally different in behavior, where gas travels through the soil and forms an area emission at the surface, these sources require different methods for estimating the emission, e.g. mass balance or eddy covariance. Area emissions could form if a plugged well leaks from corrosion of the borehole casing, but this will not be discussed in this study.

Several methods are being used to measure emissions from these smaller point sources, i.e. less than 180 g CH_4 $hour^{-1}$. The chosen measurement approach depends on how close an observer can get to the source, instrumentation availability and the meteorological/micrometeorological conditions at the measurement site. Measurement methods can be classed as direct, i.e. touching/enclosing the source, and downwind measurements where access is not possible. Direct methods include static chambers (Livingston and Hutchinson, 1995), dynamic flux chambers (Riddick et al., 2019a, 2020b; Aneja et al., 2006) and Hi Flow sampling (Pekney et al., 2018; Allen et al., 2013; Brantley et al., 2015). While downwind methods include Gaussian-based plume models (Baillie et al., 2019; Caulton et al., 2014; Riddick et al., 2019b, 2020a; Edie et al., 2020; Bell et al., 2017) and Lagrangian dispersion models (Riddick et al., 2019b, 2017; Denmead, 2008; Flesch et al., 1995). Emissions calculated using the majority of these methods have not been comprehensively compared using to controlled emission source rates.

Other quantification methods are generally unsuitable for measuring emissions from abandoned wells. Infra-red cameras, such as FLIR cameras, cannot be used to quantify emission and have difficultly detecting plume smaller emissions (Zimmerle et al., 2020). Mass balance approaches are unlikely to detect the small and narrow plume from the abandoned well. Tracer release is technically demanding, takes a long time to make a single measurement and requires road access for measurement. Remote sensing has typical detection limits of 10+ kg CH_4 h^{-1} for aircraft (Duren et al., 2019), 100+ kg CH_4 h^{-1} for satellites (Cooper et al., 2022) and unsuitable for these types of emission source. As such, these other quantification methods will not be investigated in this study.

In general, as access becomes more restricted, emission rates larger, or safety concerns increase (such as the co-emission of harmful gases), the method used to estimate the CH_4 emission rate of a source must be carefully considered. From experience and the response of a 4-gas monitor, working close enough to measure emissions greater than 200 g CH_4 h^{-1} for many of these methods (especially the chambers and Hi Flow) can be unsafe, therefore this study is limited to quantifying CH_4 emissions between the lowest flow METEC can produce (40 g CH_4 h^{-1}) and the highest flow we feel comfortable measuring with these methods (200 g CH_4 h^{-1}). Putting these emission ranges into real-word context, the maximum emission from unplugged and abandoned wells was measured at 177 g CH_4 h^{-1} in West Virginia (Riddick et al., 2019a), 175 g CH_4 h^{-1} in Pennsylvania (Pekney et al., 2018), 146 g CH_4 h^{-1} across the US (Townsend-Small et al., 2016) and 35 g CH_4 h^{-1} in the UK (Boothroyd et al., 2016). As most of the methods presented here require access to the source, we considered 200 g CH_4 h^{-1} to be a sensible limit to the emission rate and is larger than the emissions observed by many previous studies. Therefore, the scope of this study is limited to estimating CH_4 emissions from a single point source that we would realistically be able to approach and measure, i.e. between 40 and 200 g CH_4 h^{-1}."

Reviewer 2 General comment 2:

There is a lack of context elaborating on the specific needs of the industry (e.g. buried sources are ignored or implicitly included as the paper seem to focus on aerial point sources), and possibly introducing some sort of statistical distribution of leak size would be useful.

Response to reviewer:

Emissions from buried oil and gas infrastructure is a different study. The authors are involved in many studies to investigate this (R-PLUME Large Pipeline Leaks - Energy Institute (colostate.edu); Upstream Pipeline Safety, Integrity and Detection - Energy Institute (colostate.edu); Innovative Sensor Network for Subsurface Emissions - Energy Institute (colostate.edu)) and, unlike above-ground point sources, below ground emissions evolve to become area emissions at the surface with methane sinks throughout the surface as well as at the surface-atmosphere interface. To our knowledge, evaluating below-surface quantification methods using controlled releases has never been conducted before and would merit a separate publication.

The data presented in this manuscript represents the first time a controlled release has been used to evaluate and compare the performance of five above-ground point-source emission quantification methods. This presents a major step-forward in understanding utility of each these quantification methods. Additionally, we present data for researchers, regulators or private operators on which method is the most suitable for deployment in a given circumstance with operational shortcomings, such as instrument availability.

Changes to the manuscript:

To highlight why we are measuring above-ground point sources we have included the following text.

At L28:

". Quantification of CH_4 emissions from abandoned wells has recently become an area of interest as studies suggest over 200 Gg CH_4 yr^{-1} is emitted from 2.2 million abandoned wells in the US alone (US EPA, 2021). Quantifying and then plugging these wells makes them an attractive target for achieving goals set out in the Paris Agreement (Nisbet et al., 2020). Additionally, private companies are beginning initiatives to generate revenue through carbon credits gained by plugging wells and accurate quantification is essential for realizing the capital.."

At L33:

"Typically, an emission from an abandoned well can be considered as an above-ground point source that is relatively small in emission size, up to 180 g CH_4 $hour^{-1}$ (Riddick et al., 2019a; Pekney et al., 2018; Townsend-Small et al., 2016; Boothroyd et al., 2016). Other emission sources, such as emissions pipeline leakage, are fundamentally different in behavior, where gas travels through the soil and forms an area emission at the surface, these sources require different methods for estimating the emission, e.g. mass balance or eddy covariance. Area emissions could form if a plugged well leaks from corrosion of the borehole casing, but this will not be discussed in this study."

Reviewer 2 General comment 3:

Finding that accuracy calculated from 3 estimates (Ar) improves compared to the discrepancy of a first, single estimate (A) is trivial but nevertheless occupies a significant part of the results (sect 3.1) and discussion: Sect. 4.1 is dedicated to this, even discussing a particular occurrence where 3-point accuracy is lower than the single point discrepancy. In general, if Ar is available, there is little statistical sense to discuss A at all.

Response to reviewer:

Have redone the results and discussion sections to remove emphasis from *A*.

Changes to the manuscript:

At L317:

"Both the dynamic chamber (A_r = -10%, -8%, -10% at emission rates of 40, 100 and 200 g CH_4 h^{-1}, respectively) and Hi Flow (A_r = -18%, -16%, -18%) repeatedly underestimate the emission, but the dynamic chamber is the most accurate for measurement. For the far field methods, the bLs method underestimated emissions (A_r = +6%, -6%, -7%) while the GP method significantly overestimated the emissions (A_r = +86%, +57%, +29%) despite using the same meteorological and concentration data as input. These findings are consistent with another study (Bonifacio et al., 2013), however, this is the first study that has compared both to a known emission rate. In all cases the accuracy in the emission estimate increased with emission rate apart from the Hi Flow. The Bacharach Hi Flow system is designed to measure emission from 50 g CH_4 h^{-1} to 9 kg CH_4 h^{-1} to an accuracy of ± 10%. All flow rates presented here are at the lowest range that the Hi Flow can measure and it is likely that the uncertainty in the systems sensors that measures between 40 and 400 g CH_4 h^{-1} is of negligible difference.

The method that improves the most as the emission rate increases is the GP method, where accuracy increases from +87% to +29% as the emission rate increased from 40 to 200 g CH_4 h^{-1}. This improvement in emission is likely caused by the increased size of the plume and the ability of GP model to parameterize the concentration at distances from

the centerline of the plume. Although not explicitly stated, the parameterization of the lateral dispersion in the GP model is the same at 100 m as at 5 m which is unlikely. Other controlled release experiments using the GP approach show similar uncertainties, one experiment reported average emissions calculated using a GP model less than 20% (release rates were not published), with the uncertainty mainly driven by atmospheric variability (Caulton et al., 2019). Another showed uncertainties of ±50% for triplicate measurements of emissions between 90 and 970 g CH_4 h^{-1} (Caulton et al., 2018).

Data do not exist on controlled release experiments using a dynamic chamber. One study suggested a theoretical emissions uncertainty in the dynamic chamber approach of ±7% (Riddick et al., 2019a), with the largest source of uncertainty caused by the measurement of the flow rate of air through the chamber. Other sources of uncertainty for the dynamic chamber methods are relatively negligible as the methane quantification of the background gas and the gas at steady state (assuming complete mixing of the gas in the chamber) using the GC is highly accurate over a large concentration range and the volume of the chamber fixed by a plastic structure.

A controlled release has been conducted for the bLs model, but only for an emission from an area source (Ro et al., 2011) at the surface and not analogous to the emissions of this study. Ro et al. (2011) estimated the bLs uncertainty at ± 25% for a gas emitted at an unspecified rate from a 27 m^2 emission area. As with the GP approach, the bLs model's main source uncertainty is the parameterization of the atmospheric stability (Riddick et al., 2012; Flesch et al., 1995; Ro et al., 2011). The main advantage of the bLs model over the GP at these short distances is it calculates the lateral dispersion of gas for individual particles, while the GP uses an averaged dispersion parameter."

Reviewer 2 General comment 4:

The "decision making paradigm" (Sect 4.2) is limited in scope by ignoring other techniques and situations that may representative of the industry, and it seems to operates in its own limited rationality, letting the reader ignore other works.

Response to reviewer:

After thinking about this we have decided to remove the flow diagram.

Specific comments

Reviewer 2 Specific comment 1:

L16: not only for 200 g/h.

Response to reviewer:

As suggested, have included the range of emissions.

Changes to the manuscript:

At L 16:

"for emissions of 40 to 200 g CH_4 h^{-1}"

Reviewer 2 Specific comment 2:

L29: why is this 200g/h threshold important? Is there a scientific rationale? Is it specifically representative of situations or technical challenges in the industry?

Response to reviewer:

From experience, it becomes difficult to work effectively in areas near NG emissions greater than 200 g CH_4 h^{-1}. As a rule of thumb at this emission rate, we suggest access to areas closer than 10 m of the source be restricted. As most of the methods require access to the source, we considered this to be a sensible limit to emission and it is more than the upper limit to emissions measurements in the field. The maximum emission from unplugged and abandoned wells was measured at 177 g CH_4 h^{-1} in West Virginia (Riddick et al., 2019a), 175 g CH_4 h^{-1} in Pennsylvania (Pekney et al., 2018), 146 g CH_4 h^{-1} across the US (Townsend-Small et al., 2016) and 35 g CH_4 h^{-1} in the UK (Boothroyd et al., 2016).

Changes to the manuscript:

At L59:

"From experience and the response of a 4-gas monitor, working close enough to measure emissions greater than 200 g CH_4 h^{-1} for many of these methods (especially the chambers and Hi Flow) can be unsafe, therefore this study is limited to quantifying CH_4 emissions between the lowest flow METEC can produce (40 g CH_4 h^{-1}) and the highest

flow we feel comfortable measuring with these methods (200 g CH_4 h^{-1}). Putting these emission ranges into real-word context, the maximum emission from unplugged and abandoned wells was measured at 177 g CH_4 h^{-1} in West Virginia (Riddick et al., 2019a), 175 g CH_4 h^{-1} in Pennsylvania (Pekney et al., 2018), 146 g CH_4 h^{-1} across the US (Townsend-Small et al., 2016) and 35 g CH_4 h^{-1} in the UK (Boothroyd et al., 2016). As most of the methods presented here require access to the source, we considered 200 g CH_4 h^{-1} to be a sensible limit to the emission rate and is larger than the emissions observed by many previous studies. Therefore, the scope of this study is limited to estimating CH_4 emissions from a single point source that we would realistically be able to approach and measure, i.e. between 40 and 200 g CH_4 h^{-1}."

At L74:

"We add the caveat that we will only present data from measurement methodologies can be conducted safely, wearing PPE as regulated at the Colorado State University Methane Emissions Technology Evaluation Center (METEC) facility in Fort Collins, CO, USA (steel toe boot, FR overalls, hard hat, safety glasses and 4-gas monitor)."

Reviewer 2 Specific comment 3:

L31-38: A number of techniques and approaches (FLIR, mass balance, tracer release, remote sensing…) are ignored in this study. Their existence and their absence here should be acknowledged and thoroughly commented. They may not be included in this study for some (presumably good) reasons?

As presented above, we have clearly stated that this is for measuring small above-ground point source emissions and particularly focusses on abandoned oil and gas wells. Other methods that the reviewer has listed below are simply not suitable for measuring emissions typical of the majority of abandoned wells. FLIR is not a quantification approach, mass balance could be used but I have never come across it being used to measure individual point sources, tracer release is a long a technically tricky approach and ill-suited to this sort of hit-and-run measurement approach, and remote sensing has typical detection limits of 10+ kg CH_4 h^{-1} for drones and 100+ kg CH_4 h^{-1} for aircraft and completely unsuitable.

Changes to the manuscript:

At L52:

"Other quantification methods are generally unsuitable for measuring emissions from abandoned wells. Infra-red cameras, such as FLIR cameras, cannot be used to quantify emission and have difficultly detecting plume smaller emissions (Zimmerle et al., 2020). Mass balance approaches are unlikely to detect the small and narrow plume from the abandoned well. Tracer release is technically demanding, takes a long time to make a single measurement and requires road access for measurement. Remote sensing has typical detection limits of 10+ kg CH_4 h^{-1} for aircraft (Duren et al., 2019), 100+ kg CH_4 h^{-1} for satellites (Cooper et al., 2022) and unsuitable for these types of emission source. As such, these other quantification methods will not be investigated in this study."

Reviewer 2 Specific comment 4:

L37: "despite the interest in developing methods": unclear

Response to reviewer:

Have reworded the sentence

Changes to the manuscript:

At L47:

"Emissions calculated using the majority of these methods have not been comprehensively compared using to controlled emission source rates."

Reviewer 2 Specific comment 5:

L78: what is "total reflection of CH4 at the surface"?

Response to reviewer:

This reflects an amendment to the original Gaussian Plume model where gas reflected from the surface of the Earth is accounted for in the downwind plume.

Changes to the manuscript:

At L177:

"where gas reflected from the surface of the Earth is accounted for in the downwind plume."

Reviewer 2 Specific comment 6:

L79: I would argue that mass balance is the simplest method, rather than GP.

Response to reviewer:

Fair enough, however, mass balance is generally not used for abandoned wells.

Changes to the manuscript:

At L179:

"considered here"

Reviewer 2 Specific comment 7:

L81: perfect "gaussian" plumes are indeed seldom met in nature. But also it is rare to have lonely 'weak' plumes in an industrial setting, so the GP approach needs to account somehow for multiple sources.

Response to reviewer:

Yes, any background interference can be accounted for up upwind and downwind measurement. The exclusion of this was an oversight in the manuscript and should have been highlighted in the text. Here we define the enhancement as the difference between the downwind concentration and the background concentration measured upwind.

Changes to the manuscript:

At L171:

"enhancement"

At L176:

"The enhancement is defined as the difference between the downwind concentration and the background concentration measured upwind."

Reviewer 2 Specific comment 8:

L98: is it the same 38% uncertainty applicable for 4ug/h and 3kg/h?

Response to reviewer:

This is the major shortcoming of non-controlled release emission uncertainties and doesn't account for any variability in the emission size or variability in environmental conditions. This uncertainty was calculated as a desk based study and derived from the root of the sum of the individual sources of uncertainty squared.

Changes to the manuscript:

At L47:

"Emissions calculated using the majority of these methods have not been comprehensively compared using to controlled emission source rates."

Reviewer 2 Specific comment 9:

L102: if the 200g/h limit is for safety/practical reason, is it still useful in real life?

From experience, it becomes difficult to work effectively in areas near NG emissions greater than 200 g CH_4 h^{-1}. As a rule of thumb at this emission rate, we suggest access to areas closer than 10 m of the source be restricted. As most of the methods require access to the source, we considered this to be a sensible limit to emission and it is more than the upper limit to emissions measurements in the field. The maximum emission from unplugged and abandoned wells was measured at 177 g CH_4 h^{-1} in West Virginia (Riddick et al., 2019a), 175 g CH_4 h^{-1} in

Pennsylvania (Pekney et al., 2018), 146 g CH_4 h^{-1} across the US (Townsend-Small et al., 2016) and 35 g CH_4 h^{-1} in the UK (Boothroyd et al., 2016).

Changes to the manuscript:

At L59:

"From experience and the response of a 4-gas monitor, working close enough to measure emissions greater than 200 g CH_4 h^{-1} for many of these methods (especially the chambers and Hi Flow) can be unsafe, therefore this study is limited to quantifying CH_4 emissions between the lowest flow METEC can produce (40 g CH_4 h^{-1}) and the highest flow we feel comfortable measuring with these methods (200 g CH_4 h^{-1}). Putting these emission ranges into real-word context, the maximum emission from unplugged and abandoned wells was measured at 177 g CH_4 h^{-1} in West Virginia (Riddick et al., 2019a), 175 g CH_4 h^{-1} in Pennsylvania (Pekney et al., 2018), 146 g CH_4 h^{-1} across the US (Townsend-Small et al., 2016) and 35 g CH_4 h^{-1} in the UK (Boothroyd et al., 2016). As most of the methods presented here require access to the source, we considered 200 g CH_4 h^{-1} to be a sensible limit to the emission rate and is larger than the emissions observed by many previous studies. Therefore, the scope of this study is limited to estimating CH_4 emissions from a single point source that we would realistically be able to approach and measure, i.e. between 40 and 200 g CH_4 h^{-1}."

At L74:

"We add the caveat that we will only present data from measurement methodologies can be conducted safely, wearing PPE as regulated at the Colorado State University Methane Emissions Technology Evaluation Center (METEC) facility in Fort Collins, CO, USA (steel toe boot, FR overalls, hard hat, safety glasses and 4-gas monitor)."

Reviewer 2 Specific comment 10:

L108: can you please then comment on what was done at leak rates higher than 200 g/h? What are the limitations to transfer these conclusions to smaller leak rates? Why should we care about leak rates below 200g/h?

Emissions of 200 g h^{-1} are not insignificant and you know they are happening if you stand next to them. Typically emissions of this size would be measured using downwind techniques. From experience, it becomes difficult to work effectively in areas near NG emissions greater than 200 g CH_4 h^{-1}. As a rule of thumb at this emission rate, we suggest access to areas closer than 10 m of the source be restricted. As most of the methods require access to the source, we considered this to be a sensible limit to emission and it is more than the upper limit to emissions measurements in the field. The maximum emission from unplugged and abandoned wells was measured at 177 g CH_4 h^{-1} in West Virginia (Riddick et al., 2019a), 175 g CH_4 h^{-1} in Pennsylvania (Pekney et al., 2018), 146 g CH_4 h^{-1} across the US (Townsend-Small et al., 2016) and 35 g CH_4 h^{-1} in the UK (Boothroyd et al., 2016).

Changes to the manuscript:

At L59:

"From experience and the response of a 4-gas monitor, working close enough to measure emissions greater than 200 g CH_4 h^{-1} for many of these methods (especially the chambers and Hi Flow) can be unsafe, therefore this study is limited to quantifying CH_4 emissions between the lowest flow METEC can produce (40 g CH_4 h^{-1}) and the highest flow we feel comfortable measuring with these methods (200 g CH_4 h^{-1}). Putting these emission ranges into real-word context, the maximum emission from unplugged and abandoned wells was measured at 177 g CH_4 h^{-1} in West Virginia (Riddick et al., 2019a), 175 g CH_4 h^{-1} in Pennsylvania (Pekney et al., 2018), 146 g CH_4 h^{-1} across the US (Townsend-Small et al., 2016) and 35 g CH_4 h^{-1} in the UK (Boothroyd et al., 2016). As most of the methods presented here require access to the source, we considered 200 g CH_4 h^{-1} to be a sensible limit to the emission rate and is larger than the emissions observed by many previous studies. Therefore, the scope of this study is limited to estimating CH_4 emissions from a single point source that we would realistically be able to approach and measure, i.e. between 40 and 200 g CH_4 h^{-1}."

At L74:

"We add the caveat that we will only present data from measurement methodologies can be conducted safely, wearing PPE as regulated at the Colorado State University Methane Emissions Technology Evaluation Center (METEC) facility in Fort Collins, CO, USA (steel toe boot, FR overalls, hard hat, safety glasses and 4-gas monitor)."

Reviewer 2 Specific comment 11:

L110: it seems that the section 2 repeat the method description from the introduction, or at least lists again each method. The text would be more fluid and logical if all method description (including Eqs 1-5) are into the Method section and the introduction then focuses more strongly on state of the art, context and research questions.

Response to reviewer:

As suggested have moved introduction to the methods section

Reviewer 2 Specific comment 12:

L117: what are the uncertainties associated to the release rate (for example from the mass flow controllers)? How long does the releases last? What is the shape of the injection exhaust/outlet? Is there any attempt to reproduce a 'diffusive' exhaust? Or to control the gas exhaust velocity?

Response to reviewer:

At the METEC site, compressed natural gas, with methane compositions ranging from 85 to 95%vol, is supplied from two 145 L cylinders. Flown rates are controlled using a pressure regulator and precision orifices. A point source is considered to be an emission from an aperture. In this case, the hole was a 6mm diameter tube. We have included this detail in the methods section.

Changes to the manuscript:

At L86:

"(diameter 6 mm)"

At L82

"At the METEC site, compressed natural gas, with methane compositions ranging from 85 to 95%vol, is supplied from two 145 L cylinders and flow rates controlled using a pressure regulator and precision orifices."

Reviewer 2 Specific comment 13:

L122: how long is the calibration? How precise is the gas standard?

Response to reviewer: The calibration takes the same time as a sample to run through the GC ~ 20 minutes. The gas sample accuracy is rated at ± 5%.

Changes to the manuscript:

At L93

"(accuracy of standard ± 5%)"

Reviewer 2 Specific comment 14:

L127: how well do we expect the air inside the chamber to be mixed with the fan?

Response to reviewer:

We expect the air to be completely mixed, this is a major source of uncertainty with the statice chamber.

Changes to the manuscript:

L103:

"A fan was secured inside the chamber and used to circulate the air following the method of (Riddick et al. 2019a) to ensure the air inside the chamber was fully mixed."

Reviewer 2 Specific comment 15:

L128 how is the air sample drawn?

Response to reviewer:

Using a gas syringe.

Changes to the manuscript:

L106:

"When the chamber is sealed with the ground, following Riddick et al. (2019a), an air sample is drawn using a gas syringe."

Reviewer 2 Specific comment 16:

L134: why zero wind condition and not "wind speed below X m/s"? can you elaborate on this serious limitation?

Response to reviewer:

"zero wind conditions" was used to represent the absence of wind (which happens often at METEC). During any wind the chamber acted as a sail, smaller chambers are better in the wind but more dangerous as they fill with gas quickly.

Changes to the manuscript:

At L113:

"During any wind the chamber acted as a sail and larger chamber lifted from the ground, therefore, smaller chambers are better in the wind but quickly fill with gas making measurement difficult."

Reviewer 2 Specific comment 17:

L165: at what height is the gas scooter measurement made? Was any attempt made to measure CH4 across the plume (cross wind) to confirm the gaussian shape of the plume? If not, why?

Response to reviewer:

1.5 m. Measuring the shape and size of a plume is non-trivial even in neutral atmospheric stability conditions and difficult to do with plumes as small as presented here as very small changes in wind direction will change the plume. The aim here was to measure the time averaged concentration at a point downwind of the plume.

Reviewer 2 Specific comment 18:

L178: What is the impact of very small distance (5m) on the accuracy – does the model have a lower limit? Also measurements at 5m downwind suggest that relatively close access (and at the same height AGL) is possible, this should be acknowledged (also in Table 1).

Response to reviewer:

The model does not technically have any minimum limiting caveats about distance.

Changes to the manuscript:

At L192:

"Here, it assumed that the experiments are conducted as close as possible to the source without direct access to the emission point."

Table 1 caption

"(Y denotes having permission to touch/enclose the emission point and N denotes experiments are conducted as close as possible to the source without direct access)"

Reviewer 2 Specific comment 19:

L195 and following: why not automate chamber opening to avoid explosive limit? This should be easily done in a commercial context, and acknowledged in the discussion.

Response to reviewer:

This could be done, but the chamber is still collecting an unknown composition of gas. The automation of a chamber then takes away the operator's control of when this is released. The operator taking an unnecessary risk unless they are wearing self-contained breathing apparatus.

Changes to the manuscript:

At L257:

"The static chamber could be automated to release gas when CH_4 concentration inside the chamber approaches LEL to prevent chamber becoming explosive. The major shortcoming of this strategy is that the automation of a chamber takes away the operator's control of when gas is released, which could happen at an inconvenient during measurement. If an automated system is used for collecting gas of unknown composition self-contained breathing apparatus should be worn."

Reviewer 2 Specific comment 20:

L227: the accuracy of the single "snapshot" becomes irrelevant once 3 repeat measurements are available for the release. Accuracy and precision derived from 3 points supersede the single measurement discrepancy as informative numbers.

Response to reviewer:

We acknowledge this may be of limited interest and have removed the section and have discussed A_r.

Changes to manuscript

At L317:

"Both the dynamic chamber (A_r = -10%, -8%, -10% at emission rates of 40, 100 and 200 g CH_4 h^{-1}, respectively) and Hi Flow (A_r = -18%, -16%, -18%) repeatedly underestimate the emission, but the dynamic chamber is the most accurate for measurement. For the far field methods, the bLs method underestimated emissions (A_r = +6%, -6%, -7%) while the GP method significantly overestimated the emissions (A_r = +86%, +57%, +29%) despite using the same meteorological and concentration data as input. These findings are consistent with another study (Bonifacio et al., 2013), however, this is the first study that has compared both to a known emission rate. In all cases the accuracy in

the emission estimate increased with emission rate apart from the Hi Flow. The Bacharach Hi Flow system is designed to measure emission from 50 g CH_4 h^{-1} to 9 kg CH_4 h^{-1} to an accuracy of ± 10%. All flow rates presented here are at the lowest range that the Hi Flow can measure and it is likely that the uncertainty in the systems sensors that measures between 40 and 400 g CH_4 h^{-1} is of negligible difference.

The method that improves the most as the emission rate increases is the GP method, where accuracy increases from +87% to +29% as the emission rate increased from 40 to 200 g CH_4 h^{-1}. This improvement in emission is likely caused by the increased size of the plume and the ability of GP model to parameterize the concentration at distances from the centerline of the plume. Although not explicitly stated, the parameterization of the lateral dispersion in the GP model is the same at 100 m as at 5 m which is unlikely. Other controlled release experiments using the GP approach show similar uncertainties, one experiment reported average emissions calculated using a GP model less than 20% (release rates were not published), with the uncertainty mainly driven by atmospheric variability (Caulton et al., 2019). Another showed uncertainties of ±50% for triplicate measurements of emissions between 90 and 970 g CH_4 h^{-1} (Caulton et al., 2018).

Data do not exist on controlled release experiments using a dynamic chamber. One study suggested a theoretical emissions uncertainty in the dynamic chamber approach of ±7% (Riddick et al., 2019a), with the largest source of uncertainty caused by the measurement of the flow rate of air through the chamber. Other sources of uncertainty for the dynamic chamber methods are relatively negligible as the methane quantification of the background gas and the gas at steady state (assuming complete mixing of the gas in the chamber) using the GC is highly accurate over a large concentration range and the volume of the chamber fixed by a plastic structure.

A controlled release has been conducted for the bLs model, but only for an emission from an area source (Ro et al., 2011) at the surface and not analogous to the emissions of this study. Ro et al. (2011) estimated the bLs uncertainty at ± 25% for a gas emitted at an unspecified rate from a 27 m^2 emission area. As with the GP approach, the bLs model's main source uncertainty is the parameterization of the atmospheric stability (Riddick et al., 2012; Flesch et al., 1995; Ro et al., 2011). The main advantage of the bLs model over the GP at these short distances is it calculates the lateral dispersion of gas for individual particles, while the GP uses an averaged dispersion parameter."

Reviewer 2 Specific comment 21:

L232: abs(A) "decreases": this is not confirmed in Fig 2b. can you please explain where this comes from?

Response to reviewer:

Have removed the section

Reviewer 2 Specific comment 22:

L242: please quantify “generally”

Response to reviewer:

Have removed the sentence

Reviewer 2 Specific comment 23:

L245: Ar>A : to me this is meaningless. It just means that there was some luck in the first value, and therefore it carries little sense to report A once you have Ar.

Response to reviewer:

As suggested, have removed the sentence.

Reviewer 2 Specific comment 24:

L251: This is expected and may be seen as trivial. On the opposite, what is surprising is when it is not the case. Why does SD for some techniques increase with increasing emission rate?

Response to reviewer:

Have added text to discuss the S.D.

Changes to the manuscript:

At L357:

“The emission estimates quantified using direct methods, dynamic chamber and Hi Flow sampler, have a lower S.D. than the far-field methods (Figure 2B). The S.D. of direct measurement methods remain relatively constant for emissions between 40 and 200 g CH_4 h^{-1} and reflects the relative simplicity of the methods. Assuming all other

parameters are measured correctly, for direct methods the variability in emission estimate is a function of how well the CH_4 is mixed into the air in the chamber during the measurement.

Variability in the far field emission estimates is much larger an reflects the relative complexity of inferring emissions. Variability in wind speed, wind direction and atmospheric stability over the 20-minute averaging period are likely to propagate through to large variability in the emission estimate. It may be reasonable to suggest that the variability in bLs calculated emission less than for the GP method because of the added parametrization available (roughness length and gas species). In summary, the penalty of downwind measurement is a higher uncertainty in individual measurements, but this appear to be corrected for by the bLs model through repeat measurements where uncertainty is corrected for by the stochastic nature of particle movement modelling."

Reviewer 2 Specific comment 25:

Fig.3 in my opinion would deserve further comments and discussion.

Response to reviewer:

Have moved Figure to become Figure 2B, as suggested, and added text to discuss.

Changes to the manuscript:

At L346:

"The emission estimates quantified using direct methods, dynamic chamber and Hi Flow sampler, have a lower S.D. than the far-field methods (Figure 2B). The S.D. of direct measurement methods remain relatively constant for emissions between 40 and 200 g CH_4 h^{-1} and reflects the relative simplicity of the methods. Assuming all other parameters are measured correctly, for direct methods the variability in emission estimate is a function of how well the CH_4 is mixed into the air in the chamber during the measurement.

Variability in the far field emission estimates is much larger an reflects the relative complexity of inferring emissions. Variability in wind speed, wind direction and atmospheric stability over the 20-minute averaging period are likely to propagate through to large variability in the emission estimate. It may be reasonable to suggest that the variability in bLs calculated emission less than for the GP method because of the added parametrization available (roughness length and gas species). In summary, the penalty of downwind measurement is a higher uncertainty in individual measurements, but this appear to be corrected for by the bLs model through repeat measurements where uncertainty is corrected for by the stochastic nature of particle movement modelling."

Reviewer 2 Specific comment 26:

L265: Which technique has the lowest A is relatively unimportant when Ar is available. Ar>A is also fairly trivial from a very general perspective.

Response to reviewer:

As suggested, have removed the sentence.

Reviewer 2 Specific comment 27:

L280-283: These sentences are vague, complicated. There seems to be a confusion between the site scale uncertainty and the single GP uncertainty. Could you please clarify in terms of separating biases and random errors? (and all methods here seem to have consistent biases). Did you make any attempt to look at uncertainty budget in the different methods, the GP in particular, including the choice of stability class? Does it match with the 3-point accuracy?

Response to reviewer:

We have added to this section to investigate the causes of the inaccuracy of the GP model.

Changes to the manuscript:

At L363:

"It is, however, concerning that many of the methods show a bias in measurement results and in particular the GP model (Figure 3). In most studies, it is assumed that in taking multiple measurements the average uncertainty will be reduced to an aggregate, unbiased emission estimate. Taking the GP emission estimates as an example, the individual calculated emissions are all overestimates of the true emission, therefore, suggesting a fundamental shortcoming in the method (Figure 3). These measurements were taking four days apart in similar environmental conditions (all PGSC C) with wind direction being the only difference between measurements, which can be seen from the correlation between the uncertainty and horizontal distance from plume center (Figure 3B). As mentioned above, it is likely that this is due to the lateral dispersion in the GP approach being parametrized incorrectly, i.e. using values that were defined for distances of 100 m. This suggests that using the GP approach for distances less

than 100 m, it is not correct to assume that repeat measurements will remove bias in the calculated average emission."

Reviewer 2 Specific comment 28:

L293-294: Can you please explain the "meaning and balance" and provide examples of studies that "present unexpected findings"?

Response to reviewer:

Have removed this sentence

Reviewer 2 Specific comment 29:

Fig. 4: other parameters/selectors would be useful but are ignored in this study: is the source buried? What is the source intensity? Did you perform a source detection prior to quantification? Some techniques are missing; as such, the value of this diagram is very poor for decision making, although the idea is good.

Response to reviewer

On reflection we have decided to remove the decision making flow chart.

Editorial comments

Reviewer 2 Editorial comment 1:

L21: GP: expand acronym

Changes to the manuscript:

At L 11:

"(GP)"

Reviewer 2 Editorial comment 2:

L103: “copy” I suggest “reproduce”

Response to reviewer:

Changed as suggested

Reviewer 2 Editorial comment 3:

L177-178: the cited papers are not listed in the References section

Response to reviewer:

Added to Reference Section

Reviewer 2 Editorial comment 4:

Fig 3: what is y axis unit?

Response to reviewer:

Added “(%)” to Figure 3